Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

Real Time Cnn Based Detection of Face Mask Using Mobilenetv2 to Prevent Covid-19

S.Shivaprasad[1*],M.Dinesh Sai[2], U. Vignasahithi[2],G.Keerthi[3], S. Rishi[4], P. Jayanth[5],

[1*]Assistant professor, Department of CSE, SR University, Hasanparthy, Warangal.

[2,4,5]Students of 3rd year, Department of CSE, SR UNIVERSITY, Hasanparthy, Warangal

[3]Assistant professor, Department of CSE, VIGNAN UNIVERSITY, Guntur, A.P

[1*]shiva.prasad923@gmail.com

ABSTRACT

COVID-19 is also known as Corona virus, is spreading rapidly throughout the world. The World Health Organization (WHO) has released precautionary measures to reduce the virus spread among the people. Among those measures, mask is the prior precaution of all. So, to solve this various models for face detection are created using different algorithms and techniques. The approach which is put forward in the paper utilizes Deep learning, OpenCV, TenserFlow and Keras which helps in detection of faces with mask. With the help of this model we ensure the safety. The approach for face detector we have used CNN and MobileNetV2 architecture as the classifier, it is light weight, uses less parameters and it can be used in embedded systems (raspberry pi, Onion Omega2) to carry out real time mask detection. The technique used in the paper gives an accuracy of 0.96 and F1-score of 0.92.In this paper, the dataset is collected through different sources which can be put to use for further modern models by various researchers such as facial landmarks, facial features and face recognition for detection process.

KEYWORDS

Convolutional Neural Networks, MobileNetV2, Data augmentation, Bottleneck ,Finetuning, COVID 19

1. INTRODUCTION

COVID-19 is having the worst impact on all the sectors throughout the world. Many people are still being affected with COVID -19 virus and even many deaths are rising in the phase 2 of this pandemic. Many of them even lost their lives, even. COVID-19 was discovered by a doctor Dr. Li Wenliang in the Wuhan city , China in the month of December 2019 . The spread of the virus is due to the droplets of mucus or saliva of the infected person. If the COVID affected person sneezes or coughs, the droplets from their nose, mouth travels through the air and spreads further.

There is no proper medication for the cure of the disease, these are still in the development stages. So, we should probably get sanitized frequently and do not allow the virus into the body by covering the nose and mouth properly, to avoid the spread we should be very careful following all the precautions. The main precautions to be taken in order to reduce the risk of affecting with COVID-19 virus involve, wearing a proper face mask, using a hand sanitizers, maintain social distance .By following those precautions we can reduce the spread of corona virus. It is even declared by the World Health Organization (WHO). In spite of creating much awareness in the people throught the world, many of them are still not following the precautions properly and being the reason for further spread of the virus. Due to this, some of the countries even are facing the phase- 2 of this virus due to mutated virus. Now India is facing the worst situation due to COVID-19 second wave. India is the third country in the world in which most of the deaths have taken place. The model proposed in this paper, can help us a little in reducing the spread of virus i.e., Face Mask Detection using OpenCV, TensorFlow, Keras and

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

Deep learning .In this we will detect whether the person is wearing the mask or not. This model will be very useful in various places mostly in public places. This project can avoid the human involvement in checking the masks of the people. This project can be used in many organizations too, to check the persons coming in are wearing mask or not. In this pandemic we feel good to develop such kinds of projects which can save the mankind even. The proposed project can check the person that they are wearing mask or not through live camera.

The accuracy of the model is 0.9703. Firstly, the datasets are collected which contains the data of images with mask and without mask. If the collected datasets have real time images then, the accuracy of detecting faces in real time will be more accurate when compared to artificially created datasets. The image classifier MobileNetV2 architecture helps in detection of faces in image or video stream.

2.LITERATURE SURVEY

In past, many researchers worked on grey-scale by identifying patterns, non-parametric prejudice by taking trial prototype of face images [1]. Other researchers using adaboost model, some other researchers are building identification pattern model. It consists of information regarding face model [2], adaboost is the outstanding classifier which is used for training. Then comes the advanced face detection technique which is known as voila-Jones detector then it made real time face detection practicable but there came many problems like orientation, problem with intersection and self-illumination. So, basically we can say that this model has problems in low light or very high light. Later on the researchers started researching in order to create new model which can easily identify faces as well as mask which is put on face.

The past years, datasets for detection for face were designed in order to create a new models for face mask detection. Datasets which were earlier made consists of images which were taken from the surroundings, when it comes to the present datasets which were taken from online images such as Celeba [3], MALF [4], WiderFace [5], IJB-A[3].

When compared to earlier datasets to present datasets and present faces , the present dataset s are more accurate.In order to train for better accuracy and detection which is performed in real world scenario the machine need to be trained with large datasets and also we need various algorithms for deep learning which can use those datasets to detect faces and masks by using the data provided[5] .

Many models were created for detection of face masks, these models were divided into many categories. In the boosting-based classification, they used easy haar features were used from Voila-Jones face detector[6], which was explained in the above paragraph. Inspiring Voila-Jones, they came up with a detector which can detect multiple face masks. They have used decision trees algorithms for the above discussed model. These face mask detector were divided depending on the efficiency of detection of face mask.

In convolutional neural network classification, these models (i.e., face detector) use the users data to learn directly from that data and they undergo many deep learning algorithms to learn [7]. In 2007 [8] they proposed using Cascade CNN. Yanget al [4] proposed an idea on features aggregation in his model of face detection. After many research works [1] architecture was upgraded in order to fine tuning images in dataset.

The model SSDMNV2 was created, that uses DNN modules from TensorFlow and OpenCV, which have single shot Multibox detector i.e., a model for object detection. ResNet-10, such classification architecture were used as backbone architecture for the model and MobileNetV2 as image classification classifier, it is been improved

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

over MobileNetV1 classifier as it have 3x3 convolution layer followed by 13 times the previous building blocks. When compared to MobileNetV2 architecture which has 17, 3x3 convolutional layers along with 1x1 convolutions, it is the average layer for max tooling and a classification layer. MobileNetV2 classifier has a new additional residual connection.

3. PROPOSED METHODOLOGY

In order to detect the person whether they are wearing a mask or not, initially we need to train the model using the proper dataset which has been collected. Details of the dataset are discussed in below 3.1. After the classifier is trained, an accurate model is required for face detection so, CNNMNV2 model is used as classifier to classify whether a person with mask or not. The main motto of this paper is about the race of accuracy for mask detection without using too many heavy resources. In order to do this we are using CNN modules are used from OpenCV, TenserFlow and Keras. We use MobileNetV2 model for object detection as it is architecture [9]. MobileNetV2 model classifier uses pre-trained model to predict if a person is not wearing mask or with mask.

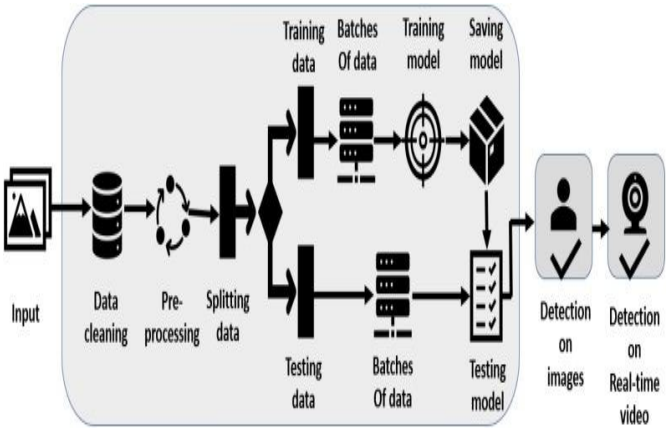


Figure.1 FLOW-DIAGRAM

3.1 DATASET

The datasets which are available for detection of face mask, mostly they are created artificially which is not suitable for real world precisely or dataset contains noise or wrong label. So, in order to choose a good dataset which is best for CNN MobileNetV2 model and it is little hard to find. For training the model the way of approach we use is the combination of different open source picture and datasets, this data is collected from Kaggles mask dataset by MikolajWitkowski and PrajnaBhandary dataset was collected at PyImageSearch and also the dataset was collected [10].

The dataset which was created artificially by PrajnaBhandary images of faces and collected facial landmarks, these landmarks are the facial features of the person likeeyes, mouth, nose, jawline and eyebrows. The artificial way of creating the dataset of a person wearing a mask by using a person image who is not wearing a mask. After creating the artificial images we are not able to use the same non-mask face sample which makes this model heavy. So, it is better to use a dataset which consists of people images with masks and without masks which will correct the errors a mentioned above. A dataset has 5521 images with labels “with_ mask “ and “without_ mask” which makes the dataset balanced.

The images in the dataset are available in the above link.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

Figure 2.DATASETS OF MASKS AND WITHOUT MASKS.

3.2 PRE PROCESSING

After the collection of datasets then, the weresize the image size to 224x224 downgrade from 1024x1024 then we convert the image to array. Then we pre- process the input image sample using MobileNetV2 and we perform hot encoding. The list is sorted then the label of images is converted to tensors. Then list is converted to NumPy arrays which helps in fast calculation .After this, data augmentation takes place to increase accuracy of the model which we have trained.

3.3 DATA AUGMENTATION

In order to train the model, we need large amount of data to train efficiently and effectively since the availability of the amount of data for training of the proposed model is available but not so accurate. So, to solve this problem we came up with the method called data augmentation. By using this technique, it performs rotation, zooming, shifting, shearing and flipping. The sample images are used for this to generate similar sample images. So, we use image augmentation for data augmentation process. For image augmentation function image data is generated, it returns test and used to train data in batches.

3.4 IMAGE CLASSIFICATION USING MOBILENETV2

For classification problems we use MobileNetV2 which is a deep neural network. ImageNet has pre-trained weights which are loaded from TensorFlow. Then we need to freeze base layers so that they are not updated in first training of the model which has already features learned. After training the model, the trained layers are added and these layers were trained by using the dataset which we have collected. So, by using the features the model can classify whether there is a mask on face or not. The model is fine-tuned and weights were saved. The reason to use pre-trained models is to avoid unnecessary computational costs and it has advantage because the weights does not cause any effect on features which are pre-learned.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

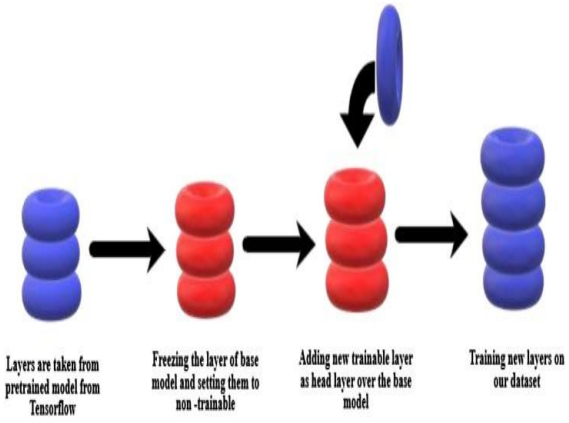


Figure 3.PIPELINE OF USING PRE-TRAINED MODEL.

3.4.1 MOBILENETV2 BUILDING BLOCKS

MobileNetV2 is a pre-trained deep learning model which has base as convolutional neural network. The following layers and functions of MobileNetV2 are shown below.

1.CONVOLUTIONAL LAYER

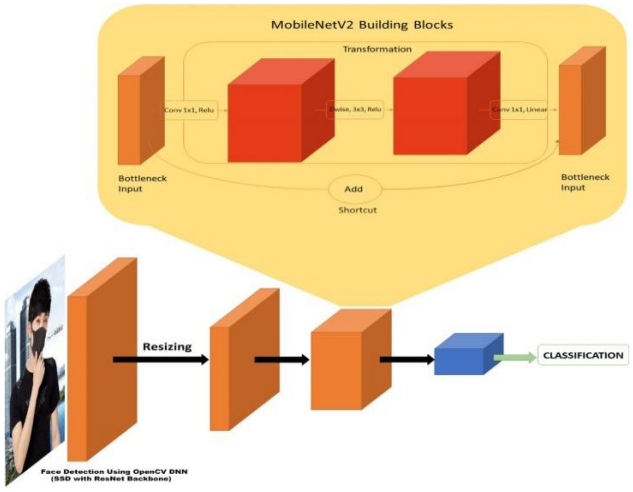


Figure.4 ARCHITECTURE OF MOBILENETV2

The above layer is fundamental block for convolutional neural network. The word convolution is a mathematical combination of two functions in order to get third function.

The features of an image are drawn out by using mechanism called sliding window. It has convolution functional matrices, one is matrix considered as the input sample of image matrix A and B convolutional kernel that gives output C.

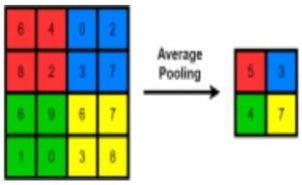
$$CT=A*Bx=\int_{-\infty}^{\infty} AT\times BT-x\, dT$$

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

2. POOLING LAYER

Pooling layer makes it easy for calculation faster by reducing input size matrix without effecting the loss of features . There are few kinds of pooling , among them two are explained below.

1. Average pooling : Average of all the values which are in the current region, the kernel takes the value of the output from the cell of matrix value.



Average-Pooling Operation.

Figure. 5 AVERAGE POOLING OPERATION

2. Max pooling : The maximum value of selected region, the kernel takes the value of output cell of matrix value.

3. Dropout layer : It is used to reduce over fitting which might occur in the process of training the model the baised neurons randomly drop. Those neurons might be a part of hidden layers and visible layers. The dropping of neuron can change by altering the ratio of dropouts.

4. Non-linear layer :Non linear layer are also convolutional layers but, without any non linearity function which includes various kinds of Rectified Linear Unit (ReLU) [11] i.e., noisy ReLU [12], Leaky ReLU[13], Exponential ReLU[14] etc sigmoid function along with tanh functions. These are the equations of different non-linear functions.

Sigmoid: $\sigma(x)= 1/1+e-x$

Leaky Relu: $f(x)=\max(0.1x, x)$

Tanh: $f(x)=\tanh(x)$

Maxout: $f(x)=\max(wT1x+ b1 , wT2x+b2)$

Relu: $f(x)=\max(0,x)$

ELU: $f(x)= \{x\ x\geq 0\ \alpha(ex-1)\ x<0)$

5 Fully-connected layer : The model is appended in these layers and they have full connection with activation layer. These layers are used to classify the sample images in multi-class or binary classification. SoftMax is example of activation function which is used in those layers and it gives predicted output for the classes.

6. Linear Bottlenecks : Multiple matrix multiplication can't convert into smaller to a single numerical operation, ReLU6 which is a nonlinear activation function is used in neural networks. So , removing several with discrepancies. We can also make multilayer neural networks. ReLU does not allow values less than zero. Reversed residual blocks, layers in the blocks are compressed and contrary. This happen at a particular point of

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

then skip connections are linked, which might effect the network In order to resolve this, linear bottleneck concept is introduced adding of blocks to initial activation, left-over block is linear output for the last convolution.

4. EXPLANATION OF ALGORTHIM

Proposed CNN mobileNetV2 methodology is clearly explained by using 2 algorithms.

Algorithm-1 : pre-processing and training with dataset.

Input : Images with pixels values.

Output : model is trained.

Step 1 : take the pixel value and images should be loaded.

Step 2 : processing of images is Done i.e., normalization, resizing and conversion into arrays.

Step 3 : File names and there labels are loaded.

Step 4 : Data augmentation process is done and splitting of data for testing and training of model.

Step 5 : Mobilenetv2 model is loaded. Training batches and Adam optim,izer is used for complication process.

Step 6 : saving the model.

Algorithm 2 : deployment process

Input : choosing files to deploy.

Output : detection of people wearing mask or not.

Step 1: classifier is loaded and face detector is loaded from opencv.

Step 2: If classification is done on image then load image.

Step 2.1: face detection model is used to detect faces in the image.

Step 2.2: If face is detected then crop face with box using coordinates from the model and get prediction from image classifier model.

Show prediction and save results.

If face is not detected show no output.

Step 3: If classification is done real-time.

Input is taken from Video stream using OpenCV.

Taking frame by frame from video stream.

Step 3.1: face detection model is used to detect faces from frames.

Step 3.2: If frame is detected then crop face with box using coordinates from the model and get prediction from image classifier model.

Output is shown in real-time video stream.

If face is not detected then show normal video stream.

Step 4: q is pressed to end video stream

5. ACCURACY

The metric for the model is explained below,

Accuracy =Tp + Tn / Tp+ Fp+ Fn + Tn

f1 score = 2* Recall * Precision / Recall + Precision

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

Where, Fp = False positive,

Fn = False negative,

Tp = True positive,

Tn = True negative.

The formulae mentioned above ,Tp value refered to images that are labeled as true and the prediction of model gives true results. Similarly, for true negative we get prediction for images as false result. Fp denotes images with false labeled and results predicted for the model is false . So, it is false positive. Fn denotes images with false labeled and result predicted by the model is true. So, it is false negative. As, classes are balanced we got good accuracy. Precision gave positive value. By the classifier we get positive results

for Recall and for test accuracy was taken from F1-score. The metrics evaluation were taken for their ability to obtain the best results.

Pre-processing of the data is done before training the data. Firstly, a sorting function is applied in order to convert the alphabetical data to binary form i.e., 0 or 1. Pre-processing function is described, it takes the folder in dataset as input, it loads all the data present in folder and images are resized to 224x224 for the model. After sorting, images are turned into Tensors. Then, all the lists is transferred into NumPy arrays for calculating fast. After using

pre-processing function accuracy is increased of the model. So, by using data augmentation process we increase the accuracy of the model.

6. RESULTS

The test results from video are given below,We have taken 4 sets of tests by using MobileNetV2. The rectangular green box indicates that the person is wearing a mask with accuracy on top of the box.These are the two test images from video stream.

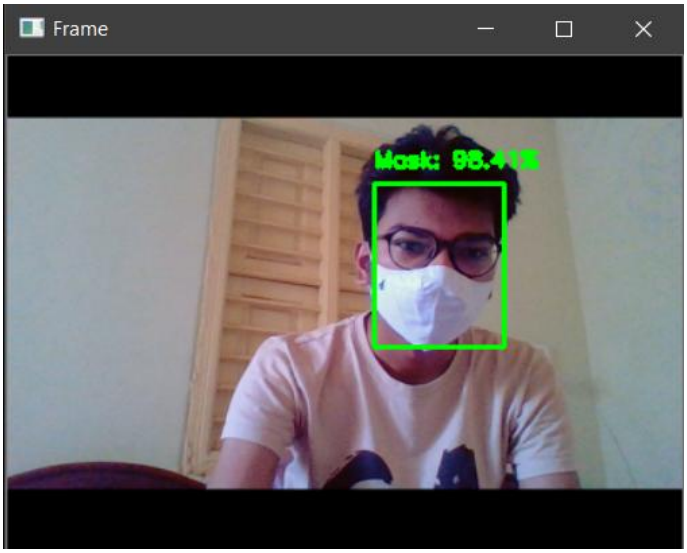


Figure.6 RESULTS OF WITH MASK

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

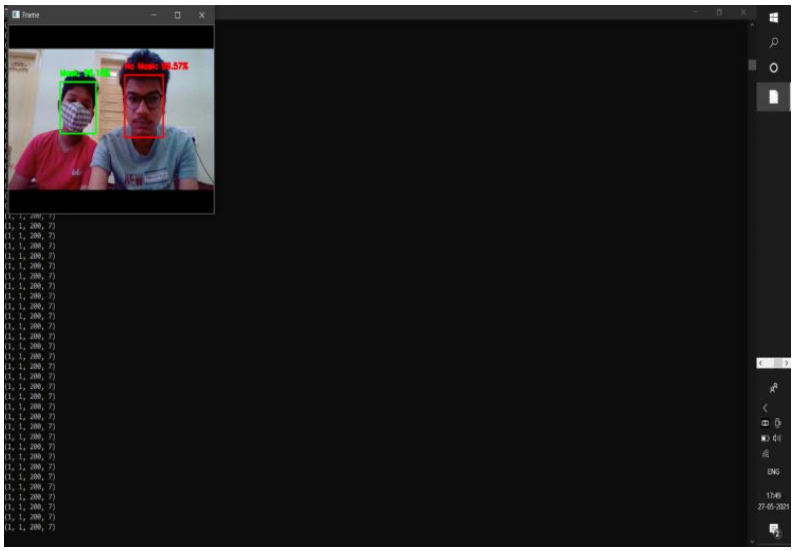


Figure.7 RESULT OF ONE WITH MASK AND OTHER WITHOUT MASK.

Similarly, red rectangular box indicates that the person is not wearing a mask or wearing incorrectly. The model predicts using the pattern from training of dataset and labels of the data.

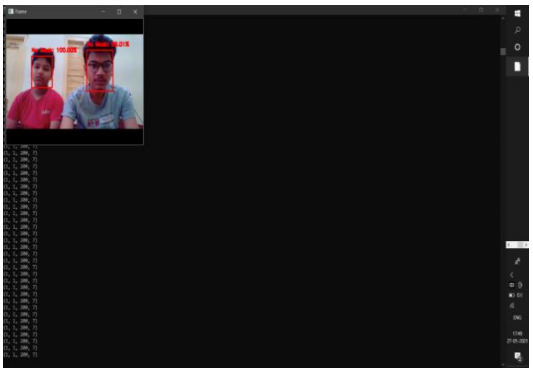


Figure.8 RESULTS OF BOTH WIHTOUT MASK.

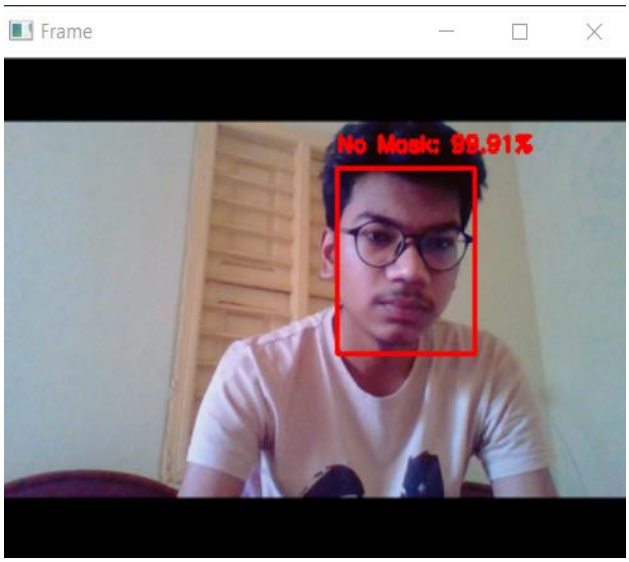


Figure 9 RESULTS OF WITHOUT MASK.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

The model is trained on 20 epochs ,with learning rate 1e-4 i.e., 0.301 and the dataset is divided 20% for testing purpose and 80% for training the model. The developed model is trained for 20 epochs since further training results cause over fitting on the training data. Overfitting is caused by making the model to learn unwanted patterns from the training sample. The model shows accuracy of 0.97 as shown in the below graph. This mean that the model has decent accuracy. The validation loss value is approximately at 0.4 and training loss is approximately 0.5 as shown in the graph.

Figure.10 LOSS AND ACCURACY GRAPH FOR TRAINING

The developed model is trained with 10 epochs, with learning rate of 1e-4 i.e., 0.0001 which gives an accuracy of 0.96, validation loss is approximately less than 0.2 and training loss is less than 0.4 .

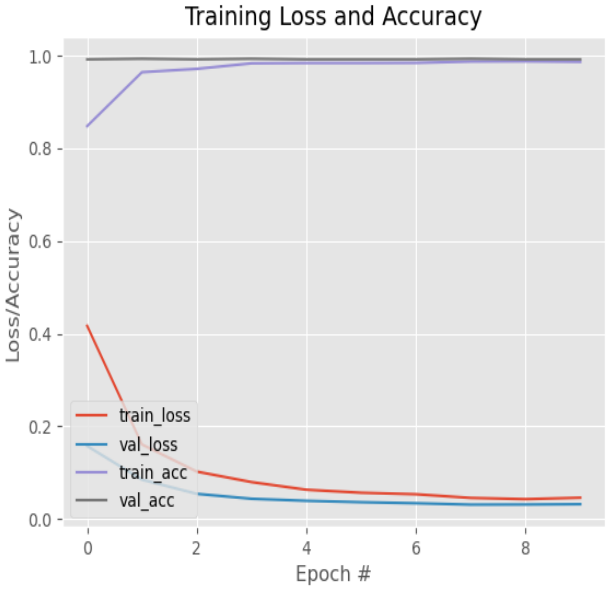


Figure.11 LOSS AND ACUURACY GRAPH FOR TRAINING

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

7. CONCLUSION

The proposed model face mask detection for training and creating image dataset, which is divided to two categories namely with mask and without mask is done successfully. Using OpenCv deep neural networks the model gets good results and by using MobileNetV2 which is an image classifier for the classification of images is processed accurately.

Many existing models are facing problems with the results in terms of accuracy with the dataset they have. Those problems are successfully removed in this model as the data is collected from different sources and images in the dataset are manually cleaned for the better and accurate results. In real time applications it is a challenging issue for the future. This model might be helpful for other researchers for to gain advancement in the model. The organizations should apply this model in real time so that, we can eliminate the involvement of human in checking masks and by taking action on people who got detected, such that it reduces the risk of infectious spread of COVID-19 by taking action on people who got detected.

REFERENCES

[1] Ojala T., Pietikainen M., Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2002;24(7):971–987.

[2] Kim T.H., Park D.C., Woo D.M., Jeong T., Min S.Y. *International conference on intelligent science and intelligent data engineering.* Springer; Berlin, Heidelberg: 2011. Multi-class classifier-based adaboost algorithm; pp. 122–127.October.

[3] Klare B.F., Klein B., Taborsky E., Blanton A., Cheney J., Allen K.…Jain A.K. Pushing the frontiers of unconstrained face detection and recognition: Iarpajanus benchmark a. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015:1931–1939

[4] Yang B., Yan J., Lei Z., Li S.Z. Fine-grained evaluation on face detection in the wild. IEEE; 2015. pp. 1–7. May.

[5] Yang S., Luo P., Loy C.C., Tang X. Wider face: A face detection benchmark. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016:5525–5533.

[6] Jones P., Viola P., Jones M. University of Rochester; Charles Rich: 2001. Rapid object detection using a boosted cascade of simple features.

[7] Ren S., He K., Girshick R., Sun J. *Advances in neural information processing systems.* 2015. Faster r-cnn: Towards real-time object detection with region proposal networks; pp. 91– 99.

[8] Li H., Lin Z., Shen X., Brandt J., Hua G. A convolutional neural network cascade for face detection. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015:5325–5334.

[9] Wang Z., Wang G., Huang B., Xiong Z., Hong Q., Wu H.…Chen H. 2020. Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093.

[10] Hahnloser R.H., Sarpeshkar R., Mahowald M.A., Douglas R.J., Seung H.S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature.* 2000;405(6789):947–951.

[11] Gulcehre C., Moczulski M., Denil M., Bengio Y. *International conference on machine learning.* 2016. Noisy activation functions; pp. 3059–3068. June. [Google Scholar]

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 17689 - 17700
Received 25 April 2021; Accepted 08 May 2021.

[12] marufurmd.Motalebhossenmd .MilonislamSaifuddinmahmud An automated system to limit covid 19 using facial mask detection in smart city network 2020 IEEE international IOT

[13] Toshanlalmeenpal,Ashutoshbalakrishnan, Amitverma, Face mask detection using semantic segmentation 2019,4th international conference on 4th Internationalcomputing, communications and security(ICCCS)

[14] Zenyunsun,Yusong,Changsheng,Face spoofing detection based on local ternary label supervision in fully convolutional networks IEE transactions on information forensis and security 2020