International Journal of Engineering, Management and Humanities (IJEMH)
Volume 4, Issue 2, Mar.-Apr. 2023 pp: 101-106 www.ijemh.com

Efficient Deep Learning model with AI strategies for face mask recognition in the time of the Coronavirus pandemic

Rahul Sawhney[1]*
[1] *Department of Computer Science and Engineering, Amity University, Uttar Pradesh, India*

Date of Submission: 24-03-2023 Date of Acceptance: 06-04-2023

ABSTRACT

The ongoing COVID-19 scenario has resulted in countless casualties and security concerns. To stop the coronavirus from spreading, people also cover themselves with masks. Face identification is highly challenging since some features of the face are hidden. To create speedy and effective solutions to this issue amid the current C-19 virus outbreak is one of the key goals of researchers. In this research paper, we offer a robust technique based on discarding the masked region and ML-based features to handle the problem of mask detection operations. We investigated the viability of combing these models with the learned model to achieve exceptionally precise outcomes with speedier inference processes. The inputs are then compared, and the model decides whether the face mask is present using predefined training datasets. In this pandemic period, software for detecting face masks offers wide spectrum of applications.

Keywords: Machine learning, Face Mask, Model, MobileNet, GPU, PyTorch, Deep learning, Convolutional neural networks

I. INTRODUCTION

Suicide risk is raised by depression, a debilitating illness. Due to Covid-19 pandemic, there has been an increase in anxiety, stress, and depression among the populace. University freshmen from countries that have been severely affected by COVID-19 are among those who are most at risk because they must adhere to strict lockdown procedures and have fewer resources to deal with it. The COVID-19 pandemic is currently viewed as one of the biggest risks to global economic and social stability as well as public health and well-being. SARS-CoV-2, a popular coronavirus that first emerged in late 2019 in Wuhan, China, subsequently spreaded itself throughout the world. Even under ideal conditions, interpreting scientific findings can be challenging. It is not surprising that the public is uncertain about the COVID-19 prevention strategies that can effectively safeguard families and communities during a pandemic (and election year). We chose to create a plain-language review of the facts on face coverings, often known as masks, because the understanding of this disease from a scientific and medical perspective is evolving so quickly.

[1] Several public health and hygiene initiatives have been launched; among the most prominent is the use of face masks. Make mask-wearing a common practice when interacting with others. To maximize their effectiveness, masks must be used, stored, cleaned, and disposed of properly. Data science, a fast-expanding area, depends heavily on ML. Algorithms are trained using statistical techniques to offer classifications or predictions and unearth critical insights in data mining operations. [2] The basic goal of ML is to create a system which can inspect from its past performance and carry out tasks on its own, which does not require outside guidance. Algorithms used to create various models are a crucial component of ML. Its algorithms will forecast an outcome for a certain input. Sample datasets are used as inputs to create ML models, and ML techniques are selected, such that they enhance the qualities in the sample data.

According to the most recent WHO report [3], which was released on June 12, 2022, more than 533 crores humans globally have been perished because of coronavirus illness (COVID-19) brought on by acute respiratory syndrome (SARS-CoV2), and over 6.3 million have died as a result. Maintaining social distance, enhancing surveillance, and fortifying health systems are the keys to controlling the COVID-19 pandemic, according to them. People are aware that Covid-19 is stepping into a new realm since it has produced a completely new frequency. Our culture is evolving quickly right now, but we must move quickly to reach the new expectations that are all around us. To make life more meaningful than before [4], a risk-free environment will be everyone's first focus but, few people because of whom the society is not safe. In our homes, we consciously take care of everything, but in public spaces like companies, malls, institutions, etc., it can be more difficult to preserve people's safety. Manually determining whether there is mask or not present on person's face, is not practical. Here, technology plays a role. It helps us to develop a tool that can recognize an image or video of human and conclude whether that person is wearing a mask. Using accessible datasets, we will train our face mask detection model, then evaluate the results on a live webcam.

II. LITERATURE REVIEW

Most projects, in general, concentrate on mask-wearing face construction identification recognition. However, the goal of our project initiative was to identify those who contribute to reducing the transmission and spread of COVID-19 by donning masks or not. Scientists have established that using a mask reduces the rate at which Covid-19 spreads.

In paper [5], Bosheng Qin and Dongxiao Li created a new technique for identifying facemask-wearing conditions. They were able to divide the use of facemasks into three groups. Correct facemask use, wrong facemask use, and no facemask use are the three categories. The suggested algorithm's face detection phase accuracy was 98.70 percent. Principal Component Analysis (PCA) technique was implemented by Sabbir in his research paper [6] , to recognize the person using both masked and unmasked facial recognition techniques. In Study, he found that the usage of masks had a profound effect on the accuracy of facial resonance using the PCA. When identified face was concealed, accuracy of recognition fell under 70%. In this study, they used PCA to remove glasses from a facial image while compensating for recursion errors [7]. The authors suggested technique to eliminate spectacles from a front-facing portrait of a person. Using PCA reconstruction and recursive error compensation, the removed portion was rebuilt.

The YOLOv3 algorithm was employed by Li, Chong in their facial recognition model research paper [8], whereby YOLOv3's Darknet-53 served as the foundation. The accuracy of the suggested method was 93.9 percent. More than 600,000 photos from the WIDER FACE and CelebA datasets were used for training. The FDDB dataset was tested. A unique GAN-based network was suggested by Nizam e in his work [9] that could remove mask from the facial part on its own and then recreate the image by creating hole. The paper, entitled - *System for Medical Mask Detection in the Operating Room Through Facial Attributes*, described a system for determining whether a mandatory medical mask is present or not in the operating room [10]. Shashank D. Joshi, Anushka G. Sandesara, in their study [14] proposed a layered 2D-Convolution model for facial mask detection that was quite effective. The suggested method, which consists of layers of 2D-Convolutional layers with ReLU activations and Max Pooling, was developed using Gradient Descent as training method and Binary Cross-Entropy as a loss function. They achieved an overall testing accuracy of 95% and a training accuracy of 97%. Suryo Adi Rakhmawan and Samuel Ady Sanjaya in their work [18], conducted research on face mask detection. The data was used to plan for coronaviruses' reduction, prevention, evaluation, and reaction. In this research, mask detection is carried out utilizing MobileNetV2, an ML system. CNN is the foundation for the image classification method MobileNetV2, which is used to categories images built on Convolution Neural Networks. The instance model's implementation achieved a 96.85 percent accuracy rate when determining whether persons are wearing masks.

A computer-vision [19] must deal with the intrinsic tasks of pattern learning and object recognition including Image-categorization and object detection. An effective object identification algorithm can be used with surveillance equipment to identify the mask hiding the face in the public space. Pipeline for recognizing objects starts by creating region proposals, then classifying each proposal into a relevant category. Human faces in a frame or image are located using an algorithm in the Viola — Jones research. Some characteristics of human faces are universal, such as eye region being less bright than the nose area and the eyes region being darker than its neighbor pixels [20].

III. METHODOLOGY

3.1. Dataset and Data Manipulation

For differentiating between masked and unmasked peoples by machine we took a dataset from Kaggle in which total images are 1,376 and from this 690 are with mask and 686 are without mask. To achieve our goal, we first loaded the images to a particular path and then pre-processing the data, [21] which include resizing of images (we are setting our target size as 224*224 pix converting images to array for storing it inside the list but for doing some operations on it in deep learning, we must convert it into NumPy array. Now, we will divide our data into two parts in testing and training dataset using sklearn: for training of the model 80% of the data has been allotted, while the remaining 20

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 4, Issue 2, Mar.-Apr. 2023 pp: 101-106 www.ijemh.com

percent is used for testing, resulting in a split ratio of 0.8:0.2.

3.2. Model Architecture

In our paper, we have used MobileNetV2 model. In practical applications, MobileNet is a CNN architecture that is effective and portable. To create lighter models, MobileNet generally replace the typical convolutions employed in earlier architectures with depth wise separable convolutions. According to their needs, model creators can trade off latency or accuracy for speed and small size using the two new global hyperparameters introduced by MobileNet, which are: width multiplier and resolution multiplier. Convolutional layers which are depth-separable provide the foundation of MobileNet. A pointwise convolution and a depth wise convolution are the two types of convolutions that make up each depth wise separable convolution layer. There are 28 levels in a MobileNet if pointwise and depth wise convolutions are counted as separate layers. By appropriately adjusting the width multiplier hyperparameter, the standard Mobile Net's 4.2 million parameters can be further decreased. The input image measures 224*224*3 pixels in size. MobileNetV2 is an open-source transfer fully convolutional framework architecture developed and maintained by google. It was introduced for efficient, on-device and fast computer vision applications. It is an enhanced version of MobileNetV1, and its pre-trained version can be easily loaded on a million images of the ImageNet dataset such as keyboard, mouse, pen etc. The initial fully convolution layer contains 32 filters, followed by 19 residual bottleneck layers making it 53 layers deep.

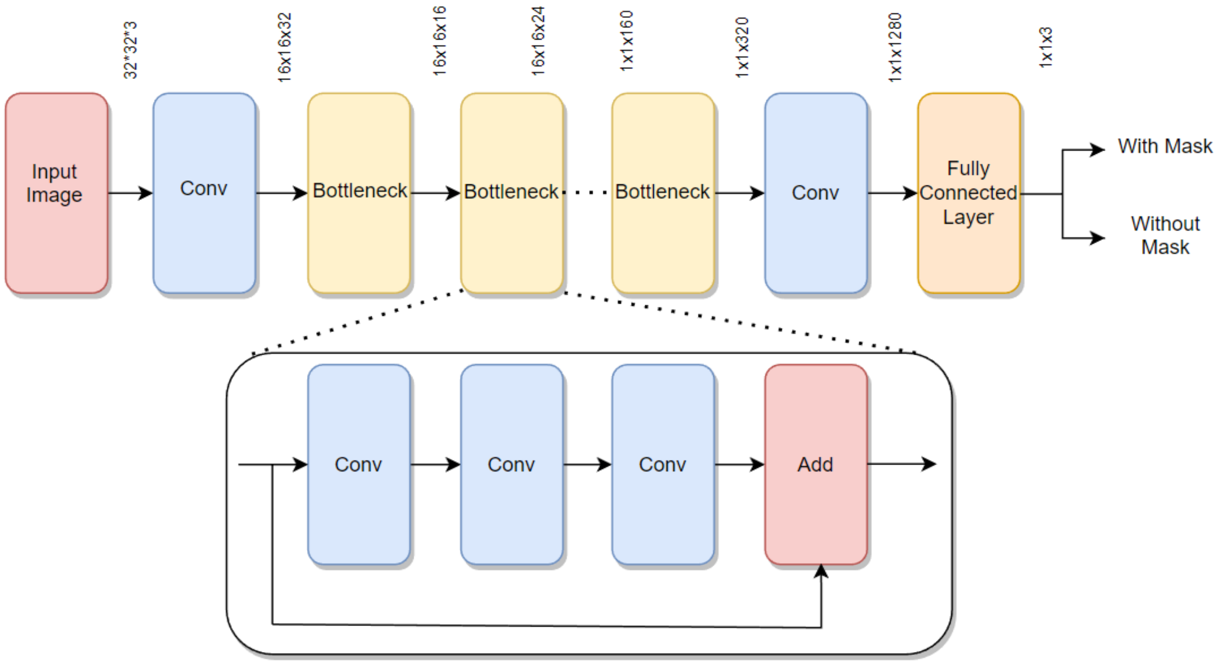

Fig 1. Model architecture

Each line in this example represents an x-times repeating succession of one or more identical layers. For all layers in the same sequence, output channels are the same. In the first layer of each sequence, stride is utilised, while stride 1 is used in the following levels. In every case, the spatial convolutions employ 3x3 kernels. Except for the initial layer, the network has been grown at a constant rate. In experiments, growth rates between 5 and 10 create performance curves that are almost identical, with smaller networks performing marginally better with lower expansion rates and larger networks performing marginally better with higher expansion rates. In order to support the forthcoming wave of mobile vision apps, MobileNetV2 is now accessible. The state-of-the-art for mobile visual identification is advanced by MobileNetV2, which performs noticeably better than MobileNetV1 in terms of classification, object detection, and semantic segmentation. To improve on the ideas from MobileNetV1, Deep-separable convolution is employed as an efficient building block in MobileNetV2. Overall, the MobileNetV2 models are quicker for the same accuracy over the whole latency range.

The new models improve accuracy over MobileNetV1 models while utilising 2x fewer processes, 30% fewer parameters, and being between 30 and 40% quicker on a Google Pixel phone. MobileNetV2 is a powerful feature extractor for object detection and segmentation. Its input is an image with a default size of 224 by 224. There are two different kinds of blocks in the MobileNetV2 architecture [22]. The residual block is in the first block, while the shrinking block is in the second. Each block in the architectural diagram below is made up of three different types of layers, the first of which is a 1 by 1 convolution layer with the activation function being the Rectified Linear Unit 6. A depth wise convolution makes up the second layer. In turn, the third layer is a 1 by 1 convolution layer again but without any non-linearity. The MobileNetV2 architecture is a very lightweight one when compared to other transfer learning models such as VGG16, Dense Net's, Inception Net's etc. and is only 14 MB in size. It has a total of 3.5 million parameters and boasts a top 5% accuracy of 90.1%. It is one of the fastest transfer learning models and is used in various applications.

IV. RESULTS AND EVALUATION

The outcomes are more in line with what the model predicted. The camera is used to implement the mask recognition, and the results are precise. Model will recognize a person's face when it is in the camera's field of view, and a green or red frame will appear over the face. The model's accuracy is 96 percent. Additionally, it is seen in the accuracy loss graph that was created below in Fig 2 that the Loss value is founded to be less than 0.1 and as one can see there is improvement in the accuracy and the loss gradually decreases with each iteration of the model training. The video feed is used to test the output wheatear the person is wearing mask or not and in percentage it tells the proper way of wearing mask which shows a green rectangle in frame with the chance of being correct is drawn over the face to signify a face wearing a mask and a red rectangle in the frame is drawn across a face that does not have a mask, along.

Deep learning models are usually trained using the stochastic gradient descent optimization algorithm. Learning rate is a crucial part of this algorithm. It is a hyperparameter which can make or break the results achieved by the model. Thus, it is crucial to monitor and investigate the effects of the learning rate on the model’s performance and to understand its nature. The learning rate plot helps us visualize how changing the learning rate of the model affects its performance, thus helping us identify the best learning rate to choose for our model training. Learning curves is one of the most used tools when it comes to deep learning and machine learning when the model learns incrementally with time or number of epochs. One of these learning curves is the Loss curve. It depicts our model’s error over the passage of time or training epochs. An ideal loss curve is one which keeps decreasing as the model learns and does not have a lot of fluctuations.

A confusion matrix is yet another graph widely used graph or matrix in machine learning and deep learning problems. It is used for classification problems and is used to define the performance of a classification algorithm across various classes in the dataset. It attempts to visualize and summaries the performance of the classification model. There are four basic components of a confusion matrix: TP (True Positive): The model’s positive prediction for a class is true or correct. FP (False Positive): The model’s positive prediction for the class is false or wrong. TN (True Negative): The model’s negative prediction for a class is true or correct and FN (False Negative): The model’s negative prediction for a class if false or wrong. The evaluation report also attempts to summaries the performance of the model using various values such as the precision, recall, f-1 score and the support. These values are of crucial significance and tell things about the model or the dataset which can never be seen using just the accuracy of the model. Precision is the true positive predictions divided by the total positives detected by the model. Recall is the true positive predictions divided by the total positives present in the dataset. F-1 Score is the weighted average of the precision and recall values. Support for a class is the total number of actual occurrences of that class in the dataset.

The technique is effective at detecting faces that are partially hidden by a mask, hair, or hand. To distinguish between a hand-covered face and an annotated mask, it considers the degree of occlusion in four regions: the nose, mouth, chin, and eye. A mask that completely covers the face, including the chin and nose, will only be regarded as a 100% mask by the model. The method's main problems are largely a variety of viewpoints and a lack of clarity. It is more challenging because of the blurry moving faces in the video stream. Following the motions of multiple video frames, however, enables one to make a better choice between "with mask" and "without mask."

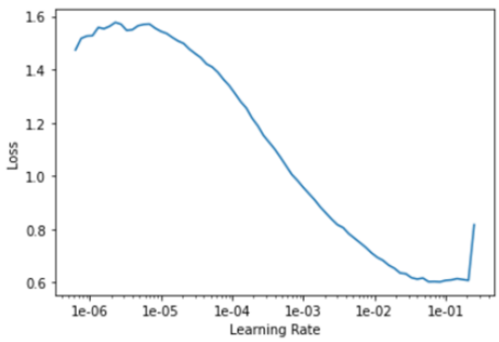

Fig 2. Model Learning Rate Plot

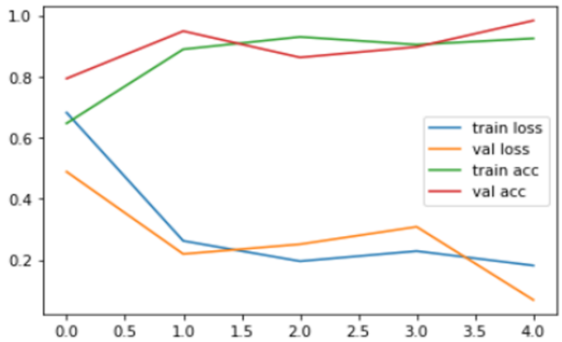

Fig 3. Model Loss Plot

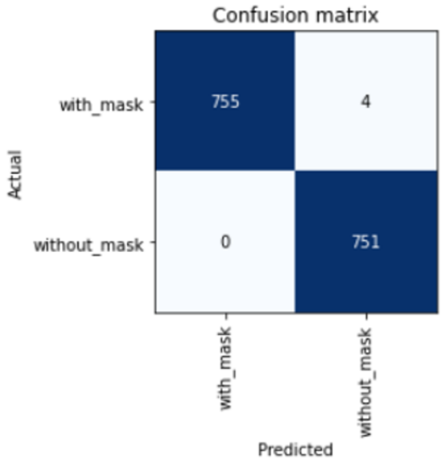

Fig 4. Confusion Matrix

	Precision	Recall	F-1 Score	Support
With Mask	0.99	1.00	0.99	759
Without Mask	0.99	1.00	0.99	751
Accuracy		0.99		1510

Table 1. Evaluation Report

V. CONCLUSION AND FUTURE WORKS

In numerous recent studies, the issue of face mask recognition in the wild had been investigated, and the associated face detectors had been tested on datasets of average faces. Some face detectors have shown exceptionally good performance on the datasets that are now accessible, and it seems difficult to further improve them. Images are used to test the two models of this system, improving precision, and optimizing each model. The database successfully classifies a person's face to produce the name tag and success likelihood. The "actual wild" scenarios, however, present even greater difficulties since they involve faces that are taken at unexpected resolutions and with unexpected illumination and occlusion. One of the best systems now in use, according to the MobileNetV2 classifier, will be integrated along with an interface for warning and alerting systems in upcoming generations. This system will be combined with the system for enforcing social distance to create a comprehensive system that can have a significant impact on the spread of. According to experts, CNN, wearing a face mask is the greatest way to stop the transmission of airborne viruses like Corona. However, this poses a serious security risk to the country because it would present a huge opportunity for those who hide their faces for evil purposes. Experts also warn that the widespread use of masks could make it more difficult to solve crimes in the future because facial recognition is crucial for locating culprits. When the COVID-19 epidemic is gone, this system kicks in for chemical plants, banks, glass manufacturers, etc.

REFERENCES

[1]. Inamdar, Madhura, and Ninad Mehendale. "Real-time face mask identification using facemasknet deep learning network." Available at SSRN 3663305 (2020).
[2]. Jiang, Mingjie, Xinqi Fan, and Hong Yan. "Retinamask: A face mask detector." arXiv preprint arXiv:2005.03950 (2020).
[3]. Covid, CDCx, et al. "Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019—United States, February 12–

March 28, 2020." Morbidity and Mortality Weekly Report 69.13 (2020): 382. - Google Search. (n.d.) https://www.paho.org/en/news/2-6-2020-social-distancing-surveillance-and-stronger-health-systems-keys-controlling-covid-19

[4]. Javed, Iram, Muhammad Atif Butt, and Samina Khalid. "Face Mask Detection and Social Distance Monitoring System for COVID-19 Pandemic." (2022).

[5]. Qin, Bosheng, and Dongxiao Li. "Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19." Sensors 20.18 (2020): 5236.

[6]. Shabbir, Shagufta H., et al. "Pattern-based recognition for the rapid determination of identity, concentration, and enantiomeric excess of subtly different threo diols." Journal of the American Chemical Society 131.36 (2009): 13125-13131.

[7]. Park, Jeong-Seon, et al. "Glasses removal from facial image using recursive PCA reconstruction." International Conference on Audio-and Video-Based Biometric Person Authentication. Springer, Berlin, Heidelberg, 2003.

[8]. Li, Chong, et al. "Face detection based on YOLOv3." Recent Trends in Intelligent Computing, Communication and Devices. Springer, Singapore, 2020. 277-284.

[9]. Din, Nizam Ud, et al. "A novel GAN-based network for unmasking of masked face." IEEE Access 8 (2020): 44276-44287.

[10]. Nieto-Rodriguez, Adrian, Manuel Mucientes, and Victor M. Brea. "System for medical mask detection in the operating room through facial attributes." Iberian Conference on Pattern Recognition and Image Analysis. Springer, Cham, 2015.

[11]. Khan, Muhammad Kamran Javed, et al. "Interactive removal of microphone object in facial images." Electronics 8.10 (2019): 1115.

[12]. Hussain, Shaik Asif, and Ahlam Salim Abdallah Al Balushi. "A real time face emotion classification and recognition using deep learning model." Journal of physics: Conference series. Vol. 1432. No. 1. IOP Publishing, 2020.

[13]. Shete, Sanket, et al. "Mask Detection and Tracing System." (2021).

[14]. Sandesara, Anushka G., Dhyey D. Joshi, and Shashank D. Joshi. "Facial Mask Detection Using Stacked CNN Model." Int. J. Sci. Res. Comput. Sci. Eng. Inform. Technol (2020).

[15]. Guillermo, Marielet, et al. "COVID-19 Risk Assessment through Multiple Face Mask Detection using MobileNetV2 DNN." The 9th International Symposium on Computational Intelligence and Industrial Applications (ISCIIA2020), Beijing, China. https://isciia2020. bit. edu. cn/docs/20201114082420135149. pdf. 2020.

[16]. Bhadani, Amrit Kumar, and Anurag Sinha. "A facemask detector using machine learning and image processing techniques." Eng. Sci. Technol. Int. J (2020): 1-8.

[17]. Rosebrock, A. "COVID-19: Face Mask Detector With Opencv, Keras/Tensorflow, And Deep Learning-Pyimage search.[online] PyImageSearch." (2020).

[18]. Sanjaya, Samuel Ady, and Suryo Adi Rakhmawan. "Face mask detection using MobileNetV2 in the era of COVID-19 pandemic." 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI). IEEE, 2020.

[19]. Qiao, Siyuan, et al. "Few-shot image recognition by predicting parameters from activations." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[20]. Nerpagar, Tejal, et al. "Face mask recognition using machine learning." International Journal of Innovative Research In Technology, IJIRT 7.7 (2020): 43-45.

[21]. Bhadani, Amrit Kumar, and Anurag Sinha. "A facemask detector using machine learning and image processing techniques." Eng. Sci. Technol. Int. J (2020): 1-8.

[22]. Kadam, Kalyani, et al. "Detection and localization of multiple image splicing using MobileNet V1." IEEE Access 9 (2021): 162499-162519.

[23]. Talahua, Jonathan S., et al. "Facial recognition system for people with and without face mask in times of the covid-19 pandemic." Sustainability 13.12 (2021): 6900.

[24]. Yang, Guanhao, et al. "Face mask recognition system with YOLOV5 based on image recognition." 2020 IEEE 6th International Conference on Computer and Communications (ICCC). IEEE, 2020.