Real World Face Mask Detection using MobileNetV2 and Raspberry Pi

IJERA Journal

Tanuja R, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 10, (Series-I) October 2021, pp. 26-32

RESEARCH ARTICLE OPEN ACCESS

Real World Face Mask Detection using MobileNetV2 and Raspberry Pi

Tanuja R.[1]*, Mahesh S.[2]*, S.H. Manjula[3] and Venugopal K. R[4].
[1]*Associate Professor, Department of Computer Science and Engineering, UVCE, Bengaluru, India*
[2]*M.Tech Scholar, Department of Computer Science and Engineering, UVCE, Bengaluru, India*
[3]*Professor, Computer Science and Engineering, UVCE, Bangalore, India*
[4]*Vice Chancellor, Bangalore University, IEEE Fellow, ACM Distinguished Educator, Bengaluru, India*

ABSTRACT
On March 12, 2020, Corona Virus Disease 2019 (COVID 19) was declared a global pandemic. Because of its quick spread from one person to another, this disease was thought to be more hazardous. Face masks have proven to be a good and effective way to stop the spread of COVID 19. Detection of Face Mask is a challenging problem. This paper proposes the method to solve this challenge by using deep learning. This work uses Multi-Task Cascaded Convolutional Neural Network (MTCNN) for detection and identification of face. MobileNetV2 is used as an object detector for mask detection. A total of 3833 images from different data sources were chosen for this work. This is later implemented using Raspberry Pi and pi cam, this setup transmits live video data from a remote location and hence the prediction of wearing mask is accomplished. The amount of information lost in the process is decreased gradually at 20th epoch is 0.0199. The accuracy by which the mask/ no mask detection is increased.
Keywords- Covid-19, Deep Learning, Face Detection, Face Mask Detection, MobileNetV2, MTCNN, Raspberry Pi

---

Date of Submission: 30-09-2021 Date of Acceptance: 13-10-2021

I. INTRODUCTION

Pandemic diseases are not new to humankind, the history of pandemics is sort of long and extensive, from the 1918's Spanish Flu to 2019's COVID-19[1]. Pandemics have always been a curse for the well-being of humans, affecting their livelihood. The SARS is an airborne virus spread through small droplets of saliva from person to person, affecting the respiratory system claimed the life of 774 people [2]. By the time, the worst memories of SARS fading away from the minds of people, the year 2019 gave one more contagious, infectious and deadly virus to the world i.e., SARS-CoV-2 (Covid-19)[3]. On December 31, 2019, the first-ever covid-19 case was recorded in Wuhan [4], China and from there it spread across the globe to almost all the countries in the world. The first upswing trend in Covid-19 cases which was described as the first wave of infection caused severe health disasters and deaths in the USA, European countries, and few Asian countries including India. Now the world is witnessing series of upswings of cases which are described as the second wave, third wave, and so on. As of May 2021, 170 million people were infected and 3.68 million people have lost their life due to Covid-19 infection [5].

Apart from creating the health emergency in the world, Covid-19 has affected every sector of society. Millions of people are pushed into poverty and dying out of hunger and starvation. Hence, it is the need of the hour to control and contain the Covid-19 virus. World Health Organization (WHO) and health experts across the world suggest to take steps such as maintaining two meter distance, not going in crowds, using sanitizers and always wearing masks and to use tissue in order to avoid the spread of the virus [6].

As per WHO, the main symptoms of Covid-19 this virus is highly contagious and spreads to one another by the millions of droplets expelled during coughing and sneezing, shaking hands, and exchanging personal accessories with others [7]. Several precautionary measures should be considered to bring down the spreading of this virus, prominent among them are wearing a face mask, frequent washing of hands, maintaining physical distance, avoiding touching of body where there can be possibility of infection, and sanitizing the work place. Among these measures, the most effective one is wearing a face mask. It's very tragic that people are ignorant and not obeying and following the measures given by health authorities and the government.

Hence the health officials and government are enforcing the public to wear masks by putting penalties for those who are not wearing the mask in the public place. But due to the insufficient number of health officials and policemen it has become a herculean task to monitor the people regarding the wearing of masks, hence technology-driven, Mask detecting system would reduce the burden of public authorities and ensure people wear masks.

Face mask discovery is a strategy for deciding if the individual is wearing a mask or not. Recognizing anything from an image is practically identical. Picture acknowledgment frameworks have been presented in an assortment of ways. In clinical applications, deep learning (DL) measures are regularly utilized [8, 9]. DL [10] have as of late exhibited critical importance in object tracking. These constructions can be utilized to identify a mask on an individual's face.Concealing that is detected by image processing and can only cover and interpret the mouth section of the body. Many novel applications of image processing techniques, particularly in the domain of convolution neural networks (CNN), have been introduced with DL and are increasingly becoming reality. Face detection and classification are also included [11]. Image processing is classified as computer graphics, descriptive image conversion, machine acquiring, and output evaluations [12].

II. RELATED WORKS

To detect masked faces, Bu et al. [13] built and implemented a CNN-based framework consisting of three CNNs. Because there aren't enough masked face training examples, a new dataset named "MASKED FACE dataset" was employed to fine-tune the CNN models. Mask -1 was a CNN with a very thin, 5-layered CNN that yields a chance of being a masked face for each target object. Mask-2, the second CNN, has seven layers and reflows the option slots as well as setting a detection rate from the first CNN. Mask-3, the third CNN, that compresses the data frames it obtains and determines based on a current threshold to detect. The anticipated detection results are the detection windows that remain after NMS. The outcomes of this method were good.

Bodan Kwolek proposed a method for detecting different regions of the face by combining a Gabor filter and a convolutional neural network (CNN). This allows the analysis to work at different rates and used in the first phase which exclusively extracts intrinsic facial features and transforms the image into four sub-images. In the second stage, CNN is applied to these four sub-images for face detection. It was concluded that the approach used yielded adequate detection [14].

The goal of Prem et al. [15] was to calculate and anticipate the impact of physical separation on the propagation of the Covid-19. They employed susceptible-exposed-infected-removed (SEIR) models to mimic the outbreak's continuous path utilising specified place way approach. It suggested that early and abrupt reduction of interpersonal isolation rules might contribute to an increase in Covid-19 infections, and proposed gradually relaxing interventions to flatten the infection curve.

Jiang et. al. [16] developed Retina Facemask, an efficient and accurate framework for detecting face masks. Retina Facemask was a single-stage detector that combined a triangle network with numerous feature maps to blend increased spatial meaning with a growing influence focus unit to focus on recognising face masks. A cross-class entity removal technique was also included in the system to discard suggestions with low levels of confidence.

Nair et al. [17] used the Viola-Jones object identification system to recognise masked men in surveillance films in less timeframe. For this, the researcher presented an approach that comprises of four multiple phases: estimating the person's located amidst from the lens, face portion identification. According to the study, identifying elements such as face masks requires substantially longer than detecting faces. To detect the faces and mouths of people, the work proposes the Viola-Jones face detection method. Unless the eye is identified first, then the face, it is assumed that no face mask was applied. If the irises are identified but not the face, it is likely that the person in question was wearing a face mask.

Jagadeeswari and Uday evaluated the performance of smart detection of face approach with various DL classifiers like MobileNet V2, ResNet 50, VGG 16, ADAM, and SGD. For each classifier, optimizers like ADAM, ADAGRAD, SGD (Stochastic Gradient Descent) were used and performance was evaluated. Authors found that, among various optimizers used, the ADAM optimizer was best when performance was considered. Among classifiers, MobileNet V2 was given highly accurate results. [18]

Sandler et al. presented MobileNet V2, a latest smartphone design that enhance the efficiency of phone versions on a variety of tasks, as well as appropriate means for deploying these models to object recognition in a new means. Through the project, they built a face mask identification system based on a ML methods. MobileNetV2 is a CNN image classification method created by Google with enhanced efficiency and additions to become more effective in picture recognition [19].

III. PROPOSED METHODOLOGY

The proposed system is implemented in 2 ways:

A. Software Implementation
B. Hardware Implementation

A. Software Implementation:
Software implementation is split into 3 parts:
1. Face Detection
2. Face Mask Detection
3. Output

1. Face Detection
The proposed methodology tries to recognize if people are wearing masks or not in video footage of a public space. To accomplish so, we first identify the person's face before determining whether or not a facial mask is visible.

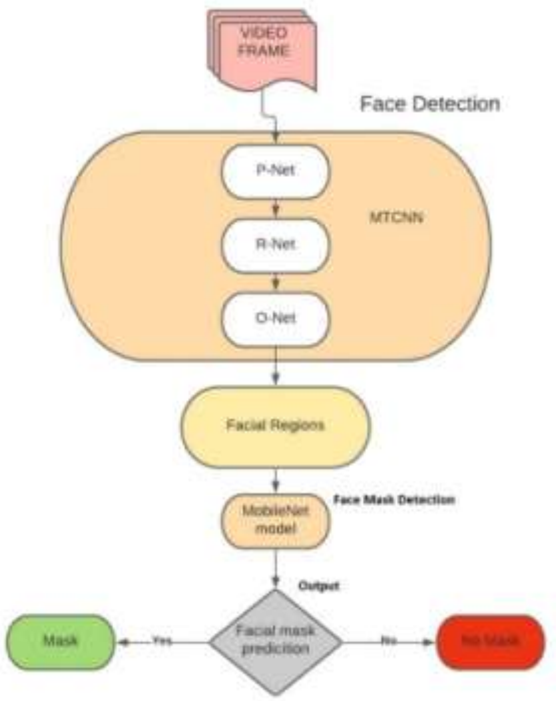

Fig 1: Framework OverView Flow.

The suggested framework seeks to determine the above mentioned activity from a public location. To accomplish so, we first identify the person's face before determining for the presence of mask. The Multitask Cascaded Convolutional Neural Network (MTCNN) was used as the method for this task.

The input image is first scaled in various dimensions to create an image stream, further fed to the network with 3 stages. A Fully Convolutional Network (FCN) is used in the first stage to identify probable face areas based on the input. A CNN called Refine Network (R-Net) is used in the second step to filter these candidate windows. R-Net eliminates a high number of incorrect candidates and calibrates the candidates found using bounding box regression. The divergence between each prospective frame and the closest subsurface is predicted and represented by Libox. The training assignment is a regressive issue, and for each sample Li, the Euclidean loss is determined as follows:

$$L_i^{box} = \left\| \hat{y}_i^{box} - y_i^{box} \right\|_2^2$$

(1)

$$L_i = \|\hat{y}_i - y_i\|_2{}^2 \quad (2)$$

Where,

$\hat{y}_i$=facial landmark coordinate predicted by the network
yi = ground truth coordinate

The third stage includes an O-Net CNN that generates facial landmark locations, such as the eyes, nose, and mouth areas of the face. The detection of face landmarks is a regression problem, similar to bounding box regression, and the following Euclidean loss is minimized:

$$L_i = \|\hat{y}_i - y_i\|_2{}^2$$

$\hat{y}_i$ = target of the network

y_i= ground truth coordinate

Face classification is a binary classification problem and cross-entropy loss for each sample Li is calculated by:

$$L_i = -(y_i \log P_i + (1 - y_i)(1 - \log P_i))$$
(3)

Pi = probability produced by the network that the sample was a face

yi = ground-truth label

The third stage includes an O-Net CNN that produces features of face.

2. Face Mask Detection
For face mask identification and segmentation, we used the MobileNetV2 architecture as shown in Figure 2 (a), which is a powerful feature extractor. Figure 2 (b) shows a single unit in MobileNet, MobileNetV2 contains 3 convolutional layers, these are in single block. The block's source and destination are near zero matrices, whereas the

block's screening is done on strong matrices. In total, there are 17 bottleneck residual blocks in the MobileNetV2 design. After that, there's a normal 1X1 convolution Layer. Following the architecture outlined previously, we construct a facial mask classifier with four layers. To create one lengthy classification pool, we down sample each 2X2 face image using the pooling layer. We utilize a SoftMax function to retrieve the probability distribution across the predicted classes after going into depth with a ReLU feature used in the activation of features[20-21].

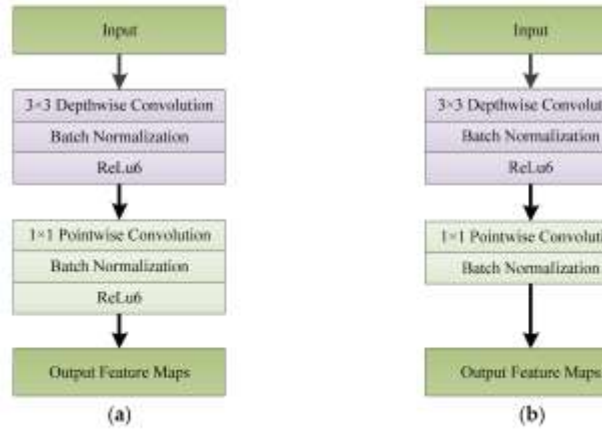

Fig 2. (a) The basic structure of MobileNet: (b) unit in MobileNet.

Dataset:

A dataset is a collection of cases that all have the same feature. In ML and DL, data is extremely crucial. The photograph depicts people of various races, as well as various forms of features of faces and to indicate if the face is worn or not. The dataset includes photos taken with various camera features, a variety of camera angles, varied lighting conditions, noise from kaggle. The data set contains total of 3833 images among which 1918 images were representing unmasked images, 1915 images were masked images. 80% of the entire dataset was used for training and remaining 20%was used for testing.

3. Output:

The described facial mask learner receives the facial regions acquired from the model as input, and the outcome is a limiting circle over each face area, with the tag "Mask" indicating the existence of a face mask and "No Mask" signifying the non-presence of a face mask. Figure 3 shows the construction on face mask classifier.

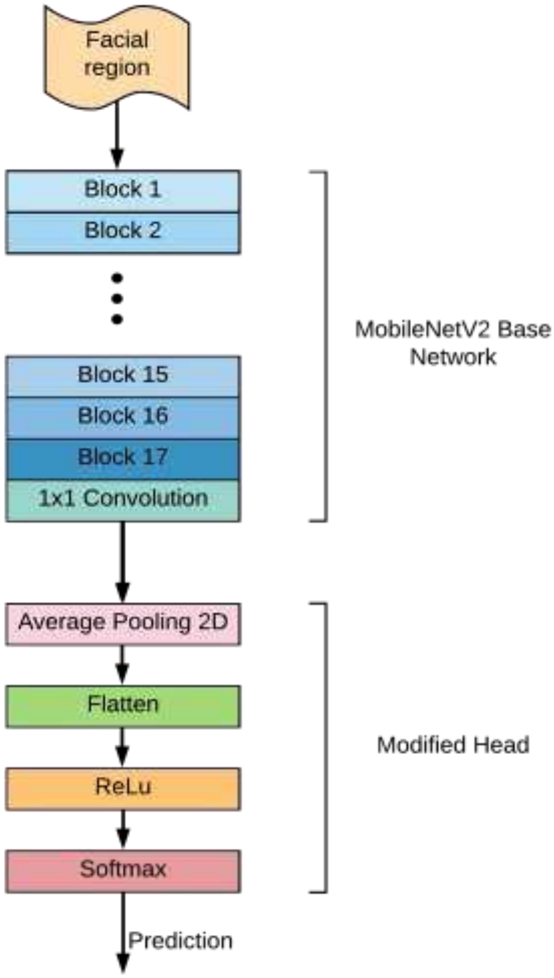

Fig 3: Facial mask classifier constructed using MobileNetV2 architecture.

B. Hardware Implementation

The Raspberry Pi is a family of small solitary processors developed by the Raspberry Pi Foundation in partnership with Broadcom in the UK, and it uses a standard keyboard and mouse. It's a little device that allows individuals to learn about computing and programme in programs such as Scratch and Python. Due to its embrace of HDMI and USB devices, it is mostly utilised by electrical and software hobbyists.

The Module can capture both video and still images in high-definition. It's simple to use for newcomers, but it has a lot of room to grow advanced users who want to learn more. People have used it for various uses. A five-megapixel remedied lens on the module captures video in 1080p30, 720p60, and VGA90 formats, and also screen grabs. It connects to the Raspberry Pi's CSI port via a 15cm ribbon wire. It is accessible via the MMAL and V4L APIs, and numerous third-party programs have been developed for it, including the Picamera Python library. Raspberry pi acts as a medium that takes the live video input from the camera, detects faces and performs the classification and produces the output.

IV. IMPLEMENTATION RESULTS

The following stage is predicting input information from the saved model carried out in the live video frame to frame.

The figure 4 depicts the results obtained in this project. The precision, recall and accuracy is calculated using true positive, false positive, true negative, false negative values. Table I shows the Model Evaluation

TABLE I: Model Evaluation

	Precision	Recall	F1-Sc	Support
With mask	0.99	0.99	0.99	383
Without mask	0.99	0.99	0.99	384
Accuracy			0.99	767
Macro avg	0.99	0.99	0.99	767
Weighted avg	0.99	0.99	0.99	767

TABLE II: Cascaded framework for mask detection correlation with proposed framework

Approach	Accuracy	Recall
Proposed Framework	99%	99%
Cascade Frame for mask detection	86.6%	87.8%

The accuracy of the training model obtained is 99% and the recall is 99%. In the cascade frame for mask detection the accuracy is 86.6% and the 87.8%. Comparing the cascade framework for mask detection with proposed framework is better accuracy.

Fig 4: The results for no mask and wearing mask.

As in the Figure 4, it can be clearly seen that the green box indicates the inclusion of the mask whereas the absence of mask is indicated as red color.

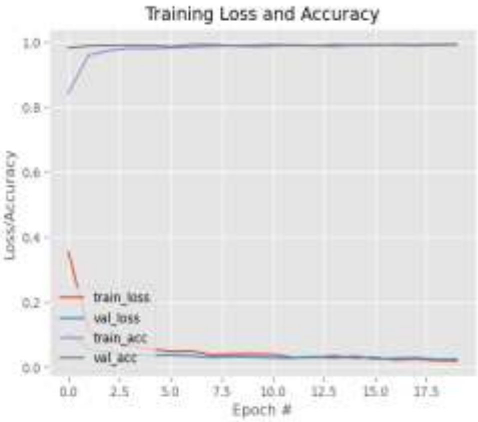

Fig 5: Training loss and accuracy.

The training graph indicates the amount of information lost in the process. It also indicates the amount of accuracy the model has exhibited in detecting the mask on the face. The epochs= 20n and batch size of 32 was used in this process. This method was purposely used, since it yields adequate results.

V. CONCLUSION AND FUTURE SCOPE

A new method for recognizing face masks in live environment is proposed in this paper. Face images and cues are obtained using a highly successful face detection model. Deep learning is used to create a distinct facial classifier for detecting the existence of a face mask in facial photos identified. The combination of Multi-Task Cascaded Convolutional Neural Network and MobileNetV2 tend to be a good fit to solve this problem In MobileNetV2, a superior module is presented with modified residual structure. Non-linearity in narrow layers are taken out this time. With MobileNetV2 as spine for include extraction, best in class exhibitions are likewise accomplished for object discovery and semantic division.Raspberry Pi and pi cam was the best hardware fit for implementing Face mask classification in real world.

REFERENCES

[1]. Newsgram.com. Available at: https://www.newsgram.com/a-brief-history-of-pandemics/?gclid=Cj0KCQjwGFBhDeARIsACH_kdanp4lFWqgF_UdCAFVHQN9CPgovRBFS3GAiv0_2IugHVwkvbU_C2gsaAgzKEALw_wcB (Accessed: August 2, 2021).

[2]. SARS (2021) Cdc.gov. Available at: https://www.cdc.gov/sars/about/fs-sars.html (Accessed: August 2, 2021).

[3]. Coronavirus disease (COVID-19): Virus Evolution Who.int. Available at: https://www.who.int/news-room/q-a-detail/sars-cov-2-evolution (Accessed: August 2, 2021).

[4]. Archived: WHO Timeline - COVID-19 Who.int. Available at: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19 (Accessed: August 2, 2021)

[5]. WHO Coronavirus (COVID-19) Dashboard Who.int. Available at: https://covid19.who.int/?gclid=Cj0KCQjwGFBhDeARIsACH_kdbmXgd1i3BBaEUkqVFowXNvUin1qqzuJ5-e5tzvKcMekydg5gV3fcYaAiiCEALw_wcB (Accessed: August 2, 2021)

[6]. When and how to use masks Who.int. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks (Accessed: August 2, 2021).

[7]. L. Li et al., "COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis," J. Med. Virol., vol. 92, no. 6, pp. 577–583, Jun. 2020

[8]. M. Z. Islam, M. M. Islam, and A. Asraf, "A Combined Deep CNNLSTM Network for the Detection of Novel Coronavirus (COVID-19) Using X-ray Images," Informatics in Medicine Unlocked, vol. 20, pp. 100412, Aug. 2020.

[9]. L. J. Muhammad, M. M. Islam, S. S. Usman, and S. I. Ayon, "Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery," SN Comput. Sci., vol. 1, no. 4, p. 206, Jun. 2020.

[10]. L. Liu et al., "Deep Learning for Generic Object Detection: A Survey," Int. J. Comput. Vis., vol. 128, no. 2, pp. 261–318, Sep. 2018.

[11]. S. Wan, Y. Liang, Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification", Comput. Electr. Eng. 72 (2018) 274–282.

[12]. M. Z. Islam, M. M. Islam, and A. Asraf, "A Combined Deep CNNLSTM Network for the Detection of Novel Coronavirus (COVID-19) Using X-ray Images," Informatics in Medicine Unlocked, vol. 20, pp. 100412, Aug. 2020.

[13]. W. Bu, J. Xiao, C. Zhou, M. Yang, and C. Peng, "A cascade framework for masked face detection," in Proc. of IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), 2017, pp. 458–462

[14]. Bogdan Kwolek, W-Pola 2 "Face Detection Using Convolutional Networks and Gabor Filters". Rzeszao University of Technology, Poland.

[15]. K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday et al., “The effect of control strategies to reduce social mixing on outcomes of the covid19 epidemic in wuhan, china: a modelling study,” The Lancet Public Health, 2020.
[16]. M. Jiang, X. Fan, and H. Yan, “Retinamask: A face mask detector,” 2020.
[17]. A. Nair and A. Potgantwar, “Masked face detection using the viola jones algorithm: A progressive approach for less time consumption,” International Journal of Recent Contributions from Engineering, Science & IT (iJES), vol. 6, pp. 4–14, 12 2018.
[18]. C. Jagadeeswari and M. Uday Theja, “Performance Evaluation of Intelligent Face Mask Detection System with Various Deep Learning Classifiers”. International Journal of Advanced Science and Technology. Vol 29, 2020, P P 3074-3082
[19]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. “Mobilenetv2: The next generation of on-device computer vision networks. URL Https://Ai. Googleblog. Com/2018/04/Mobilenetv2- next-Generation-of-on. 2020.
[20]. Yu D, Xu Q, Guo H, Zhao C, Lin Y, Li D. An Efficient and Lightweight Convolutional Neural Network for Remote Sensing Image Scene Classification. Sensors. 2020; 20(7):1999. https://doi.org/10.3390/s20071999
[21]. A. S. Joshi, S. S. Joshi, G. Kanahasabai, R. Kapil and S. Gupta, "Deep Learning Framework to Detect Face Masks from Video Footage," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), 2020, pp. 435-440, doi: 10.1109/CICN49253.2020.9242625.