Face mask detection using MobileNet and Global Pooling Block

Isunuri B Venkateswarlu
Computer Science and Engineering
IIITDM Kancheepuram
Chennai, India
coe19d001@iiitdm.ac.in

Jagadeesh Kakarla
Computer Science and Engineering
IIITDM Kancheepuram
Chennai, India
jagadeeshk@iiitdm.ac.in

Shree Prakash
Computer Science and Engineering
IIITDM Kancheepuram
Chennai, India
coe19d002@iiitdm.ac.in

***Abstract*—Coronavirus disease is the latest epidemic that forced an international health emergency. It spreads mainly from person to person through airborne transmission. Community transmission has raised the number of cases over the world. Many countries have imposed compulsory face mask policies in public areas as a preventive action. Manual observation of the face mask in crowded places is a tedious task. Thus, researchers have motivated for the automation of face mask detection system. In this paper, we have presented a MobileNet with a global pooling block for face mask detection. The proposed model employs a global pooling layer to perform a flatten of the feature vector. A fully connected dense layer associated with the softmax layer has been utilized for classification. Our proposed model outperforms existing models on two publicly available face mask datasets in terms of vital performance metrics.**

***Index Terms*—Face mask detection, Transfer learning, MobileNet, Global Pooling**

I. Introduction

Coronavirus disease (COVID-19) is the latest epidemic caused by the newly discovered coronavirus [1]. COVID-19 is an infectious disease that affects the respiratory system caused by the SARS-CoV-2 virus. It mainly spreads from person to person by airborne transmission, especially through close contact. During this pandemic time, a wide variety of artificial intelligence-based research proposals have been reported apart from clinical research. Martin *et al.* [2] have implemented a digital health assistant known as Symptoma. Zhou *et al.* [3] have presented how to use artificial intelligence for accelerating drug repurposing or repositioning. Another artificial intelligence-based tool for drug discovery has been presented by Aman *et al.* [4]. Abba *et al.* [5] have proposed DeTraC model for COVID-19 detection from X-ray images.

Recently, community transmission has raised the number of cases in most of the courtiers. According to the World Health Organization (WHO), around 49 million confirmed cases have been reported globally [6]. Due to the outbreak of COVID-19, the WHO has issued several precautionary guidelines to fight against the spread of coronavirus. Social distancing, sanitization, and wearing masks are the most noticeable guidelines. Wearing a face mask slows the community transmission of the corona. Thus, the majority of the countries have enforced compulsory face mask policies in public areas [1]. Manual observation of the face mask is a tedious task especially in crowded places such as hospitals, airports, railway stations, and shopping malls. This motivated researchers to automate face mask detection system. Jignesh *et al.* [7] have employed transfer learning of InceptionNet for face mask detection. Loey *et al.* [8] have proposed a hybrid deep transfer learning model for the detection of face mask. We also have worked in a similar direction and proposed MobileNet with a Global pooling block model for face mask detection.

The rest of the paper has organized as follows; Section 2 describes the details of the proposed MobileNet and Global Pooling model. Section 3 presents a detailed discussion of results and Section 4 concludes the paper.

II. Methodology

Deep convolution neural networks (CNN) have become a predominant tool for computer vision tasks such as image classification. There are several successful deep CNN models available as follows. Kaiming *et al.* [9] have proposed five variants of ResNet as ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. They attain a top-1 error rate of 24% and 22% with ResNet-18/34 and ResNet-50/101/152, respectively. It shows that the performance and computational time of the model are directly proportional to the number of layers. The architecture of ResNet-50 is similar to ResNet-34 and causes less computation time than other variants of ResNet.

Christian *et al.* [10] have devised an inception network to achieve better accuracy for image classification and segmentation. In general, convolution with larger spatial filters tends to high computational cost. One prominent solution to reduce this cost is inception modules. Inception reduces the cost by finding out optimal local sparse structures. The idea of the inception block is to design a layer by layer construction with the analysis of layer correlation statistics. The clusters of highly correlated layers are used to form groups of units. Each unit from an earlier layer corresponds to some region of the input image is referred to as a filter bank. This process ends up with the concatenation of huge filter banks from a single region.

Transfer learning of these models has become handy for image classification. Thus, we have proposed a MobileNet and

978-0-7381-2447-6/20/$31.00 © 2020 IEEE

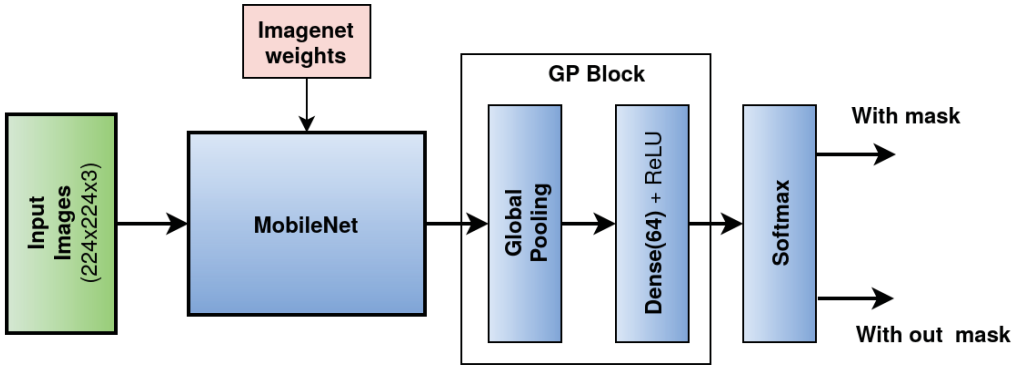


Fig. 1: Proposed MobileNet and Global Pooling

Global Pooling (MNGP) model for face mask detection. The proposed model uses transfer learning with the global pooling block and details are as follows.

A. Transfer learning

Transfer learning has been recognized as one of the common techniques for computer vision tasks such as classification and segmentation. It is the process of sharing weights or knowledge obtained while solving one problem to solve other related problems. In general, transfer learning reduces training time if the application domains are closely related. There are two common approaches to perform transfer learning as follows.

1) **Using pre-trained model**: Pre-trained model is the model that has been trained on a large-scale benchmark dataset. There are several pre-trained models including ResNet, MobileNet, GoogleNet which have trained on a large Imagenet dataset. These models accept color images as input and perform 1000 class classification. This model gives the best accuracy when we are interested to classify any class defined in the imagenet.
2) **Defining custom output layer**: In this method, the pre-trained model is considered excluding the output layer which acts as a feature extractor. These features can be used for the required classification task using a custom output layer. Finally, the custom model needs to be trained for improving the classification results.

B. Global Pooling block

Global pooling (GP) block takes the multi-dimensional feature map and converts it to a one-dimensional feature vector using global pooling. Global pooling layer transforms $(M \times M \times N)$ feature map to $(1 \times N)$ feature map where $(M \times M)$ is the image size, and N is the number of filters. There are two types of pooling namely global max pooling (GMP) and global average pooling (GAP). The global pooling layer has several benefits as follows.

1) It acts as a flatten layer for the transformation of a multidimensional feature map into a one-dimensional feature vector.
2) There is no parameter to optimize, and hence overfitting is avoided at this layer [11].

The one-dimensional feature vector is passed to a fully connected dense layer with 64 neurons.

C. Proposed MobileNet and Global Pooling

Our objective is face mask detection from the natural images and hence we have devised the transfer learning with a global pooling block as an output layer. We have selected a pre-trained MobileNet to reduce computational cost. Depthwise separable convolution is the fundamental unit of the MobileNet model. It uses the factorization of filters to reduce computational cost and model size. Depthwise separable convolution is made up of two layers as follows.

- **Depthwise convolution**: It is used to apply a single filter to each input channel and acts as layer filtering.
- **Pointwise convolution**: It is used to create a linear combination of the output with the help of 1×1 convolution.

The pre-trained MobileNet has been intended for 1000 class classification. Thus, we have replaced the output layer with the global pooling (GP) block as our objective is a binary classification. The steps for designing the proposed model are as follows.

1) Pre-trained MobileNet without the output layer accepts input image of size (224, 224, 3). and generates a feature map of size (7, 7, 1024).
2) Global pooling block (GP Block) transforms a multi-dimensional feature map into a one-dimensional vector having 64 features.
3) Finally, a softmax layer with two neurons takes 64 features and performs binary classification as shown in Fig. 1.

III. Results and Discussion

We have considered Jignesh *et al.* [7] for performance comparison. We also have selected three popular pre-trained deep networks namely RestNet50, MobileNet, GoogleNet for evaluation of proposed model.

A. Dataset description

We have selected two publicly available datasets for the evaluation of the proposed model. Details of the datasets are as follows:

- Dataset 1 (DS1) [12]: This dataset consists of 3833 color images. Out of which 1918 images are without a mask and the remaining 1915 images are with a mask. All of the images present in this dataset are real-time images.
- Dataset 2 (DS2) [13]: This dataset consists of 1650 color images. Out of which 824 images are without a mask and the remaining 826 images are with a mask. Rotation-based data augmentation has been performed to generate the dataset. Further, mask image has imposed over the face images to generate masked images. Thus, this dataset is also referred to as a simulated masked face dataset (SMFD).

B. Experimental Setup

This section presents the experimental setup of the proposed model. We have conducted our experiments using Intel Xeon processor with 16GB GPU. Our proposed model has simulated using Python and Tensorflow. Each model has trained for 50 epochs. Sparse categorical cross-entropy loss has employed while training the model. Adam is the simple and time-efficient optimizer for deep neural networks. Thus, we have utilized the adam optimizer for the compilation of our proposed model. Complete hyperparameter settings of the proposed model are listed in Table I.

TABLE I: Hyperparameter settings of the proposed model

Hyperparameter	Value
Number of epochs	50
Batch size	8
Optimizer	Adam
Initial learning rate	0.0001

C. Results analysis

There are several measures to exhibit the performance of classification results. We have considered four key metrics such as accuracy, precision, recall, and jaccard similarity. "These metrics have computed using true positive, true negative, false positive, and false negative obtained from confusion matrix". "Accuracy is the sum of true positive and true negative samples over the total number of samples".

$$Precision = \frac{\mathcal{TP}}{\mathcal{TP} + \mathcal{FP}} \tag{1}$$

$$Recall = \frac{\mathcal{TP}}{\mathcal{TP} + \mathcal{FN}} \tag{2}$$

The percentage of relevant classification results is referred to as Precision and can be computed using Eq. 1. Similarly, the percentage of total relevant results correctly classified by the model is referred to as Recall and can be computed using Eq. 2. Jaccard index denotes similarity and diversity of classification results and it can be measured using eq. 3.

$$Jaccard = \frac{\mathcal{TP}}{\mathcal{TP} + \mathcal{FP} + \mathcal{FN}} \tag{3}$$

TABLE II: Performance comparison on Dataset 1

Approach	Accuracy	Precision	Recall	Jaccard
ResNet50	84.17	84.17	84.17	72.67
MobileNet	98.96	98.96	98.96	97.93
GoogleNet	90.78	91.29	90.78	83.04
Jignesh *et al.*	98.09	98.09	98.09	96.25
MobileNet+GAP	99.48	99.48	99.48	98.96
MobileNet+GMP	99.56	99.56	99.56	99.13

TABLE III: Performance comparison on Dataset 2

Approach	Accuracy	Precision	Recall	Jaccard
ResNet50	98.38	98.39	98.38	96.82
MobileNet	99.80	99.80	99.80	99.59
GoogleNet	98.18	98.19	98.18	96.43
Jignesh *et al.*	99.80	99.80	99.80	99.60
MobileNet+GAP	100.00	100.00	100.00	100.00
MobileNet+GMP	99.39	99.39	99.39	98.79

In our experiments, we have divided the data into 70% train data and 30% test data. Train data has been used for both training and validation of the model while the test data have been used for performance computation. Fig. 2 depicts training accuracy and loss of the proposed model on DS1. It can be observed that the proposed model attains maximum performance after 10 epochs and maintains consistent learning. Table II compares the performance of the proposed model on DS1 with existing models. We have experimented with both global average pooling and global max pooling. The proposed MNGP model outperforms its competing methods with 99% accuracy. However, MobileNet with global max pooling records best perform of 99.56%. Our proposed model exhibits similar performance in other metrics.

Similarly, Fig. 3 visualizes the training accuracy and loss of the proposed model on DS2. In this case, also our proposed model achieves fast and consistent learning. The global pooling block employed in the proposed model forces fast learning. Table III lists the performance of the proposed model on DS2 its competitive models. The proposed MNGP model exhibits superior performance with 100% accuracy with global average pooling.

The confusion matrix visualizes the complete details of the classification. Fig. 4 displays confusion matrix of the proposed model with GAP and GMP layers on DS1. The proposed model has wrongly classified six images on DS1 with MNGAP as shown in Fig. 4 (a). Five images have wrongly classified with MNGMP on DS1 as shown in Fig. 4 (b). Similarly, Fig. 5 displays confusion matrix of the proposed model with GAP and GMP layers on DS2. In this case, MNGAP classifies all images correctly while three images have been wrongly classified by MNGMP. We have evaluated the performance

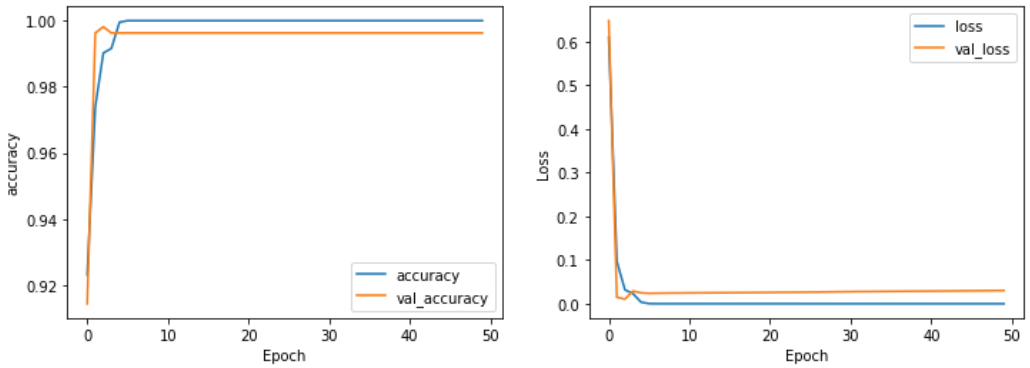


Fig. 2: Training accuracy and loss of proposed model on Dataset 1

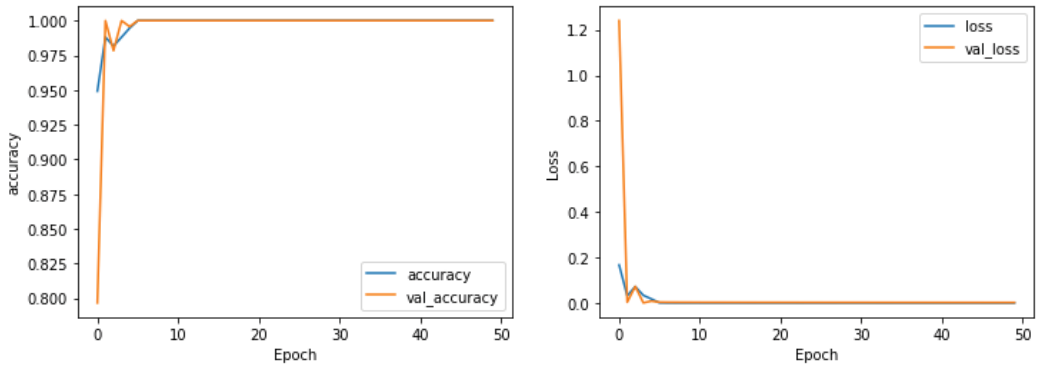
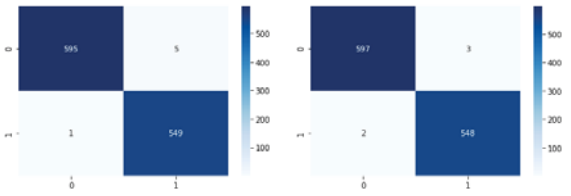
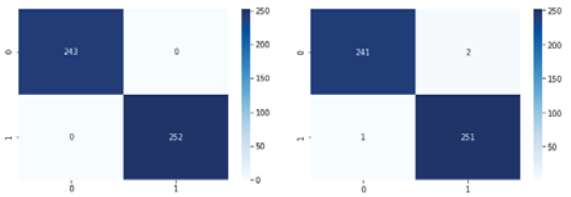


Fig. 3: Training accuracy and loss of proposed model on Dataset 2



(a) (b)

Fig. 4: Confusion matrix of proposed model on Dataset 1 (0 denotes without mask; 1 denotes with mask)



(a) (b)

Fig. 5: Confusion matrix of proposed model on Dataset 2 (0 denotes without mask; 1 denotes with mask)

of the proposed model on low-quality images. Our proposed

Fig. 6: Performance of proposed model on low quality images

model attains similar performance on low-quality images as shown in Fig. 6.

D. Time complexity

Time complexity is another important factor in deep network performance. Designing time-efficient models without compromising performance is a challenging task in deep networks. Table IV lists the number of parameters and training time of the proposed model along with its competitive models. Our proposed model takes around 3.3M trainable parameters like pre-trained MobileNet. Thus, the proposed model outperforms the existing model in the number of parameters as well as training time.

TABLE IV: Comparison of training time

Approach	# Parameters	Training time (sec.)	
		Dataset 1	Dataset 2
ResNet50	23.8 M	902	402
MobileNet	3.3 M	601	251
GoogleNet	21.9 M	902	402
Jignesh *et al.*	22.1 M	902	402
MobileNet+GP	3.3 M	601	251

E. Discussion

The proposed MNGP model outperforms the existing models. In addition to the performance, our proposed model has several advantages as follows.

- Global pooling employed in the proposed model avoids the overfitting of the model.
- Proposed model outperforms existing models on DS1 and DS2 in considered performance metrics.
- Proposed model uses only 3.3M trainable parameters as it utilizes pre-trained MobileNet features.
- Our proposed model takes less time for training and testing as compared to existing models.

IV. Conclusions

An outbreak of COVID-19 has forced the majority of the countries to enforce the compulsion of wearing a face mask. Manual observation of the face mask in crowded places is a critical task. Thus, researchers have motivated for the automation of face mask detection system. In this paper, we have proposed a pre-trained MobileNet with a global pooling block for face mask detection. The pre-trained MobileNet takes a color image and generates a multi-dimensional feature map. The global pooling block that has been utilized in the proposed model transforms the feature map into a feature vector of 64 features. Finally, the softmax layer performs binary classification using the 64 features. We have evaluated our proposed model on two publicly available datasets. Our proposed model has achieved 99% and 100% accuracy on DS1 and DS2 respectively. The global pooling block that has been used in the proposed model avoids overfitting the model. Further, the proposed model outperforms existing models in the number of parameters as well as training time. Our future work focuses on face mask detection over multi-face images.

References

[1] X. Liu and S. Zhang, "Covid-19: Face masks and human-to-human transmission," *Influenza and Other Respiratory Viruses*, 2020.

[2] A. Martin, J. Nateqi, and S. Gruarin, "An artificial intelligence-based first-line defence against covid-19: digitally screening citizens for risks via a chatbot," 2020.

[3] Z. Yadi, W. Fei, T. Jian, N. Ruth, and C. Feixiong, "Artificial intelligence in covid-19 drug repurposing," *The Lancet Digital Health*, 2020.

[4] A. C. Kaushik and U. Raj, "Ai-driven drug discovery: A boon against covid-19?" *AI Open*, vol. 1, pp. 1 – 4, 2020.

[5] A. Abbas, M. Abdelsamea, and M. Gaber, "Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network," *Applied Intelligence*, 2020.

[6] WHO, "World health organization," 2020, accessed 17 Oct 2020. [Online]. Available: https://www.who.int/health-topics/coronavirus

[7] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Face mask detection using transfer learning of inceptionv3," 2020.

[8] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic," *Measurement*, vol. 167, p. 108288, 2021.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[10] C. Szegedy, V. V. andSergey Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015.

[11] M. Lin, Q. Chen, and S. Yan, "Network in network," 2014.

[12] FMD, "Face mask detection," 2020, accessed 17 Oct 2020. [Online]. Available: https://github.com/chandrikadeb7/Face-Mask-Detection

[13] SMFD, "Simulated masked face dataset," 2020, accessed 17 Oct 2020. [Online]. Available: https://github.com/prajnasb/observations