Multimedia Tools and Applications (2023) 82:7553–7566
https://doi.org/10.1007/s11042-022-13718-x

COVID-19: Social distancing monitoring using faster-RCNN and YOLOv3 algorithms

Umang Ahuja[1] · Sunil Singh[1] · Munish Kumar[2] · Krishan Kumar[1] · Monika Sachdeva[3]

Received: 11 February 2021 / Revised: 21 May 2022 / Accepted: 24 August 2022 /
Published online: 27 August 2022

Abstract

As of March 31, 2021, the Coronavirus COVID-19 was affecting 219 countries and territories worldwide, with approximately 129,574,017 confirmed cases and 2,830,220 death cases. Social isolation is the most reliable way to deal with this pandemic situation. Motivated by this notion, this paper proposes a deep learning-based technique for automating the task of monitoring social distancing using surveillance cameras. To separate humans from the background, the proposed system employs object detection models based on F-RCNN (Faster Region-based Convolutional Neural Networks) and YOLO (You Only Look Once) algorithms. In the COVID-19 environment, these models track the percentage of people who violate social distancing norms on a daily basis. The authors compared the performance of both models in experimental work using the MS COCO dataset. Many tests were carried out, and we discovered that YOLOv3 demonstrated efficient performance with balanced FPS (frames per second).

✉ Munish Kumar
munishcse@gmail.com

Umang Ahuja
umangahuja1203@gmail.com

Sunil Singh
sunil32123singh@gmail.com

Krishan Kumar
k.salujauiet@gmail.com

Monika Sachdeva
monasach1975@gmail.com

[1] Department of Information Technology, University Institute of Engineering and Technology, Panjab University, Chandigarh, India

[2] Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

[3] Department of Computer Science and Engineering, I. K. G. Punjab Technical University, Kapurthala, Punjab, India

Keywords COVID-19 · Deep learning · Social distancing · Object detection · YOLO · F-RCNN

1 Introduction

Coronaviruses are a group of infections that can cause disease in humans and animals. Some coronaviruses cause respiratory problems in humans, such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) [23]. COVID-19 is a compelling disease caused by the most recently discovered coronavirus. This virus is thought to have originated in December 2019 in Wuhan, China. COVID-19 is currently affecting a large number of people from various countries. COVID-19 can be contracted from other people who are infected with it. The virus spreads from person to person via small dots from the nose or mouth that are expelled when a person with COVID-19 coughs, sneezes, or talks. Social distancing, also known as physical distancing [8, 10, 22], is a powerful technique for preventing the spread of infectious diseases by maintaining physical separation among individuals and reducing the number of times individuals come into close contact with one another [12, 16, 21] and avoiding crowding. Disease transmission can be controlled by limiting physical contact between uninfected people and contaminated people. Hand washing and maintaining respiratory cleanliness will benefit the social distance being maintained. Also, during the COVID-19 pandemic, the World Health Organization (WHO) proposed using the term "physical distancing" rather than "social distancing," because it is physical distancing that prevents transmission; individuals can remain socially connected through technology [3]. Figure 1 depicts the study after social distance norms were implemented in the COVID-19 environment.

An organisation called Landing AI has revealed the design of an AI device that has the screen social distancing at the workplace in order to prevent the spread of COVID-19 disease. After analysing real-time video streams from the camera, it can determine whether a safe physical distance between people has been maintained. This tool can be integrated into existing security cameras at various workplaces to ensure a safe distance between all workers. In order to monitor social distancing, the demo was presented in three steps: calibration, measurement, and detection. Landing AI as Cool Vendors has been identified in AI Core Technologies [2]. As a result of this motivation, the current work compares the performance of popular tracking schemes and object detection to monitor social distancing in the COVID-19

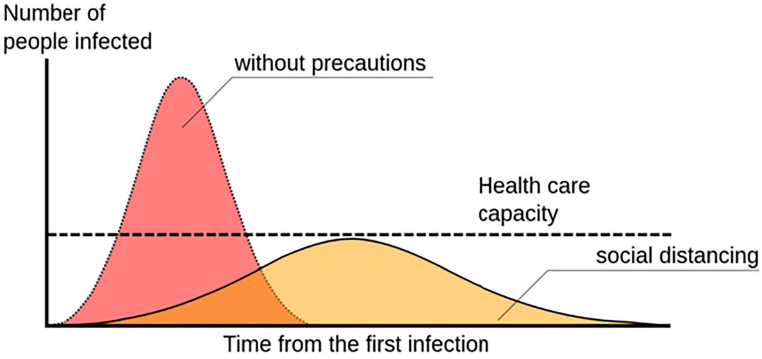


Fig. 1 The study after implementing social distance [9]

environment. The study aimed to propose a solution for detecting social distancing among people in any public place, which would aid in dealing with the coronavirus pandemic.

2 Change in travel behaviour

Because of social removal, travel requests may decrease as a result of increased telecommuting, e-learning, and fewer open activities and events. People may be dynamically arranged to practise at home with family or allies. This may result in less vehicle traffic – and less congestion during peak hours – as well as lower open vehicle ridership. Individuals may also be more inclined to return home delivery of online purchases (e.g., food, clothing), resulting in fewer shopping trips. Obviously, social isolation can influence travel mode decisions as well. Individuals should avoid open vehicles because they can be a breeding ground for infections and places where it is difficult to avoid contact with other travellers. Those who have no other option than to travel in open vehicles may attempt to avoid crowded transports by travelling during off-peak hours. Obviously, this could be a problem if open vehicle administrators decide to reduce the limit or frequency due to low ridership. Individuals who have access to a vehicle may be influenced to drive more because the vehicle "protects" them from other explorers. Given the reduced travel demand, a higher proportion of vehicle use is unlikely to result in more kilometres travelled by vehicle. In fact, less driving and lower levels of blockage can be considered normal. An increase in the use of taxicabs and ride-hailing services, particularly among those who frequently use open vehicles, may also be expected. Furthermore, strolling and cycling – if there is an occurrence of short tours – increase because social contact can (for the most part) be effectively avoided during dynamic travel. People may walk and cycle more recreationally as a result of the decrease in out-of-home exercise [5]. Out-of-home exercises may cease to exist or, in the event of reduced public vehicle services, may be restricted to those without a vehicle. This division results in lower levels of social association and personal growth, as well as increasing degrees of obesity, weakness, and debilitation. Social distancing measures have a clear direct productive effect on wellbeing because they are implemented to prevent people from becoming infected with COVID-19. However, because people frequently get physical movement from participating in certain out-of-home activities (e.g., wellness, sports, work), social distancing may result in a significant decrease in physical activity. Furthermore, less traffic may reduce air pollution, lowering the chances of respiratory illnesses, asthma, lung damage, and hypertension, as well as possibly preventing a dangerous atmospheric devotion [25].

3 Related work

As social distancing measures were initially implemented in China and then later implemented globally to control COVID-19, they are a trusted technique for stopping the spread of infectious disease. Many researchers have begun to work on this pandemic, such as Prem et al. [17], who sought to investigate the effects of social distancing measures on the spread of the COVID-19 epidemic. They used susceptible-exposed-infected-removed (SEIR) models to track the outbreak's ongoing trajectory using synthetic location-specific contact patterns [17]. Nicholas et al. [20] present a multidisciplinary approach to addressing the spread of the COVID-19. SIRNET combines epidemic modelling, physical science, and machine learning.

The benefit of epidemic modelling is that it forces our system to generate important factors from a physical standpoint, which includes a natural understanding of how the model is forecasting and provides a method to deal with conquering constrained or missing certifiable information tests. Social distancing is an excellent technique for flattening the coronavirus curve, which is increasing exponentially.

Adolph et al. [1] highlighted the state of the United States of America, where it could not be adopted in the initial stage due to a lack of consensus among all policymakers, resulting in harm to public health. Although social distancing has reduced economic productivity, many researchers are working hard to compensate. Many countries are implementing technology-driven solutions in response to the COVID-19 pandemic, and developing countries such as India are using GPS to track the movements of suspicious or infected individuals in order to monitor any possible contact with healthy people. People in India are using the Arogya Setu App, which uses GPS and Bluetooth to detect the presence of COVID-19 patients in the surrounding area. It also assists others in keeping a safe distance from the infected person [24] and assists other technology-driven products that are manually monitored, such as surveillance cameras, in detecting crowds and taking regulatory actions to disperse them. Object detection problems have been productively addressed by newly developed advanced methodologies. Convolutional neural networks (CNN), region-based CNN, and others have emerged in the last decade [27], and faster region-based CNN [13] used region proposal techniques to generate the object score prior to classification and then made the bounding boxes around the object of interest for visualisation. Because all of these CNN-based approaches use classification, these methods are efficient but have long training time requirements. Another approach YOLO considers is a regression-based method to separate bounding boxes from tardiness and explain their quadratic probabilities. Based on the preceding work, Eshel et al. [6] concentrated on crowd detection and person counting by proposing multiple height homographs to head top detection and solving the occlusions problem associated with video surveillance-related applications. Following the detection of an object, several classification techniques can be used to identify humans based on texture, shape, or motion-based characteristics. Shape-based methods use information about the shape of moving areas, such as dots, boxes, and drops, to identify humans. Due to limitations in standard template-matching schemes, this method performs poorly [26]. Further identification of a person through video surveillance can be done by using face [11]. All of the preceding work clearly shows that the application of detecting humans can be modified to serve the current situation, such as checking social distancing among humans and other hygiene-related activities.

Deep learning and object detection are also among the fastest-growing areas of medical image analysis, with a significant impact on a variety of clinical and research applications [4]. New results are paving the way for a bright future for artificial intelligence, paving the way for more accurate segmentation, classifications, detections, and predictions even at the level of expert radiologists. Deep learning techniques outperform traditional methods in medical image analysis.

Medical imaging limitations can also be quite extensive, depending on the specific problem. Nonetheless, while there are numerous limitations in DL, some common factors that limit model performance can be identified. First, it is clear that the networks that generalised well without overfitting were trained with a large amount of data, indicating that the amount of data is still an important factor in network training. However, while there is currently a large amount of medical data and organisations in charge of collecting databases, no data covers all of the heterogeneity of acquisition protocols or variation between study subjects.

3.1 Object detection using YOLOv3

Image classification is the process of labelling an image with a class label. Object localization is the process of drawing a bounding box around an object in an image. Object detection is a more difficult task that combines these two tasks by drawing a bounding box around each object of interest in the image and assigning it a class label.

Let's look at how simple binary or multi-classification algorithms can be modified to draw the object's bounding boxes.

We simply change the output labels from the previous algorithm to teach our model the class of the object as well as its position in the image. In the output layer, we add four more numbers, including the object's centroid position and the proportion of width and height of the bounding box in the image. This solves the object localization problem. We now want our algorithm to be able to classify and localise all of the objects in an image, rather than just one. So the plan is to divide the image into multiple images and run CNN on each of the cropped images.

First, create a window that is much smaller than the actual image size. Crop it and send it to CNN, who will make the predictions. Continue to slide the window and feed the cropped images into CNN. After cropping all of the image's portions with this window size, repeat all of the steps for a slightly larger window size. Pass cropped images through CNN once more and let it make predictions. Finally, you will have a set of cropped regions with some object, as well as the object's class and bounding box. Object detection with sliding windows is the name given to this solution. This concept is used by the YOLOv3 algorithm for object detection.

YOLOv3 employs successive 3 × 3 and 1 × 1 convolutional layers, as well as some shortcut connections. It has a total of 5 × 3 convolutional layers.

3.2 Object detection using faster-RCNN

The Faster R-CNN architecture consists of the RPN as a region proposal algorithm and the Fast R-CNN as a detector network.

Faster R-CNN attempts to identify potential object areas by combining similar pixels and textures into several rectangular boxes. The simple R-CNN employed 2000 search-selective proposed areas (rectangular boxes). However, fast R-CNN was a significant improvement because, rather than applying CNN to the proposed areas 2000 times, it only passes the original image to a pre-trained CNN model once. Faster R-CNN makes further progress than Fast R-CNN. Search selective process is replaced by Region Proposal Network (RPN).

To obtain the feature map, the input image is first passed through the backbone CNN (Feature size: 60, 40, and 512). Aside from test time efficiency, another important reason for using an RPN as a proposal generator is the benefits of weight sharing between the RPN backbone and the Fast R-CNN detector backbone. Following that, the RPN bounding box proposals are used to pool features from the backbone feature map. The ROI pooling layer is in charge of this. The ROI pooling layer, in essence, works by a) Taking the region corresponding to a proposal from the backbone feature map; b) Dividing this region into a fixed number of sub-windows; c) Performing max-pooling over these sub-windows to give a fixed size output. To understand the details of the ROI pooling layer and its advantages, read Fast R-CNN. The output from the ROI pooling layer has a size of (N, 7, 7, 512) where N is the number of proposals from the region proposal algorithm. The features are fed into the sibling classification and regression branches after passing through two fully connected layers. These classification and detection branches differ from the RPN's. The classification layer in this case has C units for each of the detection task's classes (including a catch-all background class). The features are passed through a SoftMax layer to get the classification scores — the probability of a proposal belonging to each class. The regression layer coefficients are used to improve the predicted bounding boxes.

4 Proposed work

We had many options for the object detection algorithm to use in our model (YOLO, F-RCNN, SPP-net, SSD, Retina-net, and so on). Some of these detectors were one-stage, while others were two-stage. To be as objective as possible, we decided to test our model on two algorithms, the first a one-stage detector and the second a two-stage detector, and to gain some insights into these algorithms by testing them on our social-distancing analysis project. For the one-stage detector, we chose YOLOv3 (though more powerful concept YOLOv4 was also introduced last month but keeping in mind that it would take lots of resources and computation power to train a new model from scratch, we decided to stick to YOLOv3 which has been trained for about ~350 epochs on COCO dataset [14] with weights available publicly). For a two-stage detector, we chose Faster-RCNN. We used Facebook's object detection API-Detectron2 Model Zoo for F-RCNN implementation. We used the C4 baseline (ResNet conv4 backbone with conv5 head) which is also the original baseline in the Faster R-CNN paper. It comes pre-trained with the 3 × schedule (~37 COCO epochs) [18].

4.1 YOLOv3

4.1.1 Model description [18]

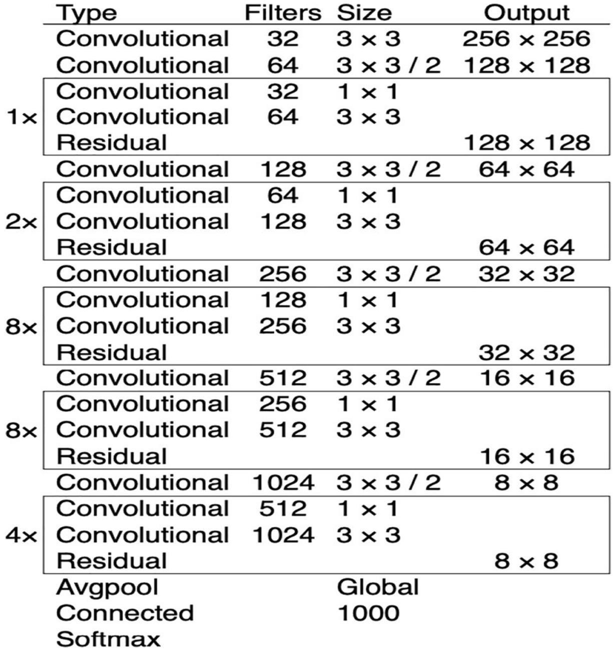
YOLOv3, proposed by Redmon et al. [18], employs a more effective network for feature extraction than its predecessors. Because this network is a hybrid of the network used in YOLOv2 and the residual network, it contains some shortcut connections. As shown in Fig. 2, it has 53 convolutional layers.

Unlike YOLOv2 which in the last layer predicts the output, whereas if we talk about YOLOv3 predicts boxes at 3 different scales as depicted in Fig. 3.

The YOLOv3 network predicts 4 coordinates for each bounding box, t_x, t_y, t_w, t_h. If the cell is balanced from the upper left corner of the image by (c_x, c_y) and the bounding box prior has width and height p_w, p_h then the forecasts compare to:

$$\begin{gathered} b_x = \sigma(t_x) + c_x \\ b_y = \sigma\left(t_y\right) + c_y \\ b_w = p_{w^{e^{t_w}}} \\ b_h = p_{h^{e^{t_h}}} \end{gathered}$$

YOLOv3 also uses logistic regression to predict an item score for each bounding box. If the bounding box already covers more ground truth objects than another bounding box, it should be 1. For example, the first bounding box (which has the most IOU) covers the primary ground truth object more than any other bounding box, and the first two bounding boxes cover the secondary ground truth object more than any other bounding box. For each ground truth



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 2 Darknet-53

object, the framework first assigns a bounding box. If a bounding box has not previously been assigned to a ground truth object, it has no effect on the organised or class prediction. If one looks at the architecture of YOLOv3, one can see that it has 53 convolutional layers that were trained on ImageNet. 53 more layers are stacked onto it for detection, giving us a 106 layer fully convolutional underlying architecture for YOLOv3. In YOLOv3, detection is accomplished by applying detection kernels to feature maps of three different sizes at three different locations throughout the network.

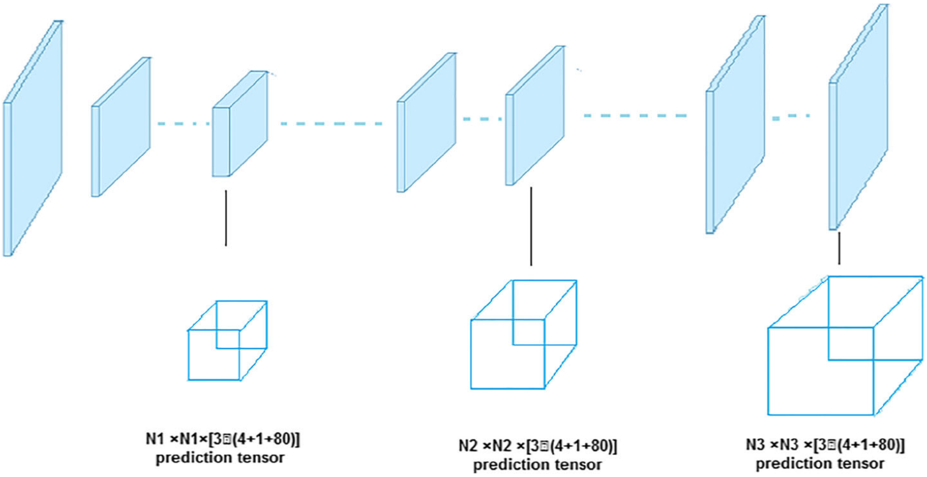


Fig. 3 YOLOv3 model

The shape of the detection kernel is 1 × 1× (B×(5 + C)). Here B is the number of bounding boxes and a cell on the feature map can predict, "5" is for the 4 bounding box attributes and one object confidence, and C is the number of classes. In YOLO v3, B = 3 and C = 80, so the kernel size is 1 × 1 × 255. The first detection is made by the 82nd layer. For the first 81 layers, the image is down-sampled by the network, such that the 81st layer has a stride of 32. If we have an image of 416 × 416, the resultant feature map would be of size 13 × 13. One detection is made here using the 1 × 1 detection kernel, giving us a detection feature map of 13 × 13 × 255. Then, the feature map from layer 79 is subjected to a few convolutional layers before being up sampled by 2 × to dimensions of 26 × 26. This feature map is then depth concatenated with the feature map from layer 61. Then the combined feature map is again subjected to a few 1 × 1 convolutional layer to fuse the features from the earlier layer (61). Then, the second detection is made by the 94th layer, yielding a detection feature map of 26 × 26 × 255. A similar procedure is followed again, where the feature map from layer 91 is subjected to few convolutional layers before being depth concatenated with a feature map from layer 36. Like before, a few 1 × 1 convolutional layers follow to fuse the information from the previous layer (36). We make the final of the 3 at the 106th layer, yielding a feature map of size 52 × 52 × 255.

4.1.2 Dataset used in YOLO algorithm

The original YOLOv3 was built, trained, and evaluated in the darknet framework (an open-source neural network framework written in C and CUDA), and OpenCV supports using darknet models directly without any explicit model conversions. As a result, we used the OpenCV framework to perform social distancing analysis using the pre-trained YOLOv3 model (trained on the COCO dataset). We chose the COCO dataset because it has already been trained on person classification (Fig. 4).

4.2 Faster R-CNN

4.2.1 Model description [19]

Ren et al. [19] proposed a faster R-CNN that consists of two modules. As shown in Fig. 5, the main module is a deep convolutional network that proposes regions, and the other module is the Fast R-CNN finder, which uses the proposed regions. The model is appropriate for object detection.

Faster R-CNN has two networks: a region proposal network (RPN) for generating region proposals and a network for detecting objects using these proposals. The main difference between Fast R-CNN and this method is that the latter employs selective search to generate region proposals. When RPN shares the most computation with the object detection network,

Fig. 4 Sample classes of COCO dataset

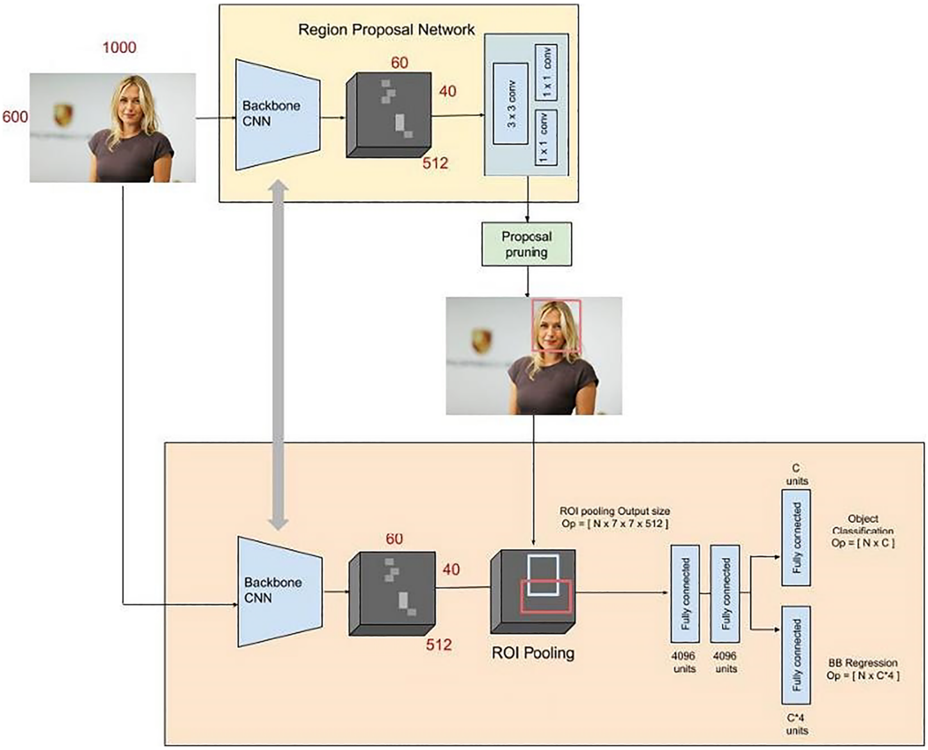


Fig. 5 Faster-RCNN Architecture

the time cost of generating region proposals is much lower than in selective search. RPN ranks region boxes (called anchors) and recommends the ones that are most likely to contain objects. The architecture is as follows [19].

Region proposal network (RPN) This region proposal network takes the backbone layer's convolution feature map as input and outputs the anchors generated by sliding window convolution applied to the input feature map.

Anchors Anchors play an important role in Faster R-CNN. An anchor is a box. In the default configuration of Faster R-CNN, there are 9 anchors at a position of an image. The following graph shows 9 anchors at the position (320, 320) of an image with size (600, 800) [19]. At each sliding-window area, simultaneously multiple regional proposals are predicted, where the quantity of greatest potential recommendations for each area is meant as k (Fig. 6). In this way, the red layer has 4 k yields encoding the coordinates of k boxes, and the cls layer yields a 2 k score that evaluates the likelihood of the object or not object to each proposal.

4.2.2 Dataset used in faster R-CNN

Detectron2 Model Zoo, a Facebook object detection API, provides many pertained models with cutting-edge object detection algorithms. As a result, we used the F-RCNN C4 model

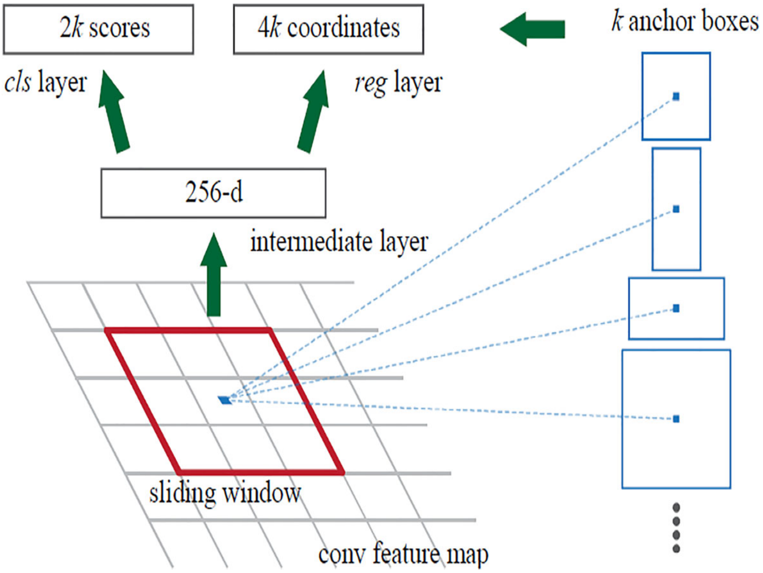


Fig. 6 Sliding-window location for multiple region proposals

provided by it to conduct social distancing analysis. The model uses a ResNet conv4 backbone with a conv5 head (The original baseline in the Faster R-CNN paper) and was pre-trained on ~37 *COCO train2017* epochs.

5 Experiments and results

Both YOLOv3 and Faster RCNN have some similarities. Both use an anchor box-based network structure and bounding box regression. YOLOv3 differs from Faster RCNN in that it performs classification and bounding box regression at the same time. For 0.5 IOU, the accuracy of both the models was very good and approximately the same but for IOU $>$ 0.5, if we take the average of results for IOU values from 0.5 to 0.95, F-RCNN performed better. However, comparing both of these algorithms solely on accuracy would be unfair because for real-time applications, speed is critical, and YOLOv3 outperforms any other state-of-the-art object detection algorithms in terms of FPS (Fig. 7).

5.1 Implementation using YOLOv3

After running our YOLOv3 model on each frame of the input video, we filtered out persons from all objects detected in the frame. We can easily calculate the distance between any two people once we've calculated the bounding box for each of them. A good approach would be to use Camera Calibration, which involves calculating the distance between two objects in an image in order to obtain the top view or bird's eye view of the video. To keep things simple, we calculated the Euclidean distance between the mid-points of the bottom edge of each bounding box, which provides a rough sense of separation between two people. We then set the proximity distance threshold (minimum separation between 2 persons according to social distancing norms). We found people within proximity distance in each frame, changed the

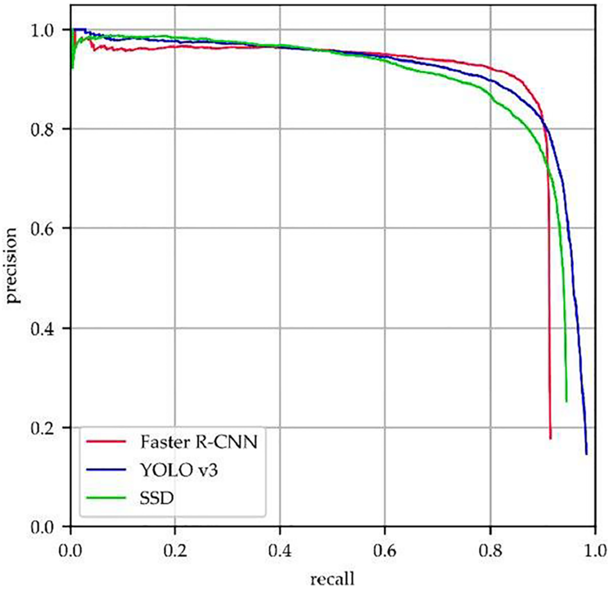


Fig. 7 Performance comparison between YOLOv3 and faster R-CNN with precision and recall as parameters [7]

colour of their bounding box to red, and recorded the proportion of people who violated social distancing. We converted the frames back to video after identifying the closest people in each frame. This gives us an idea of how many people violate social distancing rules on a daily basis.

5.2 Implementation using faster R-CNN

To avoid bias in the comparison, we implemented the F-RCNN model in the same way we implemented YOLOv3. After passing the input video through our F-RCNN model, we filtered out persons from all objects detected in the frame. Knowing each and every person's bounding box. The distance between them is easily calculated. The Euclidean distance between the mid-points of each bounding box's bottom edge was then computed. Then we set the proximity

Table 1 Results from the Focal Loss pape [15]

Method	mAP	time
[B] SSD321	28.0	61
[C] DSSD321	28.0	85
[D] R-FCN	29.9	85
[E] SSD513	31.2	125
[F] DSSD513	33.2	156
[G] FPN FRCN	36.2	172
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	37.8	198
YOLOv3–320	28.2	22
YOLOv3–416	31.0	29
YOLOv3–608	33.0	51

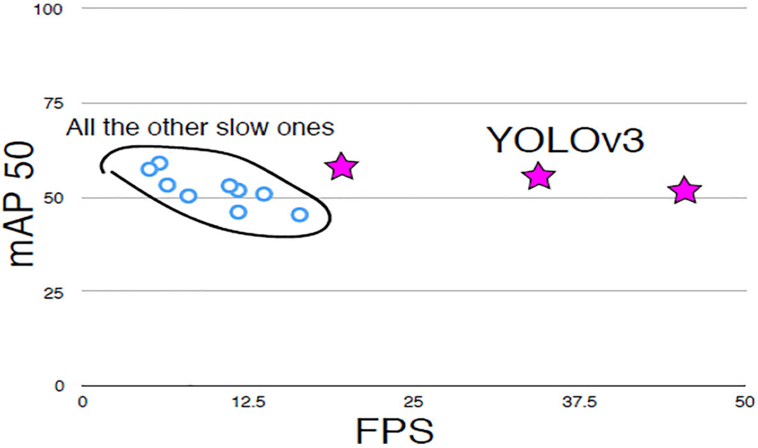


Fig. 8 Zero-axis chart [18]

distance threshold. We found people within proximity distance in each frame, changed the colour of their bounding box to red, and recorded the proportion of people who violated social distancing. We converted the frames back to video after identifying the closest people in each frame. The graph depicts a comparison of the various detection metrics of all state-of-the-art object detection algorithms [15]. For better comparison results, here are the results from the Focal Loss pape [15] (Table 1).

YOLOv3 is significantly faster than other detection methods, including F-RCNN FPN (an upgrade of the F-RCNN C4 model that we used). With comparable performance, the FPN architecture employs Feature Pyramid Network). As shown in Fig. 8, the results of a zero-axis chart look like this [18]:

Yolo is quick and precise. In the context of the.5 and.95 IOU metric, it is not as exceptional on the COCO average AP. However, it is frequently incredible on the area metric of 0.5 IOU [18]. YOLO is significantly faster (45 frames per second) than other object recognition algorithms. YOLO, or You Only Look Once, is a completely different object detection algorithm than region-based algorithms. A single convolutional network in YOLO can tell you the bounding boxes and class probabilities for these cases (Fig. 9).

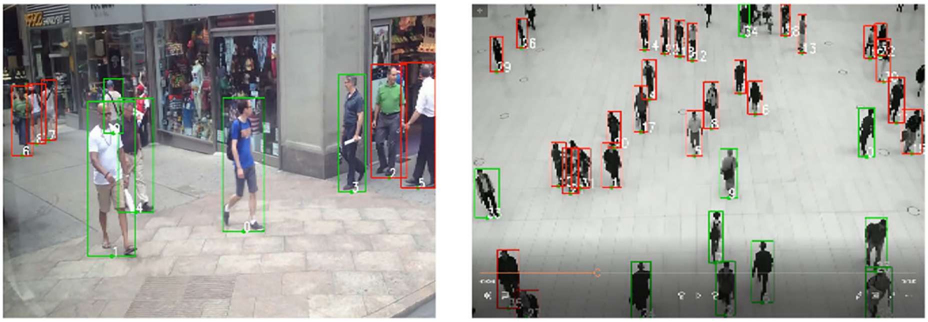


Fig. 9 Data samples showing people in bounded boxes of red colour (people violating social distancing norms) and of green colour (people following social distancing norms)

6 Conclusion

For the time being, social distancing is the only way to slow the spread of Coronavirus, and it is critical to monitor whether or not people are adhering to social distancing norms. This paper proposes a well-organized real-time deep learning technique for automating social distancing through object detection and tracking. We can build social distancing analysis tools for surveillance cameras using any of the proposed models (while keeping in mind the trade-off between speed and accuracy). Many tests were carried out, and we discovered that YOLOv3 demonstrated efficient performance with balanced FPS (frames per second).

7 Future scope

The authors presented an application that can measure the distance between pedestrians in public places and tell us how many people are breaking social distancing rules. This technique is intended to be used in any setting, whether working or not. To serve the purpose, accuracy and precision are highly desired in this application. Smart cities make extensive use of data and communication technologies to improve urban operations. Future Smart City development will place a premium on information authority and administration. Our analysis of the current situation provides an opportunity to better prepare for the next crisis or to assess the effects of large-scale social change.

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Adolph C, Amano K, Bang-Jensen B, Fullman N, Wilkerson J (2020) “Pandemic politics: Timing state-level social distancing responses to covid-19,” medRxiv
2. Ai L (2020) “Landing AI Named an April 2020 Cool Vendor in the Gartner Cool Vendors in AI Core Technologies,” https://www.prnewswire.com/news-releases/, [Online; accessed April 22, 2020]
3. Alto P (2020) “Landing AI Named an April 2020 Cool Vendor in the Gartner Cool Vendors in AI Core Technologies,” https://www.yahoo.com/lifestyle/landing-ai-named-april-2020-152100532.html, [Online; accessed April 21, 2020
4. Anaya-Isaza A, Mera-Jiménez L, Zequera-Diaz M (2021) An overview of deep learning in medical imaging, informatics in medicine unlocked. Volume 26:100723, ISSN 2352-9148. https://doi.org/10.1016/j.imu.2021.100723
5. De Vos J (2020) The effect of COVID-19 and subsequent social distancing on travel behavior. Transp Res Interdiscip Perspect 5:100121. https://doi.org/10.1016/j.trip.2020.100121
6. Eshel R, Moses Y (2008) “Homography based multiple camera detection and tracking of people in a dense crowd,” *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1–8
7. Han L, Tao P, Martin R (2019) Livestock detection in aerial images using a fully convolutional network. Computat Vis Media 5:221–228. https://doi.org/10.1007/s41095-019-0132-5
8. Harris M, Ghebreyesus A, Tedros L, Tu R, Michael “Mike” J, Vadia VK, Maria D, Diego F, Imogen O, Charles G, Corinne C (2020) "COVID-19". World Health Organization. Archived (PDF) from the original on 2020-03-25. Retrieved 2020-03-29
9. Henderson A (2020) What does social distancing and flattening the curve means” March 18, 2020
10. Hensley L (2020) "Social distancing is out, physical distancing is in – here's how to do it". Global News. Corus Entertainment Inc. Archived from the original on 2020-03-27. Retrieved 2020-03-29

11. Huang P, Hilton A, Starck J (2010) Shape similarity for 3d video sequences of people. Int J Comput Vis 89(2–3):362–381
12. Johnson CY, Sun L, Freedman A (2020) "Social distancing could buy U.S. valuable time against coronavirus". The Washington Post. Archived from the original on 2020-03-27. Retrieved 2020-03-11
13. Krizhevsky A, Sutskever I, Hinton GE (2012) "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097–1105
14. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollar P, Zitnick CL (2014) "Microsoft coco: Common objects in context," European conference on computer vision. Springer, pp. 740–755
15. Lin T-Y, Goyal P, Girshick R, He K, Dollar P (2017) Focal loss for dense object detection. arXiv preprint arXiv: 1708.02002
16. Pearce K (2020) "What is social distancing and how can it slow the spread of COVID-19?" The Hub. Johns Hopkins University. Archived from the original on 2020-03-29. Retrieved 2020-03-29
17. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CA, Munday JD et al. (2020) "The effect of control strategies to reduce social mixing on outcomes of the covid19 epidemic in Wuhan, China: a modelling study," The Lancet Public Health
18. Redmon J, Divvala S, Girshick R, Farhadi A (2016) "You only look once: Unified, real-time object detection," Proceedings of the IEEE 10 Conference on Computer Vision and Pattern Recognition, pp. 779–788
19. Ren S, He K, Girshick R, Sun J (2015) "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems, pp. 91–99
20. Soures N, Chambers D, Carmichael Z, Daram A, Shah DP, Clark K, Potter L, Kudithipudi D. SIRNET: Understanding Social Distancing Measures with Hybrid Neural Network Model for COVID-19 Infectious Spread
21. Tangermann V (2020) [20 March 2020]. "It's Officially Time to Stop Using The Phrase 'Social Distancing'". Science alert (Futurism / The Byte). Archived from the original on 2020-03-29. Retrieved 2020-03-29
22. Venske R (2020). Schwyzer, Andrea (ed.). "Die Wirkung von Sprache in Krisenzeiten" [The effect of language in times of crisis] (Interview). NDR Kultur (in German). Norddeutscher Rundfunk. Archived from the original on 2020-03-27. Retrieved 2020-03-27. (NB. Regula Venske is president of the PEN Centre Germany.)
23. W. H. Organization, "WHO corona-viruses (COVID-19)," https://www.who.int/emergencies/diseases/novel-corona-virus-2019, 2020, [Online; accessed May 02, 2020].
24. Website of Indian Government (2020) "Distribution of the novel coronavirus-infected pneumoni Aarogya Setu Mobile App," https://www.mygov.in/aarogya-setu-app/
25. WHO (World Health Organization) (n.d.) Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease (Geneva, Switzerland)
26. Wu B, Nevatia R (2007) Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. Int J Comput Vis 75(2):247–266
27. Zhao Z-Q, Zheng P, Xu S-t, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30(11):3212–3232

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.