Facemask Detection

April 24, 2021

1. Abstract

The utilization of facemasks have been introduced to the social norm due to the COVID 19 pandemic. The use of facemask helps reduce the infection rate of COVID 19. The approach of our project focuses on using convolutional neural networks(deep learning) to detect whether a person is wearing a mask or not. The YOLOv4 architecture uses a single pass to classify objects. YOLOv4 measures the accuracy through the comparison of the intersection between the original bounding box with the predicted bounding box, the more these areas intersect the higher the accuracy. When running our model it shows that the model is around 90% when we are running the YOLOv4 model.

2. Introduction

Masks have been an obvious lifestyle change since the pandemic. Many institutions and businesses require masks in order to carry out daily tasks for a good reason. According to a paper published by the National Library of Medicine, masks significantly reduce coronavirus RNA in respiratory droplets[1]. Furthermore, according to the CDC not only do masks block large droplets, but also block finer droplets from entering the respiratory system. Despite having scientific proof to back the effectiveness of masks, many people either forget to wear face coverings or decide not to adapt to the current social norms, thus increasing the rate of infection. In March, 11, 2021 the WHO declared COVID 19 as a pandemic [2], and as of April, 23, 2021 COVID-19 has a national death toll of 553,092 lives [3]. Countries around the world have been seeking ways to slow down the rate of infection of the virus. One way to reduce the spread of COVID 19 is the utilization of masks [4]. In order to enforce CDC guidelines of wearing a mask, facemask detection platforms have seen an increase in popularity. In this project we experiment facemask detection using the region based convolutional neural network architecture YOLOv4. This project explores the realm of machine learning and convoluted neural networks and applying these concepts towards facemask detection.

Previous researchers in the past used the deep neural network SSDMV2 for real time face mask detection [5]. SSDMV2 uses a single shot multibox detector in order to detect the faces, and MobilenetV2 as the object classifier for the mask. SSDMV2 turned out to yield a high f score 0.93 and an accuracy of 0.9264, while at the same time offering a lightweight solution. Similar research in the past also implemented a revised model of Resnet-50 and YOLOV2 for facemask detection [6]. This study implements a two step-stage detector. Resnet-50 is used to first identify the face, while YOLOV2 classifies if the subject is wearing a mask or not. The disadvantages of the SSDMV2 and YOLOV2 method is that it also requires an additional step to detect the face. This would lead to more inference time and a longer time to train the model.

3. Proposed Methodology

The main goal of our project is to detect whether a person is wearing a facemask or not. The first step is to collect all our data to train the model. The datasets that have been used are from Kaggle containing images of Masks and No masks. Links to the dataset have been provided below section 3.1. After processing the data we began implementing the deep learning models and began training it. To see the most accurate model, we tested multiple models: Inceptionv3, ResNet50, MobileNetV2, InceptionResNetv2, and EfficientNetB2.The reason why we chose these models because each of these models seem to perform the best on pretrained weights imagenet. With our limited dataset we would use Transfer learning.Transfer learning is where a model that is already pretrained is repurpose for another task, in our case it is used to detect masks. Each of these models ran with one of our dataset called facemask_data_v1 containing 30,729 images. Since all of those models require a detector and a classifier our goal was to have only a singular pass through of the model. Eventually, we would use YOLOv4 developed by pjreddie/darknet to demonstrate the face detection model: https://github.com/pjreddie/darknet. In fig 1, we show how the data was processed and fed to the YOLOv4 model onced trained we tested it on our images.

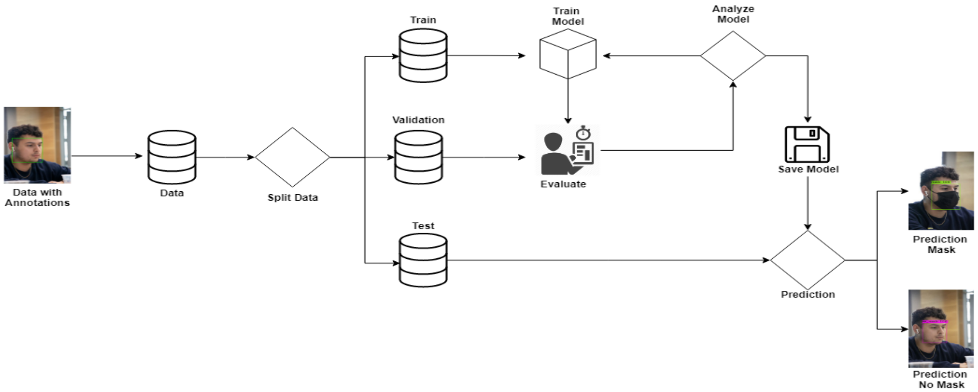


Fig 1: Showing the steps that we used to create the facemask detector.

3.1 Dataset

When it comes to working with deep learning models that deal with computer vision, to allow the machines to detect images, the model needs a lot of data to train. Each dataset came from Kaggle. For each of the datasets, there are two folders, one containing images of with_mask and the other without_mask those images are what we would use to train our model. Finally, we would merge all the data from each of the datasets by class into one folder called facemask_data_v1. Meaning, all of the with_mask images would be placed in a folder named with_mask, and all the no_mask will be placed into a folder called no_mask. The final size of the folder contains 30,729 images of mask and no_mask(15,337 images in mask and 15,392 images in no_mask).

To run the YOLOv4 model, we need to have a dataset that contains annotations. So the dataset that was used to train the model was a dataset with annotations in it (section reference). We will then process the annotation to create the required text document that is needed for training to find the IOU which is the Intersection of Union to find if the model is making the correct prediction.

3.2 Object-Recognition

After grabbing our dataset we want to start using a deep learning model for object-recognition. Each model implements computer vision in order to gather input to feed into the neural network. Image classification is the action of predicting a class of an object. Object localization involves identifying the boundaries of an object. Object recognition combines both image classification and object localization by identifying an object and drawing a box around the boundaries [7]. Object recognition is a key output towards non-regression problems in deep learning. We begin experimenting with the models such as EfficientNetB2, InceptionV3, InceptionResNetV2, MobileNetV2, MobilenetV3, and YOLOv4. For the six models we trained we implement training with and without transfer learning. For the initial weights of our models we start off with imagenet weights. Imagenet is an image database with over 14 million images, and 1,000 classes [8]. The imagenet weights are a great starting point for our neural net due to the sheer volume of images and classes imagenet was trained on. In order for a neural network to train, the network must pick out the details about the subject. Since imagenet weights have a broad variety of classes and images, many details are already recognized by the network (i.e the circular pattern of an eye in a person's face). As the trained data is passed through each layer of the neural network, the more the neural network can recognize detailed features. Out of the six models trained this study focuses on the YOLOv4 architecture. The other five models require a two step operation by first detecting the face, then detecting a mask. However, since most of the face is covered by the mask the mask detection performs poorly. YOLOv4 only requires a single pass for facemask detection. YOLOv4 also creates a bounding box around the subject, mapping out a region of interest.

3.3 Object-Localization

Image classification involves predicting the class of one object in an image. Object localization refers to identifying the location of one or more objects in an image and drawing a bounding box around their extent. Object detection combines these two tasks and localizes and classifies one or more objects in an image.In YOLOv4 the model uses object localizations, by using the images annotations the model is able to draw on a bounding box before making predictions. In YOLO after the model has made its predictions. The model will use mAP(mean average precision) to see the accuracy of the mode[9]. AP (Average precision) is a popular metric in measuring the accuracy of object detectors like Faster R-CNN, SSD, etc [9]. By displaying the precision while recalling the training. We can see how good the top prediction is and save the best weights. The model also then uses IoU which is the intersection of Union, IoU measures the overlap between 2 boundaries[9]. IOU is used to measure how much our predicted boundary overlaps with the ground truth (the real object boundary). Once the IoU calculations reach closer to 1 and cannot grow any more, this shows that the model has trained enough and will have an accuracy of between those numbers.

Fig. 3.3.1: shows how IoU work and calculates.[9]

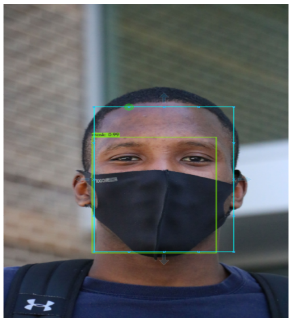
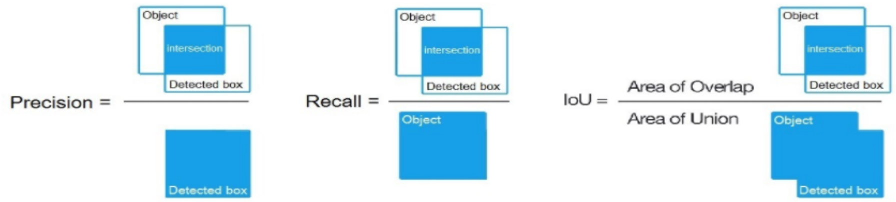


Fig. 3.3.2: shows the mathematical calculations of Precision, Recall, F1. [9]



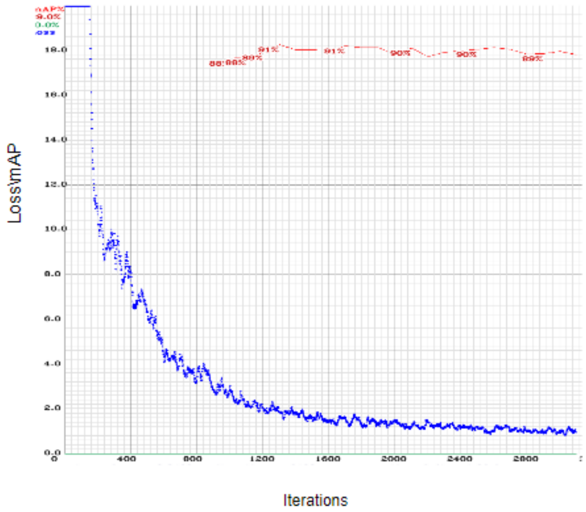
4. Experimental Result

Fig. 4.1:This is the accuracy of all of the models that we ran and how each of them perform using Imagenets for their weights.

Model Name	Test Accuracy
InceptionV3	W/O TL: 0.9965
	W TL: 0.998
ResNet50	W/O TL: 0.9975
	W TL: 0.9965
MobilenetV2	W/O TL: 0.9989
	W TL: 0.9876
Inception ResNetV2	W/O TL: 0.998
	W TL: 0.9975
Efficient NetB2	W/O TL: 0.9999
	W TL: 0.9989

In these trials we ran we noticed that the model is overfitting. Overfitting is when the model is learning too much. Meaning it is learning patterns that should not be learned and making predictions off of that.

Fig: 4.2: This is the accuracy and performance of the YOLOv4 model:



In this graph the model ran from around 3,000 epochs and at the beginning we can see the loss of the graph was all the way on top and slowly began to decrease. Around 2800 we can see the graph beginning to plateau showing that the model is already training at its best, indicating to end training. As we look at the top of the chart it displays the mAP% with a prediction of starting at 88% it slowly went up to its best at 91 percent in accuracy. As the model continues to train, the accuracy ranges between 89-90% accuracy. The weights with the highest accuracy are then used and applied for face mask detection.

5. Future Work/Conclusion

The completion of this project would allow people to use our project and see how machine learning works. We would like to develop a software where people can mess around and play with the project. In the future, we would like to see the software to be implemented into identifying other objects. By changing the dataset, the software would be able to change objects that it would be detecting. So, if someone was to identify a person, the user would have to feed in a dataset that has pictures of the object/person, and the software would be able to find that said object(s).

References

1.) "COVID-19 Provisional Counts - Weekly Updates by Select Demographic and Geographic Characteristics." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 21 Apr. 2021, www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm.
2.) M;, Cucinotta D;Vanelli. "WHO Declares COVID-19 a Pandemic." *Acta Bio-Medica : Atenei Parmensis*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/32191675/ .
3.) "Provisional Death Counts for Coronavirus Disease 2019 (COVID-19)." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 27 Apr. 2021, www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm.
4.) "Still Confused About Masks? Here's the Science Behind How Face Masks Prevent Coronavirus." *Still Confused About Masks? Here's the Science Behind How Face Masks Prevent Coronavirus | UC San Francisco*, 22 Apr. 2021, www.ucsf.edu/news/2020/06/417906/still-confused-about-masks-heres-science-behind-how-face-masks-prevent.
5.) Nagrath, Preeti, et al. "SSDMNV2: A Real Time DNN-Based Face Mask Detection System Using Single Shot Multibox Detector and MobileNetV2." *Sustainable Cities and Society*, Elsevier Ltd., Mar. 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC7775036/ .
6.) Loey, Mohamed, et al. "Fighting against COVID-19: A Novel Deep Learning Model Based on YOLO-v2 with ResNet-50 for Medical Face Mask Detection." *Sustainable Cities and Society*, Elsevier Ltd., Feb. 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC7658565/.
7.) Brownlee, Jason. "A Gentle Introduction to Object Recognition With Deep Learning." *Machine Learning Mastery*, 26 Jan. 2021, machinelearningmastery.com/object-recognition-with-deep-learning/.
8.) *ImageNet*, www.image-net.org/index.
9.) Hui, Jonathan. "MAP (Mean Average Precision) for Object Detection." *Medium*, Medium, 3 Apr. 2019, jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31**173.**

<u>Datasets:</u>

10.) https://www.kaggle.com/andrewmvd/face-mask-detection

11.) https://www.kaggle.com/sumansid/facemask-dataset?select=No+Mask

12.) https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset

13.) https://www.kaggle.com/omkargurav/face-mask-dataset

14.) Merged dataset:

15.) https://mslivesfasu-my.sharepoint.com/:f:/g/personal/chieney_jacks_sfasu_edu/EjtPQcpJNn9Ji_zviLw8V3wBnqs4SMl5lrNmN89V0ewgPA?e=H4fgyG

16.) YOLOv4 dataset:

17.) https://github.com/TheSSJ2612/Real-Time-Medical-Mask-Detection/releases/download/v0.1/Dataset.zip