**Face Mask Detection**

Ori Kleinhauz, Michael Goncharov, Andrey Bakhmat, Jonathan Bitton

ITC Data Science Fellows June 2020

horizontal line

## Problem Definition

Following the worldwide Covid-19 outbreak in early 2020, virtually all governments have imposed new rules of conduct on their citizens, among which is the obligation to wear a face mask at all times in public. While this obligation is trivially easy to follow, it is surprisingly hard to enforce. So far, the detection and penalization of disobedient individuals has been contingent upon a law enforcement officer (LEO) being present at the time and place a violation has occurred. This poses several problems among which are:

1. LEOs can only be present at some places some of the time, so a lot of violations go unnoticed.
2. Even in cases where an LEO is present, a group/individual that violated the rules may get reprimanded, only to continue their misconduct when the LEO has left.
3. While LEOs may patrol a street, park, or any other public space freely, they cannot feasibly be placed in private spaces such as shops, synagogues, and event venues, to name a few.

How should governments, businesses, and other organizations deal with this issue? Clearly, despite the huge efforts made by law enforcement to combat this phenomena, relying on their supervision is simply not enough. We propose a real time face mask detector to be used in video surveillance wherever it exists, be inside or outside, in the public sector or the private one. We aim to devise a model which will accurately determine whether each and every person in a given frame is wearing their mask properly - a solution which serves three purposes:

1. Alleviating the workload from law enforcement to save resources such as time and money.
2. Allowing organizations to monitor their environment in real time so as to avoid being fined for violations taking place on their premises.
3. Encouraging individuals to remain obedient so as to minimize their and their surroundings exposure to the virus and thus reduce its spread.

## Existing Methods

Being a trendy topic nowadays, the internet abounds with examples of mask detection projects. A simple search for ‘face mask detection’ on Github yields no less than 924 repository results, and ‘face mask detector’ yields an impressive 407. Clearly, it is beyond the scope of this report to go through any or more of them in detail, but we do want to point out [one kaggle project](https://www.kaggle.com/andrewmvd/face-mask-detection) in particular, in which a ‘mask worn improperly’ class was incorporated as well - something which we intended to do but had not found the right data to enable us to do so.

## Dataset Description

The original dataset for this project was taken from a [github repository](https://github.com/prajnasb/observations/tree/master/experiements/data) which is referenced by many projects we found. It comprises two classes - with mask and without mask, where the masked photos are composed of frontal face pictures overlaid with a generic mask picture placed in different angles and sizes.



This dataset is small but balanced- it consists of 1376 files, with roughly 50% belonging to each class, which leaves less than 300 images for the test set (at 20%) to evaluate a model on, and is inherently incomprehensive as all masks look exactly the same.

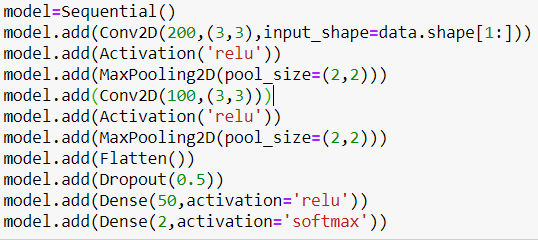
The second, and final dataset we chose to use in this project came from multiple sources including Bing Search API, Kaggle datasets, and RMFD dataset, and can be downloaded from [here](https://drive.google.com/drive/folders/1XDte2DL2Mf_hw4NsmGst7QtYoU7sMBVG). This dataset consists of 3835 images, again almost perfectly balanced. Some examples are:



The reason why we decided to switch datasets will be described in the following section

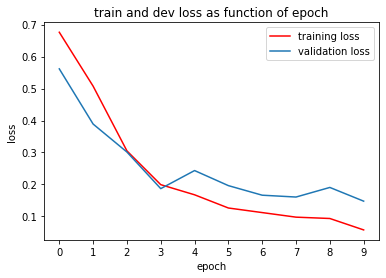
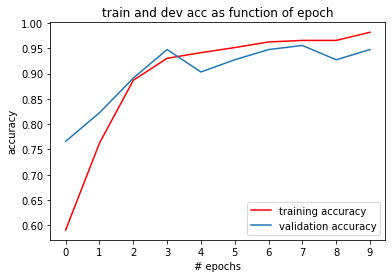
## Model Comparison

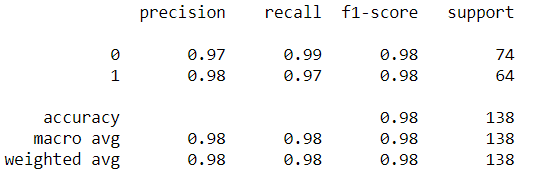
For this project we chose to apply transfer learning with a simple but powerful sequential CNN:



The images were resized, converted to grayscale, and normalized for optimal performance.

When trained on the original dataset, it achieved 97.8% accuracy on the test set following 10 epochs of training.



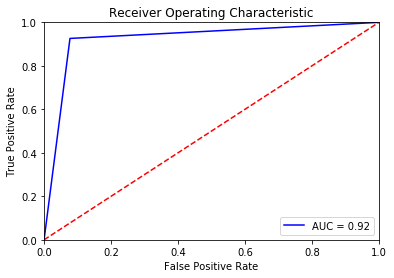
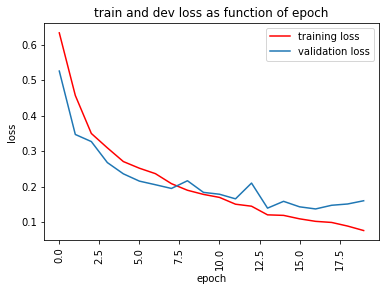
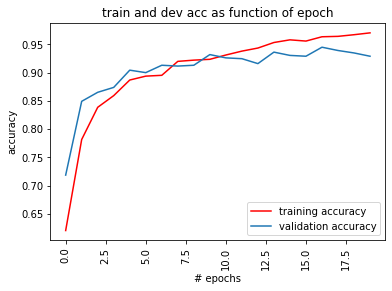


Although these results seemed astounding at first, it turned out that when making predictions on real life examples namely, selfies, it always outputted ‘without mask’ - rendering it completely useless for video detection, which is performed on a frame-by-frame basis using OpenCV Haar cascade object detection.



This implied that the model has somehow overfitted the data - be it due to the homogeneity of the masked images, the mask’s brightness, or some other reason we are unaware of.

Upon reaching this dead end, we decided to look for another dataset and examine the alternative results. We used checkpoints and early stopping, and tested the model’s video performance after each run (these typically took 12-20 epochs). Surprisingly, it weren’t necessarily the models that achieved the best accuracy on the test images that performed best in video (these ranged from 92%-99%).



We saved our best performing model (video-wise) and settled for that, although it should be noted that we are certain of the possibility for further improvement by testing more and more trained models.

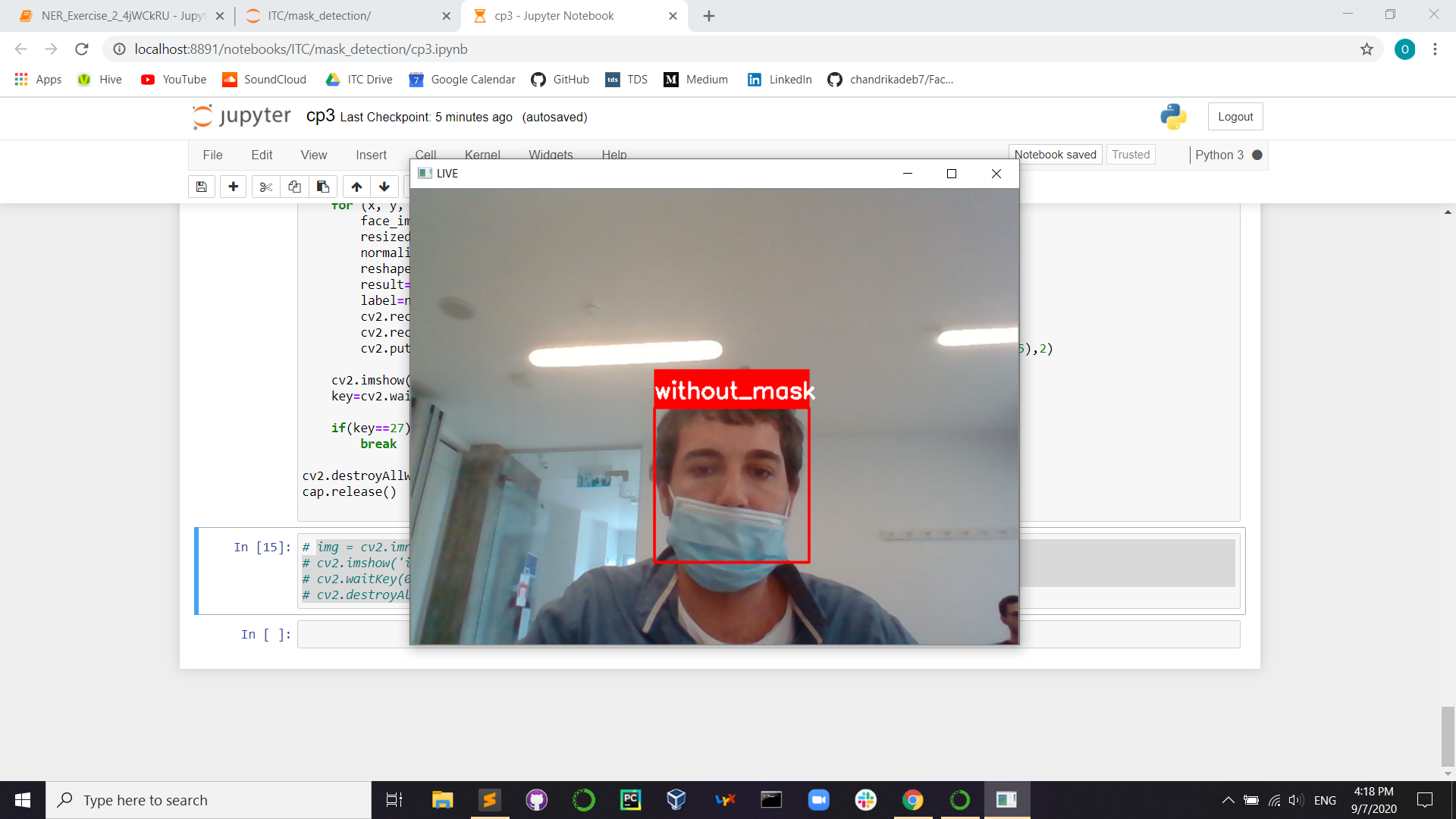
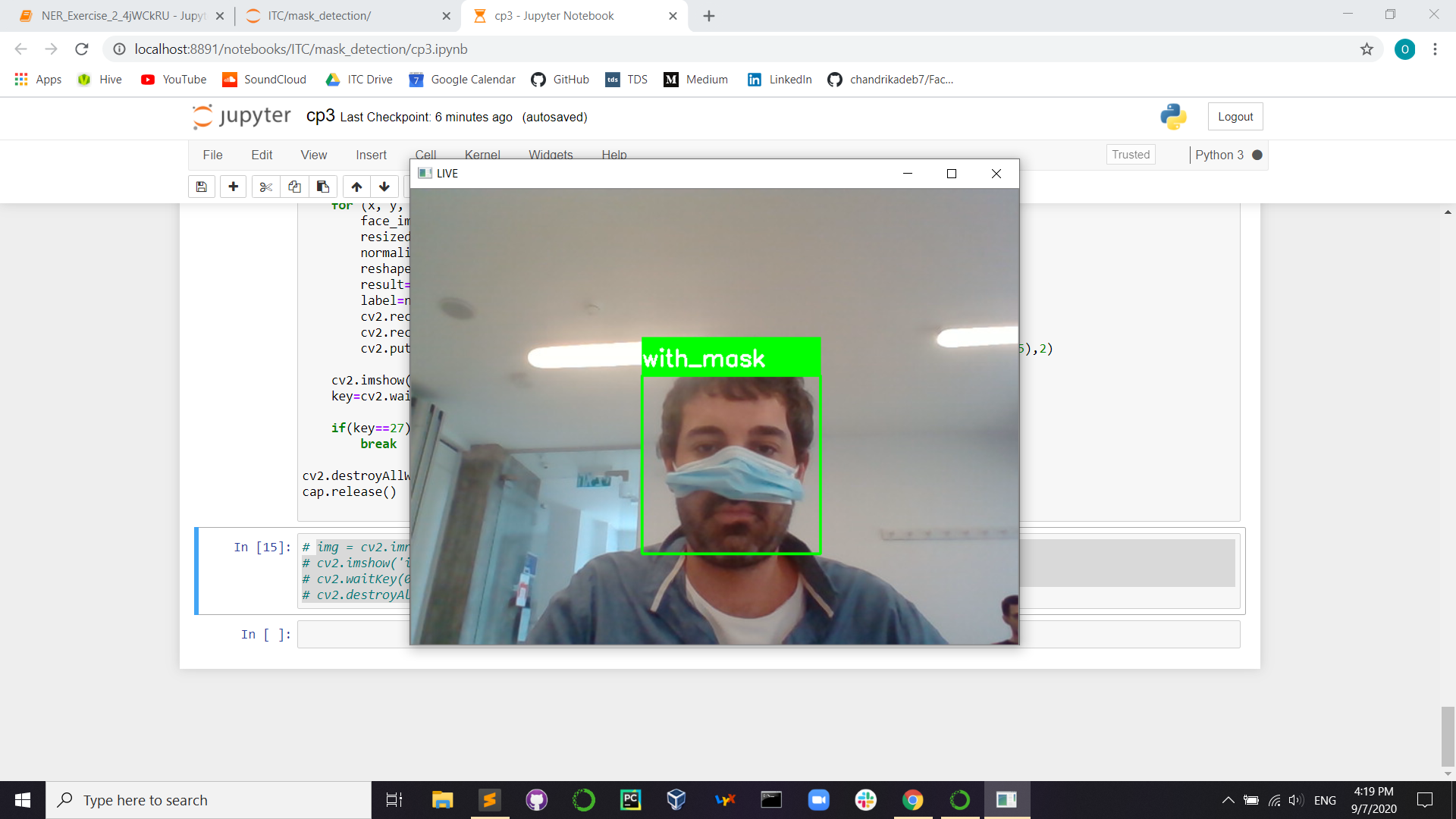
## API

As it turns out, deploying a webcam based model on a remote server to be used with a local machine isn’t that straightforward, and cannot be accomplished using python alone. In order to do so, a video must be sent in real time from the client’s webcam to the inference server, which, to our best knowledge, requires JavaScript based browser integration. Thus, we regrettably could not implement such a mechanism given the time and knowledge constraints. We did write a simple local API using flask, however we are not certain what added value this brings, as it is completely analogous to running the inference code.

## Discussion

The first challenge we encountered, as described earlier, was the inability of our model to generalize to new data when trained on the original dataset. And though this seemed intimidating at first, it wasn’t so hard to overcome - we simply had to train on another dataset.

The second challenge we encountered has to do with the model’s performance, and is more formidable. In the narrow scope, it has the implication that the model places to much (or to little) importance on certain features, e.g the nose:



This might seem somewhat disturbing, but actually the chances of someone wearing their masks in the way that is seen in the left images are extremely low. If anything, we should be concerned with what happens in the right picture - should this correctly be labeled as ‘without mask’? We are not health professionals, so we don’t know. But in a broader scope, the fact that different models trained on the same data with similar performance on the test set gave such radically different results in video (for some of them the predictions displayed a seemingly random behaviour) is something worth reflecting on. Is there a way to achieve consistency in quality? And if so, how?

The last challenge we encountered was serving our model. This has been discussed in detail in the API section of this report.

How could we further improve this solution? A few suggestions come to mind:

* Manually sifting through the dataset to prune it of ambiguous or otherwise anomalous images that may negatively affect the model’s performance
* Apply a moving average on a fixed time window to reduce the prediction’s sensitivity to error (some frames might get misclassified, but when averaging over a 3s window they would have no impact on the prediction)
* Enrich the dataset with images of partially worn masks and add a third class ‘mask worn improperly’
* Attempt to detect faces and determine mask status non-frontally (we suspect the second part would be highly challenging due to lack of detail)

In summary, working on this project was fun and enlightening - a development process carried out from A-Z with minimal supervision, which was successful for the most part and enjoyable throughout. We all feel very grateful for this opportunity, and can’t wait to apply our newly acquired skills in tackling more real-life problems over the course of our internships, and later in our careers.