International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Dec-23)

http://dx.doi.org/10.12785/ijcds/1401115

Social Distance Detection using Customized YOLOv4 Model

Vishal Parikh[1], Yash Chelani[1] and Parita Oza[1]

[1]*Department of Computer Science and Engineering, Institute of Technology, Nirma University, Gujarat, India*

Received 24 Feb. 2023, Revised 1 Nov. 2023, Accepted 30 Nov. 2023, Published 28 Dec. 2023

Abstract: This work proposes a technique to detect social distancing using deep learning to ensure the social distance between people in order to reduce the impact of the coronavirus pandemic. By analysing a video stream, the social distance detecting tool is developed to inform people to keep a safe distance from each other. In this work, we have used a customized YOLOv4 model, which was created by using some layers of YOLO from the initial model. Then, we trained those layers on a labelled person image dataset so that it can be used only for person detection. Input is given in the form of video or can be taken from a webcam, which is then converted to a 2-D frame for measuring the distance between two people. Any person not following social distancing will be displayed in a red bounding box, while those who are following will be displayed in a green bounding box. A pre-recorded video was used to test the suggested strategy. We have named our work Social Distance Detection using Customized YOLOv4 (SDDYv4).

Keywords: YOLOv4, Person Detection, Social Distance, Depth Estimation

I. Introduction and Overview

Since there was no effective treatment for the coronavirus (Covid-19) when the pandemic first started, the general public was worried about the virus's spread. Covid-19 has been classified as a pandemic by the World Health Organization (WHO) due to an increase in the number of cases that have been documented around the globe. Some nations have implemented lockdowns, requiring citizens to remain indoors during this critical time in order to stop the virus from spreading [1].

The best way to stop the transmission of Covid-19 is to avoid direct contact with other people, according to public health organisations like the Centers for Disease Control and Prevention (CDC). As there was no viable cure that time. By adhering to government directives, which included strategies like social separation, citizens contributed to flattening the Covid-19 pandemic's curve [1][2].

To achieve social distance during the quarantine, group activities and congregations like gatherings, meetings, workshops, and prayer were outlawed. To minimise face-to-face interaction, people were advised to plan and carry out events as much as possible using various internet channels [3] [4]. People were urged to adopt proper hygiene, such as routine hand washing, using masks, and avoiding close contact with those who were exhibiting symptoms, in order to stop the virus from spreading further. However, there is a difference between knowing what to do to stop the infection from spreading and actually doing it [2]. As these nations cautiously resumed their economic activities in the new post-Covid-19 environment, worries regarding worker safety have emerged. People were urged to maintain a distance of at least 2 meters between them and refrain from any direct contact to lessen the risk of infection [4].

Currently our world is experiencing many variants of COVID across the globe and yet every person had not taken vaccine. So to ensure that the pandemic doesn't spread among the people social distancing should be followed and to ensure this various techniques of computer vision and deep learning can be taken in consideration to keep an eye on people. This paper present a deep learning based method to ensure social distancing between people.

In modern times, deep learning algorithms have become the preferred choice among researchers for object detection. There is a wide array of pre-trained deep learning models available for both object recognition and object detection. Notable examples include the Region-based Convolutional Neural Networks (R-CNN) [5], Fast R-CNN [6], Faster R-CNN [7], Mobilenet [8], and Single Shot Detector (SSD) [9]. These models have significantly advanced the field of computer vision and continue to be instrumental in various applications. Parikh et al. in [10], [11], [12], [13] generated object-based multi-view summary.

For object detection, You Only Look Once (YOLO) [14]

*E-mail address: *vishalparikh@nirmauni.ac.in, 18BCE263@nirmauni.ac.in, parita.prajapati@nirmauni.ac.in*

has been implemented and reviewed. YOLO can process 45 frames per second which is very fast. A CNN based model YOLOv4 [15] is customized and used for person detection and customization of model is done using transfer learning. In this work we have used some layers of pre-trained model of YOLOv4 on MS-COCO dataset[16], which was used for object detection and then we train it on custom labelled dataset of person class so that it can be used only for person detection[17][18]. Distance between two person is measured by euclidean distance formula with depth estimation added in it. The work is evaluated on different video streams by converting it into frames[19].

II. Related Works

This section shows some research employing Deep Learning (DL) and Convolution Nueral Networks (CNN) for social distance detection and object recognition. Prem et al.[20] from Wuhan studied and analyzed the consequences and effects of social distancing on the outbreak of the deadly virus. In order to simulate an ongoing trajectory outbreak using artificial location-specific interaction patterns, they used age structure susceptible-exposed-infected removed (SEIR) models for many social distancing techniques. They anticipated that a rapid uptick in interventions would cause an early secondary peak that would gradually level down. As we all know, maintaining social distance is vital to deal with the current situation, but economically flattening the curve against communicable diseases is a harsh measure.

Adolph et al.[2] has also highlighted the situation of USA, including the state reports about the scenario including social distancing and came to a conclusion that delaying the imposition of the concept of social distancing may have an adverse effect on person health. Moreover the social distancing concept also affected the financial conditions.

Various Asian countries including India have started the use of technology based solutions. The concept used here is the tracking of mobile devices and based on various diagnostic tests available on the applications[2]. The data regarding the infected and healthy people around you can be estimated based on the distance. The application implemented and used in India is the Aarogya Setu[21]. Al-Sa'd et al. [22] recognize the anomalies using Social distance and crowd monitoring systems. V V et al. [23] uses CCTV cameras for social detection and risk management.

Similar to this, the UK has released an app called C9 Corona Symptom Tracker that allows users to register their symptoms. In a similar vein, South Korea introduced Corona 100m, which shows the affected individuals within a 100m radius. Thermal cameras powered by AI are being used by China to spot people with high body temperatures in groups. In this dire circumstance, such technologies may aid in flattening the curve, but they also pose a threat to personal information[19].

Although it has certain limitations, the subject of study on person detection in video surveillance systems is well-established and depends on manual techniques for spotting unusual behaviours [24]. Different models throughout the time span has been used for Person detection and are trained and tested on different data sets[15].

These models are classified mainly under two categories: one-stage detector and two-stage detector[15]. Main difference between them is the steps involved in object detection, in one stage detector classification and identification of object is done in single step while in two stage detector a complex pipeline structure is followed for doing classification and identification of object.

Another difference between single stage and two stage is in single stage searching and specification of the class region is not required while in two stage it is mandatory to do the specification of area[15]. Table1 shows basic architecture of one stage and two stage detector.

In addition to it Joseph Redmon et al.[14] has provided an analysis of these techniques when trained and tested on different data sets[14]. Models considered for comparison are R-CNN[20], Fast R-CNN[25], Faster R-CNN[26][27] and YOLO[14] and the comparison is done on the basis of type of detector, object detection approach and limitations of the models.

Fig.I shows the detail comparison of different models.

The most advanced object detectors now available that use DL have benefits and drawbacks in terms of accuracy and speed. The object may have multiple spatial placements and aspect ratios inside the images. As a result, real-time object recognition algorithms based on the CNN model, including R-CNN and YOLO, have been developed to detect multi-classes in various sections of images. YOLO is one of the most popular deep CNN-based object detection algorithm in terms of speed and accuracy[14].

Adapting some of the techniques from the above work we present a computer vision technique which could be used for social distance detection .Detecting person, their count and drawing bounding boxes are done with the help of YOLO. For detecting distance between two person euclidean distance along with depth estimation is used and according to the predefined distance they are classified whether they are following social distancing or not. As the work carried

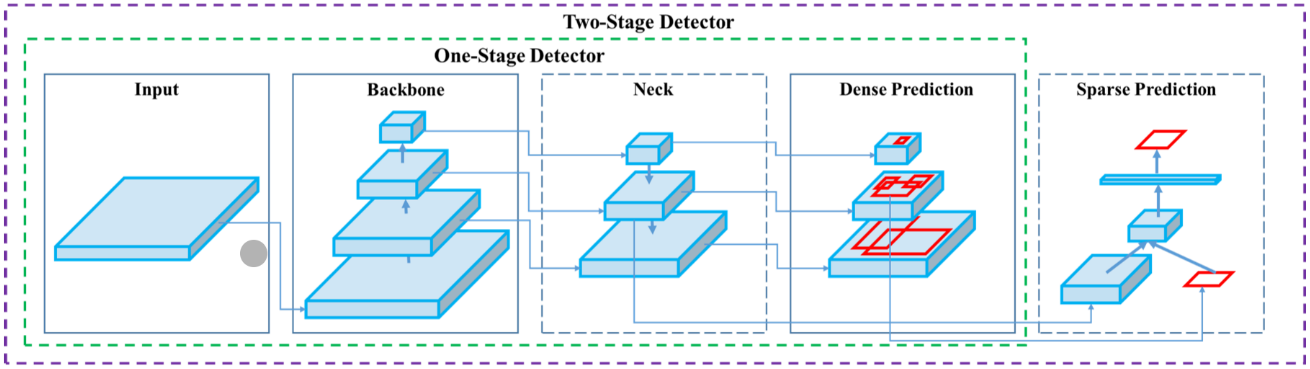


Figure 1. One stage and two stage detector[8]

Sr. No	Criteria	R-CNN	Fast R-CNN	Faster R-CNN	YOLO
1	Type of Detector	Two-stage Detector	Two-Stage Detector	Two-stage Detector	One-stage Detector
2	Approach used for object detection	Region proposal along with selective search	Same as R-CNN but the classification stage is boosted	Sharing computation along with use of neural networks to propose bounding boxes	Divide Image into Grid cells and assigns class with scores to detect potential bounding Box
3	Limitation	Huge processing time and difficult to apply	Huge processing time and difficult to apply in real-time	Real time application is difficult	Initial version of YOLO have detection problem when group of objects are detected, and their bounding box overlap

TABLE I. Comparision of different models

out is for fixed camera.

Fig.2 shows how our custom YOLOv4 detect person class only when applied on an image.

Figure 2. Illustration of YOLO algorithm for person detection

As per above discussion, several research works have been undertaken in order to create a better and more effective social distance monitoring system, but no one has used customized YOLOv4 model for person detection for detecting person.

III. Methodology

We proposed a method to detect a safe distance between people. The CNN methodologies and computer vision techniques are used in this work. Customized YOLOv4 is used for person detection and localization.

Input is given in the form of video or by webcam which is then converted to individual 2-D frames and then computed. If persons are detected in a frame then the computation process will continue otherwise it will jump to next frame. If more than one persons are detected then distance between them is calculated between person using euclidean distance with depth factor considered in it. Depth estimation is done by using the luminosity factor as it varies with respect to distance of person with camera.

Depending on the pre-defined minimum distance any two person who are not following social distancing will be indicated with red bounding box and people who are following social distancing are indicated with green bounding box.

A. person detection

Custom YOLO algorithm is used for person detection approach which in turn reduce the computational complexity by detecting the person class in multiple images in single stage. YOLO is considered one of the best algorithms of CNN used for object detection in images or videos. It has an incomparable advantage in terms of speed an accuracy when implemented in real time[28]. In this work we have used YOLOv4 algorithm over YOLOv3 as it is more accurate and faster by almost 10% to 12% then YOLOv3[29]. Its architecture basically consist of three parts which are mentioned below:-

- Backbone:- This part is used for extracting low features like edges and colours and it can learn from trained network. In YOLOv4 we use CSPDarknet-53(Cross stage partial Darknet-53)[29].
- Neck:- It is used to improving the feature extraction i.e. extraction of both low and high level features and processing capabilities of the model. In YOLOv4 we use modified SPP(Spatial pyramid pooling) and PANet aggregation for improving the feature extraction phase because they help to extract all the features in lesser amount of layers[29].
- Head:- It is mainly responsible for giving Output based on the model's requirements, such as classifier, detection frame, picture segmentation, and so on. It is same in YOLOv3 and YOLOv4[29].

The speed and accuracy difference between YOLOV3 and YOLO V4 is due to additional methods which are included in YOLOv4[29]. These methods are basically of two parts:-

- Bag of Freebies (BoF):- It includes changes that are inherited during the training phase without increasing the inference cost.
  - BoF for Backbone: for data augmentation using CutMix and Mosaic [29].
  - BoF for detector: incorporates random training shapes, co-sine annealing scheduler, regularisation, augmentation of mosaic data, self-adversarial training, reduction of grid sensitivity, use of many anchors for a single ground truth, and regularisation of the data. [29].

No. of Iterations	MAP	Class Confidence
1000	36	89
2000	38	94
3000	39	98
last.weights	37	92

TABLE II. MAP class confidence on different iteration

- Bag of Specials (BoS):- The insertion module is used to increase the accuracy of target detection by enhancing some features[29].
  - BoS for backbone: Cross-stage partial connections, Mish activation (CSP) [29].
  - Mish activation, SPP-block, SAM-block, PAN path-aggregation block, and DIoU-NMS are the detector's BoS [29].

In this work customization of YOLOv4 model is done using following steps:-

- In first step we take the weight file of 137 layers of pre-trained model of YOLOv4.
- In second step we took custom labelled dataset and use that dataset to modify and train that model to detect person class specifically.
- We saved the modified weights at every 1000 iteration to check the mean average precision(MAP) and class confidence .

For customization we modify various hyper-parameters in the configuration file, these changes are as follows:-

- Size of batch is set to 64.
- subdivision is set to 16
- max batches is set to 6000
- As we have to detect only one class steps are set to 4800,5400.
- Width and height for an input image is set to 416 X 416.
- As we have to detect only a single class filter size is set to 18.

customized weight file is considered good if avg loss comes in the range of 2 to 5. Fig.3shows relation between avg loss and number of iterations.

After training the customized weights file will only detect person class and will not detect any other class within that frame.

B. Bounding Box predictions

As YOLO detects multiple bounding boxes for a single class also, so to resolve the problem of overlapping

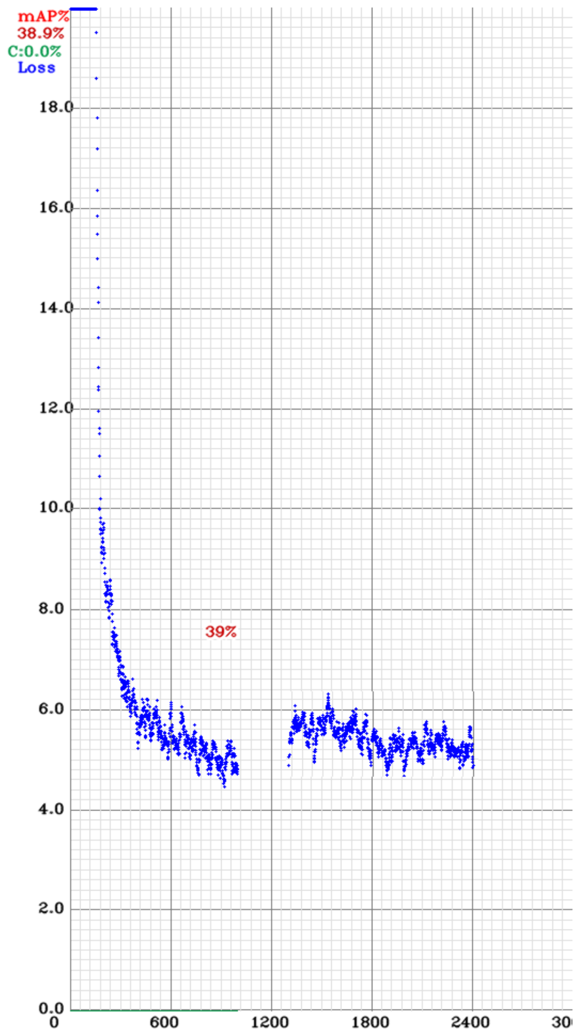


Figure 3. Average loss vs no of iterations from 0-2400 iterations

bounding boxes resulting in repeated detection for the same item is solved with non-maximum suppression (NMS). A confidence value of 0.5 and an NMS threshold of 0.3, respectively, were used to determine the final bounding boxes. This entails removing any bounding boxes that have more than a 30% overlap with another bounding box and only keeping classes with a confidence level of more than 50% [24]. fig.4shows application of NMS over an image.

Figure 4. Non Maximal suppression application[30]

C. Custom model creation

The custom model takes image as an input and gives the probability of class Person, their x min and y min coordinates along with its centroids, width and height of image. During the training and testing YOLOv4 model we have used Google open image dataset which consist over 600 classes and 1.9 million labelled images[31].

Fig.5 shows customized YOLOv4 model where we used first 137 layers weights for customizing the weight file of YOLOv4 along with some changes in batch size, subdivision and size of filter above the detection layer. A short process is represented, which will demonstrated how input is taken and in which form output is represented.

D. Depth Estimation based on surveillance camera footage

Once the people are identified, the distance between two people is calculated. The video frame available to us from the surveillance camera is only helpful for calculating the distance from 2-Dimensions, but when we convert video to 2-D frame one condition gets violated that is if two people standing on the same axis as of the surveillance camera, having different heights where the far one having a greater height than the close one's. In this case due to the height difference and the axis of capture, they both will be identified standing next to one another.

For that reason we have used a model monodepth2 which will identify the depth of the whole image and provide a disparity prediction based in the form of a heat map. Using which we can find the difference of pixel intensity and calculate more exact distance between two people.

Disparity is calculate by the formula

$$disp = disp * 416 * 0.58 \tag{1}$$

Here 416 is the width of frame taken from video of surveillance camera and 0.58 is scale factor and by using this disparity we calculate the third factor depth which we will use along the x and y coordinates of centroids to calculate the distance between two persons.

$$depth = focal * baseline/disp \tag{2}$$

In this depth estimation method focal length and baseline of surveillance camera is fixed primarily before applying.

E. Distance estimation

In this depth estimation method focal length and baseline of surveillance camera is fixed primarily before applying.

F. Distance estimation

This stage involved receiving all of the 2-D frame's bounding boxes and their centroids for the human class. Therefore, we employed euclidean distance combined with

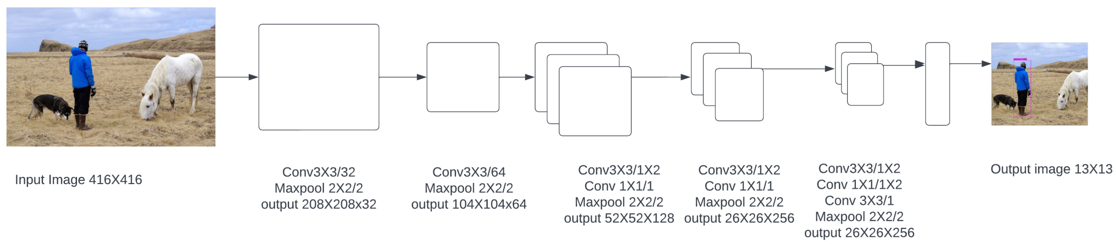


Figure 5. Customized YOLOv4 architecture

depth estimation, which is accomplished with the aid of brightness factor, to determine the distance between two or more people. Final distance between two person will be computed as:

Given the position of two persons in an image as (x1,y1) and (x2,y2) and their luminance is l1 and l2 respectively, hence the distance between them can be computed as :

$$dist = \sqrt{(l_1 - l_2)^2 + (x_1 - x_2)^2 + (y_1 - y_2)^2} \qquad (3)$$

For the pair of people who are not following the preset social distance are marked in red and the people who are following the predefined social distance are detected in green.

IV. Results and Discussion

The input video shows persons walking along the street where the camera is fixed at a particular angle. Individual frames are grabbed and resized for ease of implementation of YOLO. Fig.6, and fig.7 shows the output of social distance detection in an individual frame along with total count of people. People who are not following social distancing are depicted in red bounding box along with red centroids and people who are following social distancing are depicted in green bounding box with green centroids. Fig.8 shows the heat map generated during the processing of the video.

The custom video is used to test the working of Social distance detector. Fig.9, fig.10, and fig.11 shows social distance detection between two person when one person is moving in same axis and other person is at fixed position.

For training our custom model, initial weights file of 137 layers of YOLOv4 which is trained over labelled person dataset is used. During training the number of iterations is inversely proportional to average loss and directly proportional to the MAP and for checking this, modified weights was saved at the multiple of 1000 iterations and when checked the MAP at 1000 iterations is 32% at 2000 iterations it is 36% and at 3000 iteration it becomes 39% on the contrary average loss decreases from 1800 to 3.8.

Figure 6. Social distance detection on oxford town dataset

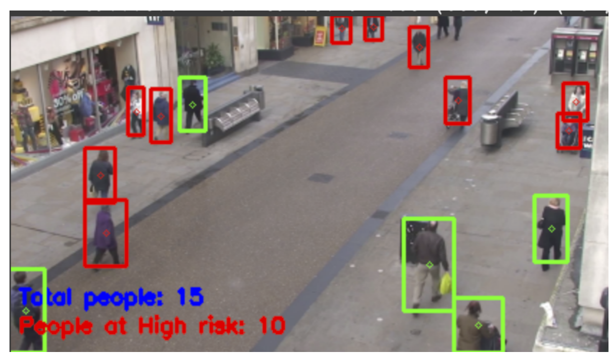


Figure 7. Social distance detection on oxford town dataset

Table III showcase the comparative analysis of the proposed work and other similar work available in the literature with respect to distance estimation method, approaches utilized for person detection, limitations and model performance (in terms of accuracy).

V. Conclusion and future work

A deep learning model is utilised to identify social distance. Any non-compliant pair of persons will have a red frame and centroid, and distance measurements between individuals may be performed using computer vision. A video of individuals strolling down a street was presented as proof for the suggested tactic. The results of the visualisation

Sr No.	Method	MAP for human detection and dataset used	Distance calculation approach	Limitations	Accuracy of the approach used
1.	A. H. Ahamad et al. [32]	Model used SSD Mobilnet V1 trained on MS COCO dataset has a MAP of 19%.	Euclidean distance	Very less MAP and doesn't detect all humans accurately Only detect social distancing in the region of interest which is masked, the rest of the frame remains unused.	62.5% on Oxford town dataset
2	Serigo Saponara et al. [33]	Model used YOLOv4 tiny trained on FLRI thermal image and custom made dataset for detection.	Euclidean distance along with bird eye view	Only detect social distancing in the region of interest which is masked, the rest of the frame remains unused.	97.48% on self taken video
3	Rujula Singh R et al. [28]	Model used YOLOv3-416 trained on COCO dataset	DBSCAN clustering algorithm	Less accuracy during person detection	89.79% on a self taken video
4	SDDYv4	Model used customize YOLO4 which is trained on Google open image dataset and has MAP of 39%	Euclidean distance along with depth estimation	Can be used only for surveillance purposes	90.3% on Oxford town dataset

TABLE III. Comparison of our method with different methods

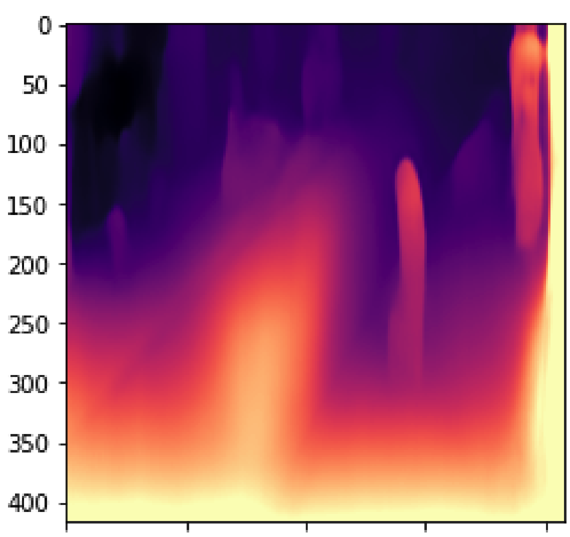


Figure 8. Heat map generated from oxford town dataset

Figure 9. Social distance detection on webcam

demonstrated the ability of the proposed technique to identify the whole population, the total number of people, and the total number of people who violate social distance. For usage in a number of settings, such as the office, a restaurant,

Figure 10. Social distance detection on webcam

Figure 11. Social distance detection on webcam

and a school, this work can be enhanced and refined. This work may be refined and modified to be used in a number of settings, such as the workplace, a restaurant, and a school.

Additionally, the work may be enhanced while optimising the processing capability of the hardware by enhancing the pedestrian detection algorithm and utilising other detection techniques including mask detection and human body temperature detection.

References

[1] "Coronavirus disease (covid-19)." [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[2] K. E. Ainslie, C. E. Walters, H. Fu, S. Bhatia, H. Wang, X. Xi, M. Baguelin, S. Bhatt, A. Boonyasiri, O. Boyd *et al.*, "Evidence of initial success for china exiting covid-19 social distancing policy after achieving containment," *Wellcome Open Research*, vol. 5, 2020.

[3] "community-mitigation-strategy." [Online]. Available: https://www.cdc.gov/coronavirus

[4] "Covid-19 (garis panduan)." [Online]. Available: https://www.moh.gov.my/index.php/pages/view/2019-ncov-wuhan-guidelines

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

[6] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[8] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[9] A. Womg, M. J. Shafiee, F. Li, and B. Chwyl, "Tiny ssd: A tiny single-shot detection deep convolutional neural network for real-time embedded object detection," in *2018 15th Conference on Computer and Robot Vision (CRV)*. IEEE, 2018, pp. 95–101.

[10] V. Parikh and j. v. n. y. Nathwani, Rutvi, "Multi-view video summarization: A review."

[11] V. Parikh and P. Sharma, "Intelligent video analytic based framework for multi-view video summarization," *International Journal Of Computing and Digital System*, pp. 619–628, 2022.

[12] V. Parikh, J. Mehta, S. Shah, and P. Sharma, "Comparative analysis of keyframe extraction techniques for video summarization," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 14, no. 9, pp. 2761–2771, 2021.

[13] V. Parikh, M. Kumhar, N. Wagh, N. Sutariya, and A. Shah, "Automatic helmet detection and gate opening using yolov3." *Grenze International Journal of Engineering & Technology (GIJET)*, vol. 8, no. 2, 2022.

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

[15] J. Li and Z. Wu, "The application of yolov4 and a new pedestrian clustering algorithm to implement social distance monitoring during the covid-19 pandemic," in *Journal of Physics: Conference Series*, vol. 1865, no. 4. IOP Publishing, 2021, p. 042019.

[16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[17] M. Yu, "Yolov4 transfer learning for scanned document structure recognition," Jul 2020. [Online]. Available: https://mingzhi2.medium.com/yolov4-transfer-learning-for-scanned-document-structure-recognition-dc3fc8bfe426

[18] R. Khandelwal, "Implementing yolo on a custom dataset," Nov 2019. [Online]. Available: https://towardsdatascience.com/implementing-yolo-on-a-custom-dataset-20101473ce53

[19] N. S. Punn, S. K. Sonbhadra, S. Agarwal, and G. Rai, "Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques," *arXiv preprint arXiv:2005.01385*, 2020.

[20] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday *et al.*, "The effect of control strategies to reduce social mixing on outcomes

of the covid-19 epidemic in wuhan, china, a modelling study," *The Lancet Public Health*, vol. 5, no. 5, pp. e261–e270, 2020.

[21] "Aarogya setu mobile app." [Online]. Available: https://www.mygov.in/aarogya-setu-app

[22] 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*, vol. 22, no. 2, p. 418, 2022.

[23] J. A. V V *et al.*, "Social distance detection using deep learning and risk management system," *arXiv preprint arXiv:2304.10259*, 2023.

[24] Y. C. Hou, M. Z. Baharuddin, S. Yussof, and S. Dzulkifly, "Social distancing detection with deep learning model," in *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*. IEEE, 2020, pp. 334–338.

[25] J. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4507–4515.

[26] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[28] S. Srinivasan, R. R. Singh, R. R. Biradar, and S. Revathi, "Covid-19 monitoring system using social distancing and face mask detection on surveillance video datasets," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE, 2021, pp. 449–455.

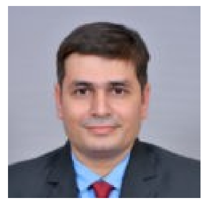
[29] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[30] J. Prakash, "Non maximum suppression: Theory and implementation in pytorch," Oct 2021. [Online]. Available: https://learnopencv.com/non-maximum-suppression-theory-and-implementation-in-pytorch/

[31] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, A. Kolesnikov *et al.*, "The open images dataset v4," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1956–1981, 2020.

[32] A. H. Ahamad, N. Zaini, and M. F. A. Latip, "Person detection for social distancing and safety violation alert based on segmented roi," in *2020 10th IEEE international conference on control system, computing and engineering (ICCSCE)*. IEEE, 2020, pp. 113–118.

[33] S. Saponara, A. Elhanashi, and Q. Zheng, "Developing a real-time social distancing detection system based on yolov4-tiny and bird-eye view for covid-19," *Journal of Real-Time Image Processing*, pp. 1–13, 2022.



Vishal Parikh Dr Vishal Parikh is working as an Assistant Professor in CSE Department. His research interests include Machine Learning, and Multimedia Communication. He has many publications to his credit. Prof. Parikh has conducted various workshops and lecture series at the department. He has also been invited for expert talk at various institutions. He also works as a reviewer for indexed and peer-reviewed journals.

Yash Chelani Mr Yash Chelani has completed his B.Tech from Computer Science and Engineering Department from Institute of Technology, Nirma University. Currently he is Market analyst trainee at Future first. His research interests include Machine Learning, and Deep Learning.

Parita Oza Dr Parita Oza is working as an Assistant Professor in CSE Department. Her research area includes Image Processing and Medical Imaging. She has several publications in international journals and conferences. She also works as a reviewer for indexed and peer-reviewed journals. She is also a member of the program committees and session chair for international conferences.