Introduction to Computer Vision

Alvin Lin

August 2018 - December 2018

Face Detection

The basic idea of face detection is to slide a window across the image and try to evaluate a face model at every location. This technique must evaluate tens of thousands of location/scale combinations. Since faces are relatively rare in the space of an image, we should try to spend as little time as possible on non-face windows for computational efficiency.

A megapixel image has 10^6 pixels and a comparable number of candidate face locations. To avoid having a false positive in every image, our false positive rate has to be less than 10^{-6}.

Viola/Jones Face Detector

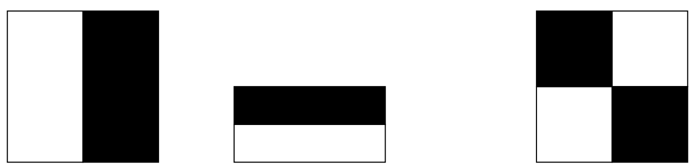
The Viola/Jones face detector is a seminal approach to real time object detection. While training is slow, detection is very fast. It involves several key ideas:

- **Integral images** for fast feature evaluation
- **Boosting** for feature selection
- **Attentional cascades** for fast rejection of non-face windows

Integral images

An integral image contains, at each pixel (x, y), the sum of all pixel values above and to the left of (x, y). This can be precomputed and used to quickly compute the the sum of the pixel values within a rectangle. Suppose A, B, C, D are the values of the integral image at the corners of an image.

The sum of the pixel values can be computed as $A - B - C + D$. Only 3 addition operations are required for any size of rectangle. This allows us to quickly compute image features using rectangle filters. These rectangle filters usually consist of a dark regions and a light regions to match corresponding gradient changes in the target image.



Feature Selection

For a 24×24 detection region, the number of possible rectangle features as described above is ~ 160000. At test time, it is impractical to evaluate the entire feature set, so we need to create a good classifier using only a small subset of all such possible features.

Boosting

Boosting is a classification scheme that works by combining weak learners into a more accurate ensemble classifier. A weak learner only needs to perform better than random classifications. Training consists of multiple boosting rounds where a weak learner that does well on examples compared to previous learners is selected.

Initially, each training example is weighted equally, but during each boosting round we find the weak learner with the lowest weighted training error. We raise the weights of the training examples misclassified by the current weak learners. The final classifier is computed as a linear combination of all weak learners where the weight of each learner is directly proportional to its accuracy. Exact formulas for re-weighting and combining weak learners depend on the particular boosting scheme.

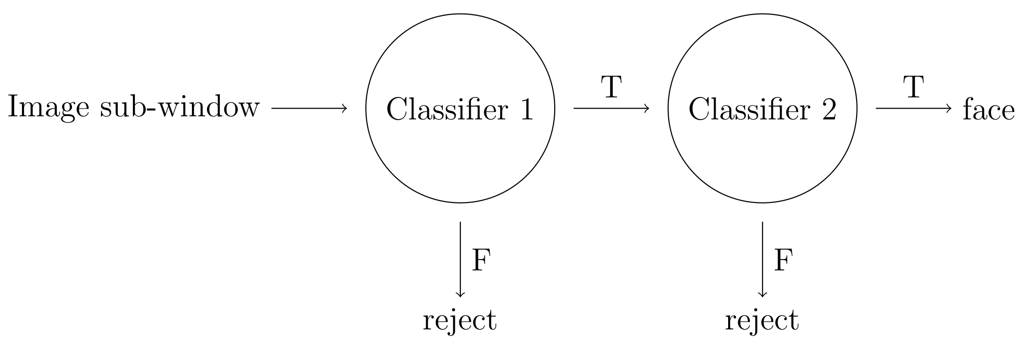
Boosting integrates classification with feature selection and is empirically not known to overfit. The complexity of training is linear in the number of training examples. Boosting is also flexible in the choice of weak learners and boosting schemes, fast, and easy to implement. However, it needs many training examples and often doesn't work as well as a support vector machine (especially for multi-class problems).

Boosting for Face Detection

We define weak learners based on rectangular features and for each round of boosting, we evaluate each rectangle filter on each example and select the best filter/threshold combination to reweight the examples. This has a complexity of $O(mnk)$ with m rounds, n examples, and k features.

Attentional Cascade

We start with simple classifiers which reject many of the negative sub-windows while detective almost all positive sub-windows. A positive response from the first classifier triggers the evaluation of a second (more complex) classifier, and so on. A negative outcome at any point leads to the immediate rejection of the sub-window.



These chain classifiers get progressively more complex and have lower false positive rates. The detection rate and the false positive rate of the cascade are found by multiplying the respective rates of the indvidual rates of the individual stages. A detection rate of 0.9 and a false positive rate on the order of 10^{-6} can be achieved by a 10-stage cascade if each stage has a detecton rate of 0.99 and a false positive rate of about 0.3 since $0.99^{1}0 \approx 0.9$ and $0.3^{1}0 \approx 6 \times 10^{-6}$.

Training the Cascade

- Set the target detection and false positive rates for each stage.
- Keep adding features to the current stage until its target rates have been met. We may need to lower the AdaBoost threshold to maximize detection. This should be tested on a validation set.
- If the overall false positive rate is not low enough, then add another stage.
- Use false positives from the current stage as the negative training examples for the next stage.

You can find all my notes at `http://omgimanerd.tech/notes`. If you have any questions, comments, or concerns, please contact me at alvin@omgimanerd.tech