VR Mini Project

Aremanda Sri Sai Suhruth
IMT2018068

Sethana Sreevardhan Reddy
IMT2018068

Shanu Kumar
MT2020030

Problem Statement:

Build a vision based automatic door entry system, to open the door only if a human wears a mask.

Approach:

1. Detect Humans using YOLOv3 and then extract the region of interest.
2. Detect Human faces in the regions of interests using VIOLA JONES.
3. Detect masks in the faces detected using CNN.

Human Detection:

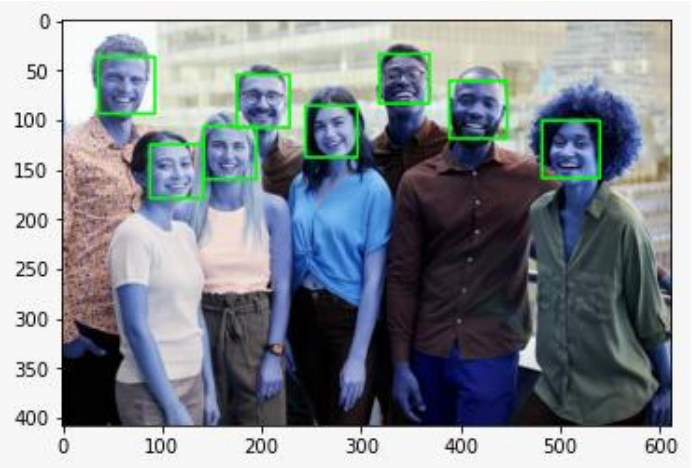
- We chose YOLO over Faster RCNN as it was more accurate over smaller regions and it was faster than Faster RCNN in several cases.
- Our initial intention was to use YOLOv5 which has shown very impressive results recently. But, due to its limited implementation (unavailable functions) we had to use YOLOv3.

Original Image

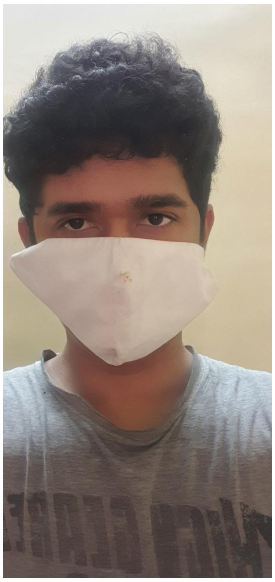
YOLO extracted image

Face Detection:

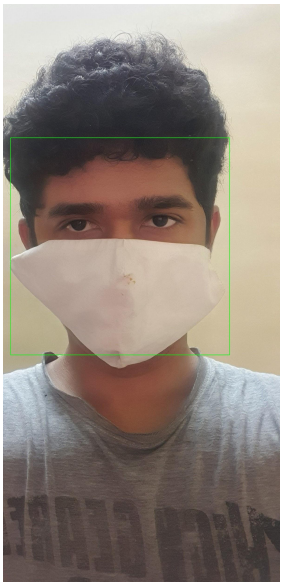
- We used Viola Jones,which basically is a Haar Classifier, a machine learning based approach where a cascade function is trained from a lot of images both positive and negative.
- Once faces are detected we send these images along with their location in the original image to the mask detector.



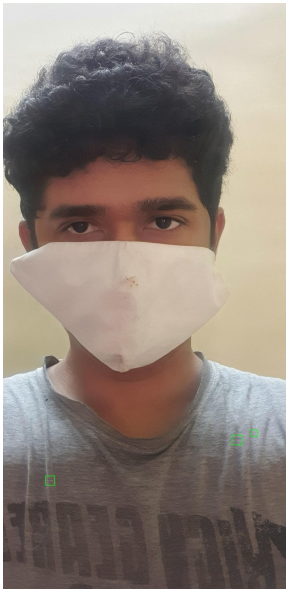
Multiple faces detected using Viola Jones



Original Image with mask (1)



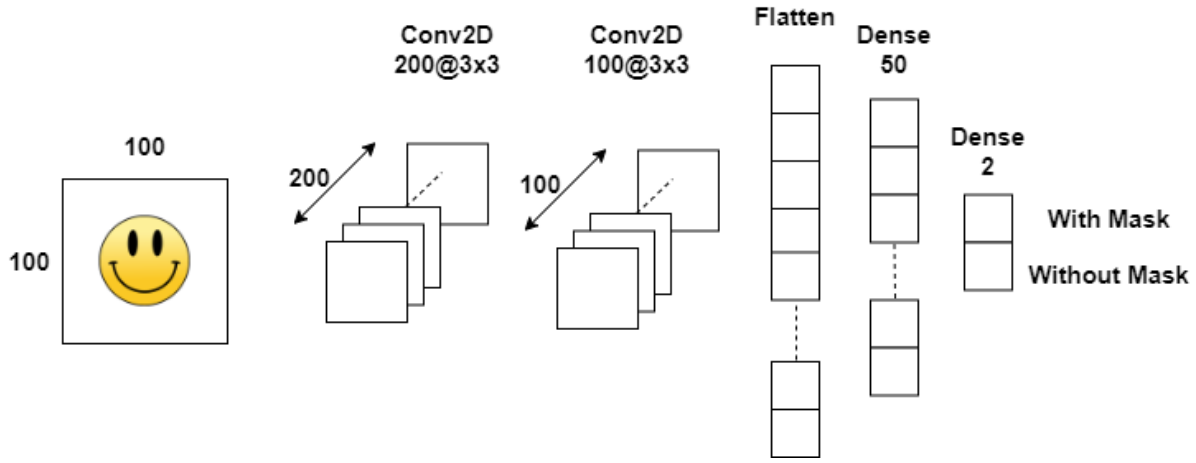
haarcascade_frontalface_alt.xml (2)



haarcascade_frontalface_default.xml(3)

As we can see in the (2) and (3) images, We found out using haarcascade_frontalface_alt.xml(2) produced the required results.

Mask Detection:



- We used a CNN with 2 convolution layers with 200, 100 kernels each and a size of 3x3, the last convolution layer was flattened and then passed to a dense layer of 50 neurons and the final output is given out by 2 neurons to detect masks.

Real Time Detection:

- First we extracted a frame from the input (image, video, camera) and applied Human and Face Detectors to find faces and then applied Mask Detector to check for masks.
- If a mask is detected a green box is displayed over the person if not a red box is displayed.
- The Human Detector extracts images of all the humans detected in the original file and they are passed to the Face detector to detect faces. The coordinates of a human and its face are saved with respect to the original image to display bounding boxes around them.

Video of real time detection:
https://drive.google.com/file/d/1QGGm5lg0OkLdXbjVh2wExBzfdqFTr53U/view?usp=sharing

References:

1. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e
2. https://towardsdatascience.com/yolov5-compared-to-faster-rcnn-who-wins-a771cd6c9fb4
3. https://jonathan-hui.medium.com/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359
4. https://github.com/ultralytics/yolov5
5. https://towardsdatascience.com/understanding-face-detection-with-the-viola-jones-object-detection-framework-c55cc2a9da14
6. https://github.com/aieml/face-mask-detection-keras