The Object Detection Challenge

COMS30121 Image Processing and Computer Vision

Carlo Alessi & Luiza Stepanyan

Subtask 1: The Viola-Jones Object Detector

Object recognition is one of the most important yet challenging areas in computer vision. To understand how the Viola-Jones face detector works and the results it produces, we used a readily-available version, containing a strong classifier trained to detect frontal-view faces to detect faces in a series of photographs. Figure 1 below shows the results, with green bounding boxes drawn around the detected faces.

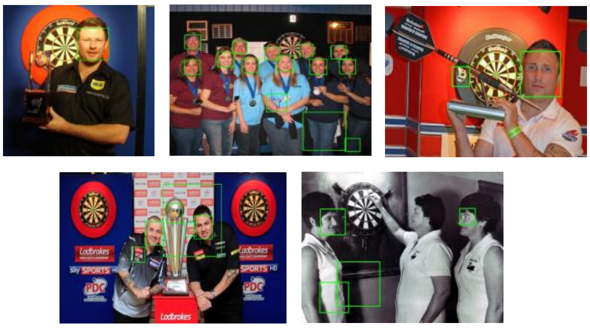

Figure 1. Results of the Viola-Jones face detector on test images dart4.jpg, dart5.jpg, dart13.jpg, dart14.jpg, and dart15.jpg with bounding boxes around detected faces.

For two of the above images, we calculated the true positive rate (TPR) to determine the percentage of successfully detected faces out of all valid faces in the image using the formula:

$$\text{TPR} = \frac{successfully\ detected\ faces}{valid\ faces}$$

For dart5.jpg, all valid faces were detected, so the TPR was 100%. For dart15.jpg, two out of three valid faces were detected, resulting in a TPR of $66.\dot{6}\%$.

There were no difficulties in determining the TPR for dart5.jpg as all the bounding boxes were clearly drawn around whole faces. It was more challenging for dart15.jpg as only the top half of one of the faces was detected and defined by the bounding box, which made it hard to say whether or not the full face was detected. Also, the fully undetected face was side-facing, yet the classifier used was only trained to detect frontal-view faces. This meant it was difficult to determine whether it should be considered a valid face, but as the other detected faces were similarly orientated, we decided it should be.

A further limitation of using the TPR as a measure of detector effectiveness is that it is always possible to achieve a TPR rate of 100%. This is because the TPR calculation does not consider false positives: There could be many bounding boxes covering the entire image and the TPR would be 100% because all valid faces or objects would be detected, along with every other object in the image. Dart5.jpg illustrates this well as although it represents a 100% TPR rate, it also shows three false positive detections that should be considered when determining the effectiveness of the detector.

Instead, the F1 score, which does consider false positives, can be calculated using the formula:

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

Where TP refers to the number of true positives (correctly detected target objects), FP to the number of false positives (erroneously detected target objects, and FN to the number of false negatives (erroneously rejected target objects).

To accurately and meaningfully calculate this for the current detector as well as subsequent detectors, we created files for each of the images that contained the number of objects to be detected, the x and y coordinates for their top left corner, and their width and height. This would serve as the ground truth for each of the images. We then loaded this data in the main program, computed the areas of intersection and union between all detectable and detected objects, and calculated the ratio between them. From this we determined the number of true and false positives and false negatives, which we then inserted into the F1 formula to calculate the F1 scores for the images.

Dart5.jpg was used to test this, resulting in an F1 score of 0.88, which provides a more accurate representation of the face detector's accuracy compared to the TPR score alone.

Subtask 2: Building & Testing your own Detector

Instead of detecting faces, the aim of this report was to develop a detector that would identify dartboards. To do so using the Viola-Jones detector, it was necessary to train a new classifier to detect dartboards instead of faces.

A dartboard image was used to generate a set of 1000 positive training images of dartboards at different angles and contrasts. This, along with a set of negative images containing no dartboards, was used to create the new dartboard classifier in three stages. After each stage, the achieved TPR and FPR (false positive rate) was displayed. Figure 2 shows how the true positive rate (TPR) and false positive rate (FPR) varied with each training stage.

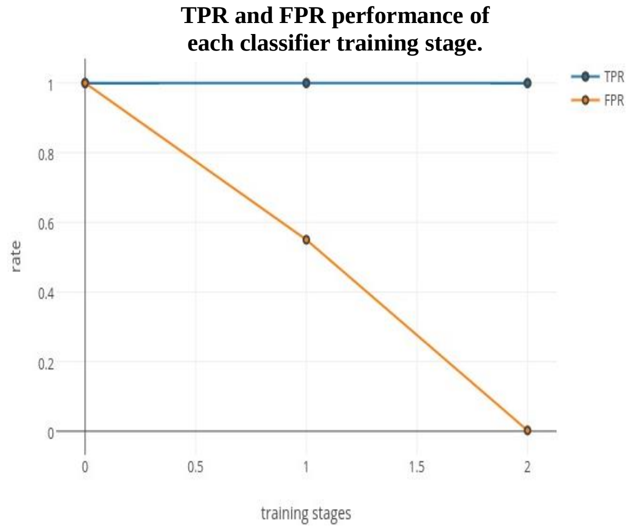

Figure 2. Graph showing the TPR and FPR performance of the classifier with each training stage.

It shows that a TPR of 1 was achieved for each training stage, which suggests that the classifier was able to correctly detect all the target objects from the beginning and that this ability did not decrease with subsequent training stages. It also shows that the FPR decreased with each training stage, which suggests that the classifier became increasingly better at correctly identifying when something was not a dartboard.

To investigate the detector's effectiveness, the classifier was used with the Viola-Jones detector to detect dartboards in 16 test images. The results of the detector are shown in Figure 3 where green bounding boxes have been drawn around detected dartboards.

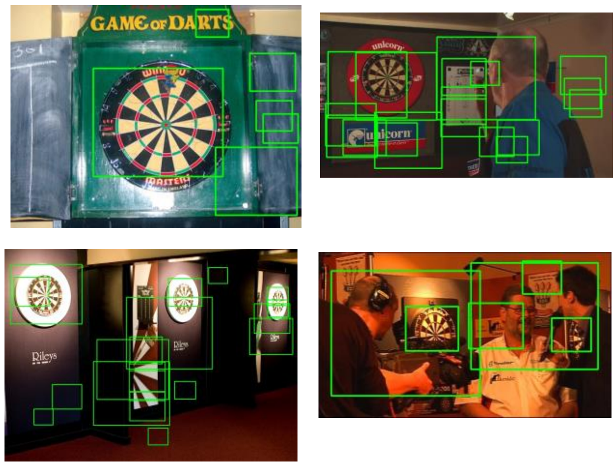

Figure 3. Results of Viola-Jones detector on detecting dartboards for test images dart1.jpg, dart2.jpg, dart10.jpg, and dart11.jpg

From these images, it is clear to see that although the dartboards have been detected, many bounding boxes have been drawn around regions where there are no dartboards, and therefore represent false positive detections. In several places, multiple bounding boxes have been drawn around single dartboards, suggesting either multiple or fragmented detections of single dartboards.

To quantify the detector's effectiveness, the overall F1 score for all 16 test images was calculated at 0.22. Since a higher F1 score represents a more accurate detector with more true positives and less false positives, a score of 0.22 suggests a relatively poor-performing detector with either few true positives, many false positives, or a combination of both. This contrasts with the results found during classifier training and displayed in Figure 2, which suggests that the detector should have a perfect TPR of 1 and an almost perfect FPR of nearly zero. In practice, this is not the case, which suggests that the plotted graph is not useful in predicting the performance of the system on test data.

Subtask 3: Integration with Shape Detectors

To improve the dartboard detector, we integrated it with a Hough transform. The Hough transform was developed as a way to detect lines and simple shapes in images by a system of voting for the most likely candidates based on a set of preassigned parameters. Since dartboards essentially consist of concentric circles and intersecting lines, using a Hough transform to detect lines and circles was appropriate. Figure 4 shows two examples of the process, from Hough transform to final detection.

Figure 4. The threshold gradient magnitude images input to the Hough transform (top), a 2D representation of the Hough space for circle detection (middle), and the final images showing final detections with bounding boxes (bottom) for dart1.jpg (left) and dart11.jpg (right).

The overall F1 score for this detector's performance was calculated at 0.67, an improvement of 0.45 compared to the Viola-Jones detector alone. This suggests that using the Hough transform to detect circles and intersecting lines was effective in improving dartboard detection. This is clear in the differences seen in dart1.jpg where the false positives detected in Figure 3 have cleared to leave just the single detected true positive dartboard in Figure 4, which suggests that the new detector is more effective at distinguishing between dartboards and non-dartboards. Similarly for dart11.jpg, the false positives detected in Figure 3 have cleared to leave the true positive dartboard detected in Figure 4. However, a false negative occurred where the new detector failed to detect the partial dartboard in dart11.jpg that was detected in the original Viola-Jones detection. This was seen in other images, too, where partially-obscured dartboards that were detected in the Viola-Jones implementation were not detected in the new one, or they were detected in fragments with many bounding boxes covering a single dartboard. This suggests that a good Hough transform cannot necessarily improve a poor Viola-Jones detection, yet it is possible for a poor Hough transform to negatively impact a good Viola-Jones detection.

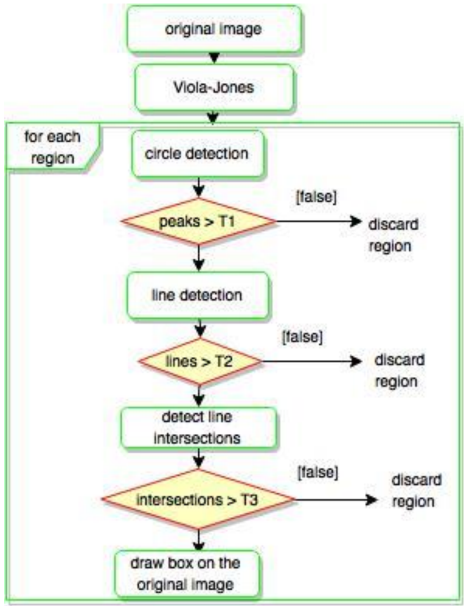

Figure 5. Flow diagram of how the evidence from the Viola-Jones detector was combined with the Hough transform. Only the regions detected by the Viola-Jones implementation (represented by bounding boxes in the displayed images above) were considered for the Hough transform as these were the areas most likely to contain dartboards; the main aim was to reduce the prevalence of false positives.

Circles and intersecting lines were chosen to be detected for as they make up some of the key features of dartboards. First, circles were detected for as they would reduce the number of regions in which to detect intersecting lines. Then, intersecting lines were detected for to distinguish dartboards from circles without intersecting lines, such as rounded letters. Any region which had votes for a circle and a number of intersecting lines above a certain threshold was regarded as containing a dartboard and a bounding box was drawn around it. Any region that did not meet the threshold target at any stage of the implementation was discarded at that stage to reduce the computational demands of the later stages and of the system as a whole.

Subtask 4: Improving your Detector

To further improve our dartboard detector and hopefully address some of its limitations, we moved away from shape detection and towards another defining feature of dartboards: colour.

The inner regions of standard dartboards are made up of a pattern of black, white, red, and green shapes, with the most distinct part being the alternating black and white sections. It seemed like a clear next step to detect for these and methods such as histogram backprojection seemed like they would enable us to do so by detecting pixels of specific hues, for example the yellow hues that the whites of dartboards tended to vary in. However, under different lighting conditions, where the whites may appear more blue-toned or in greyscale images where there were no yellow tones at all, this technique would have been ineffective.

Instead, we chose to focus on the contrast between the black and white in the inner dartboard regions. Our process involved using the greyscale versions of the test images, for ease and simplicity, and cropping the previously detected regions to eliminate background noise and focus solely on dartboard inner regions. We then computed a histogram for the cropped region and applied a threshold to detect for the black and white regions. A counter was incremented for each black or white pixel detected and the difference between the two groups calculated. If this difference was

Figure 6. Dart2.jpg and dart10.jpg after subsequent detections, from the Viola-Jones detection at the top to the final, Subtask 4 detection at the bottom.

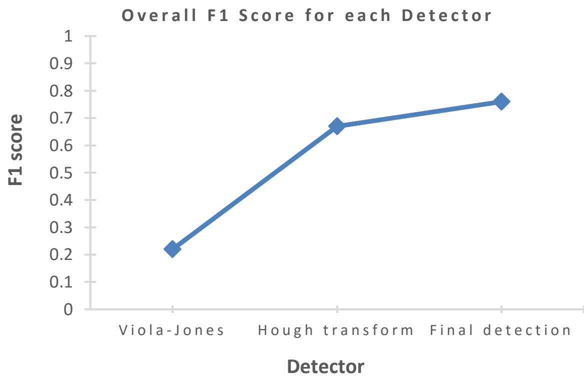

Figure 7. Graph to show improvement in F1 score with each detector

less than a specified threshold, and therefore indicative of a high-contrast black and white region, a bounding box was drawn to represent a detected dartboard.

Figure 6 shows two examples of the final image detection beneath the results of the Viola-Jones detection and the Combined Viola-Jones and Hough transform detection, and Figure 7 shows the change in F1 score between the different detectors. From this it is clear that the final detector was more effective at detecting dartboards than the Viola-Jones or the combined detectors. This is illustrated in the example images: In both dart2.jpg and dart10.jpg, the multiple detections are gone as only one bounding box surrounds each detected dartboard, thereby addressing one of the limitations identified with the previous detectors. The number of false positives has also been reduced, which addresses another previously identified problem. However, the F1 score is not perfect, which suggests that the detector is still prone to error. For example, some images still display false positive detections and the problem of not detecting partially obscured dartboards, or of detecting them with many small bounding boxes, still prevails. In order to address the latter issue, the classifier could have been retrained with images of obscured dartboards. This might have resulted in a more accurate Viola-Jones detector, which would then have had a positive effect on the Hough transform and the final implementation.