Automatic Social Distance Monitoring system using Deep Learning Algorithms

T. Shanthi *, R. Anand, K. Hareesh, M.S. Jagan, V. Baskar, and Thanneru Bhanu Prakash

Department of Electronics and Communication Engineering, Sona College of Technology, Salem- 636005, TamilNadu, India.

E-mail: *shanthi@sonatech.ac.in, anand18mgt@sonatech.ac.in

Abstract. In recent months, the worldwide outbreak of Corona Virus Disease (COVID-19) has created a panic situation for all human beings irrespective of the geographic location and economic status of the countries. The communal spread of the deadly virus can be prevented by maintaining the social distance between people and the wearing of face masks. This paper provides a prominent solution for monitoring using the Open-Source Computer Vision Library (Open-CV) and Deep Learning Algorithm to course people in public areas and to avoid crowding. This proposed project can be implemented using images taken from the Closed Circuit Television (CCTV). The distance between the people is detected with help of an object detection algorithm. The Euclidean distance is calculated and compared with the standard distance given. The persons with enough distance will be indicated in the green-colored bounding box and others with a red-colored bounding box. The total number of violations is also shown at the output feed. This project provides a solution to eradicate the spread of the deadly virus to an extent with minimum human effort and reduced risk of infection to the local authorities.

1. Introduction

The worldwide outbreak of the Corona Virus since December 2019 has created a panic situation among people with its high fatality rate. As there is no vaccine for this viral disease till now, this disease leads people to critical conditions and fatal sometimes. One of the precautionary methods is to maintain social distancing in the crowded area. World Health Organization (WHO) made a regulation to maintain distancing of 1 meter among each pair of count and the organization passed a strict law in accordance to this regulation. Since it is a communicable disease it spreads from one person to another person when having physical contact. Social distancing prevents people from having contact with each other.

In public areas, surveillance of people to follow social distance is an utmost priority since there is a lot of chance to spread disease in crowded areas by physical contact. To govern and organize people in this pandemic situation using CCTVs will be very helpful. With our proposed paper we can be able to pinpoint the person who is violating the law. Our method can be easily installed in those tracking devices and can compute the distance...
between those pairs of people. This is very much useful in large public gatherings like airports, railway stations and bus stands where people mostly gather in crowds. The authorities can be relieved with this proposed method to find the people who are violating instead of searching manually. This method is cost-efficient also because we are installing the program in existing CCTVs.

With the help of Open-CV and distance finding algorithm, the distance between the pair of people can be determined in pixel values and alert the authorities to act on the violated people. It is also helpful in tracking suspected patients who may have contact with the infected patient in-crowd.

![Image of Social Distancing](image.png)

Figure 1. Social Distancing

2. Literature Survey

In 2017, Abbas SS and co-authors of his project projected a technique for people tracking and crowd control victimization raspberry pi and Open-CV. the complete construct the plan was to documenting the scene by camera and Raspberry pi which incorporates a 4-cored CPU that analyze the video each in all frame. A model was trained to sight human heads in thronged areas. The counting of people is measured and therefore the gathering of people is managed by scrutiny the worth with the edge and if it surpasses the edge bar is done consequently [1]. In 2018, Joy JJ and co-authors of his project projected a technique for traffic density identification supported by the image processing technique. The array size and the traffic mass were documented from the pictures acquired from the traffic cameras. The input video stream was taken from cameras and logic was put in to organize the average truth conception. Results of the average truth conception may vary between fully true and fully false [2]. In 2019, Abbas SA and his co-authors proposed a method to correct any monocular image by compute a homographic matrix which transforms into a bird's eye (overhead) angle view [3].

In 2006, Yilmaz A and his co-authors reviewed tracing methods, classify them into various categories. Obstacles in tracing objects can come to abrupt object motion, appearance changing patterns of the scene and the object, non-rigid object structures, object-to-object and scene-to-object occlusions, and camera movement. They arranged the tracing methods based on the object and movement representations and provided with detailed descriptions of similar methods in each division and judge their pros and cons. They discussed the important issues related to tracking including the use of appropriate image features, selection of motion models, and detection of objects[4]. In 2020, Anand R and her co-authors projected a technique of Face Recognition and Classification by Using GoogleNET Architecture to discover pedestrians by means of their facial data and...
classifying them by their races. Face recognition is an imperative tool in Computer Vision technology. They created and trained the dataset in the GoogleNET of deep learning model with Caffe and framework Nvidia DIGITS [5].

3. Methodology And Algorithms
We projected our project supported the Open supply pc Vision (Open-CV) and YOLO algorithms to execute the functions in our project. The outline of every algorithm utilized in this project and their alternatives are explained below.

3.1 RetinaNET:
RetinaNet is a single-stage object detection model which utilizes a focal loss function to acquire class imbalance on training. The 1st subnet to do convolutional object-classification on the support network’s output; the 2nd subnet performs convolutional bounding boxed retrogression. The 2 subnet works feature a simple style that is projected specifically single-stage, intense detection. This the resulted image by RetinaNET.

![Figure 2. Result using RetinaNET](image)

3.2 Tiny YOLO:
Tiny YOLO object detection was intended to nurture the research in the detection of small objects. The main motivation of this detection is to detect the small objects in a given image with high accuracy and exact classification. This is the resulted image of Tiny YOLO and its accuracy.

![Figure 3. Result using TinyYOLO](image)

3.3 YOLO v3 Object Detection Algorithm:
YOLO (You Solely Look Once), a model of convolution neural network (CNN) for object detection task for given time period. The laws applies for one neural network to the total input image, then divides that image into regions and predicts bounding boxes and give odds for every region. These bounding boxes are unit weighted by the predictable chances.

The kernel shape is $1 \times 1 \times (B \times (5 + C))$. $B$ is number of bounding boxes in feature map which helps in prediction, 5 for the four bounding boxes and 1 detection for the target object and $C$ is for number of classes. YOLO v3 trained by COCO, the dimension of kernel is $1 \times 1 \times 255$. YOLO v3 predicts the input image by down sampling the dimensions by 32, 16 and 8.

The $1^{st}$ detection is made in $82^{nd}$ layer. The size of feature maps is resultant from the input image of size 416x416 pixels is 13x13 pixels. After the detection, the feature map is 13x13x255. The feature map from 79 is processed before 2x up sampling to dimensions of 26x26 pixels. The dimension feature map is changed to 26x26x255.

Again the $91^{th}$ layer feature map and $36^{th}$ layer feature map are fused and the final feature map is obtained in $100^{th}$ layer with the dimension of 52x52x255.

This is the accuracy output detected by YOLO v3 algorithm.
Table 1 – Comparison Table

|                               | YOLO v3            | RetinaNET         | TinyYOLO          |
|-------------------------------|--------------------|-------------------|-------------------|
| No. of Objects detected       | 11-persons, 1-cycle, 1-handbag | 10-persons, 1-cycle, 0-handbag | 10 persons, 0-cycle, 0-handbag |
| Accuracy of the detected object (Cumulative value) | 93.3482% as persons, 56.285% as cycle, 73.922% as handbag | 73.403% as persons, 53.757% as cycle | 73.375% as persons |

4. Blockdiagram

![Flow Chart of the Proposed Algorithm](image)

Figure 6. Flow Chart of the Proposed Algorithm

5. Results and Discussion:
The working of this detector is very simple, as it only needs input image or video or direct camera footage for its input. As showed in the above block diagram, initially the input images/frames are taken from the CCTV cameras, and object detection is used to detect pedestrians from the given input images using the YOLO object detector. In the next step, the bounding boxes are drawn around the detected humans with the centroids using a Python function. Then in the next step, the pair-wise distances between centroids will be computed and measured in pixel distance using the Euclidian distance method, and finally, these distances are compared with the threshold distance which should be maintained (6 feet or 2 meters). It is measured in equivalent pixels N in the frames and represented by red bounding boxes if violated otherwise green bounding box is drawn.

The frames are checked at the speed of 60 fps (frames per second), so the centroids of the pair of pedestrians are tracked in every frame and give the continued results as long the input is provided. If more people come together in a particular area, the detector analyses the distance between them and alerts the authorities or the police stations will be immediate if violations are notified. As we mentioned above the results will be represented in green and red bounding boxes around the people detected respectively to the given condition. The condition required for the detected pair of pedestrians is the distance between their centroid is greater than the N pixels then they are bounded in a green box and if violated they will be bounded by a red box. The accuracy of detecting human by this YOLO v3 is 93.3485%. With this, we can monitor the people who are violating social distancing and we can reduce the pressure on the local authorities and the police authorities assigned to those areas. This project gives us the total counts of violation that happened in the day as a cumulative figure and gives us a general idea to make necessary amendments for the people to follow social distancing. The resulting sample image is presented as figure 7 and 8.

Figure 7. Input

Figure 8. Output
6. Conclusion
As we have a tendency to picture the global post-COVID-19 pandemic conditions, the necessity for self-awareness and common sense appear irrefutably. Social Distancing would be the foremost vital precaution as COVID-19 spreads through contact with infected people. To supervise a bigger crowd, an efficient resolution is vital and this survey paper focuses on it victimization put in CCTV, authorities will track human activities and management the massive crowd and forestall violating the rules and regulation. CCTV tracks the crowd in an area, the red boxpop-up and the local authorities and therefore the police of that region are going to be notified and therefore the state of affairs will return in restraint directly. As maintaining giant crowds isn't a straightforward task, implementing this concept will scale back the on efforts of the police department and will entirely target direction conditions solely in those square measureas wherever conditions are unfavorable.

7. Future scope
This project can be implemented in the Raspberry Pi Kit for regular usage of common people. The video recorded from CCTV camera or live video stream, can be captured and detected by the program dumped in Raspberry Pi. The display can be added to the raspberry pi to show the violation count and detected people or it can be connected to monitors in surveillance rooms.

8. References

[1] Abbas SS, Anitha M, Jaini XV,” Realization of multiple human head detection and direction movement using raspberry pi. In2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) 2017 Mar 22 (pp. 1160-1164). IEEE.

[2] Joy JJ, Bhat M, Verma N, Jani M. Traffic Management Through Image Processing and Fuzzy Logic. In2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) 2018 Jun 14 (pp. 52-55). IEEE.

[3] Abbas SA, Zisserman A. A Geometric Approach to Obtain a Bird's Eye View From an Image. InICCV Workshops 2019 Oct 1 (pp. 4095-4104).

[4] Lukezic A, Vojir T, ’Cehovin Zajc L, Matas J, Kristan M. Discriminative correlation filter with channel and spatial reliability. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 6309-6318).

[5] Anand R, Shanthi T, Nithish MS, Lakshman S. Face recognition and classification using GoogleNET architecture. InSoft Computing for Problem Solving 2020 (pp. 261-269). Springer, Singapore.

[6] Shanthi T, Sabeenian RS, Anand R. Automatic diagnosis of skin diseases using convolution neural network. Microprocessors and Microsystems. 2020 Jul 1;76:103074.

[7] Anand R, Shanthi T, Sabeenian RS, Veni S. Real time noisy dataset implementation of optical character identification using CNN. International Journal of Intelligent Enterprise. 2020;7(1-3):67-80.

[8] Pranav JV, Anand R, Shanthi T, Manju K, Veni S, Nagarjun S. Detection and identification of COVID-19 based on chest medical image by using convolutional neural networks. International Journal of Intelligent Networks. 2020 Jan 1;1:112-8.