Face Mask Detection Using CNN-based Models with Face Detection Model

Dorothy Hou
Duke University
tianyun.hou@duke.edu

Abstract

The outbreak and rapid spread of COVID-19 has raised public health crisis worldwide and disrupted everyone's lives since 2020. While wearing masks in public places has become the most common yet effective measure against the transmission of the disease, it still poses challenges for tracking if the requirement of wearing masks is fulfilled in the real-time since it's usually infeasible to track people's mask status manually. However, modern deep learning techniques provides a better alternative to accomplish such tasks. This project conducts different convolutional neural networks (CNN)-based models including customized CNN model, VGG and ResNet-based transfer learning models for mask detection classification. The evaluation of these model performances demonstrates that ResNet101-based transfer learning model yields highest accuracy of 99.70% on the test set. The model is ultimately integrated with a pre-trained face detection model to provide an end-to-end illustration indicating if the person is masked or not given an image input.

Introduction

The outbreak and wide-spread of COVID-19 epidemic has brought global public health crisis and led to unprecedented changes to everyone's day-to-day life, as well as to worldwide trading and global economy. According to CDC guidelines on COVID-19 prevention, hygiene practices, improving ventilation, wearing masks, and increasing space and distance are recommended actions for protecting us from the virus. To control and reduce the transmission of the disease, wearing face masks has become one of the most important and widely adopted requirements in the general COVID safety protocols in public spaces such as school, public transportation, corporate environment, and shopping malls. As an effective protective approach during the coronavirus pandemic, people are often required to wear a mask in public places and thus, monitoring and detecting if people are wearing masks becomes an essential task to achieve. However, it's usually difficult and infeasible to track people's mask status manually. Fortunately, with the help of deep learning and modern computer vision techniques, this objective can be achieved effectively as real-time face mask detection in public places. In this report, we utilize and compare different CNN architectures, including transfer learning-based models for image classification to determine the presence or absence of face masks. By combining the object detection model for detecting human faces in the image with the CNN model for mask detection, we can draw bounding boxes around people's faces in a given image and identify if a mask is detected or not. If further extended to video streams, the model can be applied in the surveillance camera network for mask detection in real-time and thus protect public health effectively.

Related Work

Convolutional Neural Networks

Convolutional Neural Networks (CNN) (LeCun et al. 1998) plays a critical role in modern deep learning tasks, especially in the field of computer vision to be applied to analyze visual imagery. CNN is consisted of convolutional layers, pooling layers and a fully-connected layer. As the image data progresses through the layers of CNN, it starts from recognizing simple features to extracting more complex features in the image, thus achieving superior performance in computer vision tasks such as image classification.

Object Detection Models

An object detection model is trained to detect the presence and location of multiple classes of objects, which is also used extensively in computer vision-related tasks. Different from image classification which sends a whole image through a classifier and returns a classification result, object detection comes down to drawing bounding boxes around detected objects which allow us to locate them in a given scene. Some typical object detection models include R-CNN model family and YOLO model family. Face detection is one of the most common use cases of object detection models and would be useful to be incorporated with in our study.

Mask Detection

As the mask requirement becomes an essential normal in our daily life since the pandemic, there have been a number of mask detection models proposed in recent years. Militante and Dionisio (Militante and Dionisio 2020) developed a real-time facemask recognition with alarm system using deep learning. They used the VGG-16 CNN architecture and achieved an accuracy rate of 96%. Loey et al. (Loey et al. 2020) introduced a hybrid deep transfer learning model with machine learning methods for face mask detection. Their model consists of two components. ResNet50 is used for feature extraction as the first component and machine learning algorithms, i.e., Support Vector Machine (SVM), decision tree, and ensemble algorithm are used for classification process of face masks. The SVM classifier achieved the highest detection accuracies among all. Similarly, Oumina et al. (Oumina et al. 2020) combined pretrained deep learning models such as Xception, MobileNetV2, and VGG19 for the extraction of features with machine learning classifiers such as Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN) for mask detection. The best classification rate is 97.1%. which is achieved by combining SVM and the MobileNetV2 model.

Methods

Data

The data used in this project is obtained from Kaggle[1] and contains approximately 12,000 images of people with and without masks, which are split into train, test and validation set. The train set contains 10,000 images, the test set has 992 images and the validation set has 800. There are two classes presented in the data: with mask and without mask. The class distribution is balanced across the train, test and validation dataset. It is worth noting that all the images in the data are already cropped to only showing the faces. Sample images with labels are shown below.

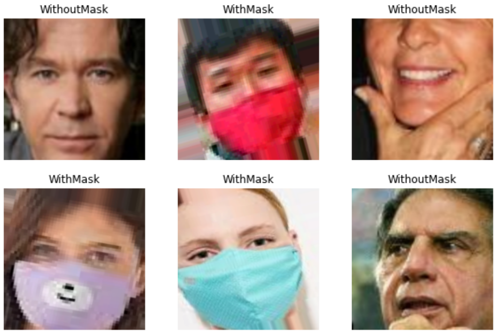

Figure 1: Sample images from dataset with labels.

Data augmentation is performed using ImageDataGenerator in Keras to improve generalization ability of the models through flipping, zooming and shearing the images. The input image size is set as 128 x 128.

Mask Detection Classifier

In this study, CNN-based models are trained to classify images as masked or unmasked using the Keras framework. We start with a simple customized CNN architecture with four stacks of convolutional layer and pooling layer, followed by a flatten layer and a dense layer, then a final layer that outputs two classes with sigmoid activation function.

Furthermore, in addition to the simple CNN architecture, deep neural nets can further represent extremely complex functions and improve the model performance. Therefore, taking the advantage of pre-trained deep neural nets, we further implement transfer learning-based models to compare the performances across the models.

VGG-19. Visual Geometry Group (VGG) is a classical deep convolutional neural network architecture with multiple layers. It is the basis of ground-breaking object recognition models. VGG-19 consists of 19 convolutional layers. To implement transfer learning, we load the VGG-19 model pre-trained on ImageNet but freeze the last layer, and add a flatten layer as well as a dense layer with 2 classes to identify if the face is masked or not.

ResNet101. Deep Residual Networks (ResNets) uses the insertion of shortcut connections, which allows the model to skip layers and also prevents the vanishing gradient problem. ResNets is another typical CNN architecture that yields outstanding performance for image classification. Compared to VGGNets, ResNets are less complex and faster to train. Similarly to VGG-19, we use a pre-trained ResNet101 model as the foundation and add a few layers at the end to tailor the model for training the mask detection classifier.

The performance of each architecture is then evaluated and compared based on their accuracy/loss plots as well as their accuracy on the test set. The best-performed model is adopted and re-trained with early stopping to prevent overfitting. The final trained model is ultimately integrated with the face detection model to provide final results.

Face Detection

In order to utilize the mask detection model in the real-world applications, detecting face is the first step required to accomplish this task. The face detection model will detect faces in the given image and extract the coordinates of the faces for the next-step mask detection classification. Haar cascade is an algorithm that can detect objects in images or real-time videos, proposed by Paul Viola and Michael Jones in 2001 (Viola and Jones 2001). In this project, a Haar cascade model trained to detect faces, which is available via OpenCV, is used to perform the face detection task and provide corresponding input for mask detection classifier. More specifically, the face detection model returns the coordinates of the faces detected in the image and only the cropped part of faces is fed into the mask detection classifier as the input to obtain final classification results.

[1] https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset

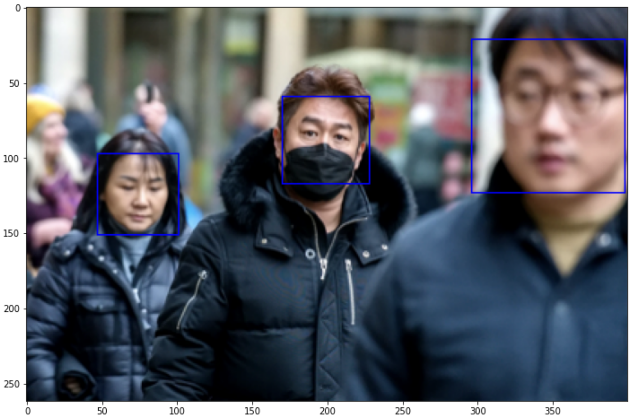

Figure 2. Sample output from haar cascade face detection model. The blue bounding boxes show the faces detected in the given image.

Experimental Results

All the CNN models are trained over 30 epochs and the figure below shows the comparison of accuracy/loss plot of the training process for customized CNN, VGG-19 and ResNet101-based transfer learning models. The ResNet101-based transfer learning model generally has lowest loss and highest accuracy on both training and validation set. The model based on VGG-19 performs slightly better than the simple customized CNN model.

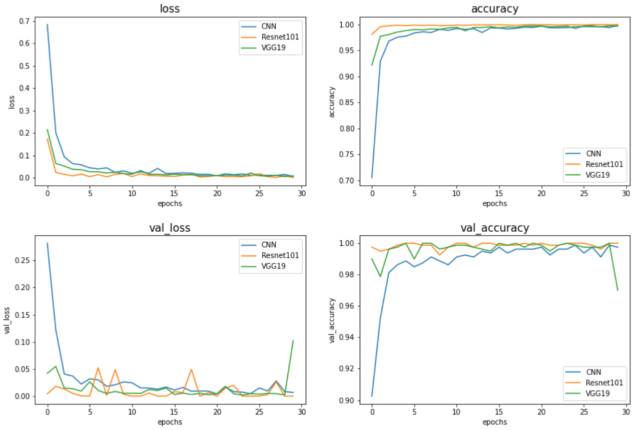

Figure 3: Comparison of accuracy/loss on train and validation set for CNN, VGG-19 and ResNet101 models.

Model	CNN	VGG19	ResNet101
Accuracy	98.99%	97.48%	99.90%

Table 1: Accuracy on test set corresponding to different model architectures.

Additionally, according to the table above comparing the accuracy on the test set from each model, the ResNet101-based transfer learning model again outperforms other CNN architectures with the highest accuracy close to 1.

As a result, ResNet101-based transfer learning model is selected as the final model trained for mask detection classifier. The architecture of the model is demonstrated below in Figure 4. After retraining the model with early stopping, the accuracy achieved on test set is 99.70%. The confusion matrix shown below shows a satisfying performance of the model with high accuracy, recall and precision. By incorporating with the face detection model from OpenCV, given an image of people as input, the final output from the face detection model and mask detection classifier is presented through drawing bounding boxes around people's faces, where the green box indicates the person is wearing a mask, while the red box indicates the opposite.

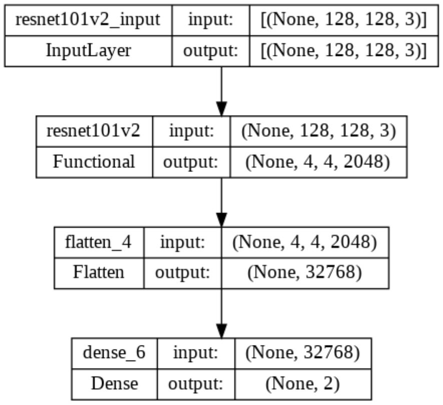

Figure 4: Architecture of the final model (ResNet101-based transfer learning model).

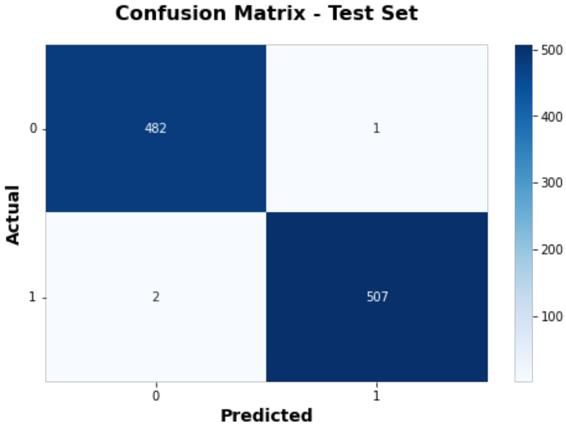

Figure 5: Confusion matrix from final model's prediction results on test set.

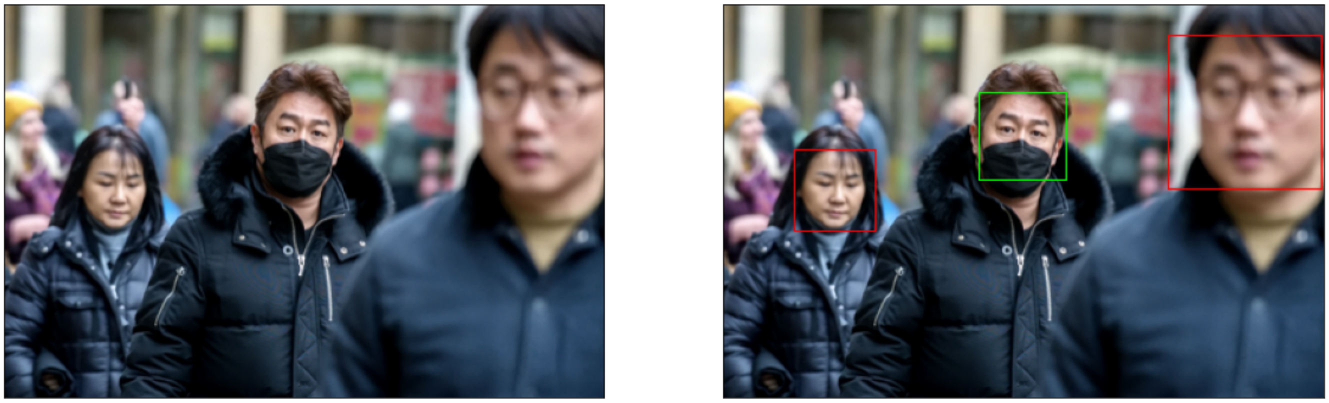

Figure 6: An illustration of the final output from face detection model and mask detection classifier given an image input. Left: Original image input. Right: Final output (green box indicates the person is masked, red box indicates the person is unmasked).

Conclusion

In this study, we implement different deep neural nets-based techniques to automate the process of face mask detection. The models we conduct in the analysis include a customized CNN architecture, transfer learning-based VGG-19 and ResNet101 architectures. Among all the models, ResNet101 yields the best performance overall and is selected as the final mask detection classifier. Since the data used only contains the scope of faces, face detection model is further integrated with the mask detection classifier to provide final results given any image that contains people instead of only faces. For future steps, these deep learning techniques can be further extended to video streams and can thus enable real-time face mask detection. Under the threat of the epidemic, this real-time mask detection model can be deployed for automatically monitoring people's mask status in public places, contributing to public health protection.

To reproduce this study, please follow the instructions in the repository for this project, which can be found at https://github.com/tianyunh/BIOSTAT823-Project.

References

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. 1998. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11):2278-2324, November 1998.

Loey, M., Manogaran, G., Taha, M. H. N., Khalifa, N. E. M. 2021. 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, article 108288, 2021.

Militante, S. V. and Dionisio, N. V. 2020. Real-Time Face-mask Recognition with Alarm System using Deep Learning. *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC),* 2020, pp. 106-110, doi: 10.1109/ICSGRC49013.2020.9232610.

Oumina, A., El Makhfi, N., Hamdi, M. 2020. Control The COVID-19 Pandemic: Face Mask Detection Using Transfer Learning. *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, doi: 10.1109/ICECOCS50124.2020.9314511.

Viola, P., Jones, M. 2001. Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition. CVPR 2001, *Proceedings of the 2001 IEEE Computer Society Conference*, vol. 1. IEEE, pp. I–I.