Mask-19: Convolutional Neural Network for Face Mask Detection

Seif Shalan[1] and Mohamed Naggar[2]

[1]900160280
Email: shalan@aucegypt.edu

[2]900170476
Email: mohamed_elnaggar@aucegypt.edu

***Abstract*—The COVID-19 pandemic has truly stopped the all nations of the world in its tracks. To limit the spread of the disease, scientists and governmental bodies have recommended the use of face masks. Although many have contributed to the elimination of the disease by wearing masks, some have failed to wear it correctly or not wear it at all. Many areas open to the public demand people to wear masks to be allowed entry. This process has been done manually by other humans. Some researchers have decided to automate this process using deep learning techniques. But, all available solutions are only capable of classifying people wearing masks and people not wearing masks. In this research, we have reached a solution that is capable of detecting if a person is wearing a mask, not wearing a mask, or is wearing the mask incorrectly with a state-of-the-art ACCR of 97.5%. This research paper outlines the architecture of our solution, our experimental results, and any further work that can be done to improve our solution.**

TABLE OF CONTENTS

I. INTRODUCTION

THIS year, a lot has changed in our lives due to the coronavirus pandemic. People have become more aware of the essence of wearing a face mask to fight the spread of the virus and stay safe. However, some people still do not wear face masks in public places or even place them wrongly on their face just because most public places nowadays demands that you enter wearing a mask. Therefore, we decided to create a deep learning software that can detect people's faces in a video and classify them into one of three classes. Each face can either be wearing a mask, not wearing a mask, or wearing a mask wrongly. The software will be very useful as it can be taken to the next level and be installed in CCTV cameras in public places especially in hospitals and crowded places to ensure that everybody is wearing their masks the right way. In this paper, we will introduce you to our approach to solve the problem. Firstly, we will discuss some of the existing solutions as well as their pros and cons. Then, we will illustrate in full detail our architecture which is briefly composed of a convolutional neural net model and a computer vision module. Afterward, we will talk about our dataset, which contains images that belong to each of our 3 classes, and how we gathered these images to generalize our solution as much as possible. Also, our best classification results compared to the results that were reached by previous solutions. Finally, conclude by giving suggestions for future improvement and exploration to scientists interested in the problem.

II. EXISTING SOLUTIONS

A. NVIDIA Real-time Face Mask Detector

NVIDIA, the famous Silicon Valley company known for its GPU and SoC designs, implemented a real-time face mask detector application specifically for the COVID-19 pandemic. Their solution uses their in-house Transfer Learning Toolkit (TLT) for the model. Their goal while tackling the issue in hand was to create a an accurate and lightweight model that can be deployed on real-time edge devices. To reduce the model size, they utilized the model pruning technique. The model was trained on two datasets. The first dataset contained 27,000 images while the second version only contained 4,000. The second dataset contains images from the first dataset that were handpicked. The latter produced a higher accuracy. The model, however, has its limitations. The non-pruned model reached an average accuracy of 86.12% while the pruned model reached an accuracy of 85.50%. A 1% decrease in accuracy is quite acceptable considering the benefits of having a smaller sized trained model. The model's accuracy was only tested on a camera placed 5 feet away from the individuals. The model may produce worse results if placed at farther distances. The positioning of the camera may affect the model's capability at detecting faces. The dataset used only contained surgical masks. So, other masks of different colors and designs may not be detected as accurately.

B. RetinaFaceMask

Created by researchers at the City University of Hong Kong, their main goal was to design a face mask detector that can differentiate between subjects wearing masks and subjects that are either not wearing a mask or have their facial features hidden in some way or another. RetinaFaceMask uses multiple feature maps and utilizes a feature pyramid network to fuse the high-level semantic information. To improve detection, a cross-class object removal algorithm was used. Moreover, to overcome dataset limitations due to its size, transfer learning was used to transfer the learned kernels from existing face detection models. The dataset used contained 7959 images with all of the faces in the images annotated with either with a mask or without a mask. Some of the faces included were masked by hands or other objects rather than physical masks. This allowed the model to classify confusing images without masks more accurately. During training, stochastic gradient descent with a learning rate of 10-3, momentum of 0.9, and 250 epochs was used as the optimization algorithm. The model yielded an accuracy of 91.9% with a recall of 96.3% for non-masked faces and an accuracy of 93.4% and a recall of 94.5% for masked faces. Overall, the average accuracy was 92.65%. The model, however, is quite biased to Asian faces and may output poor results with non-Asian individuals as a result of the undiversified dataset.

III. Proposed Method

Fig. 1: Mask Detector Solution Flow.

A. Overview

Our proposed face mask detector consists of two models. The first model is responsible for detecting the faces in the video stream at a sampling rate of 20 frames per second. The model analyzes each input frame and returns an array of bounding boxes where each bounding box contains a face. The detector then preprocesses the faces to fit the second model. The second model is responsible for classifying each face. Each face detected is inputted into the model and the model returns the detected class. The detector then displays a bounding box around the subject's face with a label and color indicating the inferred classification.

B. Face Detection Model

As achieving the highest average accuracy is our main goal, we opted for using an existing model to detect faces instead of creating our own. OpenCV's DNN face detector was our best choice for this scenario. The model is known to work well with occlusion, quick head movements, and can detect side faces. The DNN face detector is a Caffe model based on the Single Shot-Multibox Detector and utilizes the ResNet-10 architecture as its backbone. This model takes an image as an input and returns an array of X-Y coordinates that illustrate the locations of the existing faces in the image. These coordinates are known as bounding boxes.

C. Data Preprocessing

The data goes through two stages of preprocessing. Once before inputting into the Mask Classification Model and once inside the model itself. The preprocessing done inside of the Mask Classification Model will be discussed further in Section III-D. After receiving the bounding boxes from the Face Detection Model, the boxes have to resized by a factor of 1.2 to give the Mask Classification Model more context. This in return will result in a more accurate classification.

D. Mask Classification Model

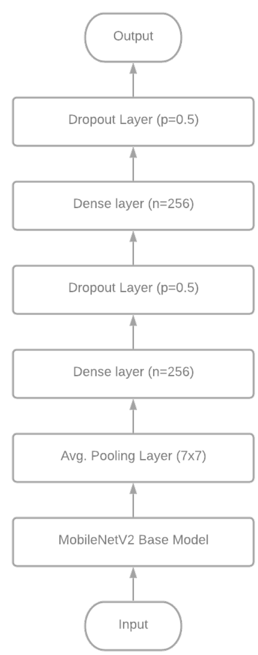

Fig. 2: Mask Classification Model Architecture.

The Mask Classification Model is based on the MobileNetV2 convolutional neural network. Our model architecture consists of the base MobileNetV2 model without its existing dense layers. An average pooling layer with a pool size of 7x7 was added. Two dense layers were added as well. Each dense layer consists of 256 neurons with a dropout layer following it. In terms of preprocessing, MobileNetV2 comes with its own preprocessing function to make the input data suit the model. During training, data augmentation was utilized to increase the dataset size. This in return improved our model accuracy and performance. The model is capable of predicting 3 classes: without mask, with mask, or incorrectly worn mask.

IV. Experiment & Results

A. Dataset

(a) Detected landmarks. (b) Detected landmarks. (c) Detected landmarks. (d) Detected landmarks.

(e) Correctly masked. (f) Incorrectly masked. (g) Incorrectly masked. (h) Incorrecly masked.

Fig. 3: MaskedFace-Net.

We collected our data from different public datasets available online. “MaskedFace-Net” is a publicly available dataset that contains images with correctly and incorrectly worn masks. However, we collected from it only the incorrectly worn mask images. The reason is that the dataset was built artificially using computer vision. To illustrate more, they used the same set of faces and placed a mask correctly and incorrectly on the same image. So, we decided not to confuse our model with faces that it has seen before in a different class. The next dataset we used is the “MaFa” dataset to collect our correctly masked faces. The dataset had a wide range of masked faces in different postures, but we focused mostly on the front view of faces. We used another dataset called ”Flickr Faces” that contains images with plain faces, and they are either 1024x1024 pixels or 128x128 pixels. So, we tried to combine a mixture of both sizes to increase the model’s ability to classify images with different qualities. Surely we did not use all the images in the mentioned datasets because each one contained tens of thousands of images. We instead collected 3303 images from each dataset to create a labeled dataset of 9909 images divided equally among our three classes and we used 80% of this data as a training set while the rest were left for validation.

B. Results

Our classification model achieved very good statistics while training. In table 1, we can safely say that the model works perfectly fine with plain faces. However, it did not perform as well in the other two classes. We can easily observe that the model’s most confusion was to predict images with the correctly worn mask as incorrect. Still, they were pretty good results as especially when we compared them to previous solutions, so we were very satisfied with our model and decided not to improve it.

TABLE I: Mask Classification Model Results

Class	Precision	Recall	F1	# of Test Images
Incorrect Mask	0.97	0.98	0.98	661
Mask	0.98	0.97	0.98	661
No_Mask	0.98	1.00	1.00	660

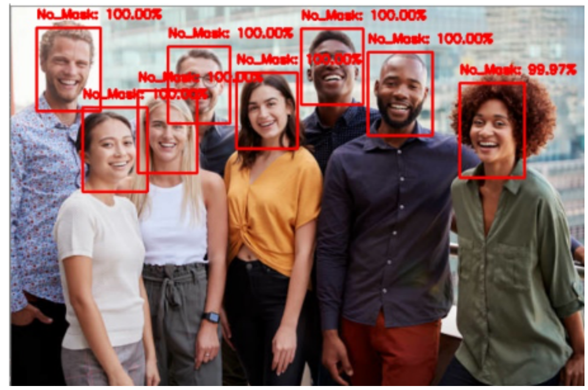

Fig. 4: A snapshot of non-masked individuals classified correctly with their respective accuracy.

Fig. 5: A snapshot of masked individuals classified correctly with their respective accuracy.

Next, with the use of OpenCV library and ‘res10’ model for facial detection, we were able to localize faces successfully in a video stream and predict each face separately and output a bounding box with our predicted values. Here are some sample images that we tested (not taken from a video) that shows two important things. Firstly, our model detected all faces in Figure 4, and almost each one of them was predicted with 100% accuracy, which shows how accurate our model is when it comes to localizing and classifying plain faces. Next, the model also was able to predict faces with a correctly placed mask with very high accuracy and that is observed in Figures 5 and 6. Though, detecting wrongly worn masks was not as accurate as predicting the other two classes. In Figure 6, we can see that the model recognized the incorrectly worn mask images but with not as high confidence as the other two classes. This probably because almost all our data in the Incorrectly worn mask class were artificially made. They were not images of people wearing real masks but wrongly. Instead, a cartoon-like mask was put on the faces using OpenCV like Figure 4 that is shown above.

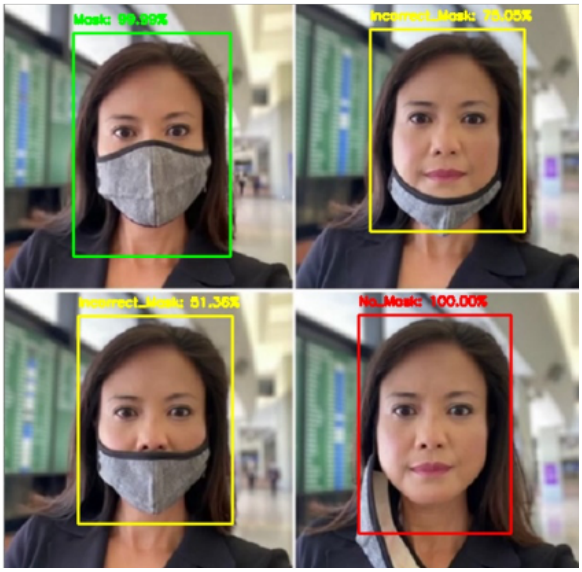

Fig. 6: The three possible classifications.

C. Benchmark Comparison

TABLE II: Accuracy Comparison between Mask-19 and benchmarks.

Model	**Class Accuracy**		
	Mask	No Mask	Incorrect Mask
NVIDIA	87.6%	84.7%	-
RetinaFaceMask	93.4%	91.9%	-
Mask-19	98%	98%	97%

Table II clearly shows that Mask-19 outperforms our benchmarks. Our model produced better results when classifying masked and non-masked individuals. It is impossible for us to compare the classification accuracy for incorrectly-worn masks as this is a new classification. However, our benchmarks were designed to be deployed on real-time edge devices. Mask-19 was not designed to be optimized on these types of devices. Nevertheless, our model is based on MobileNetV2 as its backbone which was specifically designed to perform well on mobile devices. With the help of model pruning, we can reap the benefits of our high-accuracy model and allow it to perform well in real-time. We will have to test the throughput of Mask-19 in later research to validate these claims.

V. Conclusion

It is safe to say that Mask-19 has resulted in state-of-the-art accuracies in comparison to our counterparts. We were able to introduce a new classification of detecting incorrectly worn masks. But, due to constraints in resources and health hazards, we were unable to test Mask-19 in a real-time scenario. There were also dataset limitations where most of the masks were blue surgical masks. This due to the fact that all public datasets have not taken this issue into consideration when building it. Such an issue may affect our model's accuracy when detecting non-surgical masks. We encourage future researchers to improve our model in terms of throughput and to train it on a more diverse dataset that may be available in the future.

References

[1] Kulkarni, A., Vishwanath, A. and Shah, C., 2020. Implementing A Real-Time, AI-Based, Face Mask Detector Application For COVID-19 — NVIDIA Developer Blog. [online] NVIDIA Developer Blog. Available at: https://developer.nvidia.com/blog/implementing-a-real-time-ai-based-face-mask-detector-application-for-covid-19/ [Accessed 15 December 2020].

[2] Cabani, A., Hammoudi, K., Benhabiles, H. and Melkemi, M., 2020. ”MaskedFace-Net – A Dataset of Correctly/Incorrectly Masked Face Images in the Context of COVID-19,” arXiv:2008.08016 [cs.CV], Aug 2020.

[3] Jiang, M., Fan, X. and Yan, H., 2020. ”RetinaMask: A Face Mask detector,” arXiv:2005.03950 [cs.CV], May 2020.