International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

System for Detecting Social Distance During COVID-19 using YOLOv3 and OpenCV

Barnabas Zakariya

Computer science and Engineering GH Raisoni College of Engineering, Nagpur, India Barnazaka@gmail.com

DOI: https://doi.org/10.55248/gengpi.2022.3.10.63

ABSTRACT

A virus called COVID-19 spreads between people in close proximity via minute droplets created through talking, sneezing, coughing, and most commonly by inhalation. Many people have died as a result of the pandemic's severe respiratory infection, which is still present today. You can reduce your risk of contracting COVID-19 by avoiding physical contact with others. This study suggests a real-time AI framework for people detection, monitoring social distance violations, and categorising people's social distances based on live video feeds. In this study, YOLOv3 was suggested for object detection. Its straightforward neural network architecture makes it appropriate for embedded devices that are reasonably priced. Comparing the suggested model to other real-time detection methods, it is a better choice. Additionally, with the aid of OpenCV, an open-source toolkit for computer vision, machine learning, and image processing. The major purpose of the image processing feature is to enhance the image quality so that the AI detection system would accurately recognise human movement. Computer vision is used to analyse photos and videos.The final iteration of the prototype algorithm has been put to use in low-cost CCTV Cameras made up of fixed cameras that are placed in any public area where large crowds used to congregate. The suggested method is appropriate for a surveillance system in sustainable smart cities for people detection, social distance classification, and tracking social dis- tance violations. This will make it easier for the government to understand how people who are socially isolated are doing.

Index Terms—COVID-19, AI, Machine Learning, YOLOv3

INTRODUCTION

A novel coronavirus infection is the cause of the acute respiratory infectious illness COVID-19 [1]. The major symp- toms are a fever, a dry cough, exhaustion, etc. Nasal con- gestion, runny nose, diarrhoea, and other symptoms of the upper respiratory tract and digestive system are present in a small percentage of individuals. After a week, severe pa- tients frequently experience respiratory problems, and they quickly advance to irreversible metabolic acidosis, coagula- tion malfunction, and multiple organ failure. More than 6.58 million people have died as a result of COVID-19 up to this point in several nations throughout the world. To stop the virus from spreading, numerous areas have now implemented policies including limiting traffic, Wearing face mask, using hand sanitizer frequently and cancelling significant events. The next stage is to figure out how to prevent the virus from spreading as much as possible in a regular setting. By giving consistent information from health care officials, the health system makes it simple for patients to prevent the infection. Any unexpected sharper rise and quick increase in the infection rate will result in a failure of health care services and, as a result, an increase in the number of deaths. The aim of adhering to social distancing recommendations is to limit the transmission of the virus among persons [2, 3]. Although certain vaccinations [4] have been created to combat the virus's transmission, the most effective method is to keep a safe social distance between pedestrians. Staying away from large crowds and preserving a 6 foot gap from each individual—roughly the length of a body—is what social distance entails. Isolation and quarantine are not the same as social distance. The government uses social distance as a preventative strategy for everyone. Those who have been affected or are suspected of being afflicted with infectious illnesses must be isolated in a ward and cared for by specially trained medical professionals. Persons who have been exposed to infectious people but have not yet got the disease are quarantined. This means that good pedestrian detection and distance measuring technologies can aid in the control of COVID-19 transmission. In a public setting, the most often used pedestrian detection approach is based on a computer vision solution [5]. Pedestrian recognition and social distance assessment may be accomplished easily and affordably using current public area security cameras. In comparison to systems relying on mobile devices such as GPS sensors, computer vision-based pedestrian detection approaches offer a broader variety of applications, including intelligent-assisted driving [6, 7], intelligent monitoring [8, 9], pedestrian analysis [10], and intelligent robot [11]. Furthermore, various open-source pedestrian identification data sets based on computer vision have been produced to aid in the evaluation and improvement of detection algorithms, such as the INRIA person dataset [10], the Caltech pedestrian detection benchmark [11], and the ETH dataset [12].

Previously, background modelling methods [13] were frequently used to extract foreground moving targets, after which feature extraction in the target area was performed and classifiers (e.g., multi-layer perceptron, support vector machine, and random forest) were used to classify them to determine whether pedestrians are included. In actuality, it still encounters the following issues during the application process:

(1) Lighting variations can readily produce large changes in picture grey levels, lowering detection accuracy. (2) Camera shaking can easily cause backdrop modelling to fail, affecting target position computation. (3) There may be ghost zones that impact the model's assessment. The statistical learning approach [14] automatically mines characteristics from a large number of data and builds a pedestrian detection classifier. The retrieved characteristics primarily comprise the target's grayscale, edge, texture, colour, and gradient histogram. Statis- tical learning is also confronted with the following challenges:

(1) Variable pedestrian stance, apparel, scale, and lighting environment. (2) Classifiers often require a significant number of training examples. (3) The quality of the features has a di- rect impact on the classifier's ultimate detection performance. There have been some advances in the usage of multi-feature fusion and cascaded classifiers. Haar feature [15], HOG feature [16], LBP feature [17], and Edgelet feature [18] are examples of often used features. In this research, we compare and study the processes for detecting pedestrians and watching their social distance in order to increase the efficiency of epidemic prevention. This work's contributions include:

- A unique vision-based surveillance system formonitoring social distance violations in public spaces.
- I used a strong algorithm to recognise people and quantify the distance between them. In compared to the previous techniques, I propose speedier and more accurate out- comes.
- The suggested methodology is an accurate method of translating a camera frame recorded from a perspective point of view to a top-down view. This will keep the con- version rate between pixel distance and physical distance consistent.

The goal of this project is to provide an AI-based solution to reduce the transmission of coronavirus among individuals and its economic effect. We present YOLOv3[19], a revolutionary deep learning model, together with the con- struction of an algorithm for social distance and OpenCV which is a computer vision with machine learning for effective picture processing.

METHODOLOGY

A. *The architecture of social distancing*

In this part, I will go through the actions that must be taken in order to create a sequence design that will determine and verify whether or not social distancing norms are followed by individuals.

1. Streaming the video footage captured by the camera that shows people.
2. Frame-by-frame extraction of the camera's footage.
3. YOLOv3 architecture is used to identify just the people in the camera recordings.
4. For good and accurate image processing, use the OpenCv image processing tool to count the number of individuals in the camera recordings.
5. Determine the separation between the bounding boxes' centres, which are where the people in the videos are located.
6. Last but not least, the algorithm will decide whether or not the individuals are in a violation or safe environment based on the quantity of people in the videos and the measured sep- aration between the centroid of bounding boxes. I established two distinct levels for violation with two distinct threshold set points for the measured distance between the centre points of the bounding boxes, it should be noted. Risk is the violation level, and the bounding box is coloured red to indicate this. I coloured the bounding box green to indicate the safe state.

B. *Object detection*

A computer vision technique called object detection finds the items in an image or video. The initial step in this investigation is to determine the coordinates of the people in the footage. For people detection in the Camera footage, we used YOLOv3[19]. I created a 53-layer convolutional neural network (CNN) for YOLOv3. The purpose of this research is to develop a lightweight model that takes into account the real-time application needs of convolutional neural networks (CNNs) in low-cost embedded systems, such as IoT devices. YOLOv3 is made up of two major modules: the conventional model, which has a high recognition accuracy, and the tiny model, which has a slightly reduced recognition accuracy. For primary feature extraction, the mAP (accuracy) of the standard model YOLOv3-416, which is composed of Convolutional block (Conv) and Residual networks, is used (ResNet). The YOLO v3 network seeks to forecast each object's bounding boxes (area of interest of the candidate object) as well as the probability of the class to which the object belongs. To accomplish this, the model separates each input image into a SxS grid of cells, with each grid predicting B bounding boxes and C class probabilities of objects whose centres fall within the grid cells. According to the research, each bounding box may specialise in detecting a specific type of object. The number of anchors utilised is connected with the number of bounding boxes "B." Each bounding box includes 5+C attributes, where 5 refers to the five bounding box attributes (for example, centre coordinates (bx, by), height (bh), width (bw), and confidence score) and C is the number of classes. Our output from passing this image into a forward pass convolution network is a 3-D tensor because we are working on an SxS image. The output looks like [S, S, B*(5+C)].

1) Anchor Boxes: Previously, scientists employed the slid- ing window approach and ran an image classification algorithm on each window to detect an object. They quickly recognised that this made no sense and was inefficient, so they switched to ConvNets and ran the entire image in a single shot. Because the ConvNet generates square matrices of feature values (e.g., 13x13 or 26x26 in the case of YOLO), the concept of "grid" entered the picture. The square feature matrix is defined as a grid, however the main issue arose when the objects to detect were not square in shape. These things could be of any shape (mostly rectangular). Anchor boxes were so introduced. Anchor boxes are pre-defined boxes with a specified aspect ratio. Even before training, these aspect ratios are defined by executing a K-means clustering on the full dataset. These anchor boxes are connected to the grid cells and have the same centroid. YOLO v3 employs three anchor boxes for each detection scale, for a total of nine anchor boxes.

2) Non-Maximum Suppression: There is a potential that the output expected after the single forward pass would contain numerous bounding boxes for the same object because the centroid is the same, but we only need one bounding box that is best suited for all.

For this, we can employ a technique known as non-maxim suppression (NMS), which essentially cleans up after these detections. I may specify a particular threshold that will operate as a constraint for this NMS technique, causing it to disregard all other bounding boxes whose confidence is lower than the specified threshold, so removing a few. However, this would not exclude everything, thus the following stage in the NMS would be executed, which would be to arrange all of the bounding box confidences in decreasing order and select the one with the highest score as the most appropriate one for the item. Then we discover all the other boxes that have a high Intersection over union (IOU) with the bounding box and delete them as well.

C. *OpenCV*

OpenCV (Open Source Computer Library) was first lunched in 1999 by intel[19] OpenCV (Open Source Computer Vision Library) is a free and open source software library for com- puter vision and machine learning. OpenCV was created to offer a standard foundation for computer vision applications and to speed up the adoption of machine perception.The library contains over 2500 optimised algorithms, including a complete variety of both traditional and cutting-edge computer vision and machine learning techniques. These algorithms can be used to detect and recognise faces, identify objects, classify human actions in videos[20], track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken with flash, follow eye movements, recognise scenery, and establish markers to overlay. OpenCV has around 47 thousand users and an estimated 18 million downloads[21].

D. *Dataset training procedure*

The suggested method was trained on two distinct picture datasets. The first dataset contains 1000 photos. FLIR gathered these pictures for the cameras [22]. This dataset contains the first collection of photos captured by CCTV cameras equipped with infrared radiation sensors. Dataset II features 950 photos of various persons collected under realistic conditions during surveillance and monitoring. They are from various settings,

and include individuals creeping, strolling, jogging, and in var- ious body postures. These photos were gathered from various online sources. Both datasets' photos were classified for the class of just people in the photographs. For each dataset, the photos were divided into 70 percent for training, 20 percent for validating, and 10 percent for testing the architecture. Stochastic gradient descent (sdgm) was used to train YOLOv3 [23]. To regulate the model's response to mistake, the learning rate has been tuned in the training option. The learning rate was fine-tuned to 103, and the loose curve remained stable at this value for both datasets[24].

RESULTS AND DISCUSSION

All outcome details and comparisons are presented in this section. I depicted the outcome from several angles. I ran the algorithm over the testing photos from both datasets to evaluate the performance of the suggested technique. The photographs were created using true situations captured by various cameras in outdoor settings. We picked these datasets for our tests with this in mind. I also used YOLOv3 and the approach offered for measuring social distance with OpenCV on a huge scale of films. These movies are scalable in terms of screening persons' movements as cameras measured their distance to determine whether or not they broke the social distance law. In addition to my investigation, I conducted another experiment by studying (Fast R-CNN) and you only look once (YOLOv2) detectors for persons detection, both employing the identical images from the two training datasets of images. The purpose of this is to compare these designs to YOLOv3 and suggested approaches. Using the same testing photos from both datasets and the videos database. To evaluate the suggested approach for metric computation, confusion matrix criteria were utilised. The criteria selected to evaluate the algorithm's goodness are recall, accuracy, and precision. see Eq(1)

where TP denotes the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives.

Based on the findings of these studies, YOLOv3 produced encouraging results for people detection on pictures on both testing datasets and videos database; person detection points have been exhibited in OpenCV view window for both safe and risk circumstances with assigned colours, respec- tively. Furthermore, YOLOv3 outperformed other approaches in terms of accuracy[25,26,27].

A. Equations

$$\text{precision} = TP / TP + FP$$

$$\text{Accuracy } TP + TN / TP + FN + TN + FP \text{ Recall} = TP / TP + FN \quad (1)$$

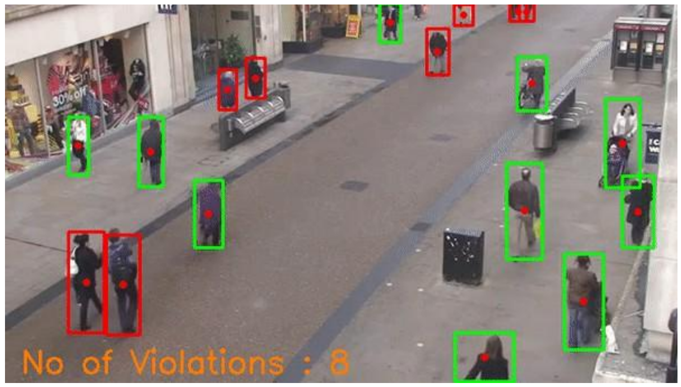

Fig. 1. Social distancing status with the proposed method, which show

8 persons violated the Risk threshold distance, and 10 persons were in safe conditions: a perspective transformation of human detection points with OpenCV view.

CONCLUSION

This study presented a deep learning-based social distance approach for people detection in movies or photos utilising OpenCV view. The obtained findings demonstrated that the designed intelligent surveillance system recognised persons who violated social distance using good picture processing. YOLOv3 performed well in terms of accuracy and precision. OpenCV and the CCTV Camera view technology have been developed to efficiently map human detection sites. The pro- posed technique is a way for the authorities to perceive pedes- trians who follow social distance norms in outdoor locations. I coloured the safe condition green for the bounding boxes, while the unsafe state is red, and the algorithm, YOLOv3, will identify and count how often individuals breached the social distancing.

REFERENCES

[1] The Visual and Data Journalism Team.: Coronavirus: a visual guide to the outbreak. 6 Mar. 2020

[2] Fong, M.W., Gao, H., Wong, J.Y., Xiao, J., Shiu, E.Y., Ryu, S.,Cowling, B.J.: Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—social distancing measures. Emerg. Infect. Dis. 26, 976 (2020)

[3] Ahmedi, F., Zviedrite, N., Uzicanin, A.: Effectiveness of workplace so- cial distancing measures in reducing influenza transmission: a systematic review. BMC Public Health 18, 518 (2018)

[4] Hotez, P.J.: COVID-19 and the antipoverty vaccines. Mol. Front. J. 4, 58–61 (2020)

[5] Mou, Q., Wei, L., Wang, C., et al.: Unsupervised domain-adaptive scene-specific pedestrian detection for static video surveillance. Pattern Recogn. 118(9), 108038 (2021)

[6] Liu, T., Du, S., Liang, C., et al.: A novel multi-sensor fusion based object detection and recognition algorithm for intelligent assisted driving. IEEE Access 9, 81564–81574 (2021)

[7] Zheng, Q., Zhao, P., Zhang, D., Wang, H.: MR-DCAE: Manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification. Int. J. Intell. Syst. (2021).

[8] Chen, Y., Ma, J., Wang, S.: Spatial regression analysis of pedestrian crashes based on point-of-interest data. J. Data Anal. Inf. Process. 08(1), 1–19 (2020)

[9] Zheng, Q., Yang, M., Tian, X., Jiang, N., Wang, D.: A full stage data augmentation method in deep convolutional neural network for natural image classification. Discrete Dyn. Nat. Soc. 2020, 1–11 (2020).

[10] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision Pattern Recognition (2005)

[11] Dollar, P., Wojek, C., Schiele, B., et al.: Pedestrian detection: an evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. 34(4), 743–761 (2011)

[12] Ess, A., Leibe, B., Schindler, K. et al.: Moving obstacle detection in highly dynamic scenes. In: IEEE Int. Conf. Robot. Autom. pp. 56–63 (2009)

[13] Rodriguez, P., Wohlberg, B.: Incremental principal component pursuit for video background modeling. J. Math. Imaging Vis. 55(1), 1–18 (2016)

[14] Zheng, J., Peng, J.: A novel pedestrian detection algorithm based on data fusion of face images. Int. J. Distrib. Sens. Netw. 15(5), 155014771984527 (2019)

[15] Park, K.Y., Hwang, S.Y.: An improved Haar-like feature for efficient object detection. Pattern Recogn. Lett. 42, 148–153 (2014)

[16] Sheng, Y., Liao, X., Borasy, U.K.: A pedestrian detection method based on the HOG-LBP feature and gentle AdaBoost. Int. J. Adv. Comput. Technol. 4(19), 553–560 (2012)

[17] Costa, Y., Oliveira, L.S., Koerich, A.L., et al.: Music genre classification using LBP textural features. Signal Process. 92(11), 2723–2737 (2012)

[18] Zhao, J.: Boundary extraction using supervised edgelet classification. Opt. Eng. 51(1), 7002 (2012)

[19] YOLOv3: An Incremental Improvement, Joseph Redmon, Ali Farhadi, Apr 2018 University of Washington

[20]I. Culjak, D. Abram, T. Pribanic, H. Dzapo and M. Cifrek, "A brief introduction to OpenCV," 2012 Proceedings of the 35th International Convention MIPRO, 2012, pp. 1725-1730.

[21] Saponara, S., Elhanashi, A., Gagliardi, A.: Implementing a real-time, AI-based, people detection and social distancing measuring system for Covid-19. J. Real-Time Image Proc. (2021).

[22] Mahamkali, Naveenkumar Ayyasamy, Vadivel. (2015). OpenCV for Computer Vision Applications.

[23] FLIR Thermal Dataset for Algorithm Training, FLIR Systems. [24]Glorot, X. et al.: Understanding the difficulty of training deep feed-forward neural networks. In: Int. Conf. on Artificial Intelligence and Statistics (2010)

[25] Sener, F., et al.: Two-person interaction recognition via spatial multiple instances embedding. J. Vis. Commun. Image Represent. 32, 63 (2015)

[26] Rinkal, K., et al.: Real-time social distancing detector using social distancingnet-19 deep learning network. SSRN Electron. J. 40, 6 (2020)

[27] Yadav, S.: Deep learning based safe social distancing and face mask detection in public areas for covid-19 safety guidelines adherence. Int. J. Res. Appl. Sci. Eng. Technol. 8, 1–10 (2020)