2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 978-1-7281-7741-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACITE51222.2021.9404731

Face Mask Detection System using CNN

Sneha Sakshi
School of Computer Science and Engineering
Galgotias University
Greater Noida, India
snehasdf@gmail.com

Ajay Kumar Gupta
School of Computer Science and Engineering
Galgotias University
Greater Noida, India
ajaykumargupta832@gmail.com

Sudeept Singh Yadav
School of Computer Science and Engineering
Galgotias university
Greater Noida, India
sudeept999@gmail.com

Uttam Kumar
School of Computer Science and Engineering
Galgotias University
Greater Noida, India
uttamsharma289@gmail.com

***Abstract*— Amid current pandemic, Covid-19 has made us realize the importance of Face Masks and we need to understand the crucial effects of not wearing one, now more than ever. Right now, there are no face mask detectors installed at the crowded places. But we believe that it is of utmost importance that at transportation junctions, densely populated residential area, markets, educational institutions and healthcare areas, it is now very important to set up face mask detectors to ensure the safety of the public. In this paper we have tried to build a two phased face mask detector which will be easy to deploy at the mentioned outlets. With the help of Computer Vision, it is now possible to detect and implement this on large scale. We have used CNN/ MobileNet V2 architecture for the implementation of our model. The implementation is done in Python, and the python script implementation will train our face mask detector on our selected dataset using TensorFlow and Keras. We have added more robust features and trained our model on various variations, we made sure to have large varied and augmented dataset so that the model is able to clearly identify and detection the face masks in real time videos. The trained model was tested on both real-time videos and static pictures and in both the cases the accuracy was more than the other designed models.**

Keywords—Face mask detection system, computer vision, deep learning.

I. INTRODUCTION

In the last few years, we have seen Science and Technology advancing so much that now we are at a stage where, we know that with the right knowledge of the technology, the humans can achieve things that seemed nearly impossible just a few decades ago. Now, we have the advancing technologies and knowledge of Machine Learning and Artificial Intelligence, which has been proven to ease our lives from the micro levels to big impossible tasks. In the last few years, there has been a rise in the onset of algorithms that have been proven to be the solution to our complex, life threatening problems. One such field is the image and object detection, which has helped us find and spot people and things with just one click.

Computer Vision plays a crucial role in our lives now. Who would have thought that while sitting in one city you can easily spot the people in the other cities? It’s almost unimaginative how Computer vision is now a very innovative aspect of the technology.

In 2019, the whole world witnessed the onset of the deadly Corona Virus, which now, still after almost a year has not left us and is still making the human race fight for its existence. In between the survival fights, we have realized how technology is very much our only life saver. From extensive internet facilities to 24/7 services online, technology has been our true companion in these hard times.

But even when we have everything present at one click, there can’t be no lives outside. In the past few months every country, every state has found its own new norms to fight the pandemic. And no matter what we do, we do need to step outside to survive. Schools, Offices, Colleges, Markets, Transportation, are the few crucial check points for any country. As much as we ask the public to be safe, the people miss their [without any restrictions lives. And so, it is now very important to closely watch the public and make them understand the importance of the tiny and small details of survival kit.

One such crucial factor is the extensive usage of face masks in our lives. Studies have proven that with the help of use of face masks, we can lower the chances of catching the Corona Virus by 80 to 85%, if it’s used properly. But, even so, it is nearly impossible to enforce the face masks completely on the human race.

With the help of AI and Computer Vision, we have the best chance at enforcing the mask policy on the humans. With the help of our system, we aim on detecting the presence of face masks on static images and real time videos. Object detection, Classification, Regression, image and object tracking and analysis are our key aspects of the paper. We are aiming at a two phased CNN face mask detector. The first phase is the training phase wherein we have trained our model and the second phase the application, where the masks are detected with “with” or “without masks” tags. Other than the images we also aim to implement this on the real time videos, where the real time faces are detected, tracked and the data about the faces with or without masks is returned.

Our paper can be of crucial help at the Stations, airports, Markets, Hospitals, Offices, Schools and many more, where the crowd can be monitored in real time.

II. LITERATURE REVIEW

The existing models have used deep learning but they lack the variation in the dataset which means that their model is not that efficient when it comes to real time images and videos. Deep learning technique has been useful for big data analysis and has its applications in computer vision, pattern and speech recognition, etc. Liu’s et al. [7] work focuses on some commonly implemented deep learning architectures and their applications. Deep learning can be used in unsupervised learning algorithms to process the unlabeled data. A CNN model for speedy face detection has been introduced by Li et al. [8] that evaluates low resolution an input image and discards non-face sections and accurately processes the regions that are at a greater resolution for precise detection.

Our model is a trained custom deep learning and computer vision model which can detect if a person is wearing a mask or not. Our model has not used morphed or unreal masked pictures in the dataset. Our model is very accurate as we have used MobileNetV2 architecture, it has made the model computationally efficient too. This made it easier to deploy the model to embedded system. We can use this face mask detection system in places that require face mask detection in view of the current pandemic. The model can be deployed at Airports, Railway Stations, Offices, Schools and other public places.

III. Scope

Our face mask detection system using deep learning and computer vision has various stages. The development of the model goes through the following stages as shown in figure 1:

A. *Collection of Data*

In this stage the data is collected as to train our model for correctly detecting the presence of face masks. This source of data is easily accessible and can be found on any open source platform like Kaggle.

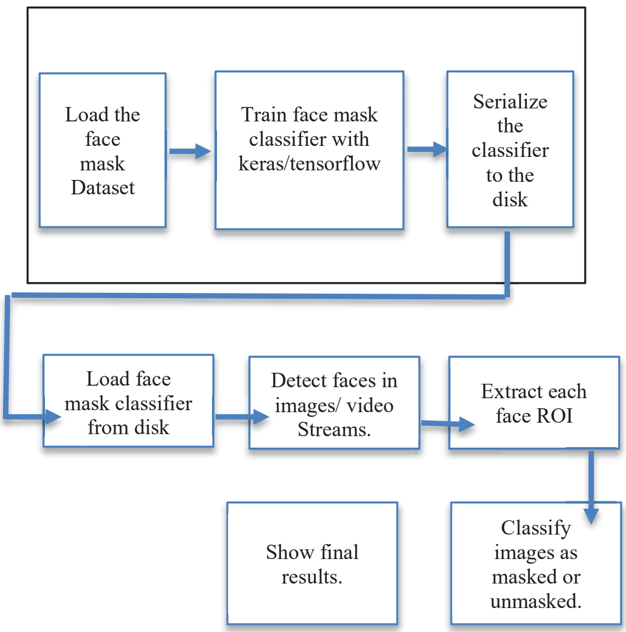
B. *Training*

In this phase we have loaded our dataset from the disk, and by using keras/TensorFlow, then we have trained our model on the selected dataset. After that we have serialized the detector to the disk.

C. *Deployment*

After our detector is trained, we will then load our face mask detector for applying the trained model on the dataset, then after detection, we will classify the faces “with_mask” or “without_mask”.

Phase 1: Training the face mask detector



Phase 2: Applying the face mask detector

Fig 1: Phases of the model

IV. Proposed Methodology

Wang et al. has made executing face mask related models an obvious task by providing three samples of masked face datasets, which comprise of Masked Face Detection Dataset (MFDD), Real world Masked Face Recognition Dataset (RMFRD) in figure 3 and Simulated Masked Face Recognition Dataset (SMFRD) [11]. Matthias et al. has done a face mask recognition model that focusses on capturing real-time images indicating whether a person has put on a face mask or not [12].

A. *Architecture Overview*

The architecture consists of two stages shown in figure 2. Stage 1 of the model is face detector, which can localize several faces in the pictures of different size and type, and it can detect faces even if there is overlapping. Convolutional Neural Network (CNN) is a structured deep learning process that plays a ground-breaking push for a variety of applications focusing on computer vision and image-based applications [14]. So we have used CNN algorithm to detect the face masks on the faces detected in stage 1. The results of this stage are sent to the further stages and the final image that is our output is classified as masked or unmasked faces, for all the faces in the input image [2]. Using 3 different image classification methods, our CNN based classifier was trained [9]. For training a custom face mask detection system, we have a two staged architecture, and each stage with several other steps.

B. *Stage 1-Training the model*

By using keras/Tensorflow, we have trained our model on the selected dataset. The dataset is loaded from the disk and then after importing all the necessary packages, an argument parser is constructed to parse the arguments from command line. All the data and the labels are then converted into numpy arrays and then a training image generator is constructed for data augmentation. Later we constructed the head of our model that is placed on the top of the base model. After compilation and training, the mask detector model is saved.

C. *Stage 2- Deployment*

After our detector is trained, we then loaded our face mask detector and applied the trained model on the dataset, then after detection, we have classified the faces “with_mask” or “without_mask”.

After the successful deployment we were able to detect the faces in real time videos and images, with tags in green and red colour, named “with_mask” and “without_mask”.

V. System Architecture

The face mask detection system is provided with an Input image as shown in figure 2, where there are people with and without masks and as the first stage, the faces in the input image are detected with the help of the facial landmarks. We start with the extraction of the regions of interest of each face and then with the help of that we apply the facial landmarks. Later, we apply on the fly mutations in order to improve generalization. After the stage 1, the ROI Extraction and Batching and Resizing phase comes under the Intermediate Processing between the Stage 1 and Stage 2.

As mentioned earlier, for ROI extraction, the images need to be cropped and expanding the width and height by about 20% to ensure that we cover the right facial region. To prevent the overlapping of faces there needs a resizing of the images and they are all batched together as per the requirements. As the process moves ahead, the extracted ROI's of each detected face is processed and sent to the classifier in batches. In the stage 2, the processed images are classified as masked or unmasked. The final output image, comes with the tags for all masked and unmasked faces in the input image. The whole process is explained in figure 2.

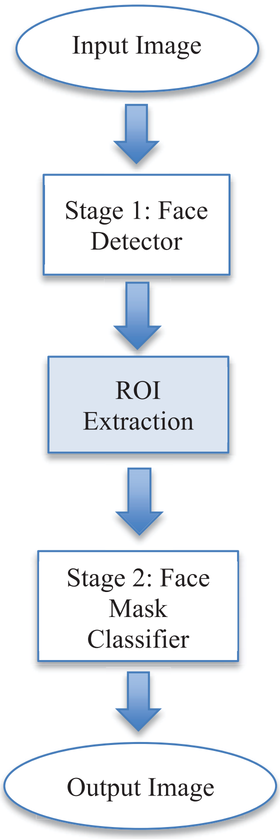


Fig 2: System Architecture

VI. DATASET

The reproductive number of COVID-19 is higher compared to the SARS [4]. Therefore, more and more people are concerned about their health, and public health is considered as the top priority for governments [5]. Fortunately, it showed that the surgical face masks could cut the spread of coronavirus [6].

The dataset of all the faces with mask and without mask have been taken from the pictorial datasets available, as well as a few of the images were scraped from the Internet in figure 4 and figure 5. Most of our masked pictures are from the Real-world Masked Face Recognition Dataset and Face Mask Detection dataset. Previously, Khandelwal et al. had stated in his work about a deep learning model that binarizes an image as a mask is used or not mask [3]. Our dataset [1] also has images of imperfectly worn masks or the one where hands are hiding the face, which we have marked as non-masked faces in figure 6. Li et al. worked on developing a HGL method for head pose classification with masks, using color texture analysis of pictures and line portraits in figure 7 [10].

Fig 3: RMFRD with mask Images[2]

Fig 4: Kaggle with mask Images[1]

Fig 5:Flickr without mask Images [13]

Fig 6: Inappropriately worn with mask pictures[13]

Fig 7: Processing of Raw Images

VII. Final Results

We have trained the model into a two-classes, one where people are wearing masks and other where people are not wearing masks. We fine-tuned the MobileNetV2 architecture on model and used a very varied dataset which helped us in obtaining a classifier that is about 99% accurate shown in figure 8.

The classifier was then applied to images and real time videos by performing the following steps:

- Face detection in images and video
- Extraction of individual faces
- Applying the face mask classifier

The use of MobileNetV2 architecture has made our model accurate and computationally efficient in figure 9, so that we can easily deploy it on embedded systems.

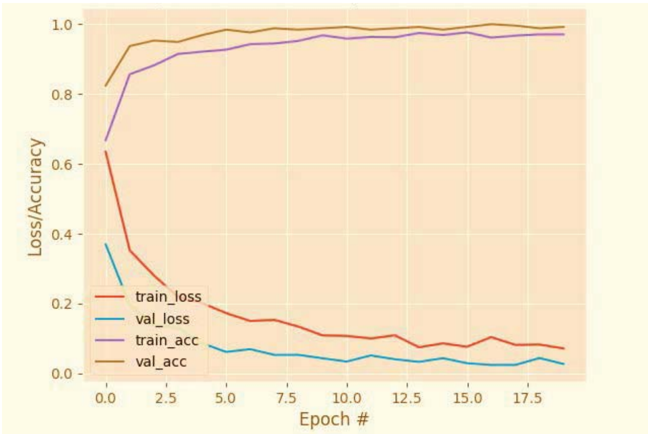


Fig 8: Our face mask detector obtained 99% accuracy

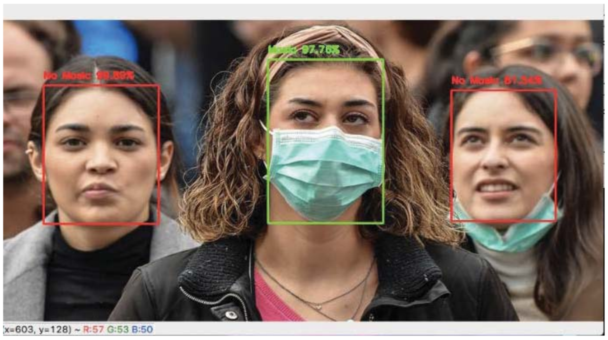


Fig 9: Detection of mask in real time

Table 1. Training Statistics of our Face Mask Classifier

Chosen Model	Training results		Validation results		Accuracy (test) (%)
	Acc(%)	*Loss*	*Acc(%)*	*Loss*	
NASNetMobile	99.65	0.00181	99.35	0.0291	99.22
DenseNet121	99.48	0.0155	98.72	0.0317	99.35
MobileNetV2	99.85	0.0013	99.47	0.0178	99.48

Based on Table 1, it is clear that all the 3 mentioned models have maintained pretty good statistics. But our MobileNetV2 model has shown overall better results as compared to the other models.

VIII. CONCLUSION

With the rise and bloom in the technology, the emerging trends and new techniques has now made it possible for us to test the presence of face mask in reality. With the help of this technology, we are able to contribute in public healthcare and welfare. Our face mask detection system is the result of the blooming technology. With the help of the MobileNetV2 architecture, we have succeeded in making the more efficient and accurate face mask detection model, which we believe can be used in various high and low computational scenarios. We have added more robust features and trained our model on various variations, we made sure to have large varied and augmented dataset so that the model is able to clearly identify and detection the face masks in real time videos. We have had several bumps in the road, but eventually we had better results as compared to many other technologies. By using Tensorflow/Keras, OpenCV, CNN and Pytorch we were able to finely detect the presence or absence of masks. The trained model was tested on both real-time videos and static pictures and in both the cases the accuracy was more than the other designed models. We tried to work more on the efficiency and optimization of the model, where in tuning the hyper parameters came to great help. This model can be of great help to our society and help the nation in fighting the pandemic as well.

IX. FUTURE SCOPE

The paper may be of important use at varied packed places like airports, railway stations, markets, offices, schools, colleges' etc. To beat the most common limitations we are able to do the subsequent changes: since there is very little clarity in the pictures from the still frames, so what we tried to do was, that we took various frames from a particular video or picture, tried to view it from more other angles, only to create a much clearer picture of the image.

Well, there are several entities in most of the frames that are merely even moving and hence we can say that they are nothing but constant in all those frames. Now for a person to encounter constants in all the videos and pictures is merely even possible. But, since we are using a machine, we can consider the possible constants and try to do this, if we perform pursuit.

Using pursuit, we are able to cluster various faces that corresponds to a constant or a single person across and in multiple frames. Using this, we can run a classifier for every single picture, and then we can mix results in a single decision, that whether there is a mask present or not. Using this, we will be able to count the number of folks and can collect a good amount of valuable statistics, that's without delay. We can work on other alternatives as well to find and estimate the total fraction of people wearing and carrying masks while they are out. We can try to sample some random frames throughout the given videos, and can classify all the detected faces to average the results into various frames. In reality this approach might not allow and permit us to count the number of folks passing by in each video, and might also be not really efficient if considered as an information, but we are going to a very small fraction of these frames, or else we would eventually end up counting the same constant people again and again.

REFERENCES

[1] Karras, T., Laine, S., and Aila, T., "A Style Based Generator Architecture for Generative Adversarial Networks," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 4396- 4405, doi: 10.1109/CVPR.2019.00453.

[2] Larxel, 2020 May, "Face Mask Detection", *Kaggle - Version 1*.

[3] P. Khandelwal, A. Khandelwal, S. Agarwal, "Using computer vision to enhance safety of workforce in manufacturing in a post covid world," *arXiv preprint arXiv*:2005.05287 (2020)

[4] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of covid-19 is arehigher compared to sars coronavirus," *Journal of travel medicine*, 2020.

[5] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis," Journal of medical virology, vol. 92, no. 6, pp. 645–659, 2020.

[6] N. H. Leung, D. K. Chu, E. Y. Shiu, K.-H. Chan, J. J. McDevitt, B. J. Hau, H.-L. Yen, Y. Li, D. KM, J. Ip et al., "Respiratory virus shedding in exhaled breath and efficacy of face masks," *Nature Medicine*, 26 pages676–680(2020).

[7] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing* 234, 11 (2017).

[8] H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, in "Proceedings of the IEEE conference on computer vision and pattern recognition (2015)," pp. 5325–5334.

[9] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L., "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition,* Salt Lake City, UT, 2018, pp. 4510- 4520, doi: 10.1109/CVPR.2018.00474.

[10] S. Li, X. Ning, L. Yu, L. Zhang, X. Dong, Y. Shi, W. He, "Multi-angle Head Pose Classification when Wearing the Mask for Face Recognition under the COVID-19 Coronavirus Epidemic" in 2020 *International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)* (IEEE, 2020), pp. 1–5.

[11] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, et al., "Masked face recognition dataset and application," *arXiv preprint arXiv:2003.09093* (2020).

[12] D. Matthias, M. Chidozie, Face mask detection application and dataset.

[13] FFHQ Dataset, Karras et al., 2019.

[14] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture.,* vol. 147, no. July 2017, pp. 70– 90, 2018.