# Face Detection & Face Mask Classification

Constantinos Makris - cm01428@surrey.ac.uk
Alexandros Constantinou - ac01173@surrey.ac.uk

**University of Surrey**
**Guildford, UK**

**ABSTRACT - Face Masks, the most efficient solution to COVID-19 disease spreading, have become an everyday item that we must wear in certain areas and spaces. In most countries, wearing a face mask indoors in public places is an obligation, and it needs to be controlled by the owners of the place. Being able to control everyone that enters a building on whether he wears a mask and if it is worn correctly, takes a lot of effort. An efficient recognition system would make this task easy, fast, and efficient. As this is an important and sensitive task, a fast and reliable tool must be undertaken. Such a system would need large datasets to be trained on, with clear data of people wearing masks correctly and incorrectly, and not wearing masks. It would then need to learn to classify new images based on the training that was carried out. This report analyses the implementation of this system with a classic implementation of faster R-CNN (faster Region-based - Convolutional Neural Network), and we also implement a new approach using the DETR (DEtection TRansformer) model. The aim is to compare the two approaches, find the one that performs better and present our findings.**

## 1. Introduction

The pandemic that we are now facing as humanity, has changed everyone's lives and habits. The main change to our lives is that it has become essential to wear a face mask, especially when entering an interior space. In most countries, the owners/workers of shops and restaurants must monitor the people entering their indoor space. This task would require at least one more employee for each entrance of the place. Small businesses with a few employees may not be able to afford another person to perform this task. Instead of an employee, this task can be handled by an Object Detection and Face Recognition system.

As this task requires reliability and immediate response, we will try to find a good solution throughout this project. To do this we will compare the performance and reliability of a current system that is using faster R-CNN, with our implementation of a new system that is using the DETR model.

We will use the same data to train both models and compare their accuracy and performance. The procedure for the implementation and research of this project is as follows:

- Adjust the dataset [7] to fit the models, and split the data into training and testing.
- Implement the DETR model for our purpose.
- Train both models with the same data
- Create graphs and outputs that visualise the performance of each model.
- Compare the models and present the findings.

## 2. Literature Review

### 2.1. Faster R-CNN

Faster RCNN is a state-of-the-art object detection system that consists of two main parts. The first part is the Region Proposal Network(RPN) which serves as the ‘attention’ mechanism, taking all the reference boxes (anchors) to output a set of good proposals for objects. Fast R-CNN is the second part of the Faster R-CNN system. It is the detector of the system, and it looks where the RPN tells.

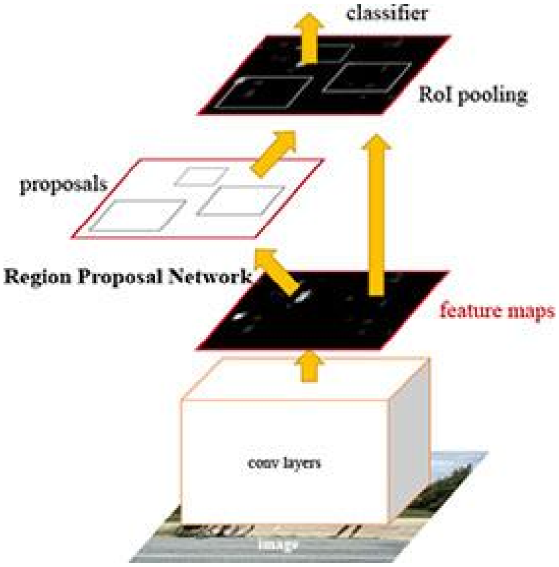


**Figure 1 - Faster R-CNN Architecture**

The steps that Figure 1 illustrates are followed by a typical Faster RCNN approach. An image is passed as an input to the convolutional layers which return the feature maps for the input image. Then the system applies the RPN on the feature maps which return the object proposals with their objectness score. Then the system applies a Region of Interest (RoI) pooling layer on the proposals and then they are passed to a classifier that will classify and output the bounding boxes for each one of them.

### 2.2. DEtection TRansformer

Two-stage detectors, (like the R-CN family) predict boxes w.r.t. proposals. In DETR, this hand-crafted process is removed and the detection process is streamlined by treating object detection as a direct set prediction problem. Two things are essential for direct set predictions:

1. A set prediction loss (bipartite matching in our case) that forces unique matching between predicted and ground-truth boxes
2. An architecture that predicts a set of objects and models their relation

DETR recognises the relations of the objects and the global image context to output a set of predictions in parallel, according to the number of queries used. Hence, DETR is very efficient due to this parallel nature.

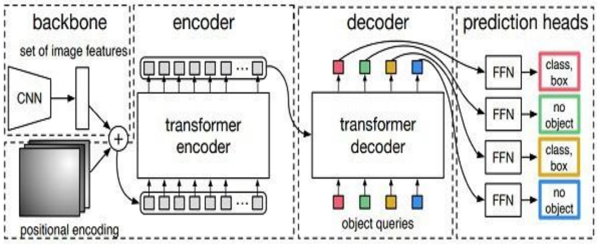


Image from the original paper

**Figure 2 - Detection Transformer Architecture**

A conventional CNN backbone extracts the image features to produce a 2D representation, which is flattened and supplemented with positional encodings to pass into the transformer encoder. Then, the decoder takes a small fixed number of object queries and the encoder output to produce the output embeddings which pass through a shared feed-forward network (FFN) to predict the final boxes and classes, one for each query.

## 3. Methodologies

### 3.1. Data and Preprocessing

The dataset used is called Medical Mask Dataset from Humans in the Loop and can be accessed by filling in a form. It consists of 6000 public images of people with diverse backgrounds, to better represent all populations. Here is a sample image:

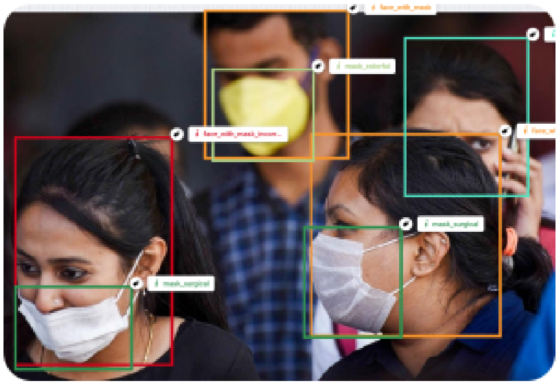


**Figure 3 - Image with bounding boxes**

The annotations included in the dataset that will be utilised in our experiments are demonstrated by the labeled boxes on the image. Each file has an entry for each bounding box (bbox), where the coordinates of the box are in the Pascal VOC format: [xmin, ymin, xmax, ymax]. The annotations are classified by the class name attribute. There are a total of 23554 bbox entries that cover 20 different classes, which can be divided into two categories: boxes of faces and head accessories. For this project we are going to use only the face entries (a total of 10853) which are distributed into 4 classes as follows:

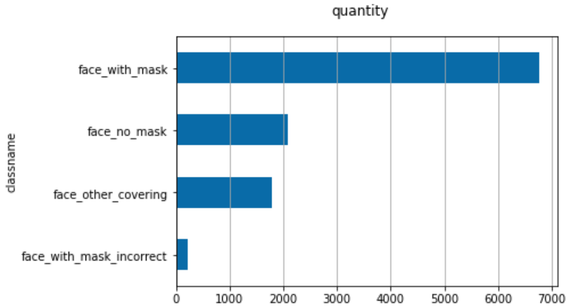


**Figure 4 - Classes of images in the dataset with their quantity**

After observing samples for each class, we found that only the face_with_mask class is considered as protected by Covid-19, which allows for a more balanced division into 2 labeled categories:

| | |
|---|---|
| 0 - protected: | 6982 samples |
| 1 - unprotected: | 3871 samples |

After the required data selection and processing, the images can be sampled through the DetrMaskDataset class, where each image is associated with a target output. Here is an input example for both architectures:

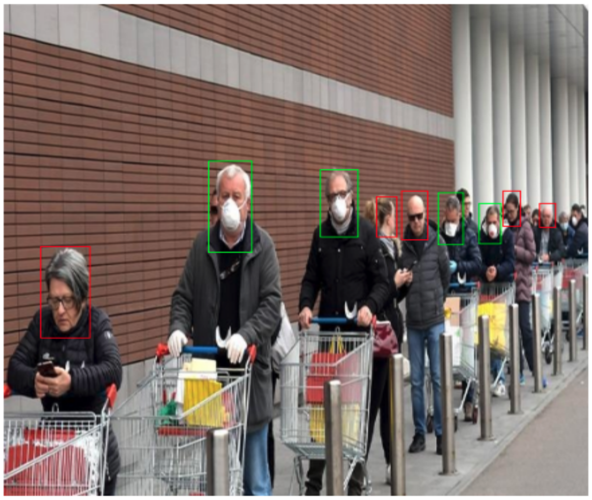


**Figure 5 - Image with bounding boxes**

where the green color in the image corresponds to the label for the bounding boxes with protected faces and the red for unprotected.

**3.2 Training and Evaluation**

**Results are shown upon running the notebook**

## 4. Future Development

This research project has a great potential for future development, which could evolve into a fully functional application for professional use.

Such an application would need much more than the model itself. An idea is for the model to be connected to a camera, that would provide a live feed of the place that needs to be monitored. The system would need to be modified to receive the live feed, and report back to the staff any cases that are classified as unprotected.

In the case that the model is used for professional purposes, it could be trained with a bigger and more diverse dataset. The dataset would need to include images from different times of the day(day/night), and with people of different ages(children/adults). Finally, the system would need to be trained with various types and colors of face masks to make it more reliable and accurate.

## 5. References

[1] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, Heling Chen, Yu Miao, Zhibing Huang, and Jinbi Liang, “Masked Face Recognition Dataset and Application”, Mar, 2020.

[2] Cabani, A., Hammoudi, K., Benhabiles, H., & Melkemi, M., “MaskedFace-Net - A dataset of correctly/incorrectly masked face images in the context of COVID-19.”, Nov, 2020.

[3] Nicolas Carion and Francisco Massa and Gabriel Synnaeve and Nicolas Usunier and Alexander Kirillov and Sergey Zagoruyko, “End-to-End Object Detection with Transformers”, May, 2020.

[4] Shaoqing Ren and Kaiming He and Ross Girshick and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Jan, 2016.

[5] “A Step-by-Step Introduction to the Basic Object Detection Algorithms” [Online]. Available: https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1

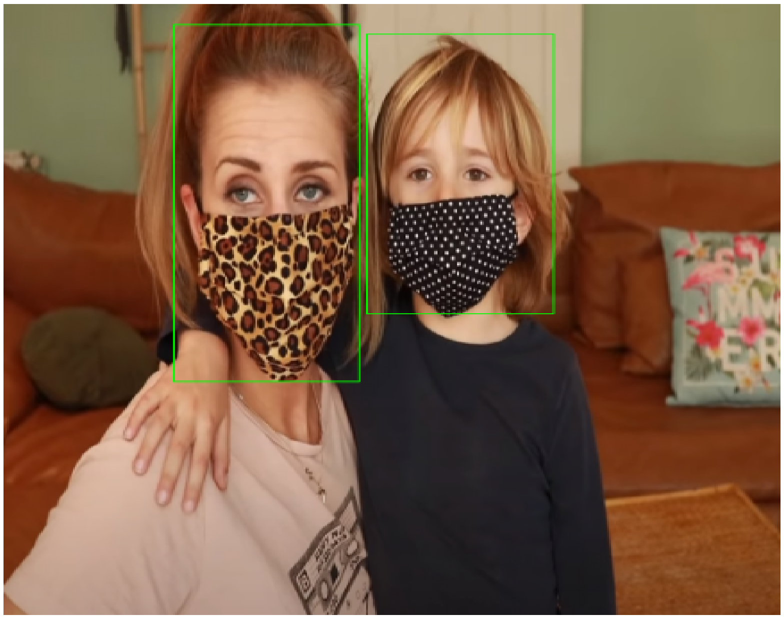
[6] “One-stage object detection” [Online]. Available: https://machinethink.net/blog/object-detection/

[7] “Medical mask dataset” [Online]. Available: https://humansintheloop.org/resources/datasets/medical-mask-dataset/

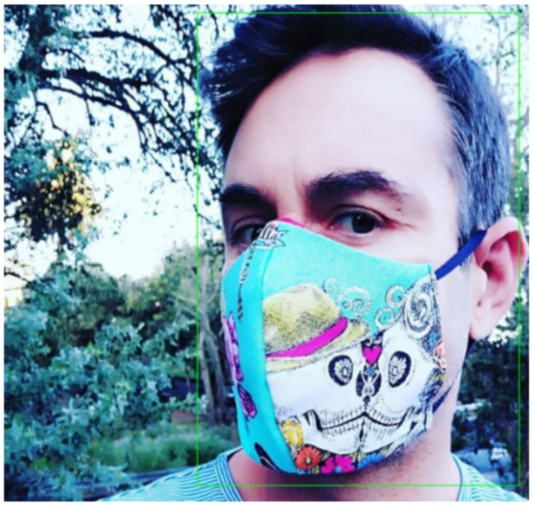
# 6. Appendices

### Appendix A: Testing Images

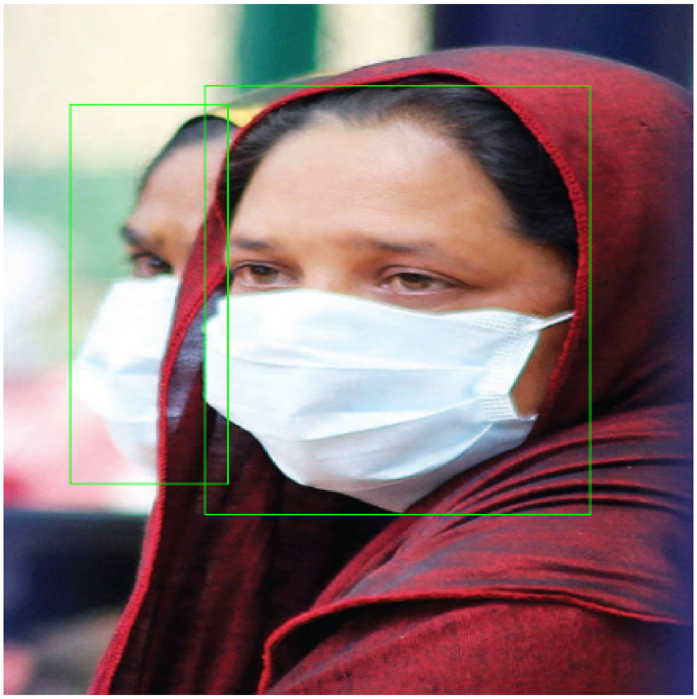
**Test image 1**



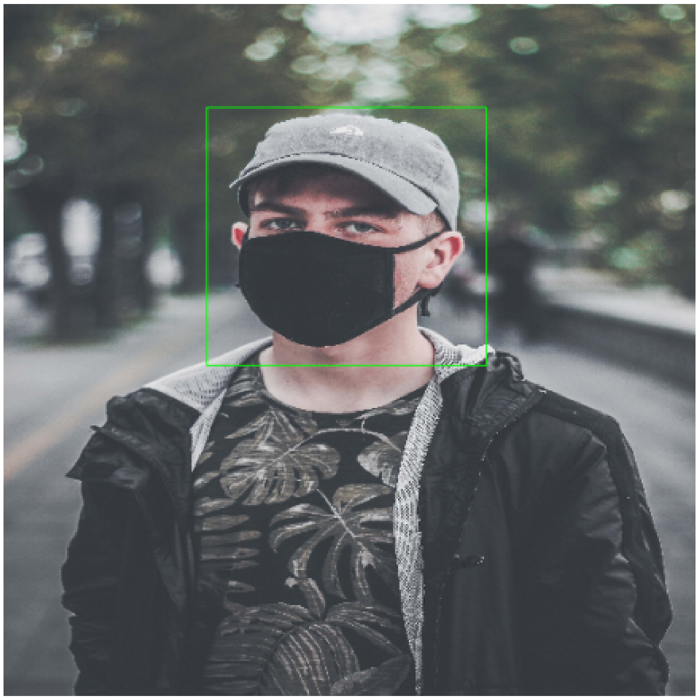
**Test image 2**



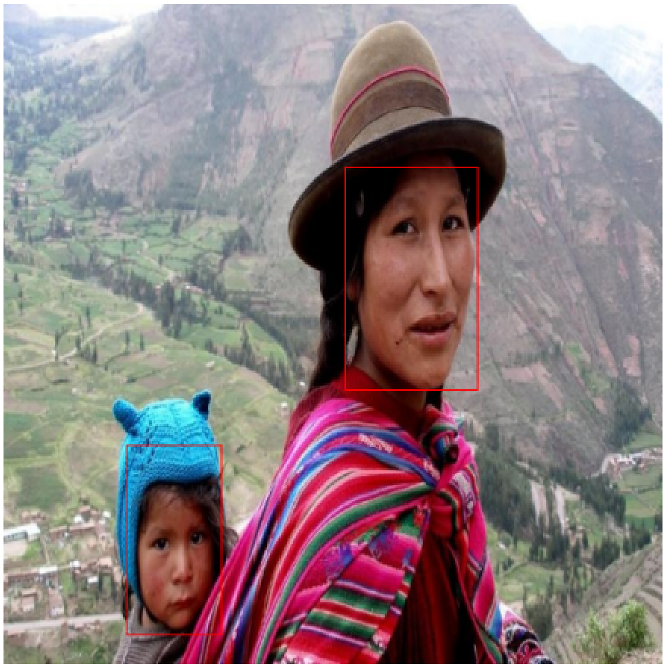
**Test image 3**



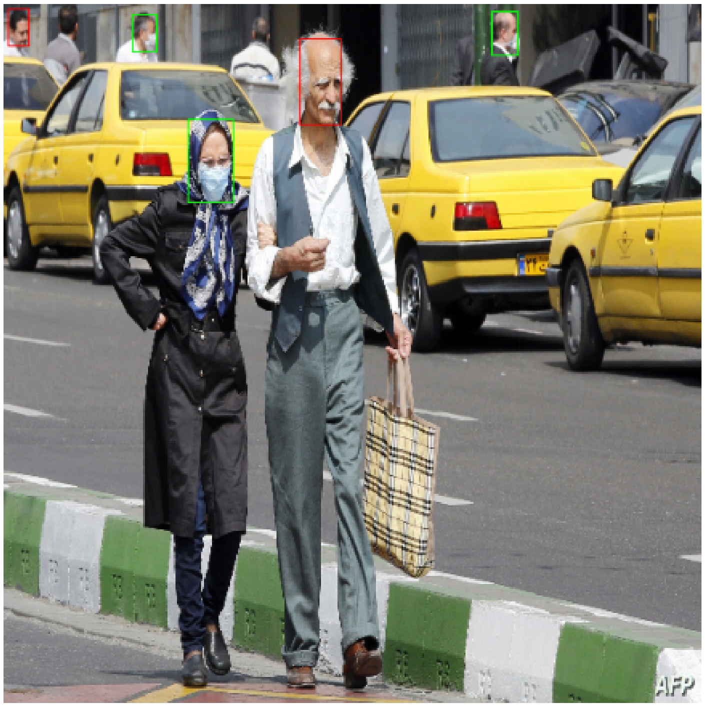
**Test image 4**



**Test image 5**



**Test image 6**


AFP