Real-time Face and Mask on-off detection

Rahul Maheshwari (MT19027)
IIIT Delhi
Drive link

Abstract

The project aims at developing a solution in the current time of pandemic going on in the world. Masks are one of the most important protection and preventive measure a person must take in order to be safe from COVID-19. A system is proposed using state-of-the-art architectures MobileNetV2 and ResNet10 to detect faces from images and real-time video stream and detect masks on faces. The final accuracy of the model is 95% with an f1 score of 0.95.

1 Introduction

Corona Virus (COVID-19) originated in the late 2019 and is seen as currently one of the biggest threats to humanity because of no cure till now and the wide spreading nature of the disease by the virus. Although the virus has low mortality rate but having no cure makes it a deadly virus and being similar to SARS and MERS which have caused many casualties in past, it is a threat to humanity and as a result has stopped the world on its toes because of its severeness and adverse effects on humans. The only preventive measure in current scenario is to wear a mask and maintain social distancing. Even existing detection mechanism for an infected human are not completely reliable. What makes the virus deadlier is that sometimes the infected person does not show any symptoms. Keeping a mask on when in public places or crowded places reduces the spread of virus and risk of getting infected exponentially and currently the only measure an individual can take is to wear masks and avoid crowded places as suggested by WHO (World Health Organization), CDC (Centers for Disease Control and Prevention) and ICMR (Indian Council of Medical Research). Keeping this in mind, a system has been proposed to detect in a video stream and images for people wearing masks or not. This idea can be used to capture CCTV footages of crowded places for the monitoring and help in preventing the spread of the virus.

2 Importance

COVID-19 brought the world into an unimaginable state where social interactions between the humans are halted and seems to be in such a condition for an indefinite time period based on the current situation in the world. It forced humans to work from home and cut interactions for a long period of time but this becomes difficult to manage for such long periods of time and people need to interact in order to make a living. As of now as there is no successful cure, the only "cure" is putting on a mask which needs to be made mandatory for everyone. The proposed system helps to achieve the same by identifying if a person is taking proper preventive measures or not and helps to prevent further spread of the virus. **In the current scenario, the proposed system is important and need of the hour for prevention of spread of COVID-19.**

3 Literature Survey

Due to the usefulness of Deep Learning techniques in computer vision and solutions based on image analysis, machine learning models are extensively used for developing novel solutions to problems of such nature. Convolutional models are used mainly when dealing with images due to its architecture, one such application is presented by Li et al. [1] for faster face detection which leveraged GPU to achieve lag free face detection in a video stream. A novel data augmentation approach was introduced for mask detection by Ristea et al. [2]. They used ResNet and GANs (Generative adversarial network) for the classification. Wang et al. [3] proposed mask detection using a similar approach. Inamdar et al. [4] proposed a novel solution using similar approach by proposing Facemasknet deep learning network. Khandelwal et al. [5] performed binary classification using MobileNetV2 for mask detection. Jiang et al. [6] proposed an efficient mask detector using ResNet and MobileNet. Qin et al. [7] used SRCNet for face mask detection and achieved a very high accuracy of 98.70 %. In most of the implementations, models gave better results with masks that are not transparent or does not reflect or illuminate any light source.

4 Dataset

The dataset used for the proposed system is created by Chandrika Deb [8]. The dataset contains a total of 3868 images out of which 1938 images are with mask and 1930 images are without mask. The images are not of fixed resolutions and are resized to 200x200 pixels for training and evaluation. Out of 3868 images, 3094 images are used for training and 774 images are used for validation and testing.

The major challenge was to prepare a dataset with real people images as most of the similar datasets available are either very small or contains images unfit for training. The dataset used is created combining data scraping from Bing, Kaggle dataset and RMFD dataset. The model thus performs much better when tested on real faces compared to the baseline model.

5 Data Analysis

Dataset contained color images of varying resolutions belonging to 2 classes “with mask” and “without mask”. Hence for faster computation, these were scaled down to 200x200 pixels and given as input for training the MobileNetV2 model.

The resultant images given as input are of 200x200x3 size. Some of the images are scraped from websites and some are included using Kaggle datasets and RMFD dataset.

Figure 1: Sample image with class label ‘with mask’.

Figure 2: Sample image ‘without mask’.

Images containing faces at different angles and views are included in the dataset for a more robust training with the deep learning model.

Images including faces wearing different types of mask and different color masks are included in order to make the model unbiased towards a specific type of mask which is also evident in results subsection.

6 Preprocessing Steps

For all the 3868 images in the dataset, data augmentation [9] is used to transform the images using random transformations in such a way so that we have robust data to train the model and no two images are exactly similar. This is certainly helpful when we have dataset containing images and we want to learn all features of those images with varied alignment. This prevents overfitting and produces a better generalized model.

The pre-processing pipeline is shown in the below flowchart.

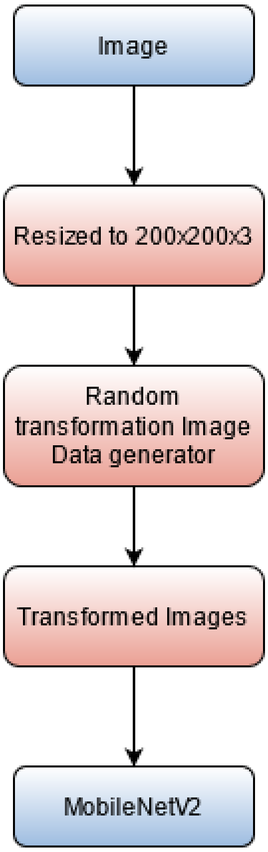

Figure 3. Pre-processing pipeline

7 Methodology/Approach

Task is divided into mainly 2 phases; first phase was to train the MobileNetV2 model using pre-trained weights to minimize training time and have a high accuracy model. The model is serialized and saved to be used for performing the detection in images and real-time video stream. Second phase is to use the trained model to detect masks on the faces detected using pre-trained ResNet10. Caffe ResNet10 is used which is highly accurate in face detection as it is trained on more than a million images for a total of 140000 iterations.

MobileNetV2 architecture is chosen for mask detection as it is a lightweight model with multiple convolutional blocks which makes it suitable to deploy on computationally less powerful device also. Both the models used are state-of-the-art models and produces a high accuracy for the given task. The model is trained using CUDA cores of GPU for efficient training. After detection of faces in the image or stream, face ROI (Region of Interest) is extracted and sent to the MobileNetV2 model. Model performs the detection of masks on all faces and highlights each face with a rectangle indicating the label "Mask" or "No Mask" with the probability with which model performed the detection. A red box is shown for "No Mask" detection and a green box is shown for "Mask" detection.

Due to use of lightweight DNN model, there is very less lag faced during real-time video stream detection with no trade-off in performance and with the help of ResNet10, the system is able to detect multiple faces with high accuracy.

7.1 Details of Deep Learning models

A lightweight DNN MobileNetV2 is used for mask detection and a pre-trained ResNet10 is used for face detection.

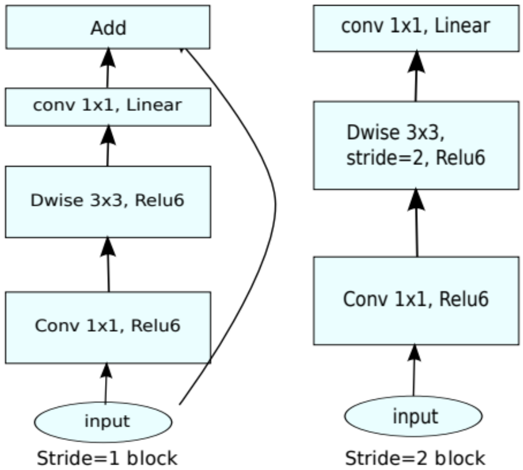

Figure 4. MobileNetV2 block architecture

MobileNetV2 contains a total of 16 blocks of the above-mentioned architecture which is combined with an Average Pooling layer, Dense layer, Dropout layer, and a softmax output layer.
The model has a total of 2,422,210 parameters out of which 164,226 parameters were trainable as pre-trained layers of MobileNetV2 architecture is implemented which helped in faster training of the task at hand.
Face detection Caffe ResNet10 model is pre-trained and used for highly accurate face detections even on images and stream containing multiple faces. All the functionalities are integrated in a web-based cross-platform GUI for ease of using the model and perform the task.

7.2 Data Augmentation

To perform more robust training, data augmentation is used which performs random transformations on the images in order to prevent overfitting and ensures that no two training samples are exactly same. This helps in generalizing the model as the model is not biased towards a certain orientation and other properties of an image.

8 Results

Faces are detected from the uploaded image in the GUI or from the real-time video stream and sent to the saved model for mask detection, Accuracy and loss were observed for 25 epochs and plotted for better visualization and analysis of the performance of the deep learning model.

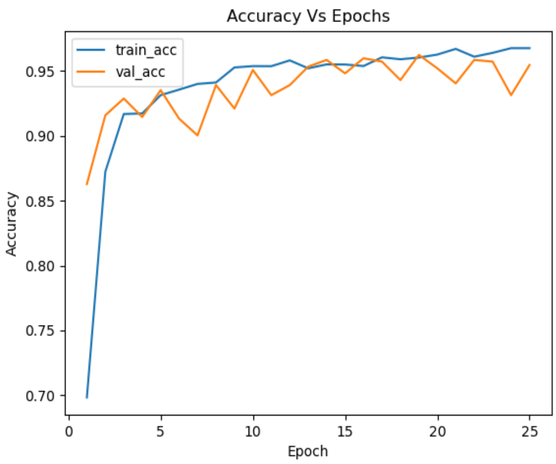

Figure 5. Accuracy Plot

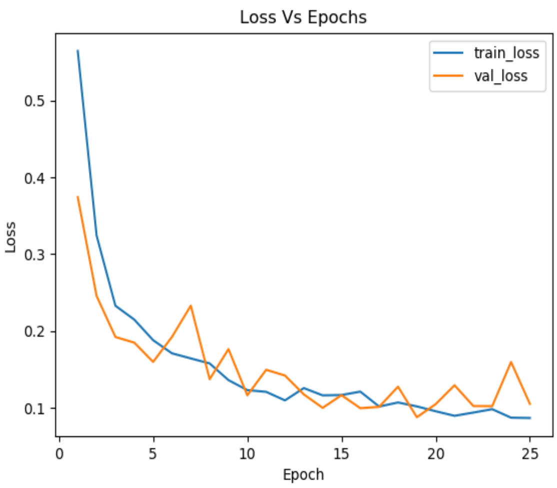

Figure 6. Loss Plot

	precision	recall	f1-score	support
with_mask	0.99	0.92	0.95	375
without_mask	0.93	0.99	0.96	398
accuracy			0.95	773
macro avg	0.96	0.95	0.95	773
weighted avg	0.96	0.95	0.95	773

Figure 7. Classification Report

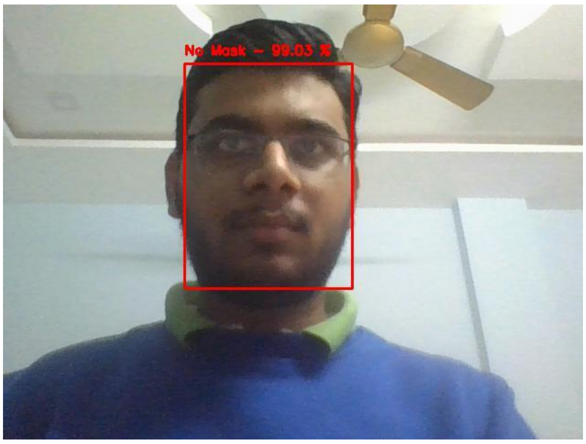

Figure 8. Real-time video stream screenshot without mask detected

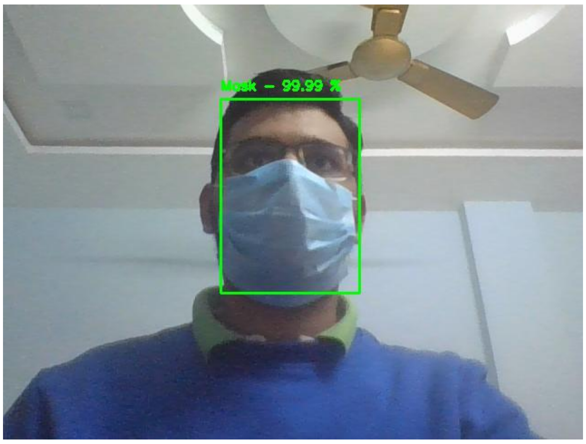

Figure 9. Real-time video stream screenshot with mask detected

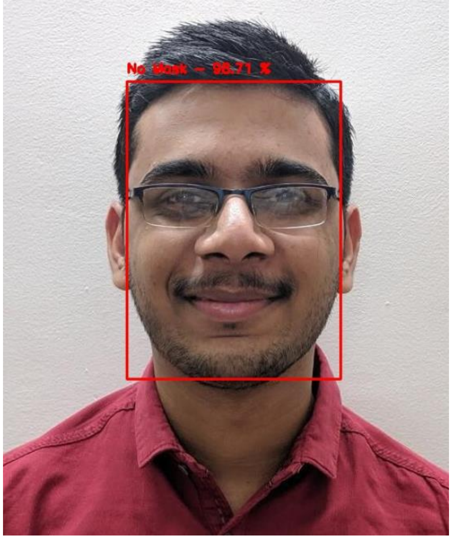

Figure 10. Mask off detection on image

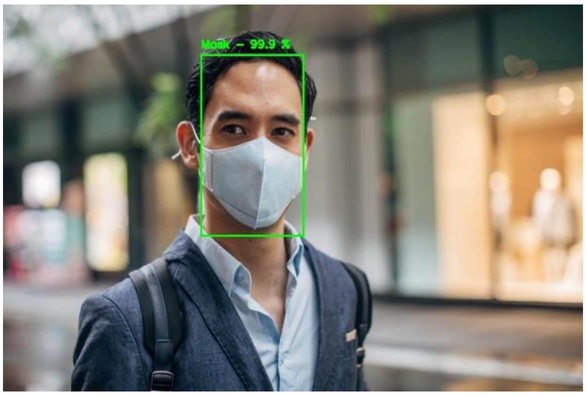

Figure 11. Mask on detection on image

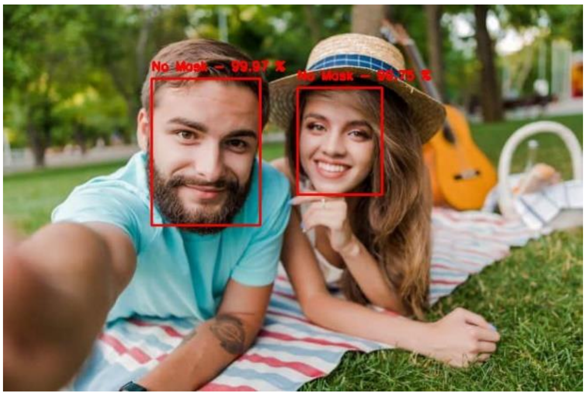

Figure 12. Mask off detection on multiple faces

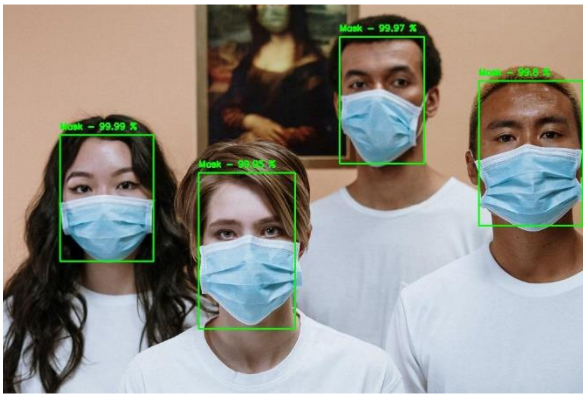

Figure 13. Mask on detection on multiple faces

9 GUI Information

The GUI is a web-based cross platform application developed using streamlit. Streamlit is used to develop GUI as it allows rapid development of front end with a modern and minimalistic design which is very usable. Another benefit of using streamlit GUI is that the same GUI is responsive for mobile devices also.

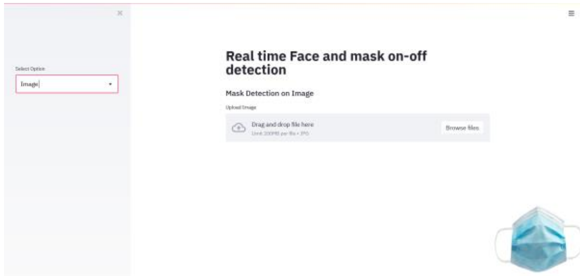

Figure 14. Main GUI screen

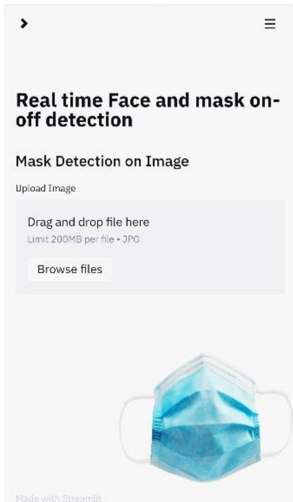

Figure 15. GUI screen on mobile device

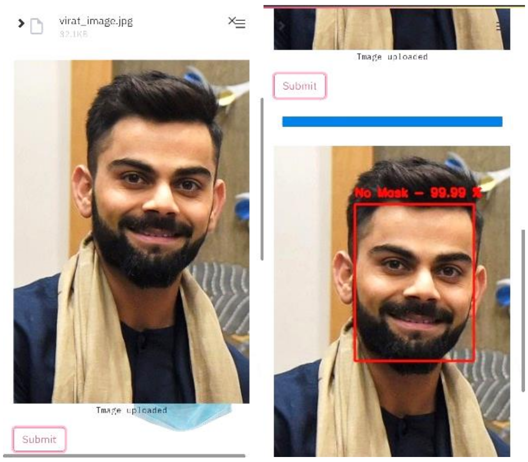

Figure 16. Detection on image using mobile device

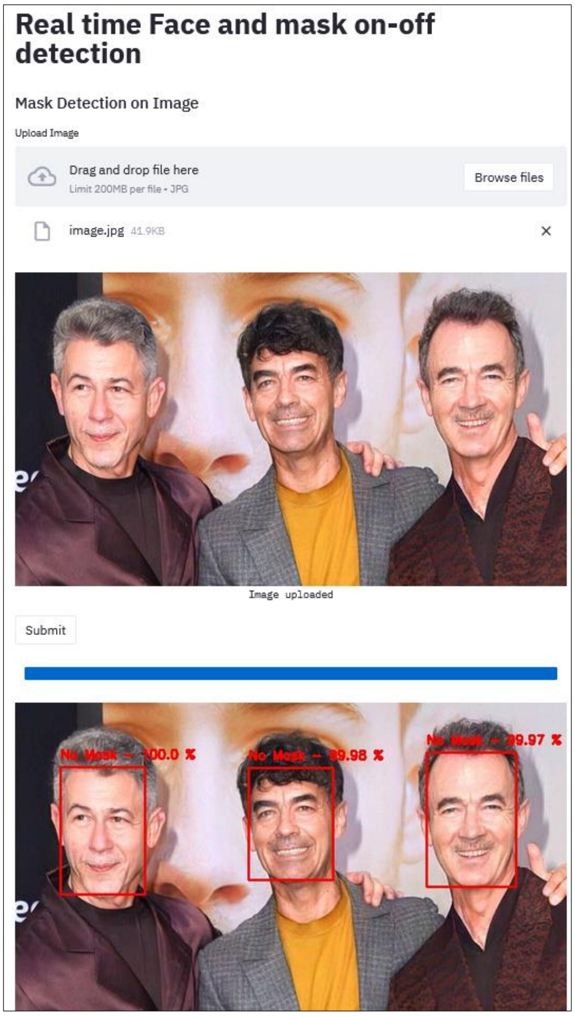

Figure 17. Face and Mask detection in GUI

GUI provides 4 choices –
Image (for mask detection on images) which requires user to upload an image on which classification is to be performed.
Video (for mask detection using webcam stream) which captures on the real-time video stream and performs the detection. From the video stream, user can save a frame by pressing 's' and display any frame at any given point in time by pressing 'Spacebar' key.
Performance – classification report and plot visualizations.
Information – Information about project.

10 Conclusion

From the results, it is clear that the model has performed well with an accuracy of 95%. Model performed well on live real-time video streams as well as on images shown figures mentioned in result subsection. It also works well with multiple faces and faces detected with spectacles. The model MobileNetV2 is lightweight, hence can be easily integrated with video surveillance systems such as CCTV cameras or mobile devices as a complete mask surveillance system under COVID-19 Protocols.

Another interesting observation from the results was found out that the model is free from any biases based on ethnicity. From the samples presented in the result subsection, we can see that the model is able to perform well for people with different ethnicity with similar accuracy. Problems such as gender and race discrimination is completely absent from the system.

References

1. H. Li, et al., in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 5325–5334.
2. N.C. Ristea, R.T. Ionescu, *Are you wearing a mask? improving mask detection from speech using augmentation by cycle-consistent gans*, arXiv:2006.10147 (2020).
3. Z. Wang, et al., *Masked face recognition dataset and application*, arXiv:2003.09093 (2020).
4. M. Inamdar, N. Mehendale, *Real-time face mask identification using Facemasknet deep learning network,* SSRN (2020)
5. P. Khandelwal, et al., *Using computer vision to enhance safety of workforce in manufacturing in a post covid world*, arXiv:2005.05287 (2020).
6. M. Jiang, X. Fan, Retinamask: *A face mask detector*, arXiv:2005.03950 (2020)
7. B. QIN, D. LI, *Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19* (2020).
8. Dataset, by Chandrika Deb.
9. *Building powerful image classification models using very little data,* The keras blog (2016).
10. Face detection using Caffe ResNet10.
11. Face mask detection using MobileNetV2.
12. Face detection identification.