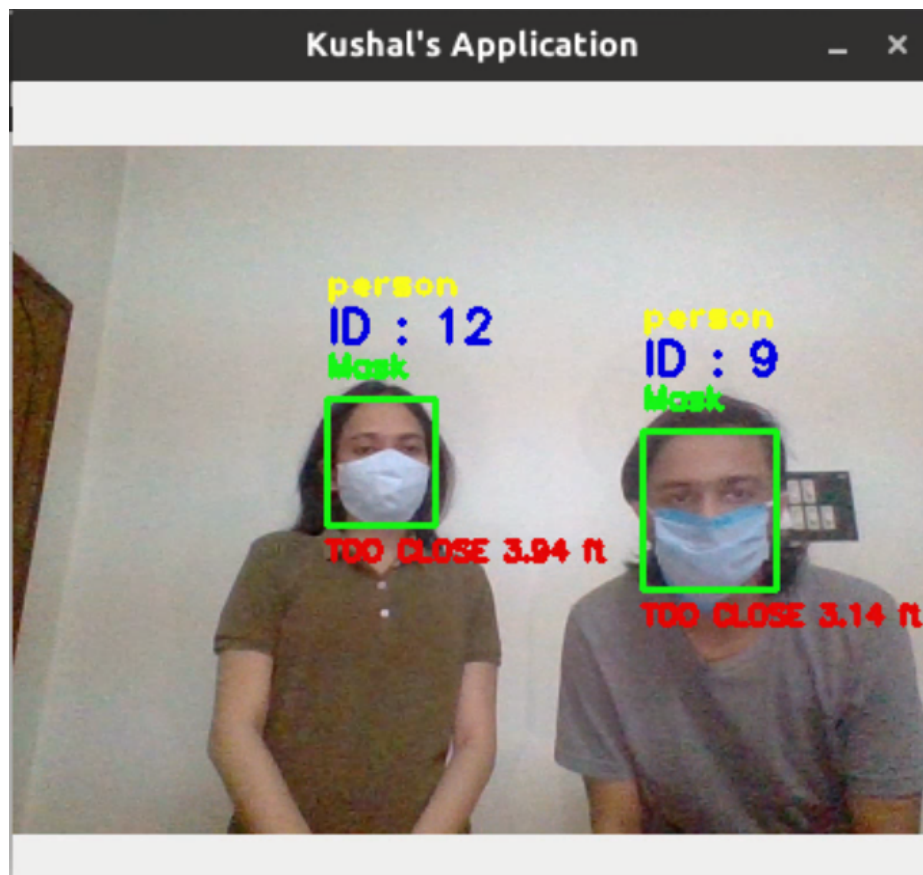


# Assignment Report

Kushal Kumar Jain

---



## Introduction

This is a report on the application made for the following problem statement :

1. If no person is in the video, it should alert "No Person".
  2. If the person walks in, detect the person and assign a unique ID.
  3. Detect the person wearing a mask or not with alerts like "Mask Detected" and "No Mask Detected".
-

- 
4. If a new person walks in, repeat steps 2 and 3 for him/her.
  5. If the two persons are standing close enough, it should alert "Maintain Social Distancing".

## Approach

Detecting the person was done by OpenCV's open source pre trained models like the mobilenet SSD. The mobilenet-ssd model is a Single-Shot multibox Detection (SSD) network intended to perform object detection. This model is implemented using the Caffe framework. This model detects objects in the video feed and classifies them into 1 out of 80 classes. We move forward once a person is detected. This solves the first part of our problem.

To assign the person a unique ID we need to track its bounding box across all frames. This was performed by the Euclidean distance tracker which keeps track of small movements in the bounding box and assigns a unique ID to all the bounding boxes.

For mask detection we first detect face by a pretrained faceNet model, and then apply maskNet (Chandrika Deb Face-Mask-Detection <https://github.com/chandrikadeb7/Face-Mask-Detection>) model to it. This part has decent accuracy and could be a subject of future improvement.

The distance has been approximated using bounding boxes and their sizes (assuming size of the face remains the same for everyone hence same size bounding boxes are formed at equal distance from the camera). Depth is calculated by the size of the bounding box. To detect the distance of people from the camera, triangle similarity technique was used. Let us assume that a person is at a distance D (in centimetres) from the camera and the person's actual height is H (I have assumed that the average height of humans is 165 centimetres). Using the object detection code above, we can identify the pixel height P of the person using the bounding box coordinates. Using these values, the focal length of the camera can be calculated using the below formula

$$F = (P \times D) / H$$

---

Then we can use this :

$$D' = (H \times F) / P$$

For depth estimation. Now that we know the depth of the person from the camera, we can move on to calculate the distance between two people in a video. There can be a number of people detected in a video. So the Euclidean distance is calculated between the midpoint of the bounding boxes of all the people detected.

Link to video : [Application2.mp4](#)