

No Entry Sign Challenge Report

1 The Viola-Jones object detector

1.1 Ground truth and visualisation

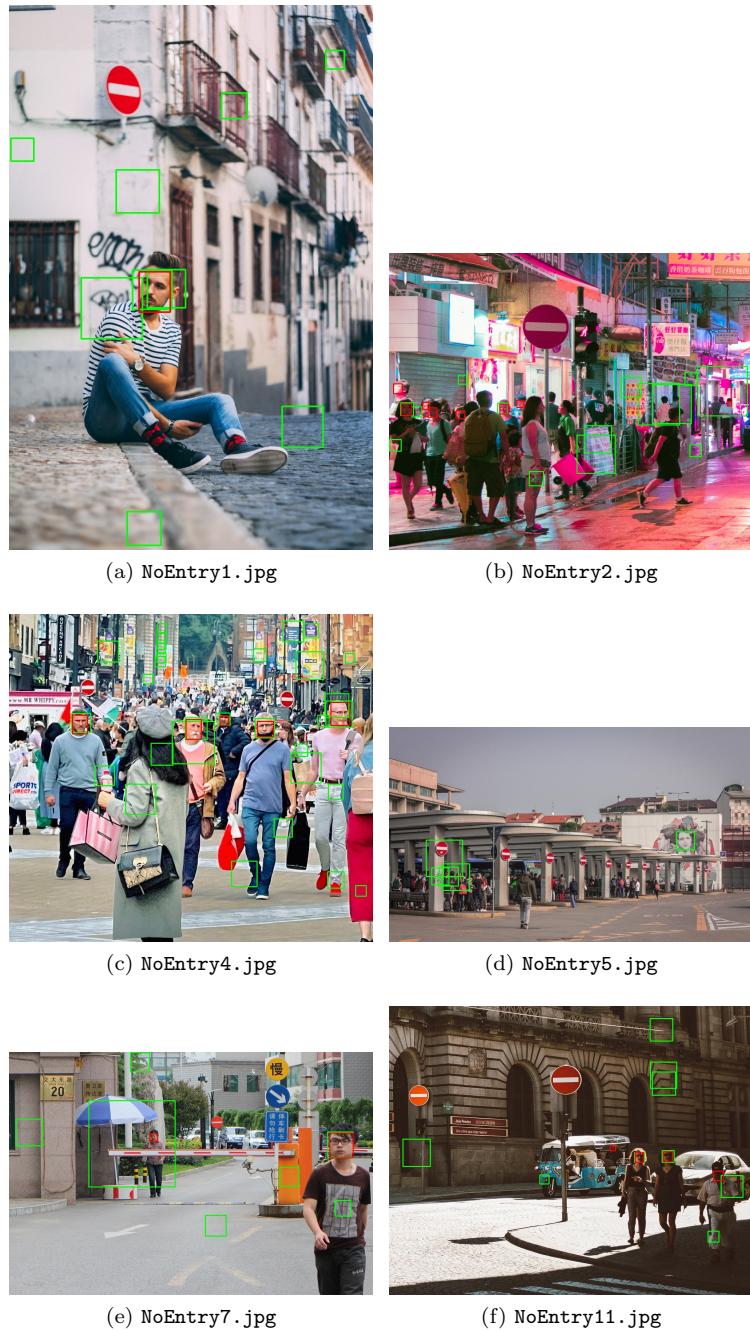


Figure 1: Six images with the ground truth bounding boxes in red and actually detected instances from `face.cpp` in green

Ground truth bounding boxes assist in determining the accuracy of a detection algorithm, and help to visualise how well it performs. Figure 1 displays six images, with their ground truth bounding boxes in red, and the frontal faces detected by `face.cpp` in green. I defined the ground truth frontal faces as those whereby the majority of the face was visible, the person was facing the camera (i.e. both eyes visible) and the person was close enough for the major facial characteristics distinguishable.

1.2 Intersection-over-union, true positive rate and F₁ score

True positive rate (TPR) is a popular metric used to assess the performance of an object detection algorithm. The TPR is the

proportion of objects that an algorithm is attempting to detect that are actually detected, and is given by the formula:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

where TP is the number of true positives, and FN is the number of false negatives.

The first practical difficulty that arises in calculating the TPR is whether or not to define a predicted bounding box as a being a true positive or a false positive. I opted to define a detected bounding box as being a true positive if it had an intersection-over-union (IOU) value greater than 0.5 with a ground truth bounding box, as this is considered a good score [1]. For two bounding boxes A and B , the IOU is calculated as follows:

$$\text{IOU}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

where $|A \cap B|$ is the area of intersection, and $|A \cup B|$ is the area of union. Moreover, for each ground truth bounding box, I only defined the detected bounding box with the largest IOU value as being a true positive if there was more than one intersecting detected bounding box.

In any detection task, TPR can be a flawed metric to use to define how well a given model performs. This is because, firstly, the TPR can be usually be artificially increased simply by lowering the IOU threshold, without the model actually performing any better in reality. Secondly, it is possible to achieve a TPR of 100% by detecting every possible pixel region as being the object being detected, despite having a large number of false positives.

As a result of this, the F₁ score is often used as a measure of a model's accuracy. F₁ score is the harmonic mean of recall (i.e. TPR) and precision (i.e. the proportion of predicted positives that are actually positive), and is calculated as follows:

$$F_1 = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})}$$

where FP is the number of false positives.

Table 1: TPRs and F₁ scores of the frontal face detector

Image	TPR	F ₁ Score
NoEntry0.jpg	Undefined	0.00
NoEntry1.jpg	1.00	0.20
NoEntry2.jpg	0.25	0.18
NoEntry3.jpg	Undefined	Undefined
NoEntry4.jpg	1.00	0.28
NoEntry5.jpg	Undefined	0.00
NoEntry6.jpg	Undefined	0.00
NoEntry7.jpg	0.50	0.22
NoEntry8.jpg	Undefined	0.00
NoEntry9.jpg	Undefined	0.00
NoEntry10.jpg	Undefined	0.00
NoEntry11.jpg	0.50	0.31
NoEntry12.jpg	0.00	0.00
NoEntry13.jpg	Undefined	0.00
NoEntry14.jpg	Undefined	0.00
NoEntry15.jpg	Undefined	0.00
All images	0.52	0.15

Table 1 displays the TPR and F₁ score that `face.cpp` achieved on each of the `NoEntry*.jpg` images. Due to the definitions of TPR and F₁ score, many of the values are undefined since division by zero yields an undefined result.

2 Building and testing my detector

2.1 Training performance

To train my own detector, I initially generated a collection of 500 artificial 30×30 images of no entry signs by applying different perspective transformations to a base image. I then used a variant of AdaBoost to train a no entry sign detector with these 500 samples.

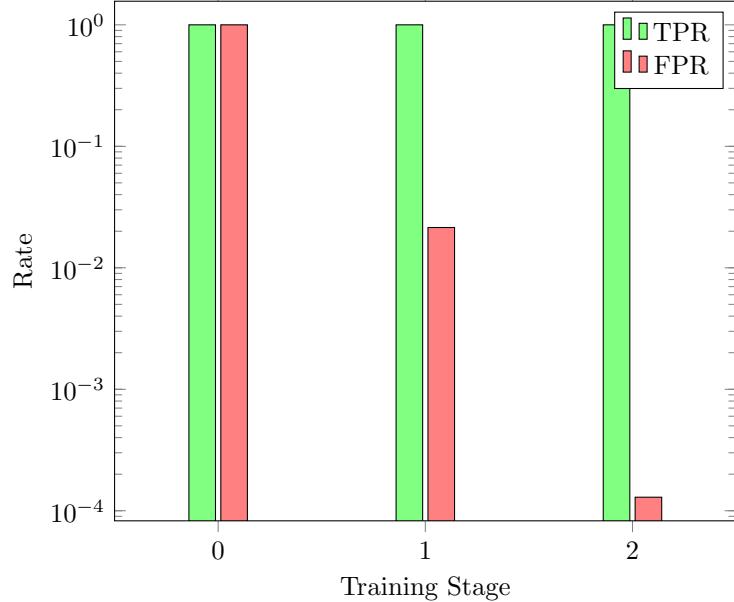


Figure 2: TPR and FPR of stages in the Viola-Jones no entry sign training process

During the training process, the false positive rate (FPR) is the number of negatives incorrectly predicted as being a positive no entry sign. Figure 2 displays the TPR and FPR at each stage of the training process. At each successive stage, the TPR remains at 100%, while the FPR decreases drastically. This is because a cascade of more complex classifiers are produced at each successive stage. Therefore, the early stages involve only the less complex (but faster) classifiers and therefore have a larger FPR.

2.2 Testing performance

Figure 3 contains three images with ground truth bounding boxes in red, and detected bounding boxes from the Viola-Jones implementation in green. I defined the ground truth no entry signs as those whereby more than 50% of the sign was visible. The implementation works relatively well, achieving a TPR of 0.44 and an F_1 score of 0.48, as shown in Table 2. However, the implementation has its shortcomings. When tested, the Viola-Jones implementation detected a large number of false positives. Upon inspection, this occurs in regions containing a slightly lighter strip horizontally between two darker strips. For example, the turn left sign in Figure 3c has been incorrectly detected in this manner.

The TPR on the test data is significantly lower than on the training data. One reason for this is likely due to the differing appearances of no entry signs in the test data. The Viola-Jones implementation was trained with a collection of artificial positives. This approach works best for completely rigid objects, which, due to the slightly differing appearances of the no entry signs between countries, the positives in our test images are not [2]. Additionally, we are only testing our implementation on 16 images; to get a more accurate idea of the TPR and F_1 score, we would ideally test our implementation on a larger test set.



Figure 3: Three images with the ground truth bounding boxes in red and actually detected instances from the no entry sign detector

Table 2: TPRs and F_1 scores of Viola-Jones no entry sign detector

Image	TPR	F_1 Score
NoEntry0.jpg	1.00	0.67
NoEntry1.jpg	1.00	0.40
NoEntry2.jpg	1.00	0.40
NoEntry3.jpg	1.00	0.80
NoEntry4.jpg	0.50	0.67
NoEntry5.jpg	0.40	0.47
NoEntry6.jpg	0.00	0.00
NoEntry7.jpg	0.00	0.00
NoEntry8.jpg	0.57	0.72
NoEntry9.jpg	0.00	0.00
NoEntry10.jpg	0.67	0.67
NoEntry11.jpg	0.50	0.25
NoEntry12.jpg	0.25	0.40
NoEntry13.jpg	0.00	0.00
NoEntry14.jpg	1.00	1.00
NoEntry15.jpg	0.50	0.67
All images	0.44	0.48

3 Integration with shape detectors

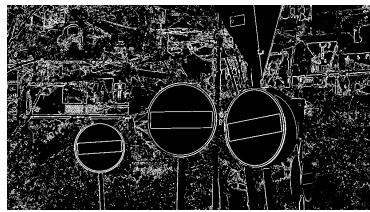
3.1 Hough Details



(a) NoEntry6.jpg with detected bounding boxes



(b) NoEntry2.jpg with detected bounding boxes



(c) NoEntry6.jpg gradient magnitude threshold



(d) NoEntry2.jpg gradient magnitude threshold



Figure 4: Two no entry sign images at various stages of the integrated implementation

3.2 Evaluation

Table 3 displays the TPR and F₁ score of the integrated implementation, as well as the differences compared with the Viola-Jones implementation. The key merits and shortcomings include:

- A significantly larger F₁ score, primarily as a result of increasing the precision; achieved by only defining an object detected by the Viola-Jones detector as positive if it has an IOU value with a bounding box from the circle Hough transform greater than 50%
- A small improvement to the TPR score, by reducing the minimum number of neighbours allowing for more true positives to be identified (the extra false positives were ‘filtered out’ by utilising the circle Hough transform)
- If a no entry sign is positively detected by one of the detectors, but not the other, it is not positively detected by the integrated implementation
- A significantly slower implementation, since the circle Hough transform takes longer to calculate

3.3 Detection pipeline

Figure 5 outlines the way I combined evidence in my algorithm. The rationale behind this was as follows:

- The Viola-Jones detector detected 44% of no entry signs, but had a low level of precision
- Upon inspection of the detected bounding boxes, it became clear the Viola-Jones detector appeared to be detecting regions with a light bar horizontally between two dark bars
- Since no entry signs are circles, we could choose to detect the circles from the circle Hough transform that have an IOU with a Viola-Jones bounding box of greater than 50% (i.e. they are likely detecting the same object)

It is important to note that if multiple circles had an IOU greater than 50% with a Viola-Jones detected bounding box, only the circle with the largest IOU was positively identified to avoid duplicate identifications of the same no entry sign.

Table 3: TPRs and F₁ scores of the integrated implementation, and the differences compared to the Viola-Jones implementation

Image	Result		Difference	
	TPR	F ₁ Score	TPR	F ₁ Score
NoEntry0.jpg	1.00	1.00	±0.00	+0.33
NoEntry1.jpg	1.00	1.00	±0.00	+0.60
NoEntry2.jpg	1.00	1.00	±0.00	+0.60
NoEntry3.jpg	1.00	1.00	±0.00	+0.20
NoEntry4.jpg	1.00	1.00	+0.50	+0.33
NoEntry5.jpg	0.20	0.31	-0.20	-0.16
NoEntry6.jpg	0.00	0.00	±0.00	±0.00
NoEntry7.jpg	0.00	0.00	±0.00	±0.00
NoEntry8.jpg	0.43	0.60	-0.14	-0.13
NoEntry9.jpg	1.00	1.00	+1.00	+1.00
NoEntry10.jpg	1.00	1.00	+0.33	+0.33
NoEntry11.jpg	0.50	0.67	±0.00	+0.42
NoEntry12.jpg	0.38	0.55	+0.13	+0.16
NoEntry13.jpg	0.00	0.00	±0.00	±0.00
NoEntry14.jpg	1.00	1.00	±0.00	±0.00
NoEntry15.jpg	1.00	1.00	+0.50	+0.33
All images	0.50	0.66	+0.04	+0.17

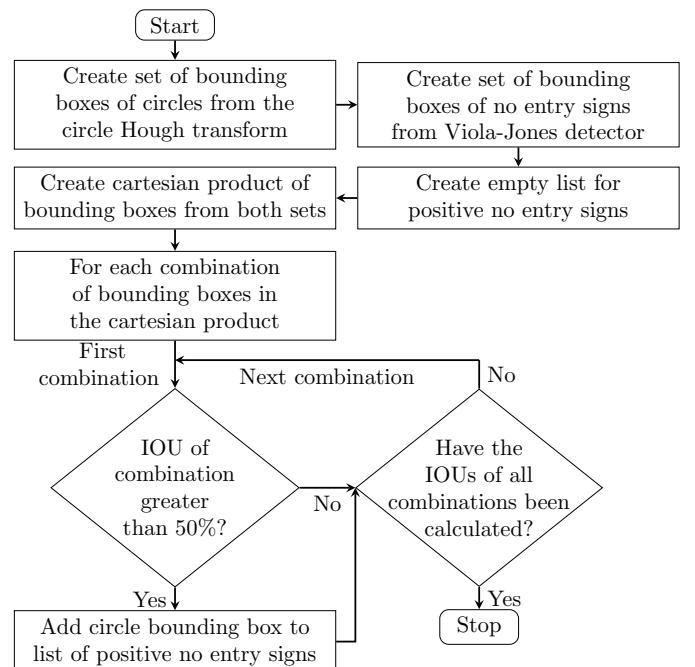


Figure 5: Flow chart detailing my algorithm that integrates Viola-Jones with the circle Hough transform

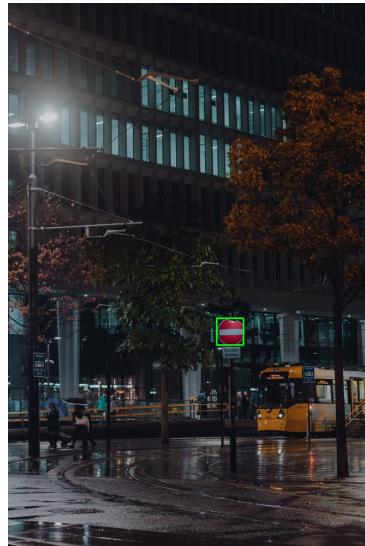
4 Improving my detector

4.1 Idea

My improved approach puts every circle detected by the circle Hough transform but not by the Viola-Jones detector through an extra stage of processing. The extra stage of processing begins by converting the colour space of the pixels within the circle from RGB to CIELAB. K -means is then performed with $k = 2$ over the green-red axis a^* and blue-yellow axis b^* . The Euclidean distance is then calculated from the cluster centres to pure red and pure white, to identify the two main colours in the circle. If the distances are less than a threshold (i.e. the two main colours are red and white), the line Hough transform is conducted on the circle to identify horizontal parallel lines. If horizontal parallel lines are detected, the circle is positively identified as a no entry sign. The rationale behind this was as follows:

- The pixels are transformed from RGB to CIELAB so the L^* axis (i.e. the perceptual brightness axis) can be ignored when doing k -means, so that no entry signs under a variety of lighting conditions can be detected
- Since CIELAB is perceptually uniform, by calculating the Euclidean distance, the algorithm can confirm whether the primary colours in the circle are red and white
- By performing the line Hough transform, we can filter out all circles that contain red and white but are not no entry signs

4.2 Visualise



(a) NoEntry6.jpg with detected bounding boxes (b) NoEntry13.jpg with detected bounding boxes

Figure 6: Two no entry sign images with detected bounding boxes from my improved detector

Figure 6 exhibits the merit of my improved approach. All three detected bounding boxes were not identified by the previous detector.

The two bounding boxes in Figure 6a very clearly contain no entry signs. They are detected by the circle Hough transform but not the Viola-Jones detector. Therefore, my improved algorithm puts them through the extra stage of processing. A merit of using k -means is very evident here: it is able to detect the two main colours in the circles by finding the values of the cluster centres. A merit of using CIELAB is also very clear: since CIELAB is perceptually uniform, my improved detector is able to identify the two main colours as red and white by calculating the Euclidean distance. A merit of using the line Hough transform is that my algorithm is able to confirm that the circles are no

entry signs by detecting two horizontal parallel lines. By using the line Hough transform we are able to eliminate false positives. Without it, my algorithm could identify any circle containing red and white as a no entry sign.

The no entry sign in Figure 6b is clearly a no entry sign, but potentially due to the very low lighting it is not identified by the Viola-Jones detector. However, it is detected by the circle Hough transform and is therefore put through the extra stage of processing. Another merit of converting the pixels from RGB to CIELAB is evident here: by focusing purely on the a^* and b^* axes, no entry sign is able to be identified, despite the poor lighting conditions.

4.3 Evaluate

Table 4: TPRs and F_1 scores of the improved implementation and the differences compared to the previous implementation

Image	Result		Difference	
	TPR	F_1 Score	TPR	F_1 Score
NoEntry0.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry1.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry2.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry3.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry4.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry5.jpg	0.20	0.31	± 0.00	± 0.00
NoEntry6.jpg	0.50	0.67	+0.50	+0.67
NoEntry7.jpg	0.00	0.00	± 0.00	± 0.00
NoEntry8.jpg	0.43	0.60	± 0.00	± 0.00
NoEntry9.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry10.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry11.jpg	1.00	1.00	+0.50	+0.33
NoEntry12.jpg	0.38	0.55	± 0.00	± 0.00
NoEntry13.jpg	1.00	1.00	+1.00	+1.00
NoEntry14.jpg	1.00	1.00	± 0.00	± 0.00
NoEntry15.jpg	1.00	1.00	± 0.00	± 0.00
All images	0.58	0.73	+0.08	+0.07

Table 4 contains the TPR and F_1 scores of the improved detector. By putting detected circles that were not detected by the Viola-Jones detector through an extra stage of processing, the improved implementation was able to detect 8% more no entry signs, and achieved a 7% higher F_1 score. The key merits and shortcomings of the improved implementation include:

- No entry signs that are facing the camera (i.e. circles) are identified almost always by the circle Hough transform, and by putting them through the extra stage of processing they are almost always correctly identified by looking for key characteristics
- No entry signs that are not detected by the circle Hough transform cannot be detected no matter what; this includes no entry signs that are at an angle (and are therefore an ellipse) and no entry signs with a significant amount of occlusion

4.4 Further Extensions

To further improve my detector, I could do the following:

- Using the ellipse Hough transform, no entry signs not directly facing the camera could be identified
- I could train the Viola-Jones detector on real no entry signs rather than artificial positives; this would likely improve the TPR

References

- [1] Adrian Rosebrock. *Intersection over Union (IoU) for object detection.* URL: <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>.
- [2] *Cascade Classifier Training.* URL: https://docs.opencv.org/3.4/dc/d88/tutorial_traincascade.html.