iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: VII Month of publication: July 2020

DOI: https://doi.org/10.22214/ijraset.2020.30560

Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence

Shashi Yadav
Department of Computer Science and Engineering, Goel Institute of Technology and Management, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India

Abstract: According to data obtained by the World Health Organization, the global pandemic of COVID-19 has severely impacted the world and has now infected more than eight million people worldwide. Wearing face masks and following safe social distancing are two of the enhanced safety protocols need to be followed in public places in order to prevent the spread of the virus. To create safe environment that contributes to public safety, we propose an efficient computer vision based approach focused on the real-time automated monitoring of people to detect both safe social distancing and face masks in public places by implementing the model on raspberry pi4 to monitor activity and detect violations through camera. After detection of breach, the raspberry pi4 sends alert signal to control center at state police headquarters and also give alarm to public. In this proposed system modern deep learning algorithm have been mixed with geometric techniques for building a robust modal which covers three aspects of detection, tracking, and validation. Thus, the proposed system favors the society by saving time and helps in lowering the spread of corona virus. It can be implemented effectively in current situation when lockdown is eased to inspect persons in public gatherings, shopping malls, etc. Automated inspection reduces manpower to inspect the public and also can be used in any place.
Keywords: Deep Learning, Computer Vision, Convolutional Neural Networks (CNNs), Single Shot Detector, Transfer Learning, public Safety, OpenCV, COVID-19.

I. INTRODUCTION

The spread of COVID-19[1] Pandemic Disease has created a most crucial global health crisis of the world that has had a deep impact on humanity and the way we perceive our world and our everyday lives. In December 2019 the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a new severe infectious respiratory disease emerged in Wuhan, China and has infected 7,711 people and 170 reported deaths in China before coronavirus was declared as a global pandemic, was named by the World Health Organization as COVID-19 (coronavirus disease 2019). According to the World Health Organization (WHO as of July 12, 2020) report , the current outbreak of COVID-19, has infected over 13,039,853 people and more than 571,659 deaths in more than 200 countries around the world, carrying a mortality of approximately 3·7%, compared with a mortality rate of less than 1% from influenza. A novel coronavirus has resulted in person-to-person transmission but as far as we know, the transmission of the novel coronavirus causing coronavirus disease 2019 (COVID-19) can also be from an asymptomatic carrier with no covid symptoms. Till now there is no report about any clinically approved antiviral medicine or vaccines that are effective against COVID-19. It has spread rapidly across the world, bringing massive health, economic, environmental and social challenges to the entire human population. At the moment, WHO recommends that people should wear face masks to avoid the risk of virus transmission and also recommends that a social distance of at least 2m [2] be maintained between individuals to prevent person-to-person spread of disease. Furthermore, many public service providers require customers to use the service only if they wear masks and follow safe social distancing. Therefore, face mask detection and safe social distance monitoring has become a crucial computer vision [3] task to help the global society. This paper describes approach to prevent the spread of the virus by monitoring in real time if person is following safe social distancing and wearing face masks in public places.

This paper adopts a combination of lightweight neural network MobileNetV2[4] and Single Shot Detector(SSD)[5] with transfer learning technique to achieve the balance of resource limitations and recognition accuracy so that it can be used on real-time video surveillance to monitor public places to detect if persons wearing face mask and maintaining safe social distancing. Our solution uses neural networking models to analyze Real-Time Streaming Protocol (RTSP) video streams using OpenCV and MobileNet V2. We mix the approach of modern-day deep learning and classic projective geometry techniques which not only helps to meet the

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

real-time requirements, but also keeps high prediction accuracy. If the person detected as not following the covid-19 safety guidelines, violation alerts will be send to the control center at state police headquarters for taking further action. It allows automating the solution and enforces the wearing of the mask and follows the guidelines of social distancing. This model was created to run on raspberry pi4 and the accuracy obtained was between 85% and 95%.

II. RELATED WORK

In recent years, object detection techniques using deep models [6] are potentially more capable than shallow models in handling complex tasks and they have achieved spectacular progress in computer vision. Deep models for person detection focus on feature learning [7] contextual information learning, and occlusion handling. Deep learning object detection models [8] can now mainly be divided into two families: (i) two-stage detectors such as R-CNN[9], Fast R-CNN[10] and Faster R-CNN[11] and their variants and (ii) one-stage detectors such as YOLO[12] and SSD. In two-stage detectors detection is performed in stages, in the first stage, computed proposals and classified in the second stage into object categories. However, some methods, such as YOLO, SSD MultiBox, consider detection as a regression issue and look at the image once for detection.

In proposed system we are using Single Shot Detector MultiBox(SSD) which seems to be a good choice for real-time object detection and the accuracy trade-off is also very little. SSD uses the VGG-16 model pre-trained on ImageNet as its basic model to extract useful image feature. At the top of VGG16, SSD adds several convolutional feature layers of decreasing sizes.

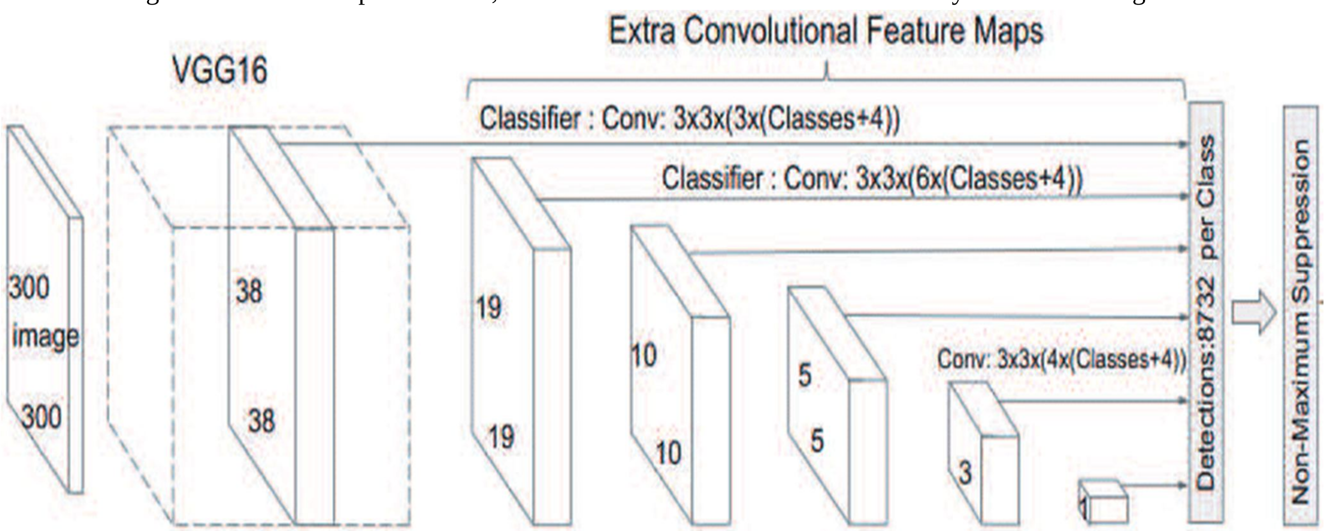


Fig 1: Model architecture of SSD.

The Viola – Jones[13] object detection system can be trained to detect any object, but is especially common for facial detection and is more accurate and faster. The Viola and Jones process is an example of supervised learning. Zhu[14] also shared another very widespread facial detection algorithm is a neural network-based detector.

It only works well with the front, upright face. Li et al. [15, 16], suggested another model for facial detection which was a Multi-View Face Detector with surf capabilities. Oro et al. [17] also proposed a haar-like feature based face detection algorithm for HD video on the GTX470 and obtained an improved speed of 2.5 times. However, they only used CUDA which is a GPU programming tool for NVIDIA GPUs.

Compared to OpenCL which is used in a number of computed components, it is unable to resolve the imbalanced workload issue experienced during the implementation of the viola-jones face detection algorithm in GPUs. Glass et al. (2006)[18] addressed the importance of social differencing and how the risk of pandemic growth can be slowly decreased by successfully preserving social distance without the use of vaccines or antiviral drugs.

The authors have carried out an exhaustive study on this in both rural and urban communities in order to demonstrate a reduction in the growth rate. Z., Luo[19] studies the identification of people with full-face or partial occlusion. This approach categorizes into way, people with hand over their faces or occluded with objects.

This approach is not suited to our scenario, which requires, in essentially, to detect faces that have their mouths covered with masks such as scarves, mufflers, handkerchiefs, etc.

III. METHODOLOGY

The proposed system helps to ensure the safety of the people at public places by automatically monitoring them whether they maintain a safe social distance, and also by detecting whether or not and individual wears face mask. This section briefly describes the solution architecture and how the proposed system will automatically functions in an automatic manner to prevent the coronavirus spread.

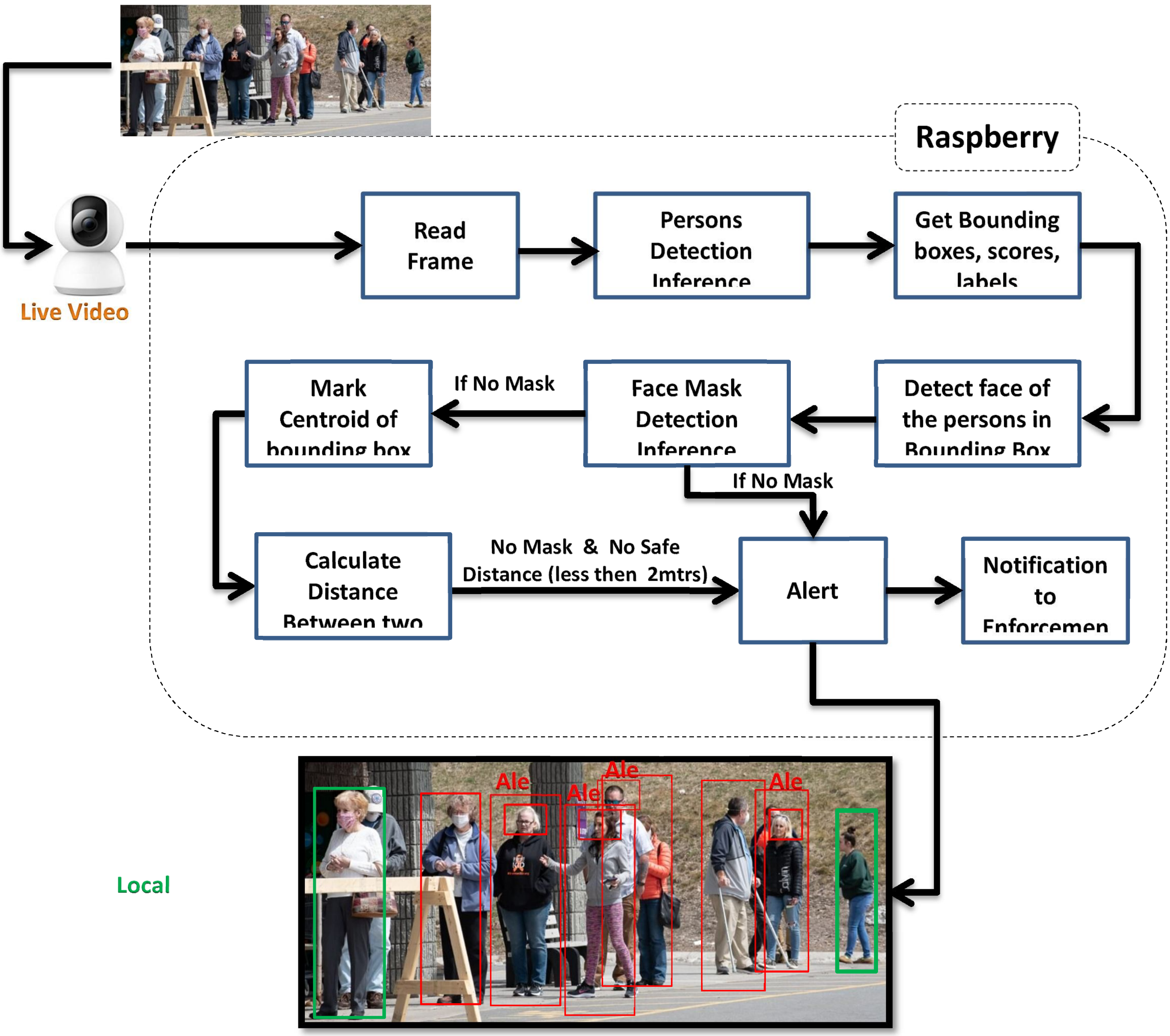


Fig 2: Solution architecture of proposed system

The proposed system uses a transfer learning approach to performance optimization with a deep learning algorithm and a computer vision to automatically monitor people in public places with a camera integrated with a raspberry pi4 and to detect people with mask or no mask. We also do fine tuning, which is another form of transfer learning, more powerful than just the feature extraction.

In this process camera video feeds from the Network Video Recorder (NVR) are streamed using RTSP and then these frames are converted to grayscale to improve speed and accuracy and are send to the model for further processing inside raspberry pi4. We have used the MobileNetV2 architecture as the core model for detection as MobileNetV2 provides a huge cost advantage compared to the normal 2D CNN model. The process also involves the SSD MultiBox Detector, a neural network architecture that has already been trained on a large collection of images such as ImageNet and PascalVOC for high quality image classification.

We are loading the MobileNet V2 with pre-trained ImageNet weights, leaving the network head off and constructing a new FC head, attaching it to the base instead of the old head, and freezing the base layers of the network. The weights of these base layers will not be changed during the fine tuning phase of the backpropagation, while the head layer weights will be adjusted. After data is prepared and the model architecture is set up for fine tuning, then the model is compiled and trained. A very small learning rate is used during the retraining of the architecture to ensure that the convolutional filters already learned do not deviate dramatically and experiments have been carried out with OpenCV, TensorFlow using Deep Learning and Computer Vision in order to inspect the safe social distance between detected persons and face masks detection in real-time video streams. The main contribution of the proposed system is three components: person detection, safe distance measurement between detected persons, face mask detection. Real-time person detection is done with the help of Single Shot object Detection (SSD) using MobileNet V2 and OpenCV, achieves 91.2% mAP, outperforming the comparable state-of-the-art Faster R-CNN model. A bounding box will be displayed around every person detected. Although SSD is capable of detecting multiple objects in a frame, it is limited to the detection of a single person in this system. To calculate the distance between two persons first the distance of person from camera is calculated using triangle similarity technique, we calculate perceived focal length of camera, we assumed person distance D from camera and person's actual height H=165cms and with SSD person detection pixel height P of the person is identified using the bounding box coordinates. Using these values, the focal length of the camera can be calculated using the formula below:

$F = (P \times D) / H$

Then we use the real person's height H, the person's pixel height P, and the camera's focal length F to measure the person's distance from the camera. The distance from the camera can be determined using the following:

$D1 = (H \times F) / P$

After calculating the depth of the person in the camera, we calculate the distance between two people in the video. A number of people can be detected in a video. Thus, the Euclidean distance is measured between the mid-point of the bounding boxes of all detected individuals. By doing this, we got x and y values, and these pixel values are converted into centimeters. We have the x, y and z (the person's distance from the camera) coordinates for each person in cms. The Euclidean distance between each person detected is calculated using (x, y, z) coordinates. If the distance between two people is less than 2 meters or 200 centimeters, a red bounding box is shown around them, indicating that they do not maintain a social distance.

In the proposed system transfer learning is used on top of the high performing pre-trained SSD model for face detection with mobileNet V2 architecture as backbone to create a lightweight model that is accurate and computationally efficient, making it easier to deploy the model to raspberry pi. We used custom face crop datasets of about 3165 images annotated in mask and no mask. Annotated images are used to train a deep learning binary classification model that classifies the input image into the mask and no mask categories using the output class confidence. The result of the SSD model extracts a person mask and displays a bounding box. The proposed system monitors public places continuously and when a person without a mask is detected his or her face is captured and an alert is sent to the authorities with face image and at the same time the distance between individuals is measured in real time, if more than 20 persons have been identified continuously breaching safe social distance standards at the threshold time, then alert is sent to the control center at the State Police Headquarters to take further action. This system can be used in real-time applications requiring a secure monitoring of social distance between people and the detection of face masks for safety purposes due to the outbreak of Covid-19. Deploying our model to edge devices for automatic monitoring of public places could reduce the burden of physical monitoring, which is why we choose to use this architecture. This system can be integrated with edge device for use in airports, railway stations, offices, schools and public places to ensure that public safety guidelines are followed.

IV. EXPERIMENTAL RESULTS

The proposed system is a deep learning solution that uses OpenCV and TensorFlow, to train the model. We combine the deep learning MobileNetV2 modal with the SSD framework for a fast and efficient deep learning solution for real-time human detection in video streams and use a triangular similarity technique to measure distance between persons detected by camera in real time in public places and comprises customized data collection to resolve a face mask detection model with variance in the types of face masks worn by the public in real time by means of a transfer of learning[20] to a pre-trained SSD face detector.

This model combine's social distance detection and face mask detection. In our approach, the Raspberry Pi 4 Model-B with the ARMv8 1.5 GHz processor and 4 GB of RAM is used as the preferred edge device.

In the proposed system, four steps are followed, such as:
1) Data collection and pre-processing
2) Model development and training
3) Model testing
4) Model implementation

A. Data Collection and Pre-processing

The proposed system uses a custom data set consisting of face images with different types of face masks which are labeled and used for the training of our models. We use the existing background subtraction [21,22] algorithm in a pre-processing step. The real-time automated detection of social distance maintenance and the verification of persons wearing masks or not are performed by the SSD algorithm. The dataset used to train our proposed face mask detector consists of 3165 images. Before the custom face mask image dataset is labelled, the data set is divided into the training data set and the testing data set. The Training data set should consist of 80% images to train the algorithm effectively and for prediction accuracy and the Testing data set should consist of 20% images to test the prediction accuracy of the algorithm. The images in the training data collection are classified into two categories: mask and no mask.

Fig 3: Sample Images of custom dataset

B. Model building and Training

Our proposed framework uses the transfer learning approach[20] and will fine-tune the MobileNetV2 model, which is a highly efficient architecture that can be applied to edge devices with limited computing power, such as raspberry pi4 to detect people in real time. We used 80% of our total custom data set to train our model with a single shot detector, which takes only one shot to detect multiple objects that are present in an image using multibox. The custom data set is loaded into the project directory and the algorithm is trained on the basis of the labeled images. In pre-processing steps, the image is resized to 224×224 pixels, converted to numpy array format and the corresponding labels are added to the images in the dataset before using our SSD model as input to build our custom model with MobileNetV2 as the backbone and train our model using the TensorFlow Object Detection API. Before model training begins, tensorflow helps in Data augmentation and download pre-trained ImageNet weights to make the algorithm's prediction efficiency accurate. After downloading the pre-trained weights and creating a new fully-connected (FC) head, the SSD algorithm is trained with both the pre-trained ImageNet weights and the annotated images in the custom data set by tuning the head layer weights without updating weights of base layers. We trained our model for 1000 steps using the Adam optimizing algorithm, the learning decay rate for updating network weights, and the binary cross-entropy for mask type classification. Parameters were initialized for the initial learning rate of INIT_LR = 1e-4, number of epoch EPOCHS = 20 and batch size BS = 32.

We used webcam for social distance monitoring using cv2 and after a person has been identified, we start with bounding box coordinates and computing the midpoint between the top-left and the bottom-left along with the top-right and bottom-right points. We measure the Euclidean distance between the points in order to determine the distance between the people in the frame.

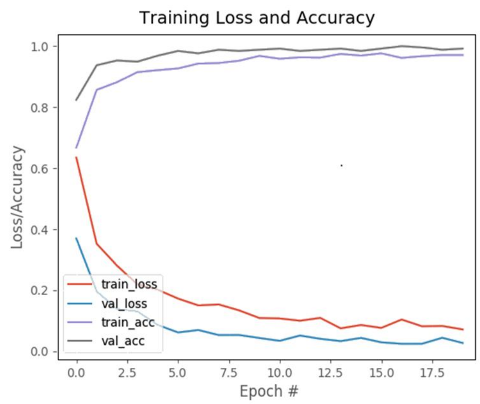


Fig 4: Model training accuracy/loss curves

C. Model Testing

The proposed system operates in an automated way and helps to automatically perform the social distance inspection process. Once the model is trained with the custom data set and the pre-trained weights given, we check the accuracy of the model on the test dataset by showing the bounding box with the name of the tag and the confidence score at the top of the box. The proposed model first detects all persons in the range of cameras and shows a green bounding box around each person who is far from each other after that model conducts a test on the identification of social distances maintained in a public place, if persons breaching social distance norms bounding box color changes to red for those persons and simultaneously face mask detection is achieved by showing bounding boxes on the identified person's face with mask or non-mask labeled and also confidence scores. If the mask is not visible in the faces, and if the social distance is not preserved, the system generates a warning and send alert to monitoring authorities with face image. The system detects the social distancing and masks with a precision score of 91.7% with confidence score 0.7, precision value 0.91 and the recall value 0.91 with FPS = 28.07.

Fig 5: Test result of model

D. Model Implementation

The proposed system uses raspberry pi4 with camera to automatically track public spaces in real-time to prevent the spread of Covid-19. The trained model with the custom data set is installed in the raspberry pi4, and the camera is attached to it. The camera feeds real-time videos of public places to the model in the raspberry pi4, which continuously and automatically monitors public places and detects whether people keep safe social distances and also checks whether or not those people wear masks.

Our solution operates in two stages: first, when a person identified without a mask his photo is taken and sent to a control center at the State Police Headquarters; and second, when the detection of a social distance violation by individuals is detected continuously in threshold time, there will be an alarm that instructs people to maintain social distance and a critical alert is sent to the control center of the State Police Headquarters for further action.

V. CONCLUSION

In this paper, we proposed an approach that uses computer vision and MobileNet V2 architecture to help maintain a secure environment and ensure individuals protection by automatically monitoring public places to avoid the spread of the COVID-19 virus and assist police by minimizing their physical surveillance work in containment zones and public areas where surveillance is required by means of camera feeds with raspberry pi4 in real-time.

Thus, this proposed system will operate in an efficient manner in the current situation when the lockout is eased and helps to track public places easily in an automated manner. We have addressed in depth the tracking of social distancing and the identification of face masks that help to ensure human health. The implementation of this solution was successfully tested in real-time by deploying model in raspberry pi4. The solution has the potential to significantly reduce violations by real-time interventions, so the proposed system would improve public safety through saving time and helping to reduce the spread of coronavirus. This solution can be used in places like temples, shopping complex, metro stations, airports, etc.

VI. FUTURE WORK

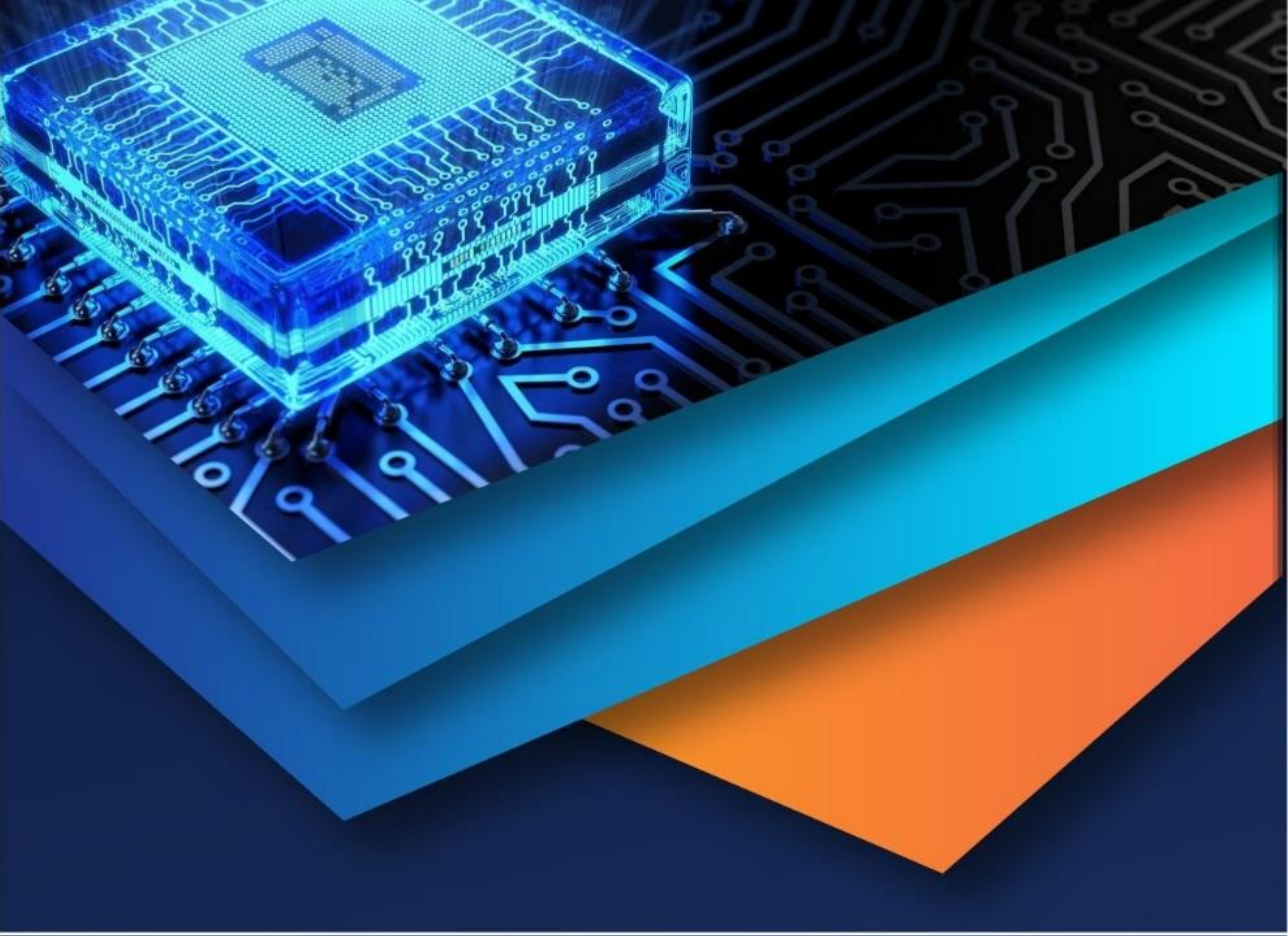
The above mentioned use cases are only some of the many features that were incorporated as part of this solution. We assume there are several other cases of usage that can be included in this solution to offer a more detailed sense of safety. Several of the currently under development features are listed below in brief:

1) *Coughing and Sneezing Detection:* Chronic coughing and sneezing is one of the key symptoms of COVID-19 infection as per WHO guidelines and also one of the major route of disease spread to non-infected public. Deep learning based approach can be proved handy here to detect & limit the disease spread by enhancing our proposed solution with body gesture analysis to understand if an individual is coughing and sneezing in public places while breaching facial mask and social distancing guidelines and based on outcome enforcement agencies can be alerted.
2) *Temperature Screening:* Elevated body temperature is another key symptom of COVID-19 infection, at present scenario thermal screening is done using handheld contactless IR thermometers where health worker need to come in close proximity with the person need to be screened which makes the health workers vulnerable to get infected and also its practically impossible to capture temperature for each and every person in public places, the proposed use-case can be equipped with thermal cameras based screening to analyze body temperature of the peoples in public places that can add another helping hand to enforcement agencies to tackle the pandemic effectively.

REFERENCES

[1] S. Wang Chen, Horby Peter W, Hayden Frederick G, Gao George F. A novel coronavirus epidemic of global concern for health. It's the Lancet. 2020;395(10223):470-473. 10.1016 / S0140-6736(20)30185-9.
[2] Matrajt L, Leung T. Evaluating the efficacy of social distancing strategies to postpone or flatten the curve of coronavirus disease. Emerg Infect Dis, man. 2020:
[3] Chen, S., Zhang, C., Dong, M., Le, J., R., M., 2017b. Using rankingcnn for age estimates, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[4] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2: Inverted residues and linear bottlenecks," IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510-4520. 2018.
[5] C.Fu, W.Liu, A.Ranga, A. Tyagi, A. Berg, "DSSD: deconvolutional single shot detector model," arXiv preprint arXiv:1701.06659, (2017)
[6] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Type Pyramid Networks for Object Detection," IEEE Conference Proceedings on Computer Vision and Pattern Recognition, pp. 2117-2125. 2017.
[7] GOODFELLOW, I., BENGIO, Y., & COURVILLE, A. (2016). Deep learning process. Chapter 6.
[8] S. S. Farfade, M. J. Saberian, and L. Li. Multi-view face recognition using deep convolutional neural networks. In ACM ICMR, pages 643–650, 2015
[9] Masita, K. L., Hasan, A. N., and Satyakama, P., 2018. Pedestrian identification by mean of R-CNN object detector. IEEE Latin American Conference on Computational Intelligence (Nov. 2018). DOI=10.1109/LA-CCI.2018.8625210.
[10] R. Girshick, "Fast R-CNN," in Proc. IEEE International Confrence Computer Vision, Dec. 2015, pp. 1440–1448.
[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with area proposal networks," in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 91–99.
[12] Liu, H., Chen, Z., Li, Z., and Hu, W. 2018. An powerful method of pedestrian detection based on YOLOv2. Mathematical Engineering Issues(Dec. 2018). DOI=https://doi.org/10.1155/2018/3518959.
[13] P. Viola and M. Jones, "Fast object detection using an enhanced cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I.
[14] X. Zhu and D. Ramanan. Face detection, pose estimation,and landmark position in the wild. In IEEE CVPR, pages 2879–2886, 2012
[15] Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Effective Convolutional Neural Networks for Mobile Vision Applications.
[16] Zhu, X. 2006. Semi-supervised study of learning literature survey. Computer Science, University of Wisconsin-Madison 2(3):4.

[17] David Oro ; Carles Fernández ; Javier Rodríguez Saeta ; Xavier Martorell ; Javier Hernando. Real-time GPU-based face detection in HD video streams. IEEE Int. Conf. Comp Vis 2011
[18] Glass RJ, Glass LM, Beyeler WE, Min HJ. Targeted social distancing architecture for pandemic influenza. Emerging Infectious Diseases. 2006;12:1671–1681.
[19] S. Ge, J. Li, Q. Ye and Z. Luo, "Detection of Masked Faces in the Wild with LLE-CNNs," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 426-434.
[20] S. J. Pan and Q. Yang, "A Survey of Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345- 1359, Oct. 2010.
[21] S. S. Mohamed, N. M. Tahir and R. Adnan, "Background modelling and background subtraction efficiency for object detection," 2010 6th International Colloquium on Signal Processing & its Applications, Mallaca City, 2010, pp. 1-6, doi: 10.1109/CSPA.2010.5545291.
[22] M. Piccardi, "Background subtraction techniques: a analysis," 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), The Hague, 2004, pp. 3099-3104 vol.4, doi: 10.1109/ICSMC.2004.1400815.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY