

Face Mask Detection using Mobilenet Technique

Shivam Sourav

BCA, CSE Deptt

Institute of Information Technology
& Management

Janakpuri, New Delhi

Shikhar Mattoo

BCA, CSE Deptt

Institute of Information Technology
& Management

Janakpuri, New Delhi

Chitra Nasa

A.P, CSE Deptt

Institute of Information Technology
& Management

Janakpuri, New Delhi

ABSTRACT

As the rise of Corona-virus has affected at least 188 countries in the world as of 2021 period [1], the outbreak has severely affected all kinds of age groups, so wearing a mask becomes utmost responsibility of the person to stop the outbreak from further spreading, but it is impossible to monitor all the people who are wearing mask and who are not.

So here comes the Novel Face Mask Detector which is a high accuracy mask detector. It uses Mobile-Net, which is a high accuracy and efficient face detection mask detection. Here Mobile-Net is used for pre-processing of the image, which consists of a feature pyramid network to fuse high-level semantic information with multiple feature maps, and a novel context attention module to focus on detecting face masks. In addition, a novel cross-class object removal algorithm is used to reject predictions with low confidences and the high intersection of union. When the image is processed as array, the array is forwarded to the Mobile-Net after this, the process of max pooling is done then the array is flattened and through the fully connected layer the output is given. As, Mobile-Net is the light-weighted neural network to be used in the mobile phones.

Keywords

Corona-Virus, Mobile-net, Novel face Mask Detector, CNN (Convolution Neural Network)

1. INTRODUCTION

The 2021 WHO report shows us there has been 144,358,956 confirmed cases from which 3,066,133 are dead [2], due to this wearing a mask is very important as it block 95% of airborne droplets and aerosols which can stop the spread of corona-virus to 1%. [3]. As the second wave is here and has almost affected almost all the countries it would be more deadly and will spread more quickly because of double mutant virus, Till now India has been worst hit by the second wave with no beds no oxygen supply [4]. So it becomes utmost responsibility of every person to wear a mask on their face to stop this spread. The authorities need a solution to be able to validly control the implementation of the law, which begins with the availability of the data quickly and accurately. One of the solutions is to use a regionally automated face mask recognition to differentiate between people who wear masks and those who do not [8]-[10].

The face mask detector detects whether a person wearing a mask or not and what is the location of the face. The problem is closely related to general object detection to detect the classes of objects and face detection is to detect a particular class of objects, i.e. face. Face detectors are already been used in variety of other applications like in Phone Unlock, aid forensic investigations, validate identity at ATM's and so on [5]. Face detectors are used since 2001 widely with the development of Viola-Jones Framework which detected faces

in real-time and with high accuracy, but with the drawbacks of the Viola-Jones Framework such as no detection of face in mask/scarf, convolution neural network (CNN) and Single Shot Detector (SSD) -- have been developed to help improve processes [6]. These One-stage detectors use a single neural network to detect objects. In contrast, two-stage detectors utilize two networks to perform a coarse-to-fine detection, such as region-based convolution neural network (R-CNN).

In this paper, a face mask detector is proposed, which is able to detect face masks and contribute to public healthcare. In terms of the network architecture, The Face Mask Detector uses Mobile-Net which is a CNN architecture model for Image Classification and Mobile Vision it uses very less computation power to run or apply transfer learning to. This makes it a perfect fit for embedded systems and computers without GPU or low computational efficiency with compromising significantly with the accuracy of the results [7]. Mobile-Net is a part of CNN architecture but in this it is replaced with the convolution layer the rest remained the same, after that the max pooling of the input is done, the input is then flattened and through a fully connected layer the output is given. The proposed method is tested on a face mask dataset. The 20% of the images is for testing set and rest 80% is used for training purpose, examples shown in Fig 1. The dataset covers a various masked or unmasked faces images, including faces with masks, faces without masks, faces with and without masks in one image and confusing images without masks.

The Layout of this paper is as follows: In Section II, review related works on object detection and neural networks is shown. The proposed methodology is presented in Section III. Section IV describes datasets, experiment settings, evaluation metrics, results, and ablation study. Finally, Section V concludes the paper and discusses the future work.

2. METHODOLOGY

With time Deep CNN have become important tools for many computer vision related task like classification of images. The face mask recognition system in this study is developed using machine learning algorithm through the image classification method: MobileNetV2. MobileNetV2 is a method based on CNN that developed by Google with improved performance and enhancement to be more efficient and faster than CNN.

Christian *et al.* [11] have derived an inception network to achieve better accuracy for image classification and segmentation. In general, Convolution Neural Network (CNN) with large filter tends to high computational cost. One prominent solution to reduce this cost is inception modules. Inception reduces the cost by finding out optimal local sparse structures. The idea of the inception block is to design a layer by layer construction with the analysis of layer correlation statistics. The clusters of highly correlated layers are used to form groups of units. Each unit from an earlier layer corresponds to some region of the input image is referred to as

a filter bank. This process sends up with the concatenation of huge filter banks from a single region.

This Study conducted its experiment over two original dataset. The first dataset was taken from Kaggle, . This data set were used for the training, testing, and in validating the model so, that it can be implemented on the dataset. This model can be produced by following some of the steps 1) Data Collection 2) Data Preprocessing 3) Split dataset into training and testing

dataset 4) building a model 5) testing the model and 6) Implementing the model. The detailed step process is shown in the Figure 2.

The second dataset is used to apply the model to the dataset which is collected from *Github* to cross check the model whether the model is detecting the faces with its most achieved accuracy.

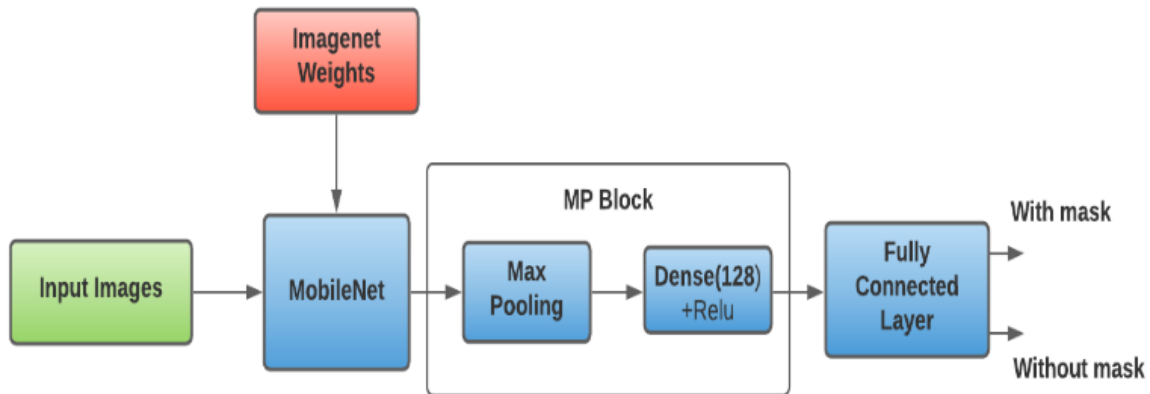


Fig1. Proposed Mobile-Net and Max-Pooling

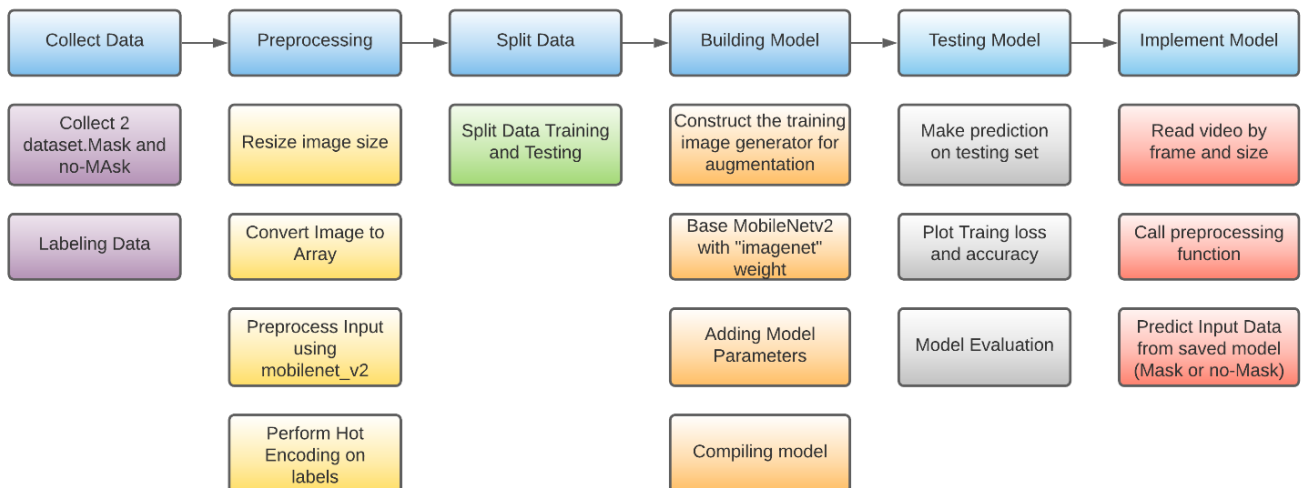


Fig2. Steps in building model

3. RESULT AND DISCUSSION

3.1 Data Collection

The Development of this model begins with the process of collecting the data. This dataset is used to train the model so, that the model can easily distinguish between the people having mask on and who is not wearing the mask.

To build this model, this study uses 3,210 images with mask and 3,325 images without the mask. All the images are cropped until the only visible thing is the face of a person. The next step is to label the data. The data which has been collected are labelled into two categories with mask and without mask. The illustration of the data is in Figure 3 labelling with mask and in Figure 4 labelling without mask.

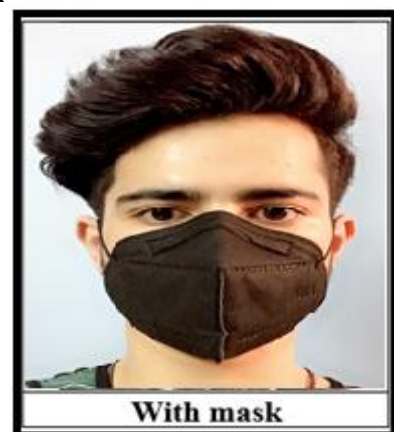


Fig3: Person With mask



Fig4: Person Without mask

3.2 Pre-Processing

The pre-processing is a phase before the training and testing of the data. In the pre-processing there are four steps which are a) Resizing of train image and test images, b) converting the images into array, c) pre-processes the data input using MobileNetV2, d) last step is labelling the data into two categories.

Resizing of the image is one of critical pre-processing step in the field of computer vision. Smaller the size of the image better it become for the training of the model, it also increases the accuracy of the model. In this study, loading of the image is done using `target_size()` the image is resized to (224 X 224) pixels.

After resizing the image into the desired size, the data input is considered size the data is to be converted into the numerical form that is into arrays by using loop function. After, converting the data into the array the data is ready to be pre-processed using the MobileNetV2.

And the last step in this phase is performing hot encoding on labels because many machine learning algorithms cannot operate on data labeling directly. They require all input variables and output variables to be numeric, including this algorithm. The labeled data will be transformed into a numerical label, so the algorithm can understand and process the data.

3.3 Splitting the Dataset

Next step after the pre-processing is splitting the dataset into testing and training dataset. In this, 80% of data is for training the model and rest of the 20% has been taken for testing the model.

3.4 Building the Model

In building the model there are two steps in first new dataset is created using the existing dataset with the help of the image data generator and in second using MobileNetV2 the model is created. In-depth knowledge of both the step is as follows:

3.4.1 Image Data Generator

Before start building the model, there is a small step which is making the every image of dataset to be understood by the model in many different ways using the Image Data Generator. Data augmentation is a technique to artificially create new training data from existing training data. This is done by applying domain-specific techniques to examples from the training data that create new and different training examples. Image data augmentation is perhaps the most well-known type of data augmentation and involves creating transformed versions of images in the training dataset that belong to the same class as the original image. Transforms include a range of operations from the field of image manipulation, such as shifts, flips, zooms, and much more.

3.4.2 MobileNetV2

After making another dataset using the image data generator, the *base model* is created using *MobileNetV2*, after all the parameters is added to the model for the compilation and training of the model. After the compilation and the training the model is saved for the future analysis to detect whether a Face Mask was wore by the person or not.

3.5 Testing of Model

After Constructing all the models and evaluating the training results of model is shown in the Table 1 which shows the result of trained dataset that includes its accuracy and loss etc. From Table 1, it is clear that the accuracy of this model is exceptionally good as the accuracy is almost equals to 1 and the loss of the model tends to decrease from the second iteration and by the 20th iteration the loss came almost equal to 0.

After training of the model, now it's time to test the model whether it is working properly or not. To do so, the test dataset is used to check the accuracy of the model. The first step is to make the prediction over test dataset.

Considering this table it is supposed that the model does not need more iteration for testing and the model is ready for evaluation. The result of the evaluation of the model is shown in Table 2.

Table1. Loss and Accuracy of the model

Epoch	Loss	Accuracy	Val Loss	Val Accuracy
1/20	0.3520	0.8435	0.1536	0.9535
2/20	0.1763	0.9302	0.1136	0.9674
3/20	0.1436	0.9441	0.1006	0.9657
4/20	0.1336	0.9503	0.0924	0.9649
5/20	0.1270	0.9511	0.0872	0.9674
6/20	0.1135	0.9577	0.0849	0.9682
7/20	0.1055	0.9585	0.0811	0.9706
8/20	0.0956	0.9624	0.0801	0.9715
9/20	0.0947	0.9630	0.0793	0.9723
10/20	0.0900	0.9639	0.0740	0.9731
11/20	0.0825	0.9671	0.0726	0.9723
12/20	0.0845	0.9645	0.0778	0.9723
13/20	0.0784	0.9698	0.0738	0.9739
14/20	0.0730	0.9706	0.0686	0.9747

15/20	0.0778	0.9690	0.0685	0.9747
16/20	0.0699	0.9725	0.0678	0.9739
17/20	0.0711	0.9708	0.0699	0.9755
18/20	0.0681	0.9745	0.0656	0.9772
19/20	0.0642	0.9733	0.0661	0.9771
20/20	0.0681	0.9762	0.0676	0.9788

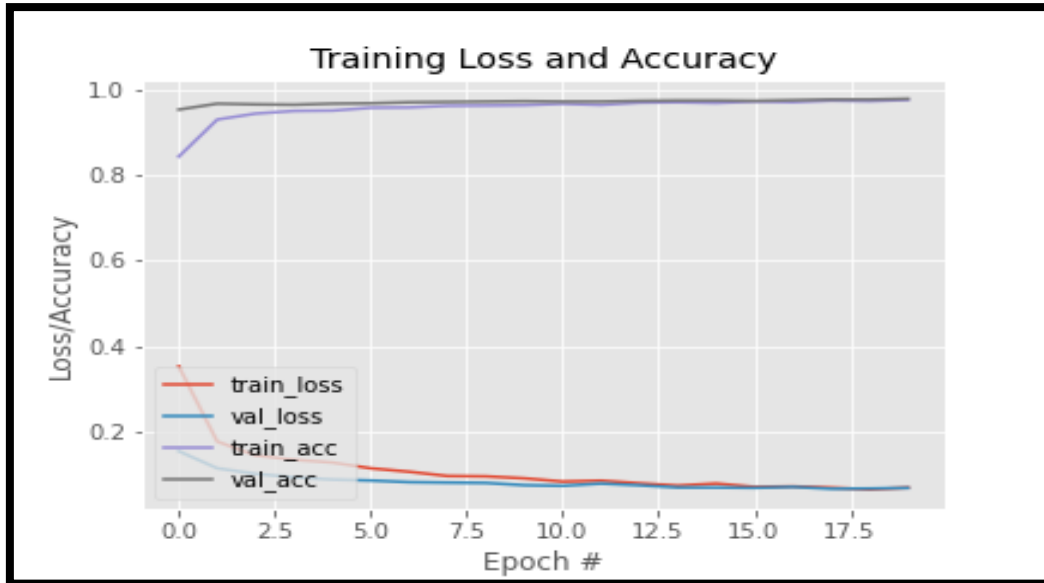


Fig 5: Graph of training loss and accuracy

Table2. Evaluation of the Model

	Precision	Recall	F1-Score	Support
With Mask	0.98	0.98	0.98	641
Without Mask	0.97	0.98	0.98	585
Accuracy	-	-	0.98	1226
Macro Avg.	0.98	0.98	0.98	1226
Weighted Avg.	0.98	0.98	0.98	1226

3.6 Implementation of the model

The model will be applicable to the real time video using the camera. So, first of all the camera will start capturing the video frame by frame, on that the face detection will start working. In that frame if the face is detected by the algorithm them it will proceed to the next step which capturing the image, resizing the image, converting image to array, pre-processing the input using data image generator and MobileNetV2.

After pre-processing the data the next step is to predict the image using the model. The model will predict after using the saved model whether the person in the input is wearing the mask or not and it did the same by labeling the ongoing-video stream. The demonstration of implementation of model is shown in Figure 6 and Figure 7.

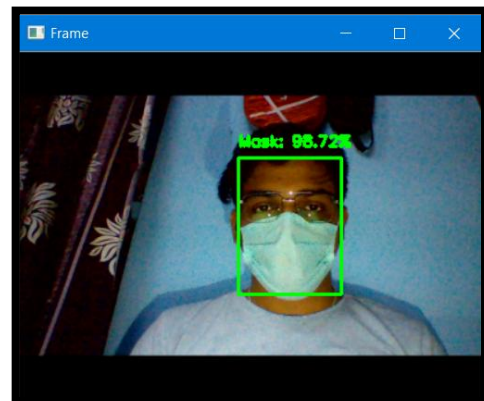


Fig 6: Implementation of Model

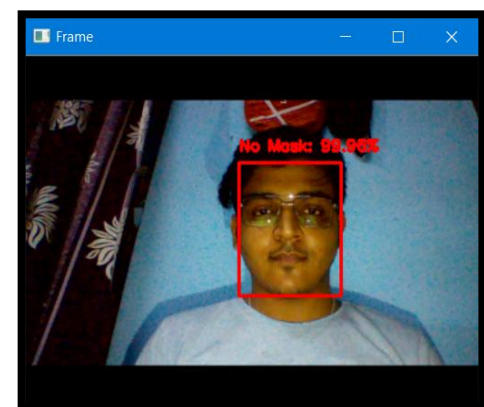


Fig 7: Implementation of Model

4. CONCLUSION

In this paper, the Novel Face Mask Detector is proposed, which can possibly contribute to public healthcare. The

architecture of Novel Face-Mask Detector consists of MobileNet as the backbone. After the training, validation, and testing phase, the model can provide the percentage of people using face mask in some cities with high accuracy. MobileNet can be used for high and low computation scenarios, respectively. Face detection and recognition method's result summary is provided in Table 1 and Table 2 respectively. The proposed method achieves state-of-the-art result accuracy of almost 98.9%.

5. FUTURE SCOPE

The proposed system can further be projected in public areas as well as in government and private departments where wearing a face mask is very important. Furthermore, this will allow us to have more control over public and ensuring that people must wear face mask. This method will help us to prepare for current pandemic circumstances where wearing a face mask is much more essential.

6. REFERENCES

- [1] Coronavirus: Which countries have confirmed cases? | Coronavirus pandemic News | Al Jazeera
- [2] WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data
- [3] How Face Masks Can Help Prevent the Spread of COVID-19 | The Scientist Magazine® (the-scientist.com)
- [4] How India's second wave became the worst COVID-19 surge in the world (nationalgeographic.com)
- [5] 21 Amazing Uses for Face Recognition – Facial Recognition Use Cases (facefirst.com)
- [6] What is Face Detection and How Does It Work? (techtargget.com)
- [7] Image Classification using MobileNet in the browser | by Mithilesh Pradhan | Analytics Vidhya | Medium
- [8] Ejaz, M. S., Islam, M. R., Sifatullah, M., & Sarker, A., "Implementation of Principal Component Analysis on Masked and Non masked Face Recognition". 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 1–5.
- [9] Hussain, S. A., & Al Balushi, A. S. A. "A real time face emotion classification and recognition using deep learning model". Journal of Physics: Conference Series, 1432(1), 12087, 2020.
- [10] Qin, B., & Li, D. "Identifying Facemask-wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19", 2020.
- [11] C. Szegedy, V. V. and Sergey Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol.abs/1512.00567, 2015