Face Mask Detection

By: Denesh Kumar Manoharan
May 6,2022

Contents

1. Abstract:

COVID-19 pandemic has rapidly affected our day-to-day life disrupting the world trade and movements. Wearing a protective face mask has become a new normal. In the near future, many public service providers will ask the customers to wear masks correctly to avail of their services. Therefore, face mask detection has become a crucial task to help global society. This project presents an approach to achieve this purpose using some Machine Learning packages like TensorFlow, Keras, OpenCV and Scikit-Learn. The steps for building the model are collecting the data, pre-processing, split the data, testing the model, and implement the model, the method detects the face from the image correctly and then identifies if it has a mask on it or not.

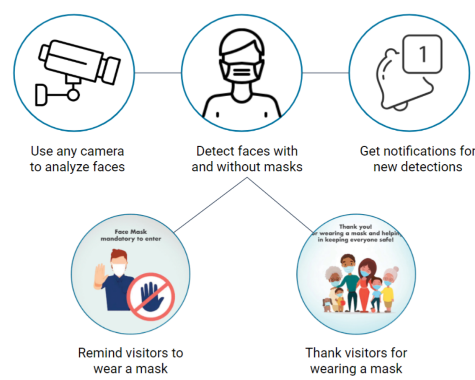
2. Introduction:

A new strain which has not previously been identified in humans is novel coronavirus (nCoV). Coronaviruses (CoV) are a wide group of viruses which cause illness that range from colds to deadly infections like Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). The first infected patient of coronavirus has been found in December 2019. From that period, COVID-19 has become a pandemic all over the world. People all over the world are facing challenging situations due to this pandemic. Every day a large number of people are being infected and died. At the time of writing this paper, almost 16,207,130 infected cases have been confirmed where 648,513 are death. This number is increasing day by day. Fever, dry cough, tiredness, diarrhea, loss of taste, and smell are the major symptoms of coronavirus which is declared by the World Health Organization (WHO). Many precautionary measures have been taken to fight against coronavirus. Among them cleaning hands, maintaining a safe distance, wearing a mask, refraining from touching eyes, nose, and mouth are the main, where wearing a mask is the simplest one. COVID-19 is a disease that spread from human to human which can be controlled by ensuring proper use of a facial mask. The spread of COVID-19 can be limited if people strictly maintain social distancing and use a facial mask. Very sadly, people are not obeying these rules properly which is speeding the spread of this virus. Detecting the people not obeying the rules and informing the corresponding authorities can be a solution in reducing the spread of coronavirus.

A face mask detection is a technique to find out whether someone is wearing a mask or not. It is similar to detect any object from a scene. Many systems have been introduced for object detection. Deep learning techniques are highly used in medical applications. Recently, deep learning architectures have shown a remarkable role in object detection. These architectures can be incorporated in detecting the mask on a face. Moreover, a smart city means an urban area that consists of many IoT sensors to collect data. These collected data are then used to perform different operations across the city. This includes monitoring traffic, utilities, water supply network, and many more. Recently, the growth of COVID-19 can be reduced by detecting the facial mask in a smart city network.
This project aims at designing a system to find out whether a person is using a mask or not and informing the corresponding authority in a smart city network. Firstly, CCTV cameras are used to capture real-time video footage of different public places in the city. From that video footage, facial images are extracted and these images are used to identify the mask on the face. The learning algorithm Convolutional Neural Network (CNN) is used for feature extraction from the images then these features are learned by multiple hidden layers. Whenever the architecture identifies people without face mask. The proposed system appraised promising output on data collected from different sources.

2.1. Goals:

In this project I will be model such than any person who comes in-front of the camera will be detected and will be classified if wearing a mask or not. As In order to achieve this I will be doing Deep learning model and I will be using a pertained model I.e MobileNetV2 to train my model and try to achieve an accuracy of more than 90 percent. Also my model will be able to classify with any color of mask the person wearing.



3. Related Work:

Face mask detection is a subset of object recognition that uses image processing algorithms. Digital image processing may be divided into two broad categories: classical image processing and deep learning-based image analysis. As opposed to classical image analysis, which uses complex formulas to recognize and interpret pictures, deep learning-based approaches. Deep Learning models have been used in the majority of past research. After correctly recognizing the face in the picture or video, the CNN-based approach by Kaur et al. [1] evaluates if the face has been disguised. It is also capable of identifying a moving face and a mask in a video as a surveillance job performance. Accuracy is great

with this method. An algorithm called YOLO-v3 was developed by Bhuiyan et al. [2] to identify face masks in public spaces. They trained the YOLO-v3 model on their own custom dataset of photos with people labeled as “mask and no-mask.” The model's performance was enhanced by Mata [3] via data augmentation. It is necessary to create a CNN model that can distinguish between ROIs with and those without a face mask in order to extract the facial area as a ROI. With the use of Mobile NetV2, Toppo et al. [4] developed a method for detecting face masks that incorporates three distinct face detector models in order to test the model's correctness and evaluate its performance. The trained model's outcome allows for implementation on low-power devices, making the mask detection method's inclusion faster than previous strategies. To recognize people who were not wearing face masks in government workplaces, Balaji et al. [5] utilized a VGG-16 CNN model developed in Keras/TensorFlow and Open-CV to detect people who were not wearing face masks. To compensate for the model's light weight, Fan et al. offered two additional methods. A unique residual contextual awareness module for crucial face mask regions Two-stage synthetic Gaussian heat map regression is used to identify better mask discrimination features. Ablation research has found that these strategies can improve feature engineering and, as a result, the effectiveness of numeric identification. For AIZOO and Moxa3K, the suggested model outperforms prior models.
Conventional deep learning algorithms for lightweight facial image classification alone do not give a good discriminating feature space, as shown by the research covered above, and they complicate the model and greatly increase the number of parameters and necessary computational resources. In this study, a Depthwise Separable Convolution Neural Network-based MobileNet.

3.1. Related Work on the dataset:

A preprocessing phase starts before the training & testing of model. The preprocessing phase comprises of four steps. Those are resizing of image size, transforming the image into to the array, preprocessing the input value with the help of Mobile_NetV2, and the last step is to do hot encoding on labels of images. The rescaling of data images is a very significant preprocessing. If images are smaller in size, then the model will definitely yield better result. In our detector model, we are resizing the images into 224 x 224 pixels.
The process after resizing images will be processed all resized images present in the dataset into an array of further process. After that conversion of images in an array, the images are ready to be used to preprocess input with the help of Mobile_NetV2.
This will be the last step for this pre-processing phase. Many machine learning require all the input and output values to be in numeric format, also including the algorithm used. So in order for the algorithm to understand and process the data accordingly, the labelled

data will have to be converted into a numerical label. So now data will have a numeric label for further process.

3.2 Related work on the Pattern Recognition Approach

In face detection method, a face is detected from an image that has several attributes in it. According to [6], research into face detection requires expression recognition, face tracking, and pose estimation. Given a solitary image, the challenge is to identify the face from the picture. Face detection is a difficult errand because the faces change in size, shape, color, etc and they are not immutable. The learning model is based on CNN which is very useful for pattern recognition from images. The network comprises an input layer, several hidden layers and an output layer. The hidden layers consist of multiple convolution layers that learn suitable filters for important feature extraction from the given samples. The features extracted by CNN are used by multiple dense neural networks for classification purposes. The architecture contains three pairs of convolution layers each followed by one max pooling layer. This layer decreases the spatial size of the representation and thereby reduces the number of parameters. As a result, the computation is simplified for the network. Then, a flatten layer reshapes the information into a vector to feed into the dense network. Three pairs of dense and dropout layers learn parameters for classification. The dense layer comprises a series of neurons each of them learn nonlinear features. The dropout layer prevents the network from overfitting by dropping out units. Finally, a dense layer containing two neurons distinguishes the classes.

4. Data

Face Mask Recognition model development phase initiates with collection of the data. The dataset is then able to train the data for people with or without a face mask. Also it distinguishes the people with or without face mask. To train any model test data and subject data is needed, the model is then trained on the accordance of testing data, and trained model is then applied on subject data. For building the Face Mask Recognition mode, data comprises of 690 images with mask and 686 images. The dimension of the data is [256x256x3]. Data augmentation will be applied such as rotation, Translation, Intensity will be applied to the data to improve the dataset. Then 80-20 rule will be applied to testing and training the model.

Fig1: Dataset to train model

5. Methods:

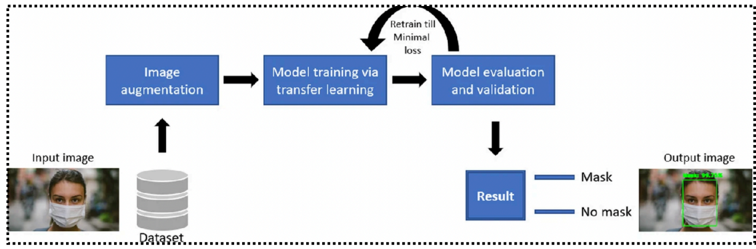


Fig2: project outline

In the initial phase the dataset of people with mask and without mask is created, and data augmentation is applied to the image set, the image is cropped according to the dimensional need for the Neural network. These images are used to train the CNN model based on Mobilenet architecture. We used Depthwise Separable Convolutional Neural Network based on MobileNet architecture, which is an effective method for lowering the computation complexity of deep learning models. It consists of a 1×1 convolution output node with spatial convolution performed independently on each pulse [7]. We utilized the convolution layer's output as a feed to the Rectified Linear Unit (ReLU) activation function, with a 1-dimensional max pooled on the result. The filtering size and depth of the first convolutional layer are both adjusted to 60, whereas the pooling layer's filter size is fixed to 20 with a step number of 2. For the fully linked layer input, the result of the succeeding convolution operation is flattened down to a stepping of six. After obtaining an intake from the max-pooling layer, the convolution layer employs a filter of various sizes, which has 10% of the max-pooling layer's complexity. For the completely connected layer input, the result is smoothed down. The hyperbolic tangent function represents non-linearity in the current layer. To calculate the probability for the corresponding target tags, the Softmax function is utilized. To reduce the potential log-likelihood objective functions, the stochastic gradient descent optimizing approach was utilized. Each functional map's matrix description is converted to a vector via the flattening layer. Several dropouts are mounted on the top of the pooling layer to eliminate the potential for overfitting. The system includes a max-pooling layer that sums the feature maps generated by the convolution layers and reduces computing costs. In order for the Depthwise Separable Convolution Neural Network (DS-CNN) to operate, the volume of the function mappings must always be reduced, along with their size. In the proposed model i.e MobileNetV2 the last layer was removed and is

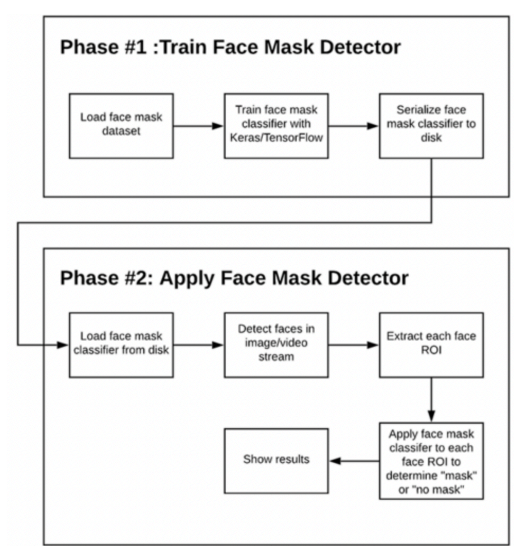


Fig3: Schematic representation of proposed work

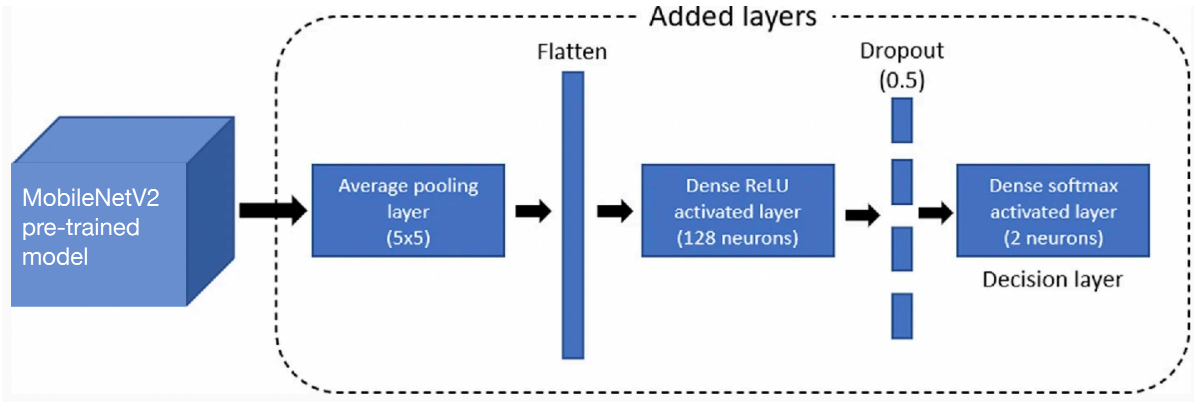


Fig4: Architecture of proposed model

fine tuned by adding 5 more layer to the network. The 5 layers that are added are average pooling layer with pool size of 7 x7, a flattening layer, followed by dense layer of 128 neuron with ReLU activation function and a dropout rate of 0.5 and a final decision layer with two neurons and softmax activation function is added to classify whether a person is wearing mask or not.

5.1 Steps to achieve the goal

Step 1: Collect the dataset required for Training model

Step 2: Implement Image Augmentation on dataset to improve the dataset

Step 3: Train CNN model to for face mask classifier

Step 4: Save model with great classification rate

Step 5: Detect Images from videos stream

Step 6: Extract each face Region of interest

Step 7: Apply Face mask classifier on Region of interest to determine mask or not

Step 8: calculate accuracy of the classifier from the test data

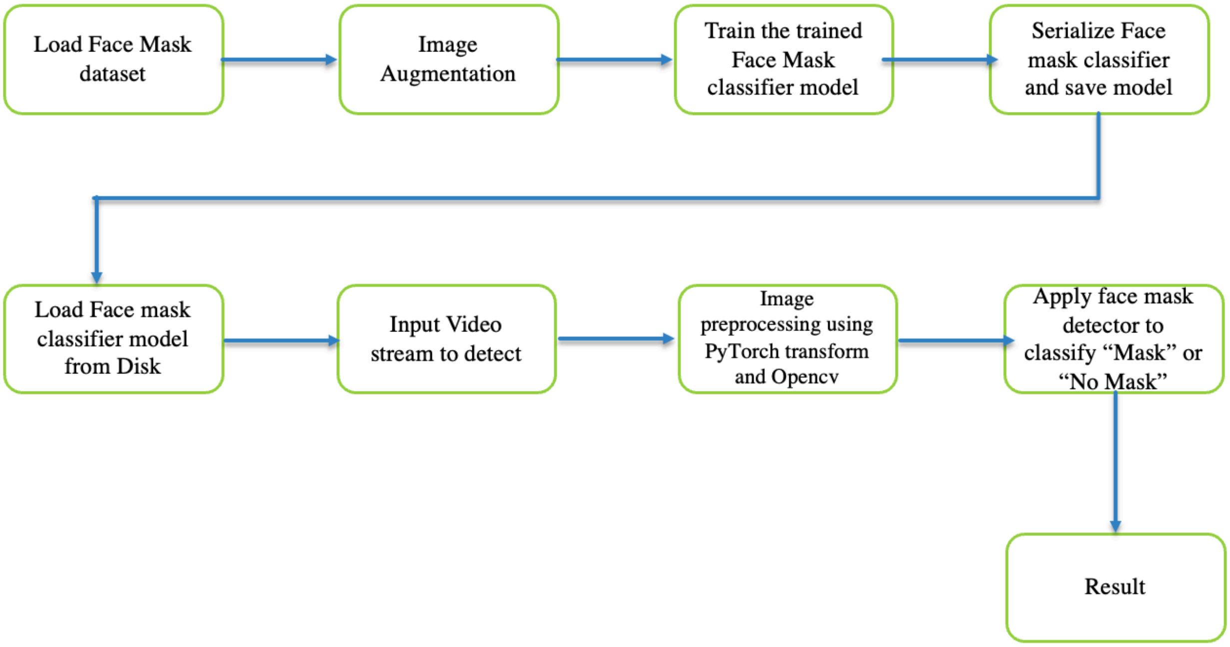


Fig5: Steps to achieve goal

6. Quantitative validation method:

As this is a classification based project classification accuracy will be quantitative validation method to evaluate the face mask detection model.

Metric	Formula	Interpretation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Overall performance of model
Precision	$\frac{TP}{TP + FP}$	How accurate the positive predictions are
Recall Sensitivity	$\frac{TP}{TP + FN}$	Coverage of actual positive sample
Specificity	$\frac{TN}{TN + FP}$	Coverage of actual negative sample
F1 score	$\frac{2TP}{2TP + FP + FN}$	Hybrid metric useful for unbalanced classes

Fig6: Validation methods

Accuracy, precision, Sensitivity, Specificity, F1 score will be calculated.

7. Expected outcome:

When a video stream is given as an input to the the image frames will be processed and region of interest will be extracted and will be passes to the model which will classify if the person is wearing a mask or not. Also I am expecting to achieve an accuracy of classification rate above 90 percent and my model will classify people with any color of mask.

The performance of the model is dependent on dataset in our case of classification, because there can be situations where the extracted face for input to classifier is not completely visible or if the person side of the face is extracted, so in order to classify all this cases correctly we should include differed views or face with and without mask, but due to unavailability of data it is hard to achieve.

Also since I am using Deep learning to achieve this, we do not know what happens inside the deep learning model so there is no conceptual learning or method to make our model work for any color of mask.

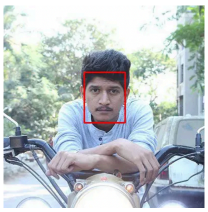
8. Estimated time table for each step:

Week	Info
week 1	Work on step 1, Collect dataset for model training
week 2	Work on step 1, Collect dataset for model training
week 3	Work on step 2, Implement Image Augmentation
week 4	Work on step 3, Train the classifier model
week 5	Work on step 3, Train the classifier model
week 6	Work on step 5, Extract region of interest
week 7	Write report and project presentation

Table1: steps for project

Midterm Update:
Extracting the face region from the given input image:

The Input image



Here I have used a deep learning method to perform face detection with OpenCV.

Face ROI

<u>Image Augmentation:</u>

Image augmentation is a technique used to increase the size of the training dataset by artificially modifying images in the dataset. In this project, the training images are augmented with shearing, contrasting, flipping horizontally, rotating, zooming, blurring. The generated dataset is then rescaled to 224×224 pixels

Fig7: Augmented Image

9. Result:

The performance of the learning model is evaluated based on several performance matrices such as Accuracy, Loss, Precision, Recall, F1 Score. The results of the model for these parameters are listed for different change in values. The best classification achieved parameter value is finalized for realtime application.

The below is the result achieved for different values of Learning rate

Epoch = 20
Batch size = 32
Loss = Binary cross-entropy

===> Training loss ===> Validation loss

===> Training accuracy ===> Validation accuracy

<u>Learning Rate = 1e-01</u>

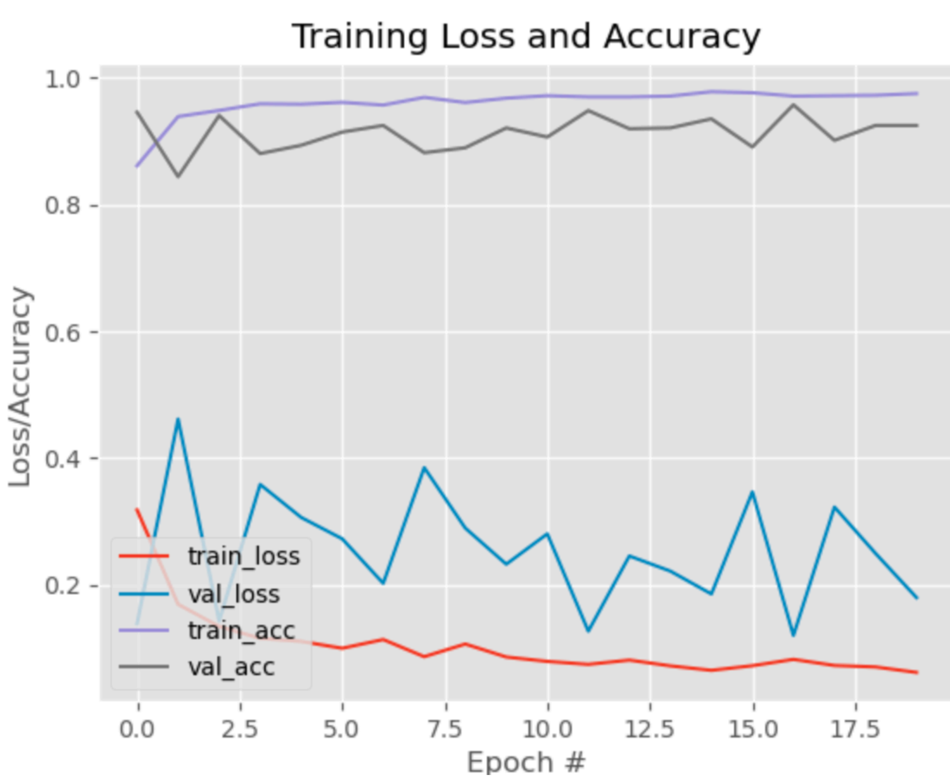
<u>Test Accuracy:</u> 0.5016
<u>Test Loss:</u> 7.6298

<u>Validation Accuracy:</u> 0.5013
<u>Validation Loss:</u> 7.6496

Learning Rate = 1e-02

Test Accuracy: 0.9609
Test Loss: 0.1324

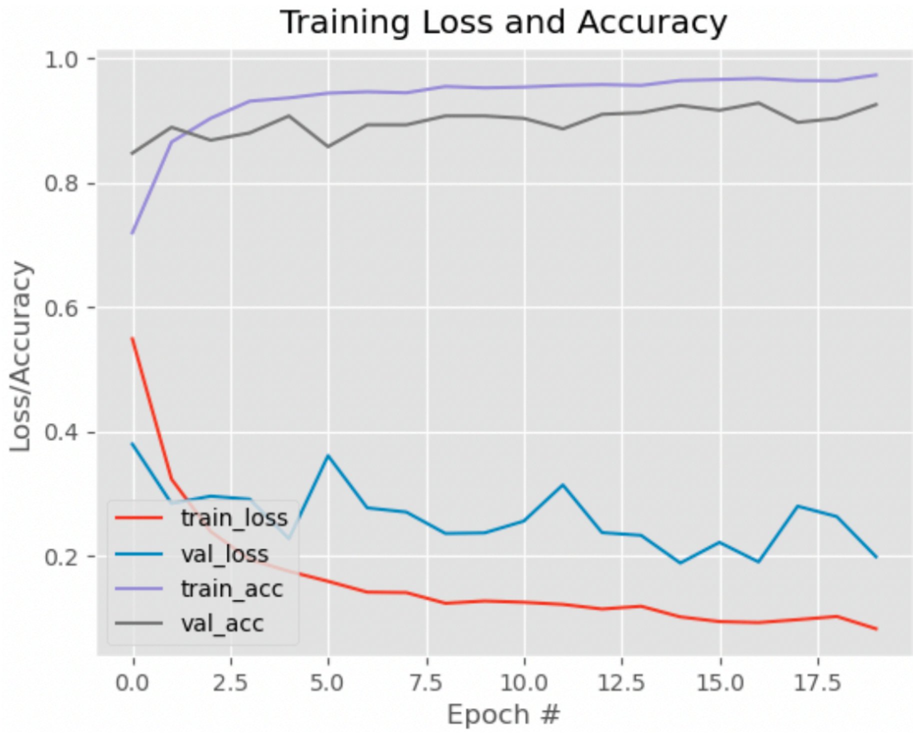
Validation Accuracy: 0.2336
Validation Loss: 0.9271



Learning Rate = 1e-03

Test Accuracy: 0.9747
Test Loss: 0.0651

Validation Accuracy: 0.9245
Validation Loss: 0.1805



Learning Rate = 1e-04

Test Accuracy: 0.9734
Test Loss: 0.0817

Validation Accuracy: 0.9258
Validation Loss: 0.1983

Learning Rate = 1e-05

Test Accuracy: 0.9064
Test Loss: 0.2550

Validation Accuracy: 0.8620
Validation Loss: 0.3116

From the above results we can Get the idea that learning rate at 1e-04 is good to achieve higher validation accuracy of 92.58% and it is also verified from the graph that the model is not overfitting or under-fitting.

In the below table-2 I have summarizing the results of other parameters such as precession, recall, F1 score.

Learning Rate		Precession	Recall	F1 Score
1e -01	With Mask	0.00	0.00	0.00
	Without Mask	0.50	1	0.67
1e -02	With Mask	0.98	0.87	0.93
	Without Mask	0.89	0.98	0.93
1e -03	With Mask	0.99	0.86	0.92
	Without Mask	0.88	0.99	0.93
1e -04	With Mask	0.99	0.86	0.92
	Without Mask	0.88	0.99	0.93
1e -05	With Mask	0.97	0.75	0.84
	Without Mask	0.80	0.97	0.88

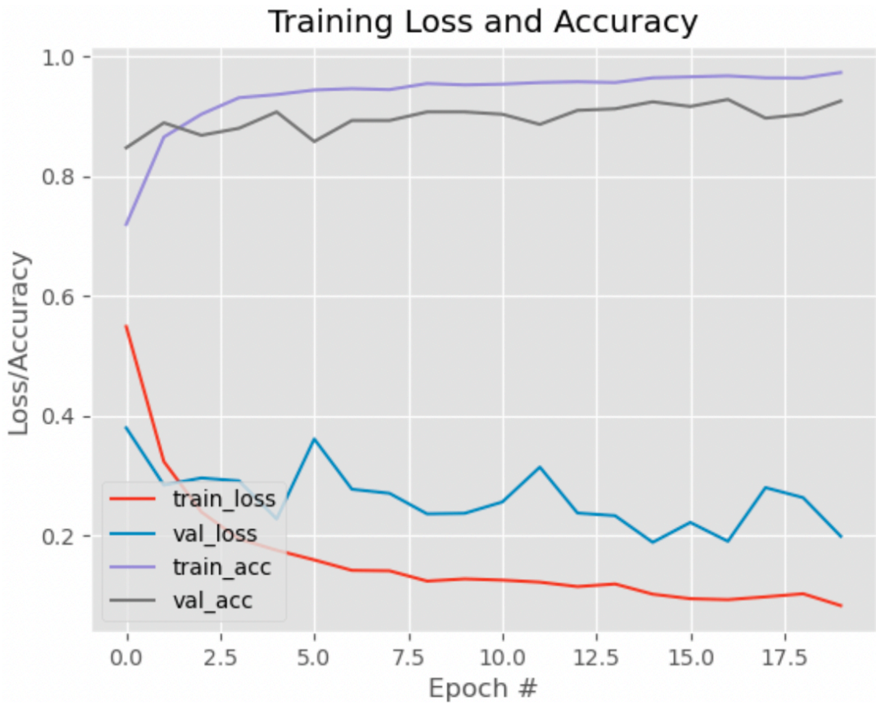
Table2: Results of Precision, Recall, F1 score

So now on fixing the Learning Rate, I have improved the Epoch to observe the change in classification rate

Epoch = 10

Test Accuracy: 0.9504
Test Loss: 0.1249

Validation Accuracy: 0.9180
Validation Loss: 0.2087



<u>Epoch = 20</u>

<u>Test Accuracy:</u> 0.9734
<u>Test Loss:</u> 0.0817

<u>Validation Accuracy:</u> 0.9258
<u>Validation Loss:</u> 0.1983

<u>Epoch = 30</u>

<u>Test Accuracy:</u> 0.9750
<u>Test Loss:</u> 0.0689

<u>Validation Accuracy:</u> 0.8737
<u>Validation Loss:</u> 0.4061

Epoch = 40

Test Accuracy: 0.9694
Test Loss: 0.0795

Validation Accuracy: 0.9180
Validation Loss: 0.2345

Epoch = 50

Test Accuracy: 0.9803
Test Loss: 0.0581

Validation Accuracy: 0.9023
Validation Loss: 0.3062

<u>Epoch = 100</u>

<u>Test Accuracy:</u> 0.9911
<u>Test Loss:</u> 0.0244

<u>Validation Accuracy:</u> 0.9154
<u>Validation Loss:</u> 0.2808

On comparing he results we observe that the maximum Validation accuracy of 0.9258 was achieved with Epoch as 20

I have also tried the loss function as Square hinge loss, the results are bellow

<u>Loss = Squared Hinge</u>

<u>Test Accuracy:</u> 0.9708
<u>Test Loss:</u> 0.5671

<u>Validation Accuracy:</u> 0.91
<u>Validation Loss:</u> 0.6592

In the below table-3 I have summarizing the results of other parameters such as precession, recall, F1 score fro change in Epoch value

Epoch		Precession	Recall	F1 Score
10	With Mask	0.97	0.86	0.91
	Without Mask	0.87	0.98	0.92
20	With Mask	0.99	0.86	0.92
	Without Mask	0.88	0.99	0.93
30	With Mask	0.99	0.76	0.86
	Without Mask	0.80	0.99	0.89
40	With Mask	0.98	0.85	0.91
	Without Mask	0.87	0.99	0.92
50	With Mask	0.99	0.81	0.89
	Without Mask	0.84	0.99	0.91
100	With Mask	0.98	0.85	0.91
	Without Mask	0.87	0.98	0.92

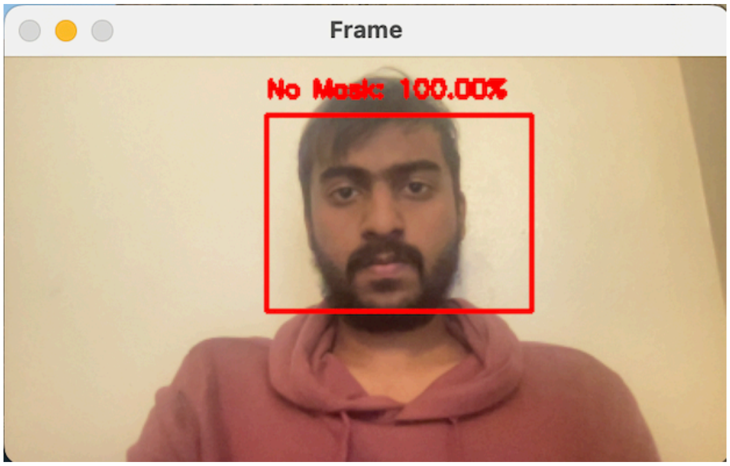
Table3: Results of Precision, Recall, F1 score

From the comparing of all the results from above by fine-tuning the parameters, I observe the best accuracy was achieved with

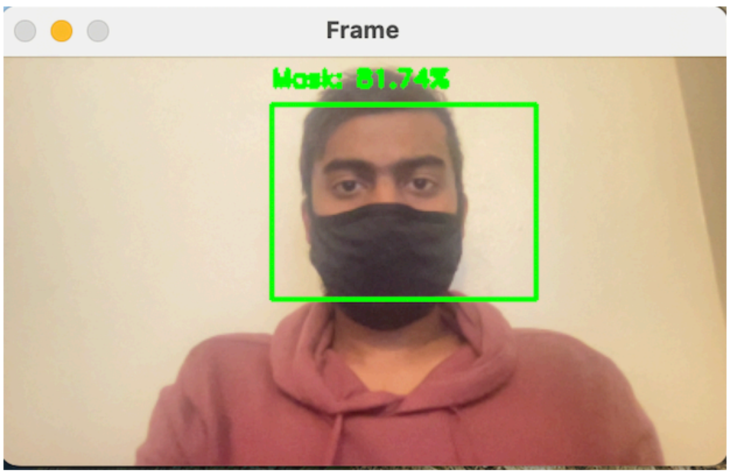
Learning Rate = 1e-04
Epoch = 20
Loss = Binary cross entropy
This model was able to achieve a best accuracy of 0.9258

10. Output:

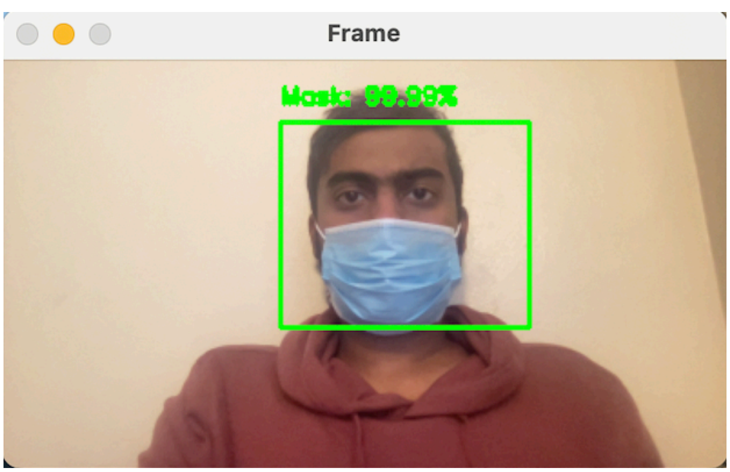
The main objective of this research is to develop an automated process to monitor their people if they are wearing masks in public. As a result, an automated system is developed that uses the transfer learning of MobileNetV2 to classify people who are not wearing masks. The screenshots of some classification is shown below



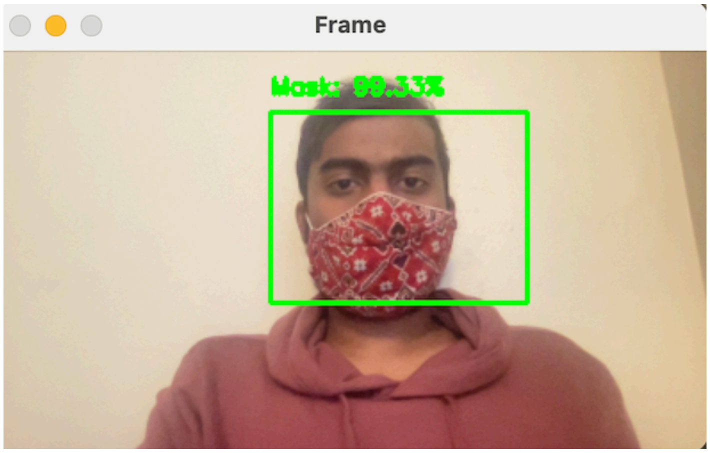
The model detects the person not wearing mask



The model detects the person wearing a mask



The model correctly classifies the person wearing mask, even if its a different color mask



The model correctly classifies the person wearing mask, even if its a different color mask

11. Conclusion:

From this project I am having a good understanding of Deep learning and implementing transfer learning in deep learning model. Initially Since this was my first machine learning course and I am from EE background I did not have a confidence of doing this project related to the coding side but at the end of this project I am having a good confidence in doing Deep learning related problems. Also on doing literature survey I have an overall idea about other pre trained model and for what kind of classification which model yields best result. I also now have a good idea about fine tuning hyper parameters of the model. In this project I was mostly struggling with OpenCV library, as I was not familiarized with it.

Me first time learning these course and when I was asked to choose my own topic for term project I was facing problems even from choosing the topic also I feel that since our project 4 i, ii was difficult there was less time to work on term project.

For Future work:

In this model, if the person is wearing a mask but not properly the model will classify it either one of the classification but in future with the availability of dataset we can train the model to classify people not wearing mask properly

The face mask Detection project is working good, It perfectly classifies people with mask ir-respective of the color of mask. It has achieved an accuracy of above 90%

12. Reference:

[1] Kaur, G., Sinha, R., Tiwari, P.K., Yadav, S.K., Pandey, P., Raj, R., Vashisth, A. and Rakhra, M., 2021. Face Mask Recognition System using CNN Model. *Neuroscience Informatics*, p.100035.
[2] Bhuiyan, M.R., Khushbu, S.A. and Islam, M.S., 2020, July. A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.
[3] Asghar, M.Z., Albogamy, F.R., Al-Rakhami, M.S., Asghar, J., Rahmat, M.K., Alam, M.M., Lajis, A. and Nasir, H.M., 2022. Facial Mask Detection Using Depthwise Separable Convolutional Neural Network Model During COVID-19 Pandemic. Frontiers in Public Health, 10.
[4] Johansson, M.A., Quandelacy, T.M., Kada, S., Prasad, P.V., Steele, M., Brooks, J.T., Slayton, R.B., Biggerstaff, M. and Butler, J.C., 2021. SARS-CoV-2 transmission from people without COVID-19 symptoms. JAMA network open, 4(1), pp.e2035057-e2035057.
[5] Balaji, S., Balamurugan, B., Kumar, T.A., Rajmohan, R. and Kumar, P.P., 2021. A brief Survey on AI Based Face Mask Detection System for Public Places. Irish Interdisciplinary Journal of Science & Research (IIJSR).
[6] Khan, A., Sohail, A., Zahoora, U. and Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. Artificial intelligence review, 53(8), pp.5455-5516.
[7] Zhou, S., Bai, L., Yang, Y., Wang, H. and Fu, K., 2019. A depthwise separable network for action recognition. DEStech Transactions on Computer Science and Engineering, (cisnrc).
[8] Junejo, I.N. and Ahmed, N., 2021. Depthwise separable convolutional neural networks for pedestrian attribute recognition. SN Computer Science, 2(2), pp.1-11