

Automatic Facemask Detection using MobileNetV2

Vasanthakumari A^[1], Mohd.Umar Peerzade^[2], Subhralina Nayak^[3], Bhagya V^[4], Shobana Padmanabhan^[5]

^{[1][2][3][4]} Students, Department of CSE, Alliance University

^[5] Assoc. Professor, Department of Information Technology, Alliance University

Vasanthakumaria96@gmail.com ^[1], peerzadeumar@gmail.com ^[2], subhralina.nayak.14@gmail.com ^[3], bhagyashreevenkatesh@gmail.com ^[4], shobanapadmanabhan@gmail.com^[5]

Abstract - The pandemic due to COVID-19 needs no introduction. An all-too-important and successful weapon against it has been the use of "face masks". Here, we present a novel technique for detecting automatically whether a person is wearing a face mask or not, based on machine learning. The findings can be used by policy makers to warn or fine people not wearing a face mask, especially at public places being monitored by Closed-Circuit Television (CCTV) such as at traffic signals, parks, cinema theaters etc. The information can also be used to plan and conduct awareness camps & / to distribute face masks, as needed. Our work is based on deep learning through neural networks, in particular, MobileNetV2, trained and tested on images of people's faces with and without masks, collected from various sources. The accuracy of our detection is 96% on the test dataset whereas the best result in the literature, to our knowledge, is 92% or lower

Key Words: Automatic face mask detection, deep learning, CORONAVIRUS, COVID-19, CNN, MobileNetV2

1. INTRODUCTION

A key weapon against the spread of COVID-19 has been the use of face masks. This has been mandated and emphasized by the governments of different countries, based on the guidelines by the World Health Organization (WHO). According to the WHO, face masks can be used for control of source (worn by infected individuals to inhibit further transmission) or for the protection of healthy people.

Automatic face-mask detection at real-time is emerging as a very interesting problem in image processing and computer vision. The goal has been to detect automatically whether a person is wearing a mask or not. We present here a novel model based on neural networks, specifically, convolutional neural networks, that has an accuracy of 96%.

Thanks to our work, governments, policy-makers, health-care providers and educationalists will be able to find out if there are particular places or regions and maybe even particular timings when people are not wearing face masks. Subsequently, they can plan and organize awareness campaigns, law enforcement reinforcements, free face mask distributions and such activities.

II BACKGROUND AND RELATED WORK

[1] Jinsu Lee et al. proposed a model called "An Ensemble Method of CNN models for Object Detection". Their research work mainly focuses on detecting objects of a certain class in digital images and videos. They use deep learning, specifically, CNNs. They have divided their work into two phases. In the first phase, they perform a two-stage detection where region proposals are generated. In the second phase, there is a one-stage detector that helps to detect and classify the object without generating region proposals. The authors combine the various properties of CNN models, advanced ensemble methods to detect objects and their own novel methods for model selecting and box voting. Overall, with experimental proof, they report an increased model accuracy.

[2] Sebastian Handrich et al. proposed a method called "Face Attribute Detection with MobileNetV2 and NasNet-Mobile". In their work, they come up with two simple and effective methods for estimating facial attributes in unconstrained photos or pictures, using straightforward and fast face- alignment techniques for pre-processing. MobileNetV2 and NasNet-Mobile were utilized to estimate face attributes. Two lightweight CNN architectures are presented and these both perform similarly, in terms of accuracy and speed. They also compare the model, with respect to processing time and accuracy, and show that the approach performs faster than the state-of-the-art model. This model was easy to use and deploy in mobile devices.

[3] Amit Verma et al. proposed a system "Facial Mask Detection using Semantic segmentation". The main aim of this paper is to design a binary classifier that will be able to detect faces in a frame (image) without considering alignment. The resultant method generates a correct face segmentation mask irrespective of the image size. VGG-16 architecture is being used for feature extraction and FCN for segmentation of faces from image. Gradient descent is used for training the model. Overall, the model provides good accuracy.

[4] Md.Sabbir Ejaz et al. present "Implementation of principal component Analysis on masked and unmasked face recognition", considering security as the main aim of

their work. They focus on providing security in biometric systems, using faces as the input. They implement this using PCA for recognition of faces with and without masks. They report that faces without masks get recognised better using PCA than faces with a mask.

[5] Md.Sanzidul Islam et al. proposed “A Deep learning based assistive system to classify a face for human safety” using YOLOv3 architecture. Their research focuses on building a custom object detection and a deep learning framework to detect face masks from video footage.

[6] has been presented by Aniruddha Srinivasa Jsohi et al. The framework capitalizes on the MTCNN face detection model which identifies faces and facial landmarks present in video footage. MobileNetV2 architecture is utilized for object detection on the face. The model is tested with a dataset of videos that contain movement of people in different places.

[7] Pathasu Doungmala et al. has presented a “Helmet detection in Thailand” using image processing. Two methods are used here: like- feature, which is used for face detection circle, and Hough transform, for no and half helmets. Their work uses a fast algorithm for detecting helmets in colour pictures by focussing on features for detecting helmet regions like face, nose, eyes etc.

[8] Zhelin Li et al. target “Ship detection and classification based on SSD_MobilenetV2”. The model proposed by these authors focus on ship image detection. Feature extraction of the images is carried out using MobilenetV2. “coco” dataset is used for training. A constructed dataset is used for fine-tuning. The result of the model shows R-CNN_InceptionV2 algorithm has better recognition accuracy.

[9] Rohith CA et al. have proposed a system “An Efficient Helmet Detection For MVD using Deep learning”. Transfer learning and fine-tuning techniques were utilized to build a model for identifying vehicles based on videos and to capture pictures from a frame, to verify whether a person is wearing a helmet or not. Using these techniques, the resultant model provides up to 76 % accuracy.

[10] Marielet Guillermo et al. have proposed a model “COVID-19 Risk Assessment through Face

Mask Detection using MobileNetV2 DNN” This study aims to promote the importance of disease control and preventive measures such as the use of face masks in crowded places. The implementation of this study was made possible through three major phases artificial face mask dataset creation, face mask detector training, and face mask detector testing. They achieved 92% of accuracy as a result.

III METHODOLOGY

Our system makes use of Convolutional Neural Networks (CNN) to classify images as having a mask or not. A Convolutional Neural Network is a type of deep learning algorithm which takes an image as the input. It then maps the various aspects of the image to weights and biases, hence, they become differentiable from each other. The reason we chose CNN for classification over other classification models is because the amount of pre-processing required for CNN is much less.

We have proposed a model that uses MobileNetV2 for image processing. It is a CNN (Fig. 1) that is 53 layers deep. It can classify images into 1000 categories. It uses depth-wise separable convolutions which constitute as the building blocks. It is the state-of-the-art network for mobile visual recognition which includes classification, object detection and semantic segmentation. MobileNetV2 (Fig. 2) is launched as a section of TensorFlow-Slim Image Classification Library.

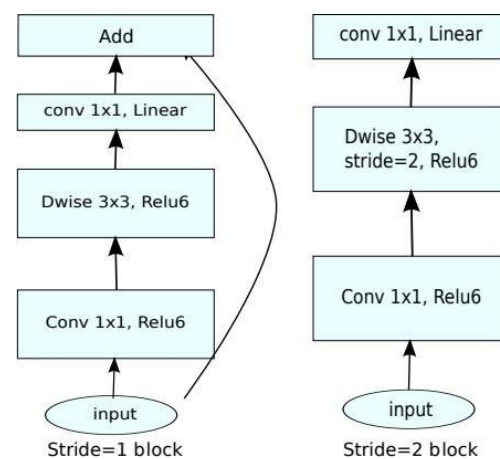


Fig -1: Comparison of MobileNetV1 and MobileNetV2

The intuition behind MobilNetV2 is based on the fact that the bottleneck layers (Fig. 2) encode the model’s input and output, whereas the inner layers encapsulate the model’s ability to evolve from low-level concepts to high-level concepts. In our case, it is the conversion from pixels to image categories.

It is faster when compared with MobileNetV1 because it uses two times fewer operations and it needs less number of parameters. As a result, it has higher accuracy for the same latency value as shown in Fig. 4.

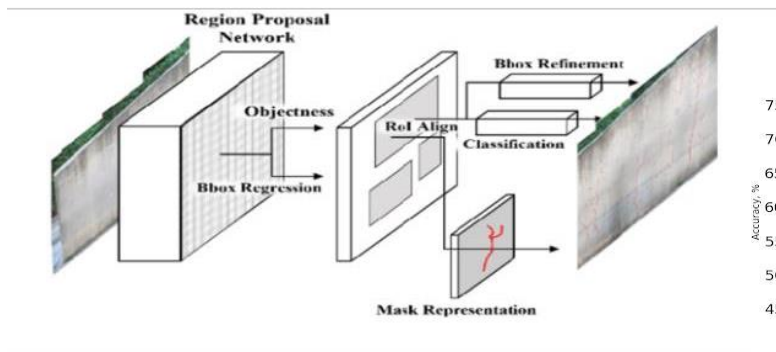


Fig -2: CNN Model

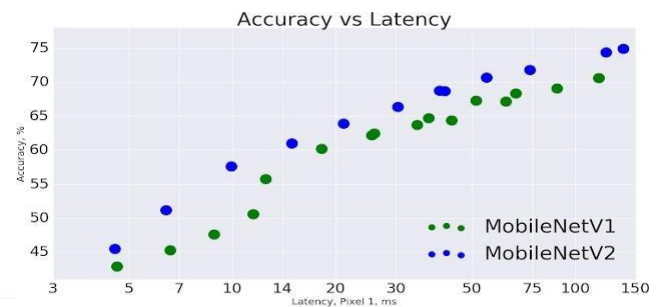


Fig -4: Accuracy Vs Latency

The following layers were added to model:

- **Average Pooling 2D:** This layer calculates the average value for all the patches in a feature map. It then creates a feature map that is down sampled.
- **Flatten:** This layer converts the pooled feature map to a single column.
- **Dense:** This layer feeds all outputs from the previous layer to all its neurons. Then each neuron returns one output to the next corresponding layer.
- **Dropout:** This layer helps to prevent overfitting. It sets the input value to 0 on a random basis during the training of the model.

IV IMPLEMENTATION AND RESULTS

The proposed model focuses on detecting people in images as wearing a mask and not, using TensorFlow, Keras and mobilnetv2. These are open-source libraries implemented in Python. CNNs are a type of Neural Networks that have proven to be very powerful and effective in areas of image recognition, processing, and classification. In machine learning, a convolutional neural network is a type of deep, feed-forward artificial neural networks (referred to as multilayer perceptron), oftentimes used to analyse visual imagery. They have been in existence for a long time now but have become very useful with the availability of massive labelled datasets and high processing capacity. It is a deep learning algorithm in which it takes input as an image, uses weights and biases and outputs the class of the images.

Our approach consists of the following phases:

- Data preprocessing
- Training and testing of the MobileNetV2 model using the corresponding dataset of images
- Classifying new (unseen) images
Classifying images at real-time when an image or a video is streamed.

A. DATA PREPROCESSING

The dataset used [Kaggle] contains a total of 8226 images out of which 3972 images are people’s faces without a mask and 4254, with a mask each. 75% of the images from the dataset are used for training and 25% of the images are used for testing the model. The images in our dataset are preprocessed as follows, before being fed into the MobileNetV2.

- Resize the input images and centre-crop the image with the pixel value of 224 x 224 x 3 via augmentation.

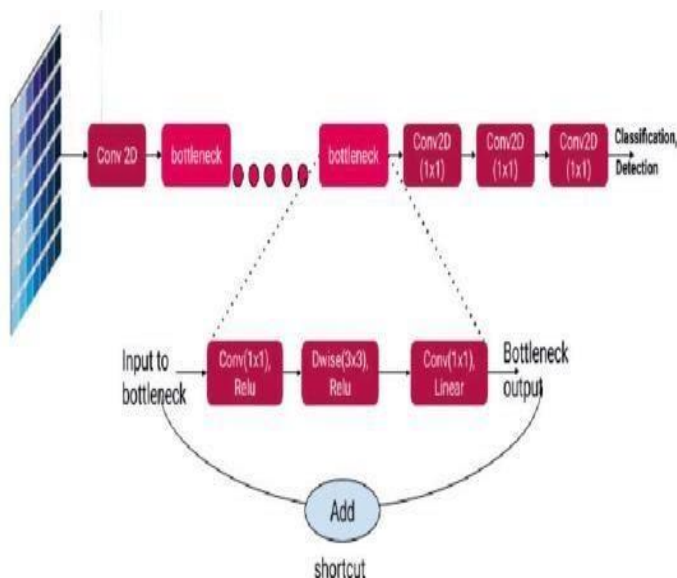


Fig -3: Architecture of MobileNetV2

- Apply colour filtering (RGB) over the channels (our model MobileNetV2 supports 2-dimensional three-channel images).
- Scale/ Normalize images using ImageDataGenerator of the Keras library.
- Downsample the images to a fixed resolution of 256*256 by extracting random 224*224 patches from 256*256 images.
- Finally, converting them into tensors, similar to NumPy arrays.

Our goal is to take a new image that falls into a category we have trained and run it through a command that will tell us the category in which the image fits--“mask” or “no mask” category.

B. REAL-TIME CLASSIFICATION OF STREAMED IMAGES

This is achieved by creating separate classes. The steps involved are as follows.

- Get the image stream from a webcam.
- Detect faces with OpenCV and add bounding boxes.
- Convert the faces to grayscale, rescale them and send them to our pre-trained network.
- Get the predictions back from our network and add the label to the webcam image. Figures 5, 6, and 7 illustrate it.
- Return the final image stream.

The accuracy on the validation dataset is 96%.

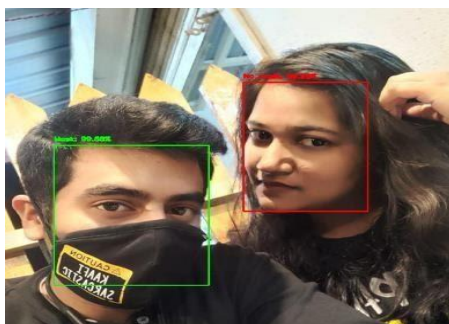


Fig -5: Result on an image with and without a mask



Fig -6: Result on a real-time image with no mask



Fig -7: Result on a real-time image with a mask

V CONCLUSION AND FUTURE WORK

As the world is fighting the COVID-19 pandemic, we have developed a novel solution for detecting whether a given image of a person has a facemask on or not. Our solution detects this even on streaming images in real-time. The accuracy on the test dataset is 96%, the highest to our knowledge. This will greatly aid public and clinical administrations. Our solution uses Mobilnetv2, OpenCV, TensorFlow, Keras and CNN. This can be used especially at public places where we can identify automatically if an individual is not wearing a face mask and may prevent their entrance. In ongoing work, we extend this work to images containing more than one face.

REFERENCES

1. Jinsu Lee, Sang-Kwang Lee, seong- Yang."An Ensemble Method of CNN Modelsfor Object Detection",IEEE,2018.
2. Sebastian Handrich," „Face Attribute Detection with MobileNetV2 and NasNet- Mobile", IEEE,2018.
3. Amith verma, Toshanal Meenpal, Asthosh Balakrishnan, "Facial Mask detection Using Semantic /segmentation", ICCCs, 2019.

4. Md.saabbir Ejaz, Md. Rabiul Islam, Md. Sifutullah, ananya sarker, "Implementation of Principal Component analysis on Masked and Non-Masked Face Recognition", IEEE, 2019.
5. Rafiuzzaman Bhuiyam, Sharun akter Khsushbbu, Md.Sanzidul Islam,"Adeep Learning Based assestive system to classify COVID-19 Face Mask for Human Safety with YOLOv3,IEEE,2019
6. Aniruddha Srinivas Joshi, shreya Srinivas Joshi, Goutham Kanahasabai, Rudraksh Kapil,"Deep Learning Framework to detect Face Mask from video footage",International Conference on computer Inteligence and communication Networks,2020
7. Pathasu doungmala, Katanyoo Klubsuwan,"Half an dFull Hlmet wearing detection in Thailand using Haar Lke Feature and Circle Hough Transformation on Image Processing",IEEE, 2016.
8. Yiyang Zou,Lining Zhao, shanxng Qin, Mingyag pan,"ship target detection and identification based on SSD_Mobilenet V2",IEEE,2016.
9. Rohith C A, shilpa A Nair, Parvathi sani Nair,Nithin Prince John,"an efficient Helmet detection For MVD using Deep learning",IEEE, 2019.
10. Athena Rosz Ann Pascua,Marielet Guillermo,"COVID-19 Risk Assessment through Face Mask Detection using MobileNetV2 DNN", International Symposium on Computational Intelligence and Industrial Applications, 2020.