Image Processing and Computer Vision Report

The No Entry Sign Challenge

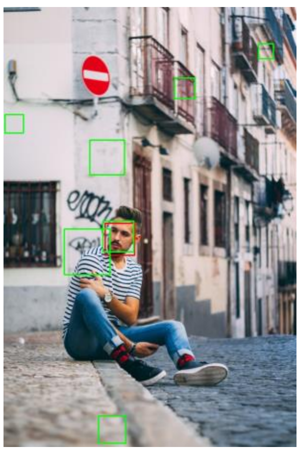
Pragya Gurung (wq19451)

1. Viola-Jones Object Detector

Ground Truth and Visualisation

The Viola-Jones object detector, used the given classifier *frontalface.xml* to detect the faces in the images below. Green bounding boxes are the faces that were detected and the red boxes are the ground truths used to visualise the true values. The true values for each input image were determined manually. Doing so, caused some ambiguity as a face does not have a clear bounding region.

Since the Viola-Jones detector detects frontal faces, any side profiles would not be true values. This can be seen *in Figure 1 NoEntry2.bmp* where the two profile faces near the centre of the image were not found by the detector.

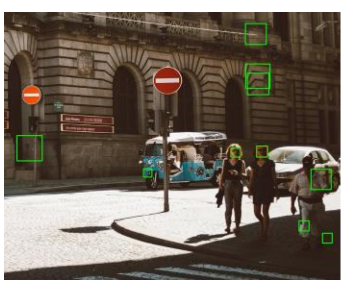


NoEntry1.bmp

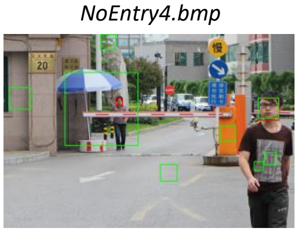
NoEntry2.bmp

NoEntry5.bmp

NoEntry4.bmp



NoEntry11.bmp



NoEntry7.bmp

Figure 1: Viola-Jones face detections (Green), Ground Truth (Red)

IOU, TPR, F1-Score

Image	TPR	F1-Score
NoEntry0.bmp	NaN	NaN
NoEntry1.bmp	1	0.25
NoEntry2.bmp	1	0.24
NoEntry3.bmp	NaN	NaN
NoEntry4.bmp	1	0.31
NoEntry5.bmp	1	0.33
NoEntry6.bmp	NaN	NaN
NoEntry7.bmp	0.5	0.2
NoEntry8.bmp	NaN	NaN
NoEntry9.bmp	NaN	NaN
NoEntry10.bmp	NaN	NaN
NoEntry11.bmp	1	0.33
NoEntry12.bmp	NaN	NaN
NoEntry13.bmp	NaN	NaN
NoEntry14.bmp	NaN	NaN
NoEntry15.bmp	NaN	NaN
Average	0.92	0.28

Table 1: Viola-Jones face detections results

Table 1 shows the True Positive Rate (**TPR**) and the F1-Score of the Viola-Jones detector on the images from *Figure 1*. Assessing the TPR meaningfully is difficult when images do not contain the object that's being detected. Therefore, images that did not contain frontal faces and so did not yield relevant outputs, have a TPR and F1-Score of NaN. For most, TPR was 1, which means that every valid frontal face was detected in those images. Mathematically, TPR is the number of correct detections over the number of total ground truths for an image.

$$TPR = \frac{TP}{FN + TP} \qquad IOU = \frac{Area\ of\ intersection}{Area\ of\ union}$$

The number of correctly detected faces is calculated with the Intersection over Union formula (**IOU**) in conjunction with a threshold. IOU is used to determine the relative overlap of a truth box and a region classified by the detector. The threshold of 0.4 was imposed manually and dictates whether the overlap is considered large enough. However, this means that it is entirely possible to achieve a TPR of 1 for every image if the threshold is relatively low. Including a threshold allowed for slight variations in the true values caused by manually defining the ground truths.

Finally, the F1-score is a measurement of the accuracy of the detector and is calculated using the formula below.

$$F1 = 2 * \frac{precision * recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

As seen in the table, the F1-score for each image was relatively low. This resulted from a lot of false positives which are shown in the images in *Figure 1* by numerous green boxes.

2. Building & Testing a Detector

Training Performance

To use the Viola-Jones framework to detect no-entry signs, a new cascade was needed. 500 positive images were produced using a single image of a no-entry sign and 500 negatives were provided. Using both sets of images, the classifier was trained. The training process was separated into 3 stages, as shown in *Figure 2*. For each stage, the TPR remains at 1, however, the False Positive Rate (**FPR**) reduces dramatically.

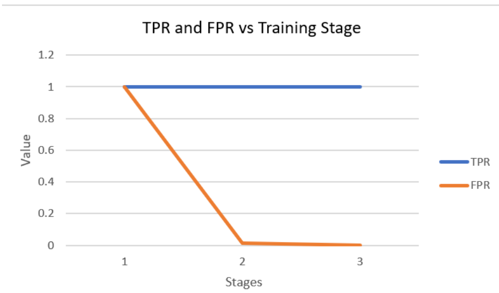


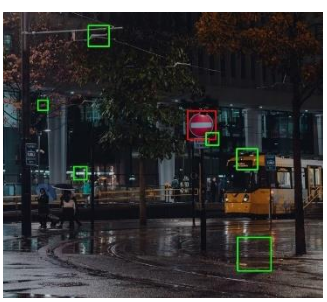
Figure 2: TPR and FPR vs Training stages

At the beginning of stage 1, all regions of the image were being accepted and so, both the TPR and FPR were 1. More classifiers were added to the cascade in the later stages, which allowed for regions to be rejected. As a result, the FPR reduced sharply. Stage 3 then shows a further reduction in the FPR for the same reason, but it was much less. This is due to there being fewer regions to look at because regions that were rejected by the previous classifiers do not get passed down.

Testing Performance

NoEntry5.bmp

NoEntry0.bmp



NoEntry13.bmp

NoEntry15.bmp

Figure 3: Viola-Jones No Entry detections (Green), Ground Truth (Red)

The images in *Figure 3* show the detected No Entry signs produced from running the Viola-Jones detector with the newly trained *cascade.xml*.

Similar to before, the TPR and the F1-Score is recorded in the table below and the IOU threshold was manually set to 0.4.

Image	TPR	F1-Score
NoEntry0.bmp	1	0.18
NoEntry1.bmp	1	0.29
NoEntry2.bmp	1	0.22
NoEntry3.bmp	1	0.36
NoEntry4.bmp	0.5	0.095
NoEntry5.bmp	0.1	0.069
NoEntry6.bmp	0	0
NoEntry7.bmp	1	0.29
NoEntry8.bmp	0.67	0.62
NoEntry9.bmp	0.5	0.29
NoEntry10.bmp	0.67	0.5
NoEntry11.bmp	0.5	0.12
NoEntry12.bmp	0.71	0.48
NoEntry13.bmp	0	0
NoEntry14.bmp	1	0.25
NoEntry15.bmp	1	0.57
Average	0.67	0.271

Table 2: Viola-Jones results

On average, the detector correctly found 0.67 of the signs in the input images. This is slightly less compared to the average TPR for frontal face detection. A reason for this could be because faces have more distinct features than no entry signs. In addition, the *frontalface.xml* may have been trained using many images of **different** faces. Whereas the *cascade.xml* was trained using sample images that were all generated from a **single** picture of a no entry sign. In reality, images can contain no-entry signs at different angles, have varying contrast/brightness levels, causing the signs to be a different colour or the signs themselves may be obstructed by something. This aspect can be seen in *Figure 3 NoEntry5.bmp* where some signs were missed by the detector because of the angle they were facing.

The results from *Table 2* show that the average F1-Score (0.271) was quite low, which signified the detector wasn't very precise. The low score results from too many false positives, as seen by the many green bounding boxes over random regions in *Figure 3*. There are a few reasons the detector may have generated many false positives. For example, *opencv_createsamples* only produced 500 positive training samples and only 500 negative sample images were used. Possibly having more of each when training could have improved the accuracy of the detector and produced higher F1-Scores.

3. Shape Detector Integration

To possibly improve the accuracy of the detector, a Hough circle transform was implemented. Doing so removed green bounding boxes which were not within a circle region, thereby reducing false positives.

Detection Pipeline

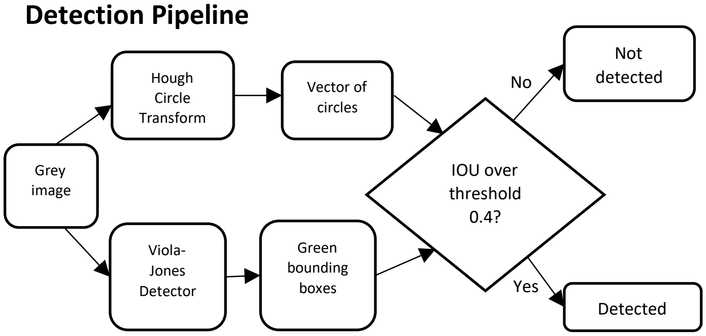


Figure 4: Flow diagram

- The program will run the Viola-Jones and Hough circle transformation separately. This will produce a vector of green bounding boxes and a vector of blue circles.
- For every circle, it computes the maximum IOU with the bounding boxes.
- If the max IOU is greater than a specified threshold, it confirms the bounding box to have detected a no entry sign.
- A threshold of 0.4 was chosen. It to be high enough so that bounding boxes that had very small intersections with circles or did not have any at all were not considered; thus reducing the false positive rate.
- It also needed to be low enough to consider the sections of the no-entry signs that were found by the Viola-Jones detector but not quite by the Hough transform.
- A Hough line transform could have also been utilised; however, it is explained in Section 4 why it was not.

Evaluation

Image	TPR	ΔTPR	F1-Score	ΔF1-Score
NoEntry0.bmp	1	0	1	+0.82
NoEntry1.bmp	1	0	1	+0.71
NoEntry2.bmp	1	0	1	+0.78
NoEntry3.bmp	1	0	0.8	+0.44
NoEntry4.bmp	0.5	0	0.67	+0.58
NoEntry5.bmp	0.1	0	0.18	+0.11
NoEntry6.bmp	0	0	0	0
NoEntry7.bmp	0	-1	0	-0.29
NoEntry8.bmp	0.67	0	0.8	+0.18
NoEntry9.bmp	0.5	0	0.67	+0.38
NoEntry10.bmp	0.67	0	0.8	+0.3
NoEntry11.bmp	0.5	0	0.67	+0.55
NoEntry12.bmp	0.57	-0.14	0.73	+0.25
NoEntry13.bmp	0	0	0	0
NoEntry14.bmp	1	0	1	+0.75
NoEntry15.bmp	0.5	-0.5	0.67	+0.1
Average	0.56	-0.11	0.62	+0.349

Table 3: Viola-Jones + Hough circle results & difference in comparison to Viola-Jones

Hough Details

The main merit of this implementation is that there are fewer false positives and so it has a higher average F1-Score.

- Average F1-Score (0.62) increases dramatically in comparison to Section 2 (0.271).
- In *Figure 5* the Hough circle transform worked well to detect the circles and so the number of green bounding boxes was greatly reduced, leaving only the ones truly detecting the signs, untouched.

Figure 5: NoEntry0.bmp

However, there were limitations to this approach, as the average TPR shows a slight decrease (-0.11).

- Some images contained distorted signs (e.g., ellipsis), or obstructed signs. Also, the contrast and or brightness of the image may not have been enough to be detected by the Hough space. Which resulted in missed circle detections.
- In *Figure 6*, the third no-entry sign on the LHS is obstructed. Therefore, the shape of the sign isn't a complete circle, which is shown in the thresholded gradient image that is then provided to the Hough circle transform function. The function only detects circles and for that reason, it could not detect that particular sign.

Figure 6: NoEntry12.bmp

- The detector only considers the bounding boxes resulting from the Viola-Jones detections. If Viola-Jones does not detect a sign in the first place, then the detector will not consider that region of the image at all. An example of this is shown in *Figure 7*.

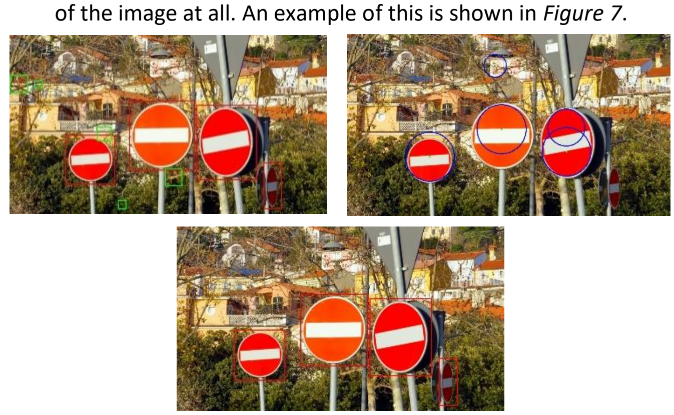


Figure 7: NoEntry6.bmp – VJ detections, Circle detections, filtered detections.

4. Improving the Detector

Idea

After looking at the previous results, the Hough circle transform was better at detecting the no-entry signs than Viola-Jones. Of the 16 images, there were 5 images in which the Hough circle could locate signs where the Viola-Jones could not. It's especially true for *Figure 8,* where none of the signs was detected by Viola-Jones. In comparison, there were only 3 images where the Hough circle could not find **some** signs that were detected by Viola-Jones.

Figure 8: NoEntry6.bmp - Viola-Jones detections. No signs detected.

There were a few approaches to improve the performance of the detector:

- Implement a Hough line transform to detect the white stripe on the signs. From the vector of lines returned, only pairs of parallel lines would be extracted. They would then be compared with the vector of circles that were left over from the IOU with the bounding boxes. New bounding boxes would be created over the circles that contain parallel lines as they would most likely be no-entry signs.
- Improve the Viola-Jones by doing additional pre-processing. Doing so could produce more green bounding boxes over signs the detector previously missed.

As the number of signs missed by Viola-Jones was greater than the number of circles missed by the Hough circle, the second method was implemented. However, combining both could lead to even better results, but it would increase the detection time.

Visualise

A median filter with kernel size 3 was used on the input images before being passed to the classifier.

Figure 9: NoEntry4.bmp - Improved Viola-Jones detections

Figure 10: NoEntry6.bmp - Improved Viola-Jones detections

Evaluation

Image	TPR	ΔTPR	F1-Score	ΔF1-Score
NoEntry0.bmp	1	0	1	+0.82
NoEntry1.bmp	1	0	1	+0.71
NoEntry2.bmp	1	0	1	+0.78
NoEntry3.bmp	1	0	0.8	+0.44
NoEntry4.bmp	1	**+0.5**	1	+0.91
NoEntry5.bmp	0.1	0	0.18	+0.11
NoEntry6.bmp	0.4	**+0.4**	0.4	+0.4
NoEntry7.bmp	0	-1	0	-0.29
NoEntry8.bmp	0.5	**-0.17**	0.67	+0.05
NoEntry9.bmp	0.5	0	0.67	+0.38
NoEntry10.bmp	0.67	0	0.8	+0.3
NoEntry11.bmp	0.5	0	0.67	+0.55
NoEntry12.bmp	0.57	-0.14	0.73	+0.25
NoEntry13.bmp	0	0	0	0
NoEntry14.bmp	1	0	1	+0.75
NoEntry15.bmp	0.5	-0.5	0.67	+0.42
Average	0.61	-0.06	0.66	+0.39

Table 4: Median filter results & difference in comparison to Viola-Jones

Overall, this model shows a better performance than any of the previous models:

- The average F1-Score increased by +0.39 compared to the vanilla Viola-Jones model and +0.04 compared to the Viola + Hough model in Section 3.
- Both the TRP and F1-Score for *NoEntry4.bmp* and *NoEntry.bmp* increased as the Viola-Jones detector could now detect the leftmost signs for both images, as shown in *Figure 9* and *Figure 10 respectively*.

There were a few shortcomings:

- The average TPR decreased slightly by -0.06 compared to the vanilla Viola-Jones model. However, this is mainly due to the limitations of the Hough circle function, as explained in Section 3.
- The number of correctly detected signs decreased from 4 to 3 for NoEntry8.bmp which is shown in *Figure 11*. Some signs overlapped in the image and so applying a median filter caused the overlapping signs to become less distinct. Therefore, it was more difficult for the Viola-Jones detector to locate those signs.

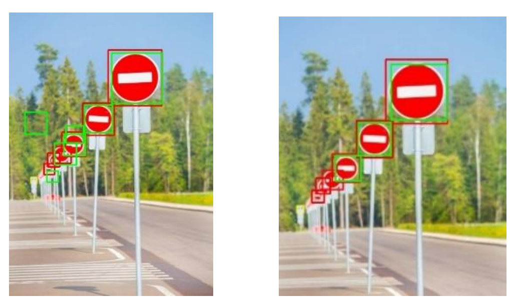


Figure 11: NoEntry8.bmp – Vanilla VJ vs Improved VJ detections

Further Improvements

As mentioned before, including a Hough line transform could improve the detector. Using this in conjunction with the Hough circle could allow for more accurate detections and reduce false negatives. Alternatively, a Hough ellipse transform could also be implemented to detect signs at different angles. However, creating these additional Hough spaces would be computationally expensive and would reduce the efficiency of the detector.