COVID-19 Face Mask and Social Distancing Detector using Machine Learning

Bhavik Limbasiya[1], Chinmay Raut[2]

[1]*UG Student, Dept. of Computer Engineering, Universal College of Engineering, Mumbai, India*
[2]*Assistant Professor, Dept. Of Computer Engineering, Universal College of Engineering, Mumbai, India*

---***---

Abstract - *Coronavirus disease that hit us in 2020 is affecting humanity drastically. The only safety measure that we may take against this pandemic is to wear "Face Mask" in public areas and maintain "Social Distancing". Furthermore, many service providers require customers to use the service only if they wear Mask correctly and maintain social distance, the places include airports, hotels, hospitals, railway stations etc. It's not possible to examine manually at all at ones to look if the rule of wearing Mask and social distance is followed as it consumes high human resources. We proposed COVID-19 Face Mask and Social Distancing Detector System which is a one-stage detector, which consists of an artificial neural network to fuse high-level semantic information with multiple feature maps, and a machine learning module to focus on detecting face Mask and social distances simultaneously. In addition, the system will use existing IP cameras and CCTV cameras combined with computer vision to detect people without Mask and violence of social distancing. This system provides tools for safety and security without any need for manual surveillance system. The system can be deployed on any infrastructure like Hospitals, Office Premises, Government Offices, Schools and Education Institutes, Construction sites, Airports etc. If deployed correctly, the face mask and social distance detector system we are building could potentially be used to help ensure people safety and the security of others.*

***Key Words*: Deep learning, ResNet50, Fully Convolutional Network, OpenCV, Face Mask Detection, Social Distance Detection, Yolov3, Image Processing, Matplotlib, Keras, Sklearn.**

1.INTRODUCTION

Coronavirus disease 2019 has affected the world seriously. One major protection method for people is to wear Mask in public areas and keep the social distance. The only way to prevent the spread of covid-19 is social distancing and wearing Mask. The goal of this project is to fight against the coronavirus, social distancing and wear Mask has proven to be a very effective measure to slow down the spread of the disease by providing the monitoring system that keeps track of the peoples by using existing IP cameras and CCTV cameras combined with Computer Vision to detect people without Mask and not following rule of social distancing. COVID-19 Face Mask and Social Distancing Detector System (COVID FSD) is an AI and Computer Vision driven image analytics solution which caters to the Covid-19 related violations. COVID FSD System uses Artificial Network to recognize if a user is not wearing a mask. The system can be connected to any existing or new IP or CCTV's cameras to detect people without a mask. It allows the application to run automatically and enforces the wearing of the mask. This system provides tools for safety and security without any need for manual surveillance system. Its artificial intelligence program detects violations like not wearing Face Mask and Social Distancing. This system can be deployed on the Hospitals, Office Premises, Government Offices, Schools and Education Institutes, Construction sites, Manufacturing units, Airports etc. COVID FSD System is simple and immediate to use, without any need for technical assistance. The system ensures total respect for customer privacy. In fact, the three applications do not record images but only release an alert signal in the specific circumstance. COVID FSD could provide an additional easy-to-access tool during the most delicate phase of the fight against the pandemic.

2. LITERATURE REVIEW

This paper adopts a combination of lightweight neural network RestNet50[1] and YOLOV3 (You Only Look Once) [2] 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 RestNet50. We mix the approach of modern-day deep learning and classic projective geometry techniques which not only helps to meet the 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 sent 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 computer local machine and the accuracy obtained was between 85% and 95%.

In recent years, object detection techniques using deep models [3] 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 [4] contextual information learning, and occlusion handling. Deep learning

object detection models [5] can now mainly be divided into two families: (i) two-stage detectors such as R-CNN [6], Fast R-CNN [7] and Faster R-CNN [8] and their variants and (ii) one-stage detectors such as YOLOv3 [9]. 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, consider detection as a regression issue and look at the image once for detection. The Viola – Jones [11] 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 [12] 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. [13, 14], suggested another model for facial detection which was a Multi- View Face Detector with surf capabilities. Oro et al. [15] also proposed a hear-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 several 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) [16] 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 to demonstrate a reduction in the growth rate. Z., Luo [17] 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 Mask such as scarves, mufflers, handkerchiefs, etc.

3. 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 by detecting whether and individual wears face mask. This section briefly describes the solution architecture and how the proposed system will automatically function in an automatic manner to prevent the coronavirus spread.

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 local machine 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 machine. We have used the RestNet50 architecture as the core model for detection as RestNet50 provides a huge cost advantage compared to the normal 2D CNN model. The process also involves the YOLOV3 Detector, a neural network architecture that has already been trained on a large collection of images such as ImageNet and Pascal for high quality image classification.

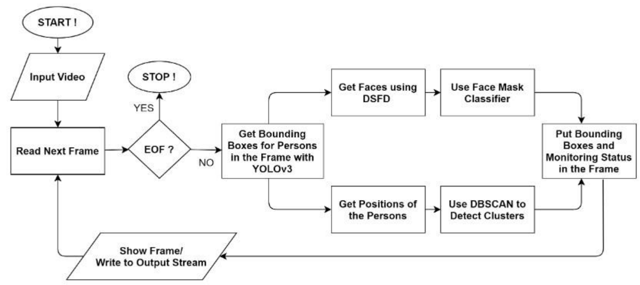

Fig -1: Pipeline Diagram

We are loading the RestNet50 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, Keras using Deep Learning and Computer Vision in order to inspect the safe social distance between detected persons and face Mask 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 YOLOV3 (You Only Look Once) using RestNet50 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 YOLOV3 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 YOLOV3 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

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

person's distance from the camera. The distance from the camera can be determined using the following:

D1 = (H x 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 YOLOV3 model for face detection with RestNet50 architecture as backbone to create a lightweight model that is accurate and computationally efficient, making it easier to deploy the model to machine. 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 YOLOV3 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 Mask 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.

4. EXPERIMENTAL RESULTS

The proposed system is a deep learning solution that uses OpenCV and Keras, to train the model. We combine the deep learning RestNet50 modal with the YOLOV3 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 Mask worn by the public in real time by means of a transfer of learning[20] to a DSFD face detector.

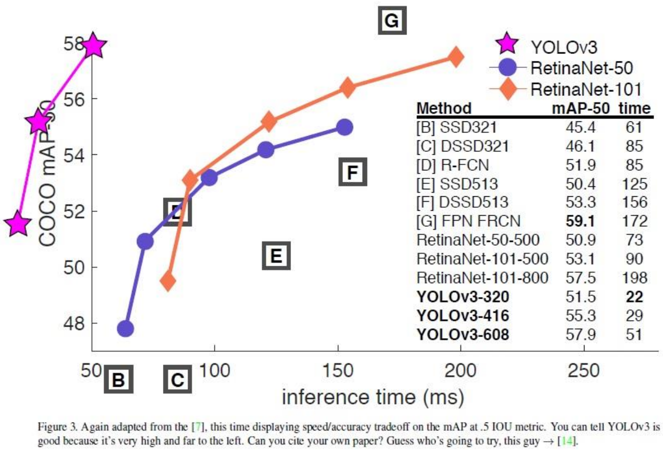

Chart -1: Yolov3 Detection Accuracy

This model combine's social distance detection and face mask detection. In our approach, the machine 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

4.1 Data Collection and Pre-processing

The proposed system uses a custom data set consisting of face images with different types of face Mask 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 Mask or not are performed by the YOLOV3 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.

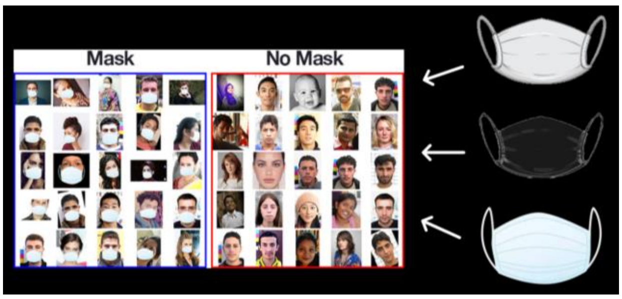

Fig -2: Data Processing

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.

4.2 Model building and Training

Our proposed framework uses the transfer learning approach [20] and will fine-tune the RestNet50 model, which is a highly efficient architecture that can be applied to edge devices with limited computing power, such as machine to detect people in real time. We used 80% of our total custom data set to train our model with a Yolov3, which takes only one shot to detect multiple objects that are present in an image using. 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 YOLOV3 model as input to build our custom model with RestNet50 as the backbone and train our model using the Keras Object Detection API.

Before model training begins, Keras 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 ResNet50 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.

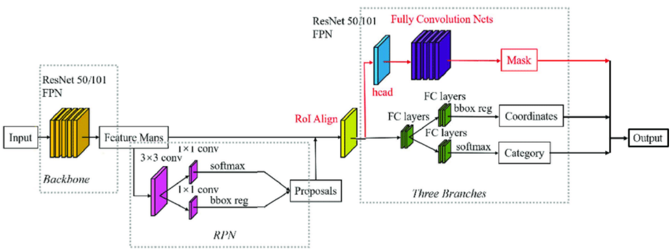

Fig -3: R-CNN

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 to determine the distance between the people in the frame.

4.3 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.

Fig -4: Detection Result

The system detects the social distancing and Mask 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: Detection Result

4.4 Model Implementation

The proposed system uses machine 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 machine, and the camera is attached to it. The camera feeds real-time videos of public places to the model in the machine, which continuously and automatically monitors public places and detects whether people keep safe social distances and also checks whether or not those people wear Mask. 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.

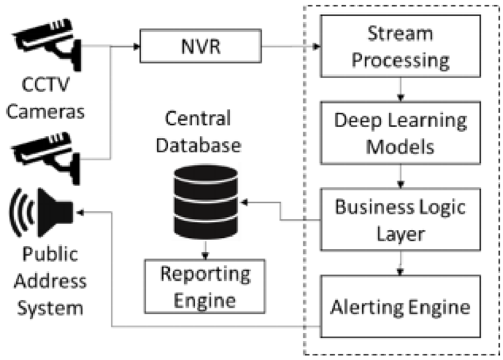

Fig -6: System Diagram

5. CONCLUSIONS

In this paper, we proposed an approach that uses computer vision and ResNet50 classifiers 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 COVID FSD 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 Mask that help to ensure human health while keeping the safety and privacy of users' data. The implementation of this solution was successfully tested in real-time by deploying model in my local Working platform (Computer). So, the face mask detection and social distancing system is going to be the leading digital solution for most industries, especially retail, healthcare, and corporate sectors. Discover how we can help you to serve the communities with the help of digital solutions. This system can be deployed on the Hospitals, Office Premises, Government Offices, Schools and Education Institutes, Construction sites, Manufacturing units, Airports etc. If deployed correctly, the COVID-19 mask detector we are building here today could potentially be used to help ensure your safety and the safety of others.

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. We believe that this approach will not only increase the safety at public places but also enhance the efficiency of plant processes in the time to come.

ACKNOWLEDGEMENT

We take this chance to precise our deep sense of feeling to our project guide and project organizer, Mr. Chinmay Raut, for his continuous steering and encouragement throughout the period of our Major-Project work. It's owing to his expertise and tremendous data, we will fulfil the necessity of finishing the Major-Project at intervals the stipulated time. We might additionally prefer to give thanks to Dr. Jitendra Saturwar, head of computer engineering department for his encouragement, whole-hearted cooperation, and support. We would additionally prefer to give thanks to our Principal, Dr. J. B. Patil and also the management of Universal faculty of Engineering, Vasai, for providing us all the facilities and also the work friendly setting. We tend to acknowledge with thanks, the help provided by division employees, library and research lab attendants.

REFERENCES

[1] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "RestNet50: Inverted residues and linear bottlenecks," IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510-4520.

[2] C.Fu, W.Liu, A.Ranga, A. Tyagi, A. Berg, "DYOLOV3: deconvolutional yolo model," arXiv preprint arXiv:1701.06659, 2017.

[3] 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, 2017, pp. 2117-2125.

[4] GOODFELLOW, I, BENGIO, Y, & COURVILLE, A. "Deep learning process", 2016 Chapter 6.

[5] S. S. Farfade, M. J. Saberian, and L. Li. Multi-view face recognition using deep convolutional neural networks. In ACM ICMR, 2015, pages 643–650.

[6] Masita, K. L, Hasan, A. N, and Satyakama, P. "Pedestrian identification by mean of R-CNN object detector". IEEE Latin American Conference on Computational Intelligence Nov. 2018.

[7] Hu, Guosheng, et al, 2015. "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition", IEEE.

[8] R. Girshick, “Fast R-CNN,” in Proc. IEEE International Confrence Computer Vision, Dec. 2015, pp. 1440–1448.

[9] 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.

[10] Liu, H, Chen, Z, Li, Z, and Hu, W. 2018. An powerful method of pedestrian detection based on YOLOv2. Mathematical Engineering Issues, Dec. 2018.

[11] 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.

[12] X. Zhu and D. Ramanan. “Face detection, pose estimation, and landmark position in the wild”. In IEEE CVPR, 2012, pages 2879–2886.

[13] Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. “Effective Convolutional Neural Networks for Mobile Vision Applications”, 2017.

[14] Zhu, X. 2006. Semi-supervised study of learning literature survey. Computer Science, University of Wisconsin-Madison 2(3):4.

[15] David Oro, Carles Fernández, Javier Rodríguez Saeta, Xavier Martorel, Javier Hernando. Real-time GPU-based face detection in HD video streams. IEEE Int. Conf. Comp Vis 2011.

[16] Glass RJ, Glass LM, Beyeler WE, Min HJ. Targeted social distancing architecture for pandemic influenza. Emerging Infectious Diseases. 2006, 12:1671–1681.

[17] 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.

[18] S. J. Pan and Q. Yang, "A Survey of Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, Oct. 2010, vol. 22, no. 10, pp. 1345- 1359.

[19] 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.

[20] 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.