SETSCI Conference Proceedings
14 ,16-19, 2022

6th International Symposium on Innovative Approaches in Smart Technologies
December 8-10, 2022, Turkey

ISAS WINTER-2022

https://doi.org/10.36287/setsci.5.2.004

FACE MASK DETECTION USING DEEP LEARNING FOR PUBLIC PLACES

Asmaallah FALLAHA, Mohamad Taj Eddin ASHOUR, Dala KRAYEM, Saed ALQARALEH*
Computer Engineering Department, Faculty of Engineering, Hassan Kalyoncu University, Gaziantep-TURKEY
saed.alqaraleh@gmail.com

Abstract – Over the past century, before and after the pandemic of COVID-19, workers in multiple sectors, such as medical, chemical, and nuclear, have been required to wear face masks during duties. However, physical 24/7 supervision is nearly impossible in public places. With the outstanding performance achieved by deep learning almost in all fields, this problem can be easily solved by building an automated mask detection system.

This paper investigates the performance of five deep learning models, particularly Convolutional neural networks(MobileNetV2, VGG19, and three sequential models) when used for mask detection, i.e., automatically distinguishing between a person wearing a face mask and a person who is not.

To ensure the results robustness of this comparison, four datasets consisting of approximately 6K, 12K, 4k, and 4k images, respectively, have been used.

Overall, the results of the experimental works showed that all models achieved a good performance when processing the first, second, and fourth datasets, with some improvement achieved by both MobileNetV2 and VGG19. However, when processing the third dataset containing low-quality images, MobileNetV2 significantly outperformed others.

Keywords – Convolution Neural Network, COVID-19, Deep Learning, Image Classification, Mask Detection.

I. INTRODUCTION

Image classification generally works by assigning classes or labels to an input image. Manually classifying is an easy task for our brain; however, it is hectic when processing a massive number of images. Hence, an automated, efficient classifier is preferred.

This topic has recently attracted researchers' attention, and some recent face mask detection studies are summarized below. In [1], a new face mask detector named SSDMNV2 was developed. This model uses Single Shot Multibox Detector and MobilenetV2 architecture. The results of [1] indicated that this model could achieve 0.9264 and 0.93 as accuracy and F1 scores, respectively.

A second face mask detection system called FMD-Yolo is proposed in [2]. This model uses Res2Net and deep residual networks in feature extraction. Next, the extracted features are fusion using the En-PAN network. Also, the Matrix NMS method was adapted at the inference stage to improve detection efficiency. Results indicated that the proposed model achieved a precision AP50 of 92.0% and 88.4% on the two used datasets.

In [3], the ensemble principle of two levels(one-stage and two-stage) detectors was used to decrease the inference time and increase the accuracy. Then, ResNet50 was used to fuse semantic information from multiple feature maps. Also, a bounding box transformation was proposed in [3] to improve mask detection localization performance. As a result, the model achieved 98.2% accuracy.

In [4], YOLOv3 and faster R-CNN were used to build an efficient mask detection system. The proposed system can detect the human face and draw bounding boxes around the face, where the red one indicates that this person is not wearing a mask. On the other hand, the green box refers to the person wearing a mask.

In [5], a mask detection based on computer vision and MobileNet was proposed. The main aim of this work was to help police and higher authorities quickly identify whether a person is wearing a mask or not. The system can record the user's photo for further confirmation, if needed. Also, this system will notify by a message everyone who is not wearing a face mask.

II. MASK DETECTION SYSTEMS

Mask detection (classification) systems consist of some common steps (shown in Figure 1) such as:

1. **Preparation of datasets**: This data will be used to train and test the desired system. In this work, four mask detection image datasets that are publicly available were used. These datasets are:

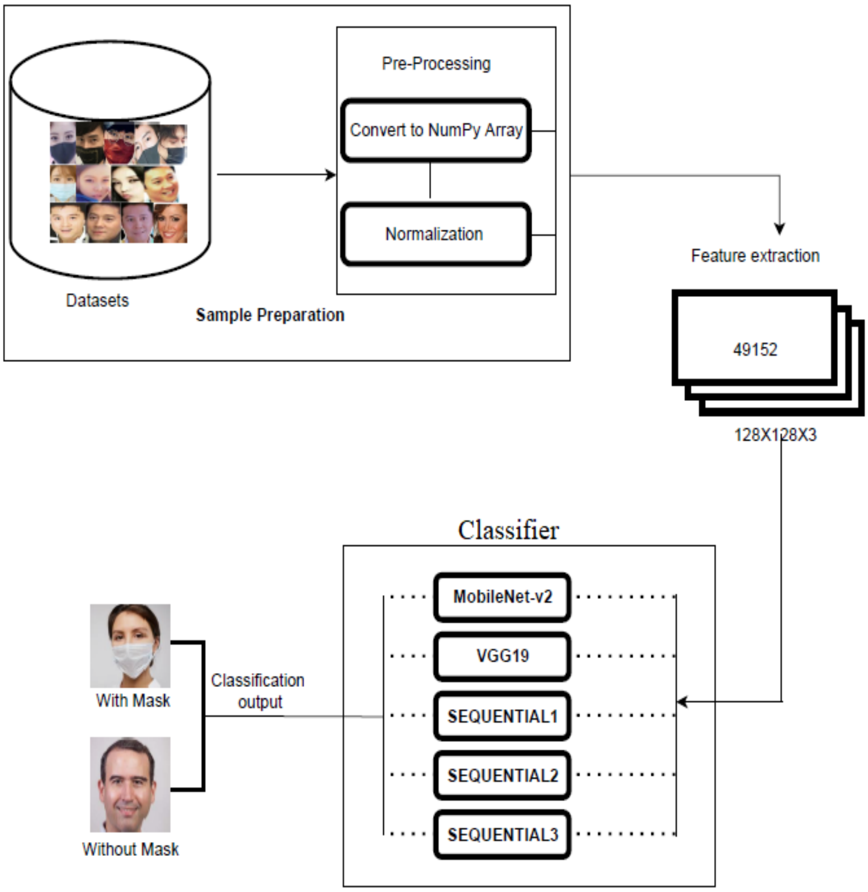

Fig. 1. Main Steps of Mask Detection System, where dot lines in the classifier box indicate the possibility of using one of the mentioned models.

Dataset 1 [6]: The total number of images in the dataset is 5988; it has three classes, 2994 images where their class is "with a mask," a class without a mask has 2994 images, and the remaining images represent a class for masks that are worn incorrectly. In this paper, we utilized only the first two categories. This dataset was created by Vijay Kumar and cleared and equally distributed across each class. The dataset contains images with a unified dimension in which each image is 128 x 128.

Dataset 2 [7]: This dataset was created by scraping images from Google and Jessica Li's CelebFace dataset. The two classes' dataset consists of almost 12K images, where the first class consists of images for people "with masks" and the other "without masks." The images' dimensions are not fixed, so the images are of different sizes, where the smallest size is 25x25, and the largest one is 563x563.

The RMFD dataset, i.e., "Dataset 3" [8]: is created by combing a Kaggle dataset and photos gathered via the Bing search API. This dataset has 4079 photos split into "with mask" and "without mask" classes. The photos' sizes vary, with the lowest being 26x37 and the biggest 4734x5412.

Face Mask Detection (FMD) dataset, i.e., "Dataset 4" [9]. This dataset is representative of the natural world, as all its images are of real people with masks(unlike the other datasets that have some images created by AI), and it has 3833 images, where 1915 images are with masks and 1918 without masks. The images in this dataset are of different dimensions and similar to "Dataset 3", the smallest size is 26x37, and the largest is 4734x5412.

The second sub-step of datasets preparation is **preprocessing**.

Mainly it works on eliminating the useless and noisy parts of the input(the images, in our case). Here, we also used manipulation functions to enhance the performance, which is done in addition to the other preprocessing steps using the "OpenCV" package. Also, to guarantee the preprocessing quality, the output array is passed into the TensorFlow preprocess function.

2. **Feature extraction**: Mainly, this step works on detecting and extracting distinguishing features that represent the input image. In this work, the array resulting from the pre-processing step will be normalized to the range [-1 to 1]. Then, a vector of 49152 features is extracted to represent the image. The samples will be divided into two parts training and testing.

3. **System classifier**: Here, we work on selecting the classification algorithm, which is an essential part of an efficient classification system. This work investigates the performance of five CNN-based mask detection models, and their details are summarized below.

VGG19: VGG is a Convolutional Neural network that stands for Visual Geometry group proposed by Simonyan and Zisserman in 2014 at the University of Oxford.

VGG19 version consisted of 19 layers, where 16 convolution layers are used for feature extraction, and three fully connected layers are used for classification. Note that 5 MaxPooling layers follow each convolution layer in VGG19. Mainly, VGG19 was pre-trained using the ImageNet database, which has 14 million images that belong to 1000 classes. In this paper, we adapted the mask detection model of [11], based on VGG19, and uses a fixed size of RGB input images (224*224).

"MobileNet-v2": The model was introduced in early 2018 by google. It is consisted of 53 layers and was also trained initially on the ImageNet database. MobileNetV2 is an improved version of MobileNetV1 that was developed to offer the possibility of developing mobile-oriented deep learning models with low computational power. It is worth mentioning that MobileNetV2 is a 35% faster and more effective extracting tool compared to V1 while still able to achieve the same accuracy. In this paper, we adapted the mask detection model of [10] based on the pre-trained MobileNetV2, and the input images are resized into 224x224.

SEQUENTIAL1: This model was created using Keras and consists of two convolution layers(Conv2D), where the first one has 200 filters of (3x3) and the second one with 100 filters of (3x3). Both convolution layers are followed by ReLU, MaxPooling, flatten, and Dropout layers. Then, two dense layers were used, where the first has 50 nodes(neuron) and ReLU activation, and the other has two nodes and SoftMax activation[12].

SEQUENTIAL2: This model consists of four convolution layers(Conv2D). The first layer has 64 filters of (3x3), and the second layer has 256 filters of (3x3). These two layers are followed by the ReLU activation layer and MaxPooling2D layer (2x2). Then, the third convolution layer has 128 filters of (3x3), followed by the ReLU activation layer and Dropout layer. The last convolution with 32 kernel/filter of (3x3) is followed by the ReLU activation layer and MaxPooling2D layer of (2x2). Next, another Dropout and flatten layers are added to this model, followed by three Dense layers, 100 nodes and ReLU activation, 16 nodes, and ReLU activation, and two nodes with SoftMax activation, respectively [13].

SEQUENTIAL3: Like the previous two models, this model was created sequentially. This model has three convolution layers, where all of these three layers have 32 filters of (3x3) followed by the ReLU layer and MaxPooling layer of (2x2). The model also has flatten layer and a dense layer of 100 nodes and ReLU activation; next is the Dropout layer, and the last layer is another dense layer with two nodes and SoftMax activation[14].

4. **Class assignment (Classification output)**: Here, each input image is assigned to one of the predefined two classes, i.e., a person wearing a face mask, whereas the second class is a person who is not.

III. EXPERIMENT AND RESULTS

This section compares the performance of the five mentioned models using four datasets. The first consisted of 5988, the second of 12K images, the third of 4079 images, and the fourth one containing 3833 images. Also, the 3-fold cross-validation and the Accuracy, Precision, Recall, and F1 score evaluation metrics were used.

Results are shown in Table 1, and their average is shown in Figure 2. Based on these results, the following can be concluded:

A. All models performed well when processing the first, second, and fourth datasets. However, related to the third dataset, which has low-quality images, the performance of the three sequential models has dramatically decreased compared to MobileNetV2.

B. Based on the performance using the four datasets and the average of achieved performance, MobileNetV2 has significantly outperformed all other models, where VGG19 achieved second-best results.

Table 1. The Accuracy, F1, Precision, and Recall for MobileNetV2, VGG19, and three Sequential models using the four used datasets.

Dataset	Metric	MobileV2	VGG19	Seq1	Seq2	Seq3
1st	Accuracy	0.94	0.91	0.96	0.85	0.93
	F1	0.94	0.91	0.96	0.85	0.93
	precision	0.94	0.91	0.96	0.85	0.93
	Recall	0.94	0.91	0.96	0.85	0.93
2nd	Accuracy	1.00	0.99	0.97	0.99	0.98
	F1	0.99	0.99	0.97	0.99	0.98
	precision	0.99	0.99	0.97	0.99	0.98
	Recall	0.99	0.99	0.97	0.99	0.98
3rd	Accuracy	0.93	0.91	0.86	0.86	0.85
	F1	0.93	0.91	0.86	0.86	0.86
	precision	0.93	0.91	0.86	0.86	0.86
	Recall	0.93	0.91	0.86	0.86	0.86
4th	Accuracy	0.97	0.97	0.93	0.93	0.94
	F1	0.97	0.97	0.93	0.93	0.94
	precision	0.97	0.97	0.93	0.93	0.94
	Recall	0.97	0.97	0.93	0.93	0.94

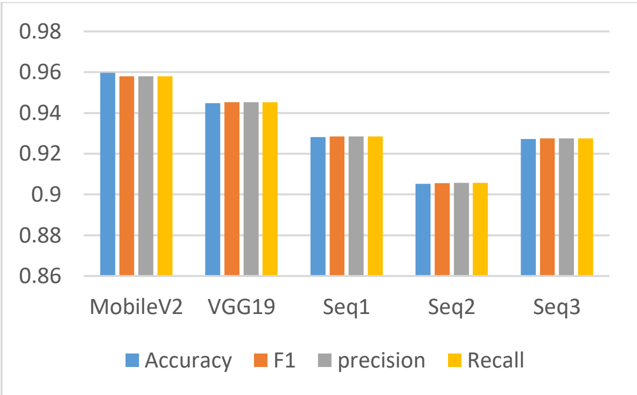

Fig. 2. The average Accuracy, F1, Precision, and Recall for the five studied models using the four used datasets.

IV. CONCLUSION

Following the pandemic caused by the COVID-19 virus, we were required to cover our noses and mouth in public places as an essential measure to keep the communities safe. The 24/7 human manual control and supervision is complicated and sometimes near impossible. As a solution, an automated monitoring system that can report people who are not complying with the rules of wearing masks is preferred.

This paper investigated the performance of five deep learning models, including MobileNetV2 and VGG19, when used for mask detection. Overall, the MobileNetV2 and the VGG19 models can be considered credible for building an efficient system. However, when we are required to classify low-quality images, MobileNetV2 can significantly outperform all other studied models.

Investigating more advanced feature extraction methodologies can be one direction for future work. Another direction is investigating the possibility of applying ensemble learning to CNN models.

REFERENCES

[1] R. J. A. M. R. A. P. K. J. H. Preeti Nagrath, "Corrigendum to "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," 2021.

[2] Wu, P., Li, H., Zeng, N., & Li, F. (2022). FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. Image and vision computing, 117, 104341.

[3] Sethi, S., Kathuria, M., & Kaushik, T. (2021). Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. Journal of biomedical informatics, 120, 103848.

[4] Singh, S., Ahuja, U., Kumar, M. et al., "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," Multimed Tools Appl, vol. 80, pp.19753–19768, https://doi.org/10.1007/s11042-021-10711-8, 2021.

[5] K. Suresh, M. Palangappa and S. Bhuvan, "Face Mask Detection by using Optimistic Convolutional Neural Network," 6th International Conference on Inventive Computation Technologies (ICICT), pp. 1084-1089, doi: 10.1109/ICICT50816.2021.9358653, 2021.

[6] VIJAY KUMAR, (2021) Face Mask Detection, [Online]. Available: https://www.kaggle.com/datasets/vijaykumar1799/face-mask-detection.

[7] ASHISH JANGRA, (2021) Face Mask Detection~12K Images Dataset, [Online]. Available: https://www.kaggle.com/datasets/ashishjangra27 /face-mask-12k-images-dataset.

[8] Real World Fasked Face Recognition Dataset (RMFRD), (2021) [Online]. Available: https://www.kaggle.com/datasets/ muhammeddalkran/masked-facerecognition.

[9] Balaji S, (2021) Face-Mask-Detection, [Online]. Available: https://github.com/balajisrinivas/Face-Mask-Detection/tree/master /dataset/ with_mask.

[10] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC., "Mobilenetv2: Inverted residuals and linear bottlenecks," InProceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520, 2018.

[11] Teboulbi S, Messaoud S, Hajjaji MA, Mtibaa A., "Face Mask Classification Based on Deep Learning Framework," InAdvanced Practical Approaches to Web Mining Techniques and Application, pp. 175-188, IGI Global, 2022.

[12] Perceptron, (2021) face-mask-detection-keras, [Online]. Available: https://github.com/aieml /face-mask-detection-keras.

[13] Real-time Face Mask Detection with OpenCV, [Online]. Available: https://projectgurukul.org/face-mask-detection/.

[14] Karan Malik, (2021) FaceMaskDetector, [Online]. Available: https://github.com/Karan-Malik/FaceMaskDetector.