International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 02 | Feb 2023 www.irjet.net p-ISSN: 2395-0072

A Survey on Person Detection for Social Distancing and Safety Violation Alert based on Segmented ROI

DEVASURITHI S[1], THULASIMANI K[2], MUTHULAKSHMI K[3]

[1,3]P. G. Student, Department of Computer science and Engineering, GCE, Tirunelveli, TamilNadu, India
[2]Professor, Department of Computer science and Engineering, GCE, Tirunelveli, TamilNadu, India

---***---

Abstract – Distancing ourselves from others protects everybody - particularly the more vulnerable in society. It is important to be clear that it is close and extended personal contact that increases our risk of transmission. Coming into close proximity to someone incidentally is unlikely to lead to transmission. Physical distancing helps limit the spread of COVID-19 – means keep a distance of at least 1m from each other and avoid spending time in crowded places or in groups. By avoiding physical contact, the risk of Covid-19 transmission can be decreased. One of the most efficient ways to stop the spread is through social distancing (SD), which calls for people to keep a certain distance from one another. The distance between the people found in the recorded video will be calculated, and the results will be compared to the values of fixed pixels. The segmented tracking area's core points and the overlapping boundary between individuals are measured for distance. Alerts or cautions can be sent out if it is determined that the space between persons is unsafe. The system's ability to detect the presence of people in restricted areas, which may also be used to trigger warnings, is another important function in addition to social distance measurement. In order to prevent physical contact between individuals, this study suggests a methodology for monitoring social distances using surveillance is based on deep learning and performance measures are evaluated by deep learning object detection method with the safety violation warning feature based on segmented ROI was shown to have higher accuracy.

Key Words: *Covid-19, Person Detection; Social Distancing, Restricted Area, Segmented ROI*

1. INTRODUCTION

Social distancing aims to decrease or interrupt transmission of COVID-19 in a population by minimizing contact between potentially infected individuals and healthy individuals, or between population groups with high rates of transmission and population groups with no or low levels of transmission. The WHO states that tiny droplets from the mouth and nose are how the corona-virus is transferred from one person to another. Try to put it another way, social distance is the best method for avoiding physical contact with suspected corona-virus carriers by maintaining a distance of at least one meter. One of the creative applications of integrated technologies that have recently experienced remarkable success and growth is object tracking and detection. Object detection is the process of identifying and categorizing objects that appear in live streams, video frames, and photographs in order to tell us of their nature and location.

Gaussian curves show a slight increase in the efficiency of the healthcare system, which enables patients to easily avoid contracting the virus by paying attention to the authorities' resolute recommendations. Any unanticipated sharpened spike and rapid rise in the infection rate will result in a breakdown of the healthcare system and an increase in the number of mortality. Fig.1 highlight how crucial it is to adhere to the recommendations for using social distance in order to reduce the spread of the virus among people.

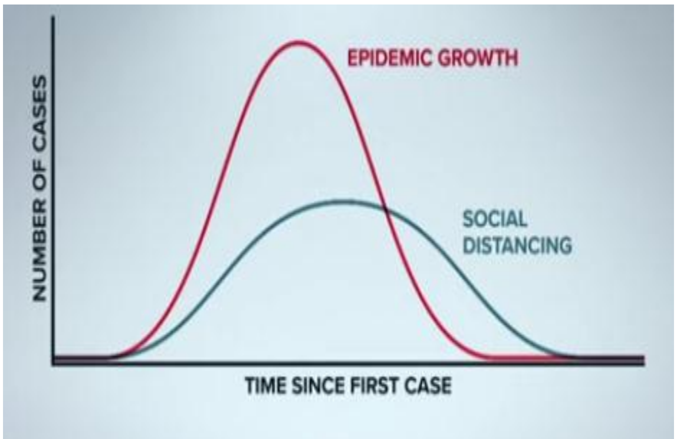

Fig. 1 Gaussian curve that illustrates the distribution virus spread rate among the individuals, with and without applying the social distancing

1.2 RELATED WORKS

Several measures are available for human detection and social distance in a crowd environment. Some articles are researched for the implementation of work. The goal of Prem et al. [1] was to investigate the impact of social distancing techniques on the transmission of the COVID-19 outbreak. Susceptible – Exposed – Infected -Removed (SEIR) models to replicate the ongoing course of the outbreak using synthetic location. Specific contact patterns are also hypothesized in early stage. Abrupt lowering of social separation might result in an earlier secondary peak, which could be leveled by gradually relaxing interventions [15].

Aslani et al. [2] suggested a spatio - temporal filter - based methodology for motion identification in which the motion parameters are discovered by analyzing three-dimensional (3D) spatio-temporal aspects of the person in motion in the picture stream. These approaches are favorable because of the simplicity and lower computer cost, but the performance is restricted due to noise and uncertainty on moving patterns.

Using a single picture acquired from a visible light camera at night, Jong Hyun Kim et al. investigated a CNN - based technique for person identification in a range of situations[5]. For evaluation procedure, the detection accuracy of a CNN with and without pre-processing, resulted in the accuracy being greater when utilizing a combined database with pre-processing.

Adolph et al. [4] emphasized that it could not be implemented at an early stage owing to a lack of common consent among all officials, resulting in ongoing harm to public health. Although social alienation has reduced economic output, many researchers are working hard to compensate. All examined literature and associated study work clearly provides a picture that the use of human detection may easily be extended to many applications to accommodate the current circumstances, such as checking mandated standards for hygiene, social distance, and work practices.

2. WORKFLOW

Controlling the spread of COVID-19 requires social distance detection among the collecton of people that is in crowd environment. The suggested approach uses any of the algorithms to detect social distance between persons. There are several stages, including video capture, video frame conversion, preprocessing, human recognition, distance computation, and SMS notifications. The algorithm is feed video obtained from CCTV. The resultant would be recognising persons in the frame, determining social distance among the people, and delivering SMS to the specific in-charge.

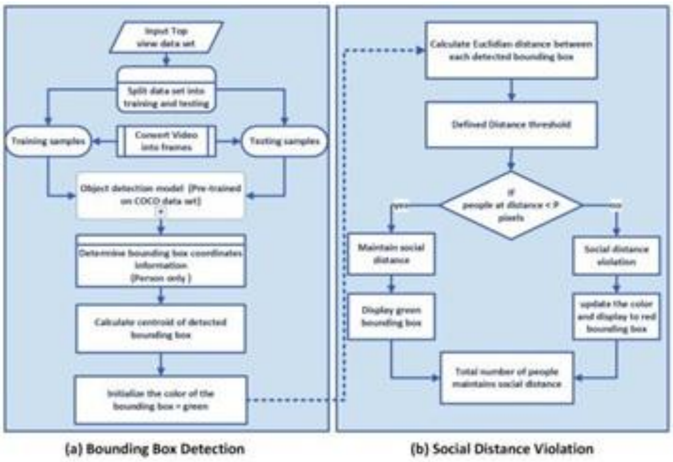

Fig.2 Workflow for person detection for social distancing and safety violation alert based on segmented ROI flowchart.

3. METHODOLOGY

Deep learning technique is an object detection strategy that reduces computer complexity by simulating tasks for predicting objects in the images. Convolutional Neural Networks (CNNs) are a form of neural network that are good at capturing patterns in multidimensional fields. CNN model is a popular for object detection algorithm in deep learning frameworks. This algorithm takes an input image and assigns biases and learnable weights to distinct classes in the image, allowing them to be distinguished from one another. CNN models of various varieties are used in many applications for object detection.

3.1 Faster Regional-CNN (Faster R-CNN)

Faster RCNN is based on an exterior region proposal technique based on Selective Search (SS) and is evolved from its predecessors. Many researchers discovered that, instead of employing the Selective Search, the benefits of convolution layers are advised for better and quicker object localization. Quicker R-CNN is a region proposal network that employs CNN models to create region proposals that are 10 times faster than RCNN. The RPN module conducts binary classification of an item as either an object or not an object, whereas the classification module provides categories to each detected object by pooling Region Of Interest (ROI) on extracted feature maps with projected areas.

The faster RCNN is a mixture of two modules, RPN and fast RCNN detector. The total multitask loss function is made up of classification loss and bounding box regression loss, both of which have L_{cls} and L_{reg} functions specified. t^u is the predicted corrections of the bounding box .t^u = {$t^u x$; $t^u y$; $t^u w$; $t^u h$ }. Here u is a true class label, (x, y) corresponds to the top-left coordinates of the bounding box with height h and width w, v is a ground-truth bounding box, p_i^* is the predicted class and p_i is the actual class.

3.2 Single Shot Detector (SSD)

A Single Shot Detector (SSD) is another object identification approach used in real time video surveillance systems to detect individuals. R-CNN works faster on region suggestions to generate boundary boxes to designate objects, resulting in higher accuracy but slower Frame Rate Processing (FRP/second). SSD enhances accuracy and FRP/second even more by combining multi-scale features and default boxes in a single operation. It employs the feed-forward convolution network approach, which creates bounding boxes of specified sizes as well as a score based on the existence of object class instances in those boxes, followed by an NMS step to generate the final detections.

3.3 YOLO V3

YOLO is another SSD contender for object detection. YOLO predicts the kind and position of an object item based on a single look at the image and treats the object detection issue as a regression task rather than a classification one in order to assign class probabilities to the anchor boxes. A single convolutional network predicts several bounding boxes and class probabilities at the same time. YOLO has three major versions: v_1, v_2, and v_3. Instead of utilizing softmax, YOLO v_3 does multi-label classification using logistic classifiers. YOLO v_3 makes three predictions for each spatial position in a picture at different sizes, removing the problem of not being able to identify small items effectively. Objectless, boundary box regresses, and classification scores are computed for each prediction.

3.3 YOLO V4

YOLO-v_4 has been investigated as a fast and accurate object detector. This model significantly improves on prior iterations. YOLO-v_4 extracts the impact of a cutting-edge Bag of Freebies (BoF) and numerous Bags of Specials (BoS). It raises the expense of training and dramatically improves object detecting accuracy. In terms of both accuracy and speed, YOLO-v_4 is rated the fastest and most accurate model. YOLO-v_4 is a simplified version of the YOLO-v_4 architecture. This model is simple to build and performs well in terms of object detection and also this approach can reduce the computational complexity on assumptions while assuring the neural network model's correctness.

4. RESULT

Object identification models are fine-tuned for binary classification with a labeled called human or non-human using the Nvidia GTX 1060 GPU with Inception v_2 as a backbone network, utilizing multiple datasets obtained from Google Open Source Community's Open Image Dataset (OID) repository. The accuracy of this system is measured to assess its effectiveness and the performance values for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) for social distance monitoring are counted for determining accuracy. The formula for determining precision is provided in:

$$Accuracy \quad \frac{(TP)+TN)}{(TP+TN+FP+FN)} \tag{1}$$

Based on the observations, the object identification model can detect the presence of a human if the camera used to capture the video is situated close to the item or in a controlled interior setting and circumstance. As a result, the social distancing system for the outside environment, particularly for recordings that capture distant landscapes, has to be upgraded. The outcomes of each model acquired at the end of the training phase, including the Training Time (TT), Number of Iterations (NoI), mAP, and Total Loss (TL).

Table -1: Performance comparison of the object detection model

Model	TT(in sec)	NoI	mAP	TL	FPS
Faster RCNN	9651	12135	0.969	0.02	3
SSD	2124	1200	0.691	0.22	10
YOLO V_3	5659	7560	0.846	0.87	23
YOLO V_4	4367	8230	0.824	0.92	30

Faster R-CNN model achieved minimal loss with maximum mAP, however, has the lowest FPS, which makes it not suitable for real-time applications. Furthermore, as compared to SSD, YOLO V_3, YOLO V_4 achieved better results with balanced mAP, training time, and FPS score.

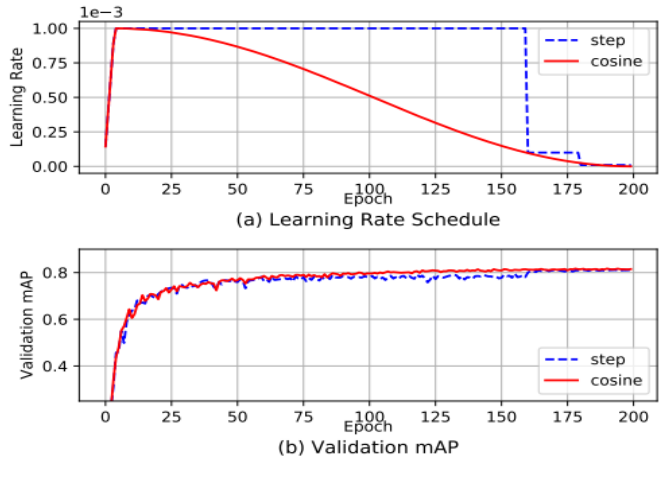

Fig.2: Losses / iteration of object detection model

Table -2: Accuracy of the object detection model

Video	TP	TN	FP	FN	Accuracy
Self Taken	2	2	0	0	100%
Town Center	11	19	14	4	62.5%
PETS 2009	14	38	19	5	68%
VIRAT_S	9	4	0	10	56,5%

During the training phase, the models' performance is continually checked using the mAP, as well as the localization, classification, and total loss in person detection, as shown in Fig.2. Sample output for object detection model is shown in Fig.3.

The processed frame with the recognised persons restricted in the bounding boxes while simultaneously simulating the statistical analysis showing the total number of social groups represented by same color encoding and a violation index term generated as the ratio of the number of people to the number of groups. The frames in Fig. 3 exhibit violation indexes of 3, 2, 2, and 2, 3, 3. Frames containing detected breaches are time stamped and saved for further examination.

Fig -3: Sample output for Object detection model

5. CONCLUSIONS

This system was built with Python and the OpenCV package to implement two suggested functionalities. The first function detects violations of social separation, while the second detects violations of accessing prohibited locations. Both features have been accuracy checked. The produced bounding boxes help in finding clusters or groups of persons that fulfill the proximity property determined using the pair-wise vectorized technique. The number of violations is validated by calculating the number of groups established and the violation index term, which is calculated as the ratio of persons to groups. Extensive testing were carried out using popular cutting-edge object detection models: RCNN, Faster RCNN, SSD, and YOLO-v_3, YOLO-v_4, But in YOLO-v_4 demonstrating efficient performance with balanced FPS and mAP score. This technique is very sensitive to the camera's spatial placement and it may be fine-tuned to better adapt with the matching field of view. Based on the overall findings, this study appears to have met all of its objectives.

REFERENCES

[1] 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 modeling study," The Lancet Public Health, 2020.

[2] S. Aslani and H. Mahdavi-Nasab, "Optical flow based moving object detection and tracking for traffic surveillance," International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, vol. 7, no. 9, pp. 1252–1256, 2013.

[3] S. A. Niyogi and E. H. Adelson, "Analyzing gait with spatiotemporal surfaces," in Proceedings of 1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects. IEEE, 1994, pp. 64–69.

[4] C. Adolph, K. Amano, B. Bang-Jensen, N. Fullman, and J. Wilkerson,"Pandemic politics: Timing state-level social distancing responses to covid-19," medRxiv, 2020.

[5] F. Ahmed, N. Zviedrite, and A. Uzicanin, "Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review," BMC public health, vol. 18, no. 1, p. 518, 2018.

[6] J. Harvey, Adam. LaPlace. (2019) Megapixels.cc: Origins, ethics,and privacy implications of publicly available face recognition imagedatasets. [Online]. Available: https://megapixels.cc/

[7] A. Agarwal, S. Gupta, and D. K. Singh, "Review of optical flow technique for moving object detection," in 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I). IEEE, 2016, pp. 409–413.

[8] World Health Organization, "Coronavirus Disease 2019,"Coronavirus disease (COVID-19) pandemic, 2020.https://www.who.int/emergencies/diseases/novel-coronavirus-2019(accessed Jun. 19, 2020).

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

[10] 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. 779788.

[11] M. Putra, Z. Yussof, K. Lim, and S. Salim, "Convolutional neural network for person and car detection using yolo framework," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 10, no. 1-7, pp. 67–71, 2018.

[12] R. Eshel and Y. Moses, "Homography based multiple camera detection and tracking of people in a dense crowd," in 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008, pp. 1–8.

[13] D.-Y. Chen, C.-W. Su, Y.-C. Zeng, S.-W. Sun, W.-R. Lai, and H.Y. M. Liao, "An online people counting system for electronic advertising machines," in 2009 IEEE International Conference on Multimedia and Expo. IEEE, 2009, pp. 1262–1265.

[14] C.-W. Su, H.-Y. M. Liao, and H.-R. Tyan, "A vision-based people counting approach based on the symmetry measure," in 2009 IEEE International Symposium on Circuits and Systems. IEEE, 2009, pp. 2617–2620.

[15] J. Yao and J.-M. Odobez, "Fast human detection from joint appearance and foreground feature subset covariances," Computer Vision and Image Understanding, vol. 115, no. 10, pp. 1414–1426, 2011.

[16] B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors," International Journal of Computer Vision, vol. 75, no. 2, pp. 247–266, 2007.

[17] F. Z. Eishita, A. Rahman, S. A. Azad, and A. Rahman, "Occlusion handling in object detection," in Multidisciplinary Computational Intelligence Techniques: Applications in Business, Engineering, and Medicine. GI Global, 2012, pp. 61–74.

[18] M. Singh, A. Basu, and M. K. Mandal, "Human activity recognition based on silhouette directionality," IEEE transactions on circuits and systems for video technology, vol. 18, no. 9, pp. 1280–1292, 2008.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1. IEEE, 2005, pp-886–893.

[20] P. Huang, A. Hilton, and J. Starck, "Shape similarity for 3d video sequences of people," International Journal of Computer Vision, vol. 89, no. 2-3, pp. 362–381, 2010.

[21] [51] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," Pattern recognition, vol. 25, no. 1, pp. 65–77, 1992.

[22] D. Cunado, M. S. Nixon, and J. N. Carter, "Using gait as a biometric, via phase-weighted magnitude spectra," in International Conference on Audio-and Video-Based Biometric Person Authentication. Springer,1997, pp. 93–102.

[23] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1. IEEE, 2005, pp. 878–885.