VISUAL RECOGNITION MINIPROJECT AUTOMATIC DOOR ENTRY SYSTEM

Jishnu V K
IMT2018033
International Institute of Information Technology Bangalore
Vinodkumar.Jishnu@iiitb.org

Karthik Hegde
IMT2018509
International Institute of Information Technology Bangalore
Karthik.Hegde@iiitb.org

Saad Patel
IMT2018514
International Institute of Information Technology Bangalore
Mohammad.Saad@iiitb.org

Saravan sriram
IMT2018521
International Institute of Information Technology Bangalore
Saravan.Sriram@iiitb.org

I. INTRODUCTION

While the study of techniques and concepts that improve the benchmark performance forms one aspect of the object detection and recognition tasks, the other dimension of the problem is to make it work in real-time scenarios. The focus then shifts towards the analysis of the time efficiency of the models as well. Techniques such as Faster RCNN, YOLO have occupied an important place when it comes to building real-time visual recognition systems. This project aims at building one such system - 'automatic door entry system under covid pandemic rules of wearing a mask'. The system will first detect a human and then his/her face and will classify whether he/she is wearing a mask or not. The entry to the person is denied if he/she is not wearing a mask, not covering the mouth and nose.

II. HUMAN DETECTION

The initial module in the pipeline is the human detection. The purpose of this module is to detect any humans in the video feed and forward the bounding boxes and cropped human images to the subsequent modules in the pipeline. The detector used was faster RCNN pretrained on the COCO 2017 dataset. Experimentation with various backbone architectures like ResNet, MobileNet and retinaNet was done to find the best suited architecture for our need. Finally, the decision to choose MobileNet was taken due to the fact that all the models had similar satisfactory performances with respect to accuracy. While comparing forward pass time and memory requirements, mobileNet 320 was a clear winner with under 0.2 second forward pass time and very low memory requirement.

III. FACE DETECTION

This module of the pipeline had the job of detecting the faces of the humans detected in the previous stage. This module worked with the Viola-Jones method using the haar-cascade classifiers provided in opencv. The pretrained classifiers provided seemed to work best when the full face is visible. Masked faces had a possibility of escaping detection. To counteract this, the hyperparameters of the classifier needed to be relaxed to ensure all faces are captured. This causes occasional false positives to be detected but none of them are persistent through-out multiple frames. The forward pass for this module was consistently under 0.07 seconds.

IV. MASK DETECTION

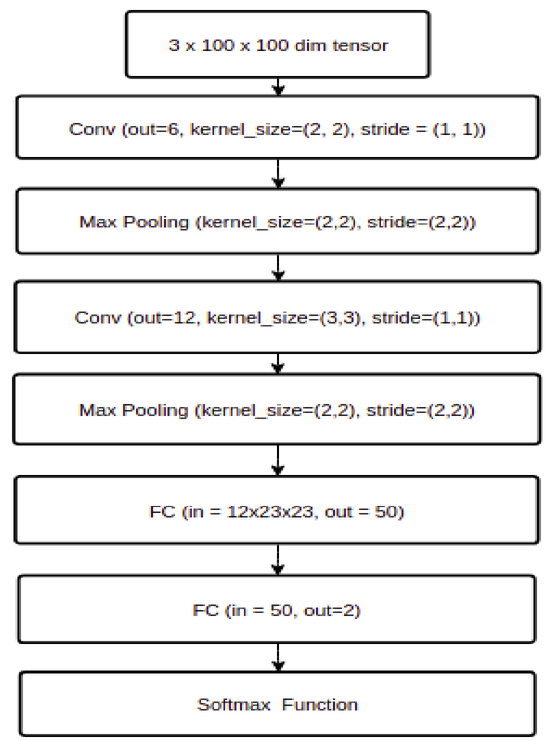

Fig. 1. CNN architecture for mask detection module

The mask detection part is the final part in the pipeline. It is concerned with classification of the face-recognized image into classes 'with mask' and 'without mask'. The flow chart depicted in the figure 1 describes the model architecture used in the mask detection module. Relu activation was used for non-linearity in all layers. Dropout layer of about $p = 0.5$ was added in between the first FC layer and the second FC layer to regularize and increase the performance. We had also tried with batch-normalization but the accuracy score was less.

Initially, the dataset was split into train and test for fine-tuning the hyper-parameters. After the fine-tuning process, the model was retrained on the entire dataset to improve the performance of the system. The loss curve while training on the entire dataset can be seen in the figure 2. The trained CNN model was then saved and loaded into the detection system's code. The standalone module takes only about 1.3 to 1.6 milliseconds to classify the face image.

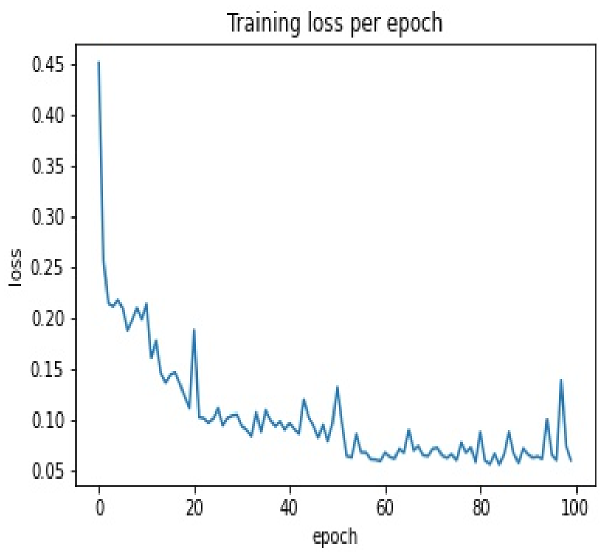

Fig. 2. Training loss curve of mask detection CNN architecture

V. Final System and Performance

The final system uses opencv to access camera frames from the device camera. It passes the image through a pipeline of the 3 modules to get the required results. It detects if there are any humans present in the view. If not, the door need not be opened. If humans are detected, their faces are detected and cropped. The mask detector checks if the faces are wearing masks. Only if all humans are wearing masks, the door will be opened. The door will not be opened if there exists a human whose face could not be detected(in case the face is out of view or the human actively tries to hide their face).

The whole system requires around 270mb of memory to run. Each frame from the video feed can be processed in 0.25 seconds on CPU. The system is fairly real time. For an automatic door opening system, processing 1 frame in a second can be sufficient for proper working.

Module	Performance
Human detection	under 0.2 sec forward pass time
Face detection	under 0.07 sec forward pass time
Mask detection	$1.3 - 1.6$ms classification time

VI. Conclusion

In conclusion, we have successfully built our first real-time object detection and classification system. The system can be installed in all public places for ensuring the safety of the citizens.

VII. References

- Face Mask Detection Dataset
- Viola Jones Method from openCV