Comparison of Deep Learning Model Performances on Real-Time Face Mask Detection

Luwei Lei
Dept of Data Science and Analytics
Georgetown University
Washington, D.C., U.S.A
ll1038@georgetown.edu

Yu Xiao
Dept of Data Science and Analytics
Georgetown University
Washington, D.C., U.S.A
yx197@georgetown.edu

Julia Zhu
Dept of Data Science and Analytics
Georgetown University
Washington, D.C., U.S.A
yz719@georgetown.edu

***Abstract*—As face covering has become an essential requirement in public activities due to coronavirus, this paper aims to develop a face mask detection algorithm by using neural network techniques and finds the best-performing model. We define and compare three deep learning algorithms trained and tested on Face Mask Dataset, and present the accuracy as the result. We also introduce the OpenCV technique combining our neural network models for real-time face mask detection.**

***Index Terms*—CNN, face detection, computer vision**

I. Introduction

Coronavirus(COVID-19) has become the headline of 2020. The undetectable symptoms and fast spreadability resulted in more than 72 million people infected globally (by December 14, 2020) and the number is still growing every day [1]. Even though we still know little about the COVID-19, scientists and doctors have found a series of protection methods to avoid infection. According to the Centers for Disease Control and Prevention(CDC) [2], wearing a mask could effectively reduce the transmission of coronavirus. Face covering has also become an essential requirement for public indoor activities such as shopping at grocery stores. Therefore, we think it is helpful to develop a face mask recognition algorithm with the best neural network techniques and implement it in a real-time face detection tool.

In order to maximize the potential of the project in terms of scalability and variability, we designed three phases. The first phase is to develop neural network models that can detect if a person is wearing a mask or not in an image. The second phase is to compare those models' performance based on the static image dataset. The third phase is to test our models using a real-time face mask detection system. In this phase, we will implement an OpenCV face mask detector to recognize if one or multiple people in front of the camera are wearing masks, which is of high practical significance.

II. Related Works

Mohamed Loey and Gunasekaran Manogaran [3] presented a hybrid deep and machine learning model proposed for face mask detection. The model consists of two parts. The first part is designed for feature extraction using the Resnet50 Model. And the second part is designed for the classification process of face masks using decision trees, Support Vector Machine, and ensemble algorithm. In this paper, three face masked datasets have been selected for investigation and the SVM classifier achieved a highest testing accuracy of 100%.

Md. Rafiuzzaman Bhuiyan and Sharun Akter Khushbu [4] applied the YOLOv3 model on face mask detection. They obtained experimental results that show average loss is 0.0730 after training 4000 epochs. After training 4000 epochs mAP score is 0.96. This unique approach of face mask visualization system attained noticeable output which has 96% classification and detection accuracy.

G. Jignesh Chowdary and Narinder Singh Punn [5] proposed a transfer learning model to automate the process of detecting the people who are not wearing masks. The proposed model is built by fine-tuning the pre-trained state-of-the-art deep learning model, InceptionV3. They used the Simulated Masked Face Dataset to train and test the proposed model. Image augmentation technique is adopted to address the limited availability of data for better training and testing of the model. The result shows that the model achieves an accuracy of 99.9% during training and 100% during testing.

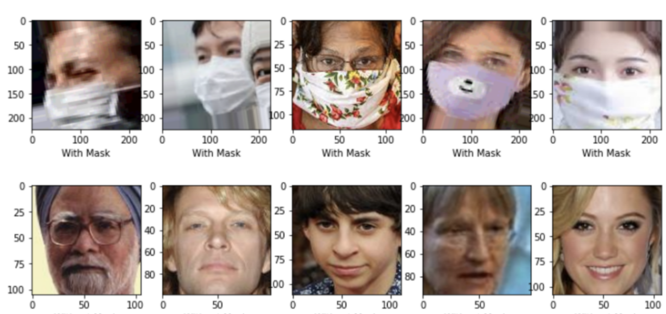


Fig. 1. Sample Images in the Dataset

III. Dataset

In the dataset we found, there's a headshot of one person in each image, either wearing a mask or not. This dataset is used to train and test our neural network models. The training dataset contains 800 files, 400 files with masks and 400 files without. The testing dataset contains 100 files in total with 50 images with masks and 50 images without. The reason that the train and test set is small is that our models get really high accuracy even with a small amount of data and that the time cost for training a larger dataset.

IV. Methods

A. Convolutional Neural Networks

One of the most popular deep neural networks is the Convolutional Neural Network (CNN). It takes this name from mathematical linear operation between matrices called convolution. CNN has multiple layers; including convolutional layer, non-linearity layer, pooling layer and fully-connected layer. The convolutional and fully-connected layers have parameters but pooling and non-linearity layers don't have parameters. CNN has an excellent performance in machine learning problems. Specially the applications that deal with image data, such as largest image classification data set (ImageNet), computer vision, and in natural language processing (NLP) and the results achieved were very amazing. [8] All CNN models follow a similar architecture, as shown in the Figure 2.

There is an input image to be worked with. A series convolution + pooling operations, followed by a number of fully connected layers are performed. If we are performing multiclass classification the output is softmax. In this paper, two CNN models with different numbers of convolutional layers are trained. The CNN-1 layer model consists of `Conv2D + MaxPooling2D + Flatten + Dropout + Dense + Dense`. The CNN-4 layer model consists of `(Conv2D+ MaxPooling2D)×4 + Flatten + Dropout + Dense + Dense`.

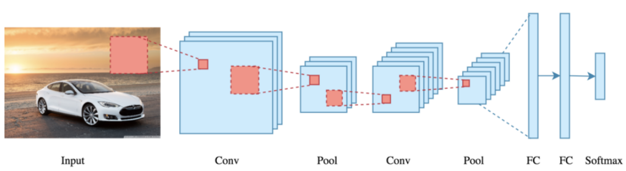
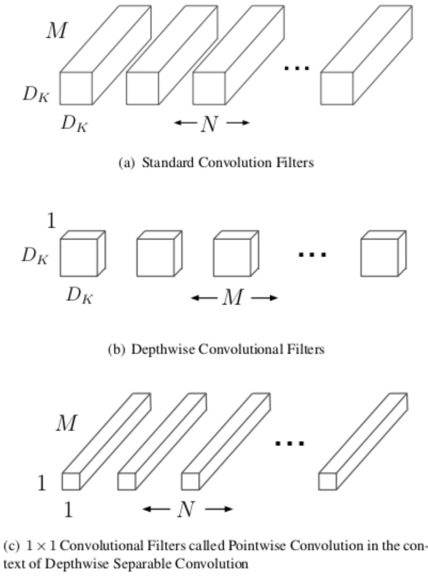


Fig. 2. A sample architecture of CNN model

B. MobileNetV2

MobileNetV2 is a small and efficient light weight model that was designed for running on personal devices without requiring a GPU. It was developed by Google in 2018 based on the MobileNetV1 model, which is a family of general purpose computer vision neural networks for mobile devices [9]. The MobileNetV2 uses depthwise separable convolution which applies a single filter to each input channel and a 1x1 pointwise convolution to combine the outputs as shown in Figure 3 [10].

This paper used the pre-trained model from Keras, which achieved 92.1% top-5 test accuracy in ImageNet database, which is a large visual database designed for visual recognition software research. A layer of flatten layer and dense layer were also applied to smooth the result.



(a) Standard Convolution Filters

(b) Depthwise Convolutional Filters

(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Fig. 3. A sample architecture of ModelNetV2

C. VGG16

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" . The model achieves 92.7% test accuracy in ImageNet. It makes improvements over AlexNet by replacing large kernel-sized filters with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks by the NVIDIA Titan Black GPU.

In VGG16 architecture, the image is passed through a stack of convolutional layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the convolutional layers. And three Fully-Connected layers follow a stack of convolutional layers. The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks. [6]

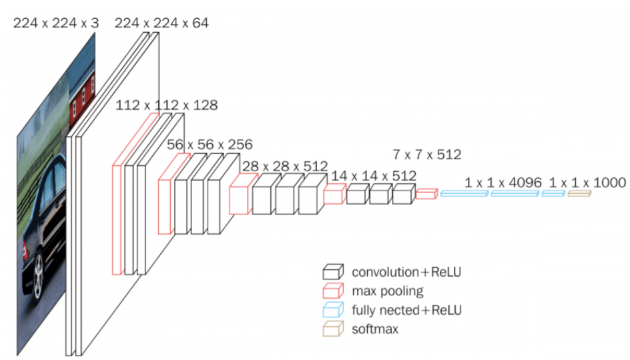


Fig. 4. A sample architecture of VGG16 model

V. RESULTS

A. Accuracy Evaluations

In the table below, we summarize the performance of four models in terms of the average train accuracy as well as the test accuracy.

Model	Loss Function	Epochs	Average Train Acc	Test Acc
CNN-1 Layer	Binary cross entropy	10	94.55	94.99
CNN-4 Layer	Binary cross entropy	10	93.30	93.99
MobileNetV2	Binary cross entropy	10	99.31	99.00
VGG16	Binary cross entropy	10	98.54	96.88

Fig. 5. Results of models. The fourth column, "Average Train Acc", denotes the mean of the train accuracies between the 2nd and 10th epoch since the values of accuracy tend to be stable on epoch 2.

All these models have similar performance (93%+) in terms of accuracy evaluations. Compared with CNN, pre-trained models such as MobileNetV2 and VGG16 from Keras could greatly improve the training accuracy. Specifically, MobileNetV2 has the best performance on our dataset, which reaches 99% test accuracy. We also observed that the 4-layer CNN has a little bit lower testing accuracy compared to the 1-layer CNN.

B. Real-time model testing with OpenCV

After training these four models on the face mask dataset, we use the OpenCV (Open Source Computer Vision Library) to conduct real-time face mask detection. Similar to the results of model accuracy, MobileNetV2 performs the best in real-time computer vision detection. Figure 6 shows that the prediction accuracy is 1.0 when running on the real time data.

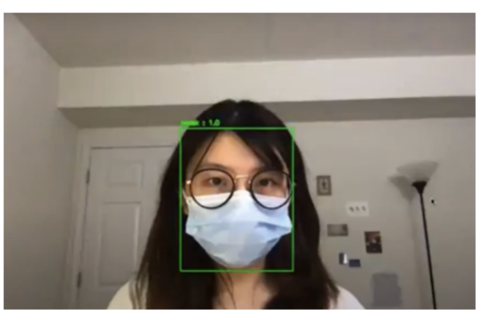


Fig. 6. Results of real-time computer vision detection using MobileNetV2. The image is a screenshot of the real-time computer vision detection window.

In real-time computer vision detection, VGG16 and MobileNetV2 can't distinguish if one's face is covered by a piece of paper or a face mask. However, CNN can distinguish between a real mask and a piece of paper as shown in Figure 7 .Based on multiple times of experiments, we also observed that the openCV is smart enough to detect multiple people in one frame. As shown in Figure 8, VGG16 can detect whether multiple people are wearing masks or not.

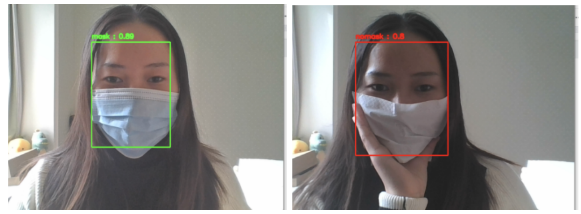


Fig. 7. Results of real-time computer vision detection using 1-layer CNN.

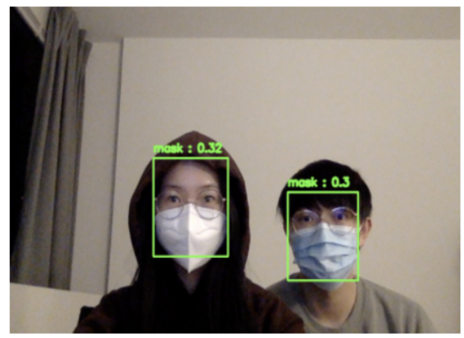


Fig. 8. Real-time computer vision detection of multiple people using VGG16.

VI. CONCLUSIONS

In this paper, we compared the performance of four deep learning models on face mask detection and applied our models to real time data using OpenCV to help maintain a secure environment and ensure individuals protection by automatically monitoring people's faces to avoid the spread of the COVID-19 virus. [7] In conclusion, MobileNetV2 has better performance than CNN and VGG16 in terms of accuracy evaluation both on the static image data and on the real-time face data. Our face mask detector can also monitor multiple people at the same time. As a next step, we plan to experiment with other deep learning models and explore how to improve the speed of our detection system.

REFERENCES

[1] Pettersson, H., Manley, B., & Hernandez, S. (2020). Tracking coronavirus' global spread. Retrieved October 02, 2020, from https://www.cnn.com/interactive/2020/health/coronavirus-maps-and-cases/

[2] How to Protect Yourself & Others. (2020, September 11). Retrieved October 02, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html

[3] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa (2020). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Measurement, Volume 167, 2021, 108288, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2020.108288.

[4] M. R. Bhuiyan, S. A. Khushbu and M. S. Islam (2020). ”A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3,” 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225384.
[5] G. Jignesh Chowdary and Narinder Singh Punn and Sanjay Kumar Sonbhadra and Sonali Agarwal (2020). Face Mask Detection using Transfer Learning of InceptionV3. Retrieved from https://arxiv.org/abs/2009.08369v2
[6] Muneeb ul Hassan (2020). VGG16 – Convolutional Network for Classification and Detection. Retrieved from https://neurohive.io/en/popular-networks/vgg16
[7] Shashi Yadav. (2020). Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID19 Safety Guidelines Adherence.In International Journal for Research in Applied Science & Engineering Technology (IJRASET) (Vol 8)
[8] Arden Dertat. (2017). Applied Deep Learning: Convolutional Neural Networks. Retrieved from https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2
[9] Sandler, M., & Howard, A. (2018, April 03). MobileNetV2: The Next Generation of On-Device Computer Vision Networks. Retrieved December 15, 2020, from https://ai.googleblog.com/2018/04/mobilenetv2-next-generation-of-on.html
[10] Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv, abs/1704.04861.