Facemask Detection to Prevent COVID -19 Using YOLOv4 Deep Learning Model

Praagna Prasad
Electronics and Telecommunication Engineering, RV College of Engineering®, Bengaluru, India

Akhil Chawla
Electronics and Telecommunication Engineering, RV College of Engineering®, Bengaluru, India

Mohana
Electronics and Telecommunication Engineering, RV College of Engineering®, Bengaluru, India

***Abstract*- The on-going global Covid-19 pandemic has impacted everyone's life and brought economies to a standstill. World Health Organization (WHO) and Governments all over the world have found that social distancing and donning a mask in public places has been instrumental in reducing the rate of COVID-19 transmission. Stepping out of homes in a face mask is a social obligation and a law mandate that is often violated by people and hence a face mask detection model that is accessible and efficient will aid in curbing the spread of disease. Detecting and identifying a face mask on an individual in real time can be a daunting and challenging task but using deep learning and computer vision, establish tech-based solutions that can help combat COVID-19 pandemic. In this paper, YOLOv4 deep learning model is designed and applied deep transfer learning approach to create a face mask detector which can be used in real time. GPU used was Google Collab to run the simulations and to draw inferences. Proposed implementation considered three types of data as input such as image dataset, video dataset and real time data for face mask detection. Performance parameters are tabulated and obtained mean average precision of 0.86, F1 score 0.77 for image dataset, 90 % accuracy for video dataset. And real time face mask detector with accuracy of 95%, it is successfully able to identify a person with and without facemask and report if they are wearing a face mask or not.**

Keywords—Convolutional Neural Network (CNN), Artificial Intelligence (AI), Machine learning (ML), Video Surveillance, Computer Vision (CV),YOLOV4, Facemask detection.

I. INTRODUCTION

The COVID-19 pandemic has instilled fear among people as this disease can transmit through the respiratory system. This virus has killed more than a million people around the globe, and it is expected to rise and continue in the same way leading to the death of millions of people. India is on the verge of a second wave despite large scale vaccination drive. Vaccinating the whole population will take a lot of time, perhaps years and there is no evidence on reinfection or on the long-term protection against Covid-19, So the best way to prevent the spread of Covid-19 is to practice safety measures and not move around. For example, maintain social distancing, wash hand regularly, and wear a mask at all times. Experts believe one of the most effective methods to control the spread of the disease is by wearing a proper mask that covers your mouth and nose. A mask of any type gives 98% protection against the virus droplets spreading through their mouth. However, we observe that a lot of people in public places don't wear their mask properly and even if they are donning a face mask, it is not worn in a way that covers both nose and mouth. Thus, endangering others around them. A lot of the carelessness has led to the subsequent second wave, third wave and so on in many countries. In India today and people still refuse to wear a mask properly while stepping out. While our healthcare system is overwhelmed and the governments not imposing lockdowns, stepping out of our homes can be a dangerous act, even if it's an essential need. We aim to contribute to assist the government and the officials in making sure that people abide by the basic rule i.e., to wear a face mask in public at all times for which we have come up with a design for a Face Mask Detection program using the state-of-the-art Deep Learning technique. The goal is to find out who is not wearing a facial mask in public places and not abiding by the rules with the help of model. As stated by the WHO, respiratory droplets and any type of physical contact are the two ways of coronavirus spread, and to prevent the virus from the spread, medical masks are the best and only solution. Mask is important because of two reasons: The virus spreads directly from the infected person's sneeze or cough and if he/she is masked, they will prevent themselves from transmitting to others. Secondly, the mask will stop the transmission of the virus in case you touch a virus- contaminated surface and then your mouth or nose. The nature of the spread of the disease is rampant, no one anticipated that we would be engulfed by a worldwide pandemic but with the power of technology, we can innovate to prevent the spread of the disease and hopefully see an end to it. With the on-going crisis, many companies and governments have used the power of technology, particularly Artificial intelligence (AI) to help combat the disease. We have seen robots delivering food, medicines etc. to infected individuals in the hospitals. AI has been used extensively to help predict the origin of the disease and also, in helping find the cure and many are using it to detect Covid-19 through X-rays and CT scans. AI even used in building tracking software and wearable tech to make sure people are abiding by their quarantine rule. Countries like Taiwan merged their national medical insurance database with records and data from their immigration and customs database, hence confronting possible infected passengers who might have contracted it while travelling from a Covid-19 hotspot. AI and technology have had significant impact in combating the disease. The use of AI and DL in healthcare applications is growing rampant but hasn't been able to eradicate human labour. However, as time progresses AI will be used in mainstream applications and an integral part in healthcare systems and governments around the world. This paper make use of deep learning to detect the use of masks to prevent the spread of coronavirus. Proposed implementation is aimed to assist the government to take action against those people who are not wearing masks in public places. We have used Deep Transfer Learning Approach to go about the problem, deep learning is a part of the family of machine learning methods based on artificial neural networks which can be supervised, semi-supervised or unsupervised, while transfer learning is an approach in deep learning where knowledge can be transferred from pre trained highly efficient models trained by researchers leading to high efficiency and reduces the computational complexity. We have also made uses of Python along with cross platform computer vision library, OpenCV which helped us develop real-time computer vision application. TensorFlow and Keras that provided a Python interface for artificial neural network.

II. Motivation, Problem Analysis and Methodology

Coronavirus disease (COVID-19) is a virus infection caused by a newfound Coronavirus. It has affected the lives of millions of people and in some cases, people have lost them lives fighting this. Economies have been shut down and unemployment rate took a sharp increase. It has impacted everyone in one way or the other and completely disrupted the way institutes, organizations and civilizations work. People with fundamental clinical issues and co-morbidities are at risk leaving them vulnerable to serious cases of illness. While following safety measures and protocols, using a face mask has proven to be a must as data shows that it can stop the spread of the virus between two people who are masked. People still tend to not follow this rule in public and endanger others. A mask detection system will help the local officials to identify the rule breakers quickly and safely without them putting themselves out there.

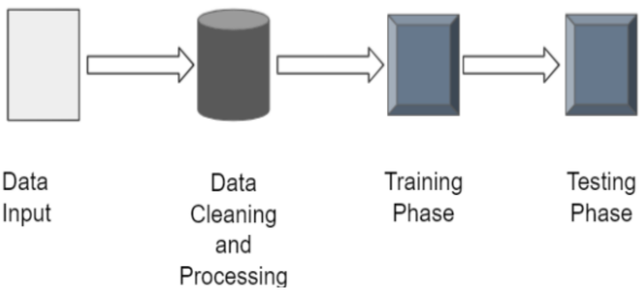

Fig.1. Block diagram of methodology

Figure 1 shows the block diagram of proposed design methodology. Three types of input data is such as image data, video data and real time data is considered for implementation. Preprocessed data was divided in to training and testing purpose on 90:10 basis.

III. YOLOV4 Deep Learning Model

A. *Architectre*

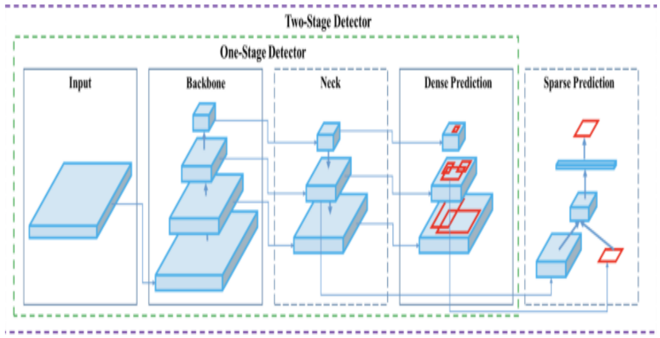

Fig.2. Architecture of YOLOV4

Figure 2 shows the architecure of YOLOv4. *Backbone* - CSPDarknet53 backbone used for feature extraction as shown in figure 3 which includes series of convolutional and residual layers. Neck – used to add extra layers between the head which is also known as the dense prediction block and the backbone. Dense Prediction finds it’s use in single stage-detection deep learning algorithms; *Sparse Prediction* — This is another type of algorith which is used in mainly in two-stage-detection algorithms.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
	Convolutional	32	1 × 1	
1×	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
	Convolutional	64	1 × 1	
2×	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
	Convolutional	128	1 × 1	
8×	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
	Convolutional	256	1 × 1	
8×	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
	Convolutional	512	1 × 1	
4×	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig.3. CSPDarknet53

Bag-Of-Freebies(BoF) - one of the most unique technique used in YOLOV4, it is a collection of intelligent techniques that is used to during training the data to help get better and more accurate outcome without adding any inference time. Bag of Freebies is widely used in data augmentation.

Primarily data augmentation is used in object detection to increase the size of our training set and it also helps in accuracy by exposing the model to semantic conditions. One of the latest application of data augmentation is mosaic data augmentation and this tiles four images together from the data set and it trains the model to identify and remember smaller object.

Bag-Of-Specials(BoS) - one of the most unique techniques used in YOLOv4. However, the difference between BoF and BoS is that BoS changes network architecture and augment the interference cost a bit. One of the most widely used technique is Mish activation, this is a type of activation function which is designed to drive signals to left and right. In drop block regularization technique, sections of the image are hidden from the first layer, which forces the network to learn features to recognize the object.

IV. DESIGN AND IMPLEMENTATION

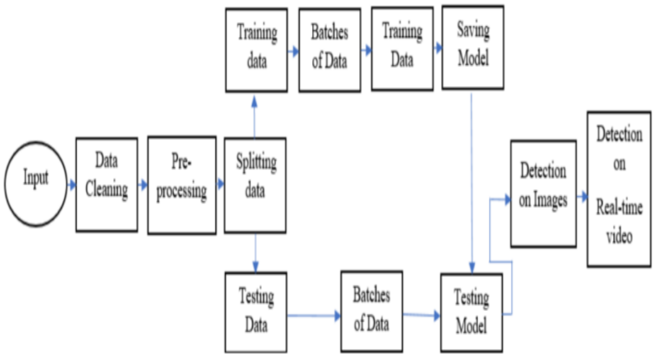

Fig.4. Block Diagram of implementation

Figure.4 shows the block diagram of face mask detection using deep learning and computer vision. The process is split in to multiple parts as follows.

A. Acquiring the Dataset

The process begins by finding or creating an appropriate dataset, which consists of multiple images of persons or people, with some people wearing a mask and some not wearing a mask. The quality of the dataset has a major contribution towards the final efficiency of the model as, in this phase, feed our model with all the possible scenarios it needs to carry out detections on while testing and later in real time implementation. The image dataset used in the implementation of our model included images of people wearing different types of masks, with multiple people in a frame with and without masks. It also included people sitting in different poses with masks on. The number of variations model covers while training directly impacts the accuracy of detection through the model.

B. Cleaning the Dataset

It is the process of ensuring data is correct, consistent and usable. This step involves keeping only the meaningful and required part of the dataset that is relevant to the final prediction that the model needs to be trained on, this is critical for the working of the model.

C. Data Pre-Processing

It involves transforming raw data into an understandable format. The major part of the processing is the data labelling process, to get the dataset ready for training.

D. Data Labelling

Data Labelling is an important tool in the design and implementation process it helps to identify raw data such as images, files, etc. and adds meaningful labels to these data so it can provide better context to the deep learning algorithm. In object detection, data labelling is referred to the process of creating a border that fully encloses a digital image, known as a bounding box around an object. Once the bounding box is drawn, its information such as center of the box, height and width of the box and the class of the object enclosed in the box is stored and provided to the model while training. The process of data labelling is carried out using the Open Labelling Tool. The tool provides a GUI for labelling the dataset easily, with multiple features such as single keys for switching between classes of the object and fast switching between different images thereby assisting in the labelling process. The tool directly saves a txt file that is the Yolov4 format for that is used for training. Another feature provided by the open labelling tool to carry out automatic data labelling is labelling using the frozen inference graph. The frozen inference graph is a graph that has been taken from a pre-trained model used for the same object detection where it is used for labelling. The graph uses saved weights from the pre-trained model it belongs to, to carry out the labelling of images, and does so with the accuracy acquired by the pre-trained model. The process saves a lot of time, but does require a manual re-check after its termination as the labelling might not be 100% accurate, but it can be re-do easily using the open labelling tool GUI features.

E. Splitting the data Set

The dataset is split into training and testing datasets, with 90% data used for training and 10% used for testing.

F. Training phase

In this phase the model is provided with the training images along with the labelling in Yolov4 txt format. The training is done by creating a random fixed size batch of the input training data and provide it to the model to do prediction and keep modifying the weights. The dataset is made to pass through the model for multiple iterations until the required accuracy is reached. The errors made by the model in each iteration are back propagated through the layers and made to evolve the parameters of the layers in order to reduce the error for the next predictions. At the end of training, we have a fine-tuned weights file which when implemented to carry out the testing should give high accuracy results.

G. Testing phase

Once the model has reached desired accuracy in training phase, the model is saved and implemented on testing dataset, to determine how well model performs on unseen data and images that would basically give an idea about its performance in real time implementation. In this phase a random fixed size batch of input testing data is created and provide to model to carry out the prediction based on saved weights from the model calculated while training, multiple performance parameters are also calculated during this phase, which suggest improvements required, if any.

H. Realtime implementation

Once fine tuning of model is done, the model is tested for real-time implementation, thereby carrying out predictions on live videos, predicting if a person in the frame is masked or not.

I. Dataset Specifications

Dataset used:

- Joseph Nelson Roboflow
- Prasoonkottarathil Kaggle

- Ashishjangra27 Kaggle
- Andrewmvd Kaggle

Dataset specifications:
Training images: 1359
Test images: 151
Total No. of images: 1510

J. Sample images from Dataset

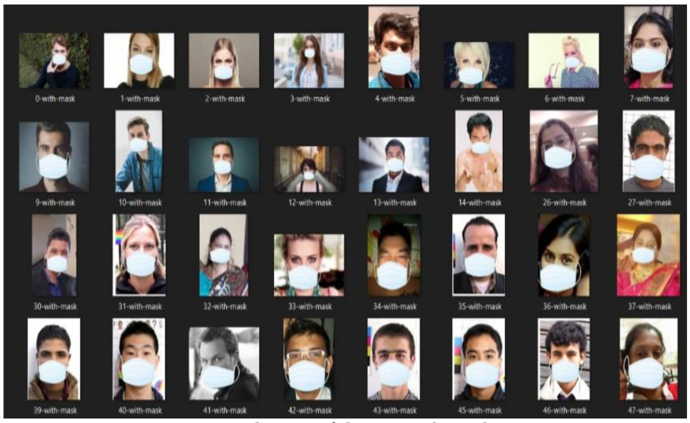

Fig.5. Glimpse of dataset with mask

Figure 5 shows a glimpse of the dataset which consist of 2000 images. We have considered a diverse dataset with people from different ethnicities, facial features and Demographic.

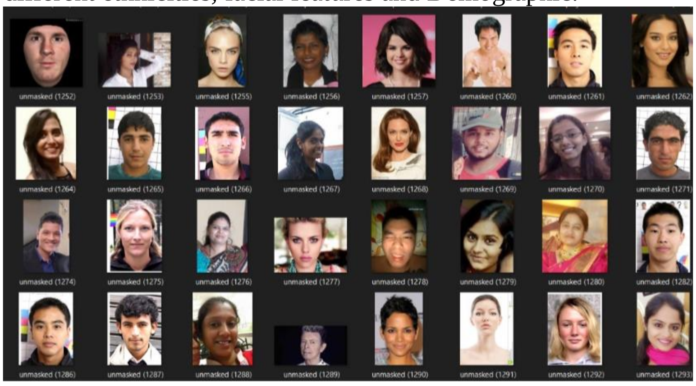

Fig.6. Glimpse of dataset without mask

Figure 6 shows a glimpse of dataset of people without wearing a mask. This has images of people unmasked in different directions, different orientations so that we can train our model accurately.

K. Performance Parameters

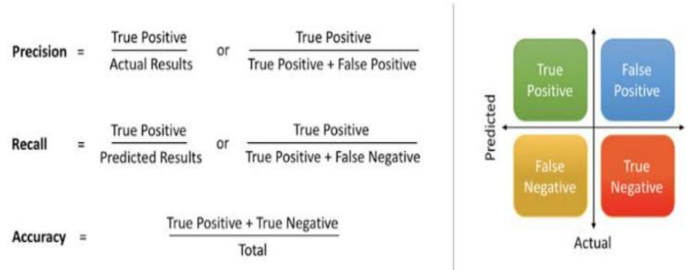

Fig.7. Performance Metrics

Figure.7 shows the mathematical formulas used evaluate the model. Precision which is also known as positive predictive value is basically relevant instances upon the retrieved instances. Recall also known as sensitivity is a fraction of relevant instances that were retrieved. Accuracy is the ratio of addition of true positive and false positive by the total number of predictions. The F1 Score is given by 2 times of ((precision*recall) divided by precision recall.

V. SIMULATION RESULTS AND ANALSIS

This section elaborates the obtained results and analysis of performance metrics on unseen data.

A. Simulation results on with mask

Results obtained after running the model on test images with mask from the dataset. The detections were done with high accuracy, even for multiple people in an image sitting in different poses.

Fig.8. Facemask detection with multiple people in frame

Figure 8 shows the result of facemask detection with multiple people in single image. Facemask detection is done on people who are clearly visible in the frame with a bounding box prediction confidence score of 0.87, 0.79 and 0.71. The image also consists of a blurred background.

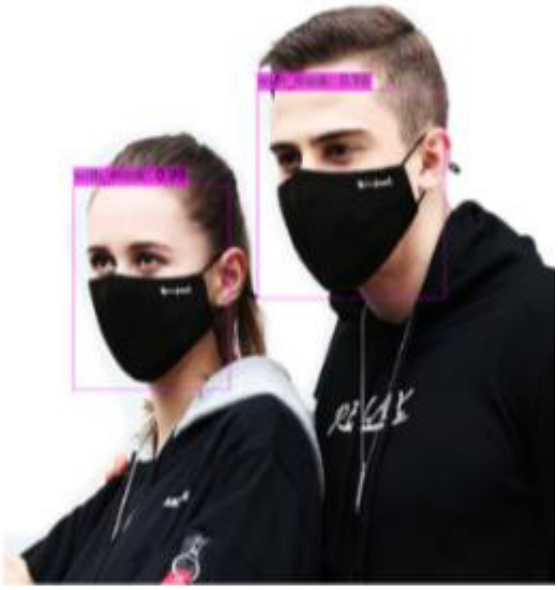

Fig.9. Facemask detection with two people standing side by side

Figure 9 shows the result of facemask detection with two people standing side by side. Two people are detected with facemasks standing in a side facing pose with a confidence score of 0.96 and 0.90. In this image the background is clear and no objects present in background.

Fig.10. shows the result of facemask detection with multiple people in single image frame. Detection is carried out on multiple people in a single frame, facing in different directions, with a confidence score of 0.97, 0.93, 0.89, 0.87.

Fig.10. Facemask detection with multiple people in a frame

.

Fig.11   Facemask is detected on a marked person

Fig.11. shows the result of facemask detection on a marked person. Here facemask is detected on a person wearing a different type of mask than other example images with a confidence score of 0.97.

B. Simulation results on without mask

Results obtained after running the model on images without mask from the dataset.

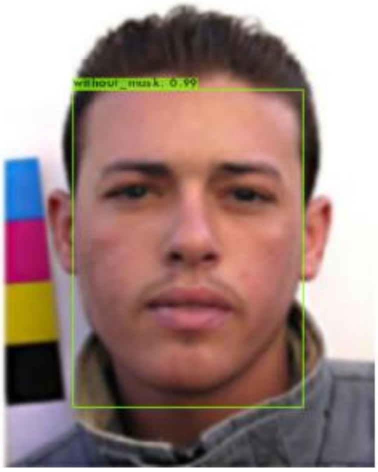

Fig.12. Facemask detection on unmasked person (Sample output image one)

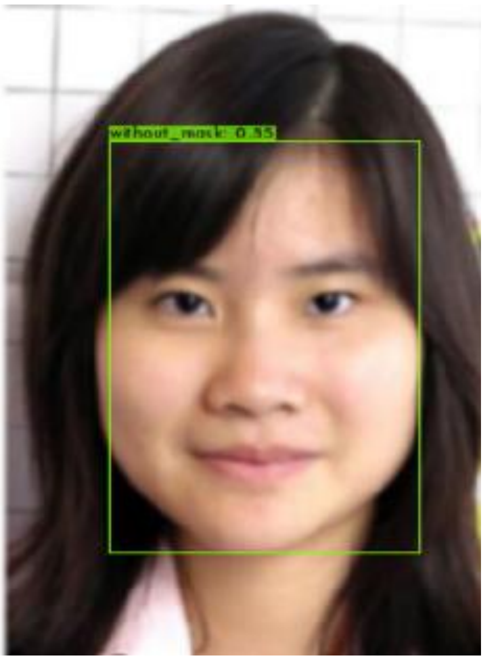

Fig.13. Facemask detection on unmasked person (Sample output image two)

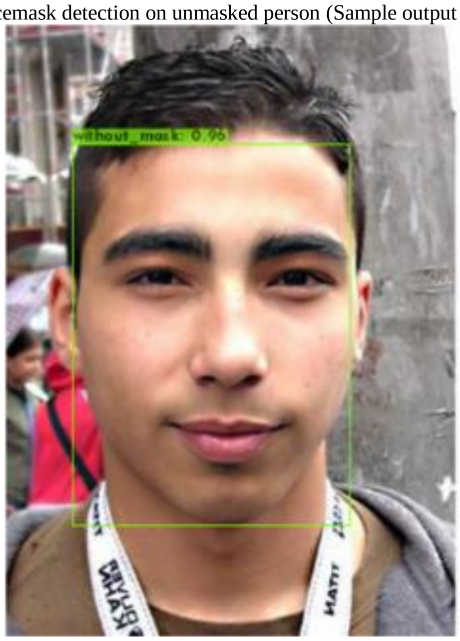

Fig.14. Facemask detection on unmasked person (Sample output image three)

Figure 12, 13 and 14 shows the results of facemask detection on unmasked person. Person is detected without facemask with a bounding box accuracy of 0.99, 0.85 and 0.96.

Fig.15. Facemask detection with multiple unmasked people in a single image frame.

Figure 15 shows the results of facemask detection with multiple unmasked people in a single image frame. Two people are detected without facemask where the first person is detected with a confidence score of 0.83 and second person in the background being a little blurred is detected with a confidence of 0.58.

C. Performance Analysis

TABLE I PERFORMANCE PARAMETERS ANALYSIS

Images used for testing	151
Detections Count	1487
Confidence Threshold	0.25
True Positive	295
False Positive	131
False Negative	44
Average IOU	53.15%
Recall	0.87
Precision	0.69
F1 Score	0.77
Mean Average Precision	0.86

Table I shows the results obtained after running detection on the test images, the images used for testing were 151, where the detection count for with and without facemask was 1487 detections. The threshold set for detection is 0.25, which basically lets the detections happen even at a confidence score of 0.25. Average IOU – It determines the average of the extent of overlap in all detections between the bounding box generated by the detector and the ground truth bounding box, which in our case came out to be 53.15%. Recall - It is the fraction of relevant instances that were retrieved, i.e., it is the ratio of true positive to the sum of true positive and false negative. Observed value for recall is 0.87. Precision - It is the fraction of retrieved detections that are relevant to the object detection. It is given by ratio of true positive to the sum of true positive and false positive. Our observed value for precision is 0.69. F1 Score – It is the harmonic mean between precision and recall. It is a statistical method that talks about the performance of the model, by taking into account the effect of false positive and false negative and conveys the balance between the precision and recall. The value of F1 score achieved by the model is 0.77. Mean Average Precision – It is a metric used to evaluate the performance of object detection by a YOLO model as it compares the ground-truth bounding box to the detected box and returns a score. Value achieved for mean average precision by the model is 0.86.

D. Real time implementation

Further designed YOLOv4 facemask detection model simulated for real time video to validate model. This section describes the model results of real time implementation of with and without mask. Figure 16 and 17 shows the sample results of facemask detection on a real time video of a masked person. Results show that the model was able to detect masked persons correctly with an accuracy of 0.95 and 0.99.

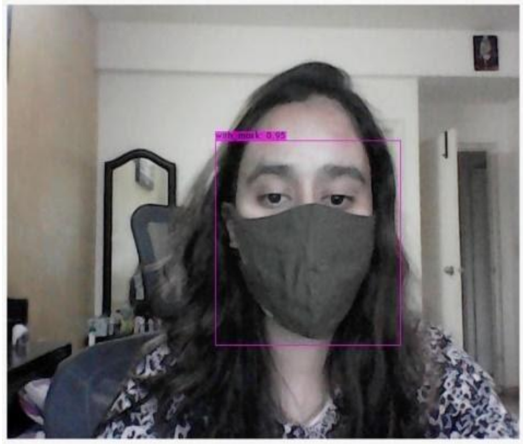

Fig.16. Facemask detection on a real time video of a masked person(sample outpu image one)

Fig.17. Facemask detection on a real time video of a masked person (sample outpu image two)

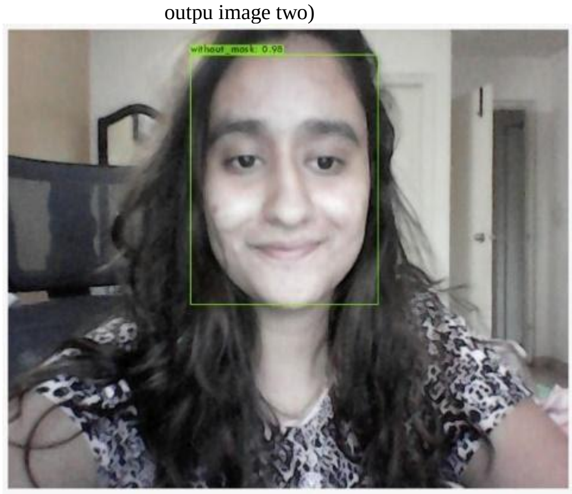

Fig.18. Facemask detection on a real time video of a unmasked person (sample outpu image one)

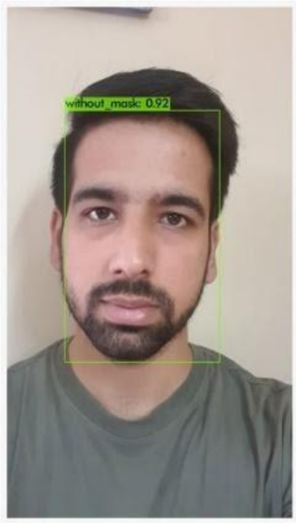

Fig.19. Facemask detection on a real time video of a unmasked person (sample outpu image two)

Figure 18 and 19 shows the sample results of Facemask detection on a real time video of a unmasked person. Results show that the model was able to detect unmasked persons correctly with an accuracy of 0.98 and 0.92.

VI. CONCLUSION AND FUTURE WORK

COVID-19 pandemic has impacted the way people interact with each other and definitely has changed the social patterns that existed before the pandemic. Mask mandate has been compulsory since the beginning of the pandemic. According to data published by WHO and various researchers, COVID-19 transmission is significantly higher in crowded places and between a groups that is not wearing face masks. We understand that mask detection is technology that would be implemented in public places, malls, workspaces, corporate offices, co-working spaces and stadiums. As the world is opening up again, businesses and offices are re-opening and wearing a mask has been compulsory and a proven method to keep us safe against the virus. In this paper, implemented a real-time facemask detection model with architecture design and pattern recognition using YOLOV4 algorithm state-of-the-art object detection model for image dataset, video dataset and real time video. Various model performance metrics tabulated to draw conclusions and to validate the model. Implemented our model to test in 3 conditions, in video form, in image form and in real time application. The real time application is the most useful application in future. Obtained an accuracy of 95% for our real time detection model and hence we can conclude that the model is successful. Key players and industry leaders like NVIDIA etc. are working on creating and developing new technology to combat the spread of covid-19. Researchers, Corporates and public and private organizations can implement smart solutions to protect their employees and ensure a safe working environment.

Development of facemask detection using mobile phones technology can be made accessible on a larger scale and public places. Having access to this technology on such a large scale can help make this popular and aid in preventing the spread of the virus. There is also scope for integrating social distancing software to detect and prevent overcrowding at public places. We also see a huge scope for integrating different algorithms such as GAN algorithm which can be used for face mask detection through speech analysis and processing, by understanding the variation in sound parameters while talking with a mask on. Industry key players and innovators are working towards tech based solutions to overcome the pandemic and can integrate the proposed work with their research.

REFERENCES

[1] Sneha Sen, Harish Patidar, “Face Mask Detection System for COVID_19 Pandemic Precautions using Deep Learning Method”,*International Journal of Emerging Technologies and Innovative Research,* Vol.7, Issue 10, pp. 16-21.

[2] https://www.researchgate.net/publication/347439579_Social_Distance_Monitoring_and_Face_Mask_Detection_Using_Deep_Neural_Network

[3] I. B. Venkateswarlu, J. Kakarla and S. Prakash, “Face mask detection using MobileNet and Global Pooling Block,” *IEEE 4th Conference on Information & Communication Technology (CICT), 2020*, pp. 1-5.

[4] https://www.researchgate.net/publication/344239546_Comparative_Study_of_Deep_Learning_Methods_in_Detection_Face_Mask_Utilization

[5] G. Jignesh Chowdary, Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal “Face Mask Detection using Transfer Learning of InceptionV3” *arXiv:2009.08369*,2020.

[6] T. Q. Vinh and N. T. N. Anh, “Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier,”*International Conference on Advanced Computing and Applications (ACOMP),* 2020, pp. 146-149.

[7] S. Susanto, F. A. Putra, R. Analia and I. K. L. N. Suciningtyas, “The Face Mask Detection For Preventing the Spread of COVID-19 at Politeknik Negeri Batam,” *International Conference on Applied Engineering (ICAE)*, 2020, pp. 1-5.

[8] https://www.researchgate.net/publication/344725412_Covid19_Face_Mask_Detection_Using_TensorFlow_Keras_and_OpenCV.

[9] W. Han, Z. Huang, A. kuerban, M. Yan and H. Fu, “A Mask Detection Method for Shoppers Under the Threat of COVID-19 Coronavirus,” *International Conference on Computer Vision, Image and Deep Learning (CVIDL),2020,* pp. 442-447.

[10] M. R. Nehashree, P. Raj S, Mohana, “Simulation and Performance Analysis of Feature Extraction and Matching Algorithms for Image Processing Applications,” *International Conference on Intelligent Sustainable Systems (ICISS), 2019,* pp. 594-598.

[11] R. K. Meghana, Y. Chitkara, A. S., Mohana, “Background-modelling techniques for foreground detection and Tracking using Gaussian Mixture Model,” *International Conference on Computing Methodologies and Communication (ICCMC),* 2019, pp. 1129-1134.