Real-Time Face Mask Detection

Aya Aridah
Department of Computer Information Systems
The University of Joordan
Amman, Jordan
aya.aridah@ieee.org

***Abstract*—Perhaps one of the most striking lifestyle changes resulting from the COVID-19 pandemic is the mandatory use of face masks in grocery stores, restaurants and other public places. According to the Centers for Disease Control and Prevention (CDC), COVID-19 infection is transmitted predominately by respiratory droplets generated when people breathe, talk, cough, or sneeze .Wearing a mask, especially when in close proximity to others, is the primary, effective, and convenient method of blocking 80% of all respiratory infections and it is imperative to slowing the spread of COVID-19. The objective of this report is to annotate and localize the medical face mask objects in real-life images using multiple models and many experiments to reach satisfy level of detection so we can use the best model in many applications.**

Keywords—Coronavirus, Masked Face, Deep Learning, COVID-19

1. Introduction

The coronavirus COVID-19 pandemic that was spread in different countries of the world has forced most people to change their lifestyles, changing their habits and behaviors, how they perform their jobs, and imposing severe restrictions on their freedoms. The governments of countries suffering from this pandemic have forced to initiate new rules for face mask-wearing to be able to control the virus and provide the treatment and vaccine needed to combat it. The goal of wearing face masks is to reduce the transmission and spreading rate. The World Health Organization (WHO) has recommended the usage of personal protective equipment (PPE) among people and in medical care. There are still a considerable number of people who forget or refuse to wear masks, or wear masks improperly. Such facts would increase the infection rate and eventually bring a heavier load of the public health care system. Therefore, many face mask monitoring systems have been developed to provide effective supervision for hospitals, airports, publication transportation systems, sports venues, and retail locations.

Today, COVID-19 is a significant public health and economy issue due to the detrimental effects of the virus on people’s quality of life, contributing to acute respiratory infections, mortality and financial crises worldwide. The main objective of this research is to detect and locate a medical face mask in an image as illustrated in Fig. 1. In this paper, the medical masked face is the main focus of research to reduce the spreading and transmission of Coronavirus specially COVID-19. Given an image, a region of the medical masked face on the input image based on DenseNet, MobileNet and VGG-16 will be illustrated in the output image.

I . Related Works

Object detection from an image is probably the deepest aspect of computer vision due to widely used in many cases. There has been supervised or unsupervised based learn in the field of computer vision to outfit the work of object detection in an image. This section conducts related work papers for applying representative works related to object detection based on deep learning for the medical face mask. Historically, most of the papers focused on performing face recognition while wearing a mask or with other obstructions, while few developed an alarm system for the wearing of masks in the operating room, combining the face detector with the mask detector, and optimized for low False Positive rate and high Recall, and obtained 95% True Positive rate [1]. Some of mask face detection focus on face construction and face recognition based on classical machine learning techniques like in (Ejaz et al., 2019)[2], authors have implemented Principal Component Analysis (PCA). From resourses reduce, proposed a system agnostic no-installation face mask detection solution to remind people who are not wearing a mask or wearing a mask improperly. As a serverless edge-computing design, it can be run locally on various edge devices, with low risk of privacy, low required network bandwidth, and low response time. The deployment scheme tackled insufficient support of deep learning from the JavaScript community, by aggregating NCNN and WASM [3]. Others who proposed a novel face mask detector, RetinaFaceMask, which is able to detect face masks and contribute to public healthcare. RetinaFaceMask is one of the first dedicated face mask detectors. In terms of the network architecture, RetinaFaceMask uses multiple feature maps and then utilizes feature pyramid network (FPN) to fuse the high-level semantic information. To achieve better detection, they propose a context attention detection head and a cross-class object removal algorithm to enhance the detection ability [4]. In other work, they acheives average precision equal to 81% using YOLOv2 with ResNet-50 based on Adam optimizer. By analyzing the performance of YOLO v2 based on ResNet-50 in handling medical masked faces with the different optimizers, they found that the performance measurement of all masked face detectors increases sharply on masked faces with strong occlusions. The proposed model improves detection performance by introducing mean IoU to estimate the best number of anchor boxes [5]. In this report, the focus is on detecting and find the human who is wearing a face mask in helpful positions to help in lessening the spreading of the COVID-19 using VGG-16 , DenseNet and MobileNet.

2. DESCRIBE THE DATA

Data set consists of 1378 RGB images (224*224*3) in 2 folders as with_mask and without_mask. Images of faces with mask are 691 and images of faces without mask are 687. Big part of the images are augmented, and they auto-generated mask [5] .

As you can see most of the images are taken from the front, there is no images were taken from the side, and that gives a false positive in many times I tried to detect the mask from side. On the other hand, if we think, the use of detecting the mask from side is not so useful. The Face Mask Detection System can be used at office premises to detect if employees are maintaining safety standards at work. It monitors employees without masks and sends them a reminder to wear a mask so, there is no big need to detect from side. If any application in future would like to detect face masks from side, you should add images that are taken from side to the dataset so the model will train on them.

3. DEVELOPMENT ENVIRONMENT

Colaboratory, or “Colab” for short, is a product from Google Research. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs. What’s really amazing about ollaborator (or Google’s generousity) is that there’s also GPU option available. GPU computing is the use of a GPU (graphics processing unit) as a co-processor to accelerate CPUs for general-purpose scientific and engineering computing and I used it for the experiments on real-time face mask detector. Google Colab already gives us about 13 GB of RAM for free. That’s quite impressive, but there are times when even that isn’t enough when we’re building heavy deep learning models, you can quickly increase the RAM in your Colab notebook.

4. DATA PREPROCESSING

First, I grabbed the list of images in our dataset directory, then initialized the list of data (i.e., images) and class images then loop over the image paths to extract the class label from the filename, load the input image (224x224) to preprocess it,and update the data and label lists,respectively. Second, convert the data and labels to NumPy arrays to perform one-hot encoding on the labels. For splitting, I partitioned the data into training and testing splits using 75% of the data for training and the remaining 25% for testing. Finally, I constructed the training image generator for data augmentation and the reason why using data augmentation is the ability to generate 'more data' from limited data (it used previously). The second one is to avoid overfitting: for a network, it is somewhat problematic to memorize a larger amount of data, as it is very important to avoid overfitting [6]. All the models were implemented; I used the same preprocessing steps. A sample of same images after preprocessing.

As we can see, some images changed and that because of data augmentation process. Horizontal flip was applied, width and height shift, zooming and rotation as the second image. Also, we can notice that the color of the images was changed. Note that, the third image was appeared once as false negative maybe that is a sign that maybe the data augmentation must be change or try different techniques reputedly.

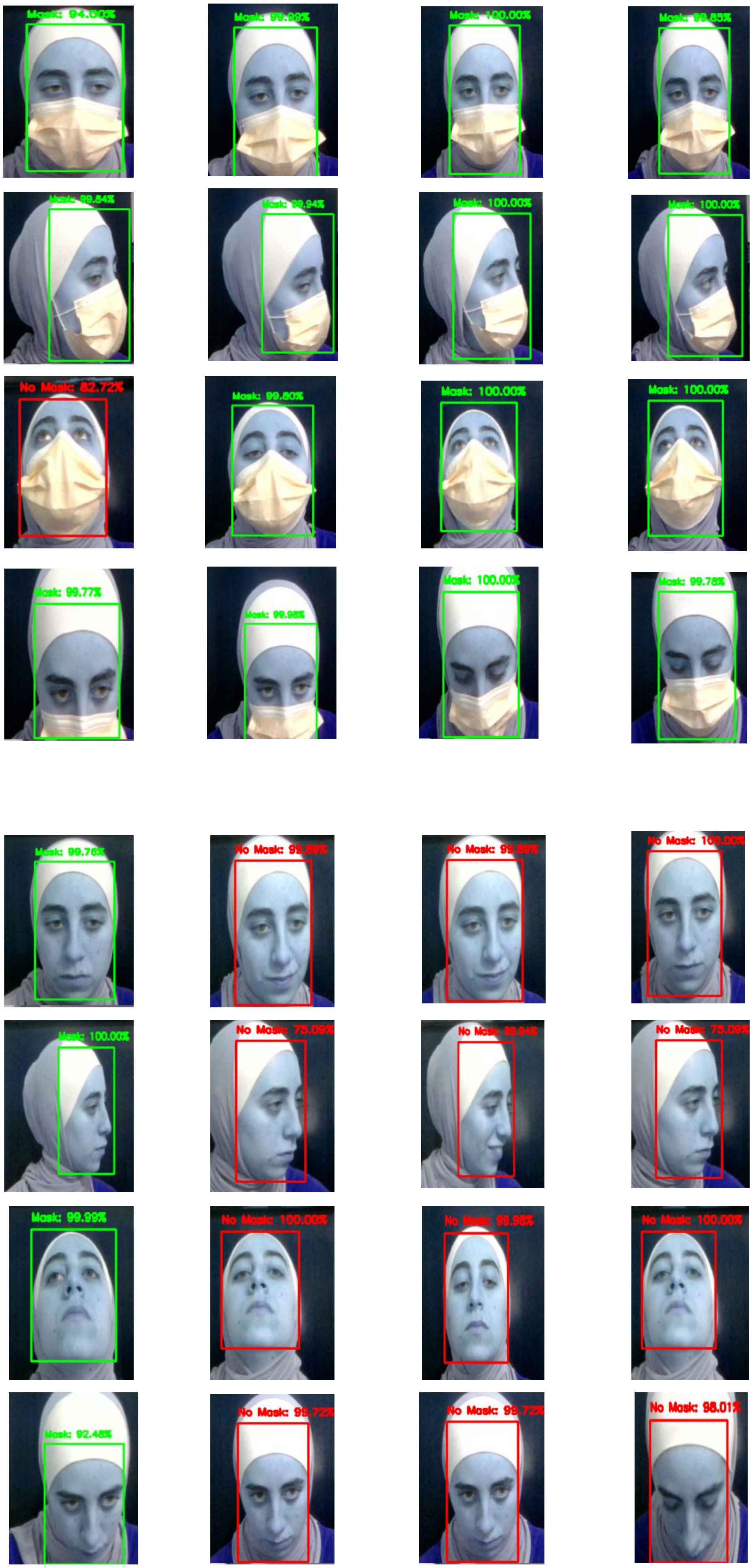
5. RESULTS

The qualitative results is presented following. It is able to correctly identify the cases when a subject is not wearing the mask properly, such as when the mouth is exposed. It also has a higher probability of recognizing that it is not a mask from further distance. The categorize are the following:

2 Convolutional Layers with 2 Maxpooling Layers | Dense | MobileNet | VGG-16

With Mask: Front Face, Side Face, Elevated view, Overhead view

Without Mask: Front Face, Side Face, Elevated view, Overhead view

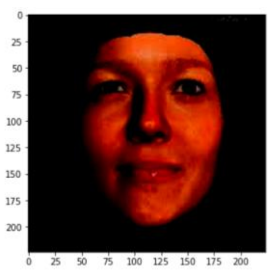


From quantitative results, the traditional convolutional model gives the most unpotential results, as we saw previously from images, when I tested it on no mask it gave me all the four experiments were false, and when the face was in elevated view. The unexpected case was in the side of the face, according to the dataset, there weren't so many images they were taken from the side. The higher accuracy was recorded for the 2 convolutional layers model was 92.1% and it's not so good comparing to the other transfer learning technique. Although, I thought that the face mask detection problem is not hard to predict and maybe there is no need to use transfer learning. Then, I implemented the DenseNet, Densenet is more efficient on some image classification benchmarks, and it's much more efficient in terms of parameters and computation for the same level of accuracy, compared with ResNet and others [7].

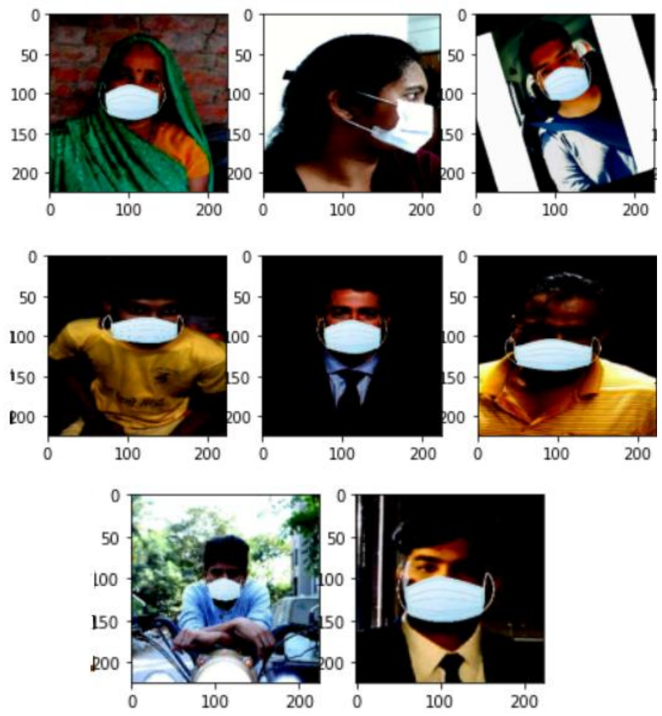
Model Hyper parameters	**Accuracy on test set**			
	2conv layers	***DenseNet***	***MobileNet***	***VGG-16***
Optimizer = adam, learning rate =0.001, epochs =10	92.1%	97%	97.2%	99.8%
Optimizer = nadam, learning rate = 0.001, epochs = 20	90%	99.2%	99.6%	99%
Optimizer = adam, learning rate =0.01, epochs = 30	51%	96.6%	98.7%	99.9%

Densenet contains a feature layer (convolutional layer) capturing low-level features from images, serveral dense blocks, and transition layers between adjacent dense blocks. When I tested it with nadam optimizer, it gave the best accuracy, and on real-time detection was predicting perfectly most of the time. The MobileNet was also from the good models to detect the face mask, the best accuracy was 99.6%. The last model I implemented was VGG-16. VGG-16 is a convolutional neural network that is 16 layers deep. When I tested it, the best accuracy I've got when I used adam optimizer, learning rate = 0.01 and number of epochs was 30 with early stopping. As the table shows, the VGG-16 model was the best to detect the face mask in my case was a binary classification, with mask or without mask with 99.9% accuracy. While testing phase I plotted the images that the model missed to detect as follows (DenseNet) :

Fasle Negative :



False Positive:



What these images tells us, maybe the data augmentation was helpful but in these cases, it wasn't. Also, the shades of the images was changed and that's maybe the main reason that the model detected them wrongly.

6. CONCLUSION

In this work, I have introduced several models with high accuracy. The best model was VGG-16 with adam optimizer , learning rate = 0.01 and 30 epochs. Through the experiments, the more hyperparameter tuning, the effective model you'll get. About hyperparameter tuning, they are important because they directly control the behaviour of the training algorithm and have a significant impact on the performance of the model is being trained. For the future work, RetinaNet will help to detect the face mask from a distance, YOLO version2 will not be helpful to detect from a further distance.

7. REFERENCES

[1] A. NIETO-RODR´IGUEZ, M. MUCIENTES, AND V. M. BREA, "SYSTEM FOR

medical mask detection in the operating room through facial attributes," in Iberian Conference on Pattern Recognition and Image Analysis.

[2] Ejaz, Md. S., Islam, Md. R., Sifatullah, M., & Sarker, A. (2019). Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition. 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology.

[3] Zekun Wang, Pengwei Wang, Peter C. Louis, Lee E. Wheless, and Yuankai Huo, Member, IEEE, “WearMask: Fast In-browser Face Mask Detection with Serverless Edge Computing for COVID-19” DECEMBER 2020.

[4] Mingjie Jiang, Department of Electrical Engineering City University of Hong Kong, Xinqi Fan, Department of Electrical Engineering, City University of Hong Kong, Hong Yan, Department of Electrical Engineering, City University of Hong Kong, “RETINAFACEMASK: A FACE MASK DETECTOR”. June 9, 2020.

[5]https://www.kaggle.com/omkargurav/face-mask-dataset

[6]https://blog.stratio.com/wild-data-part-one-augmentation-2/#:~:text=The%20benefits%20of%20data%20augmentation,very%20important%20to%20avoid%20overfitting

[7]https://medium.com/@smallfishbigsea/densenet-2b0889854a92

8. SOURCE CODE

https://colab.research.google.com/drive/1GlOSTTyUgcuKyOoVoSpwn8XEn-gnU2Rf?usp=sharing