Rapid Object Detection using a Boosted Cascade of Simple Features

Motivation for the project:

- Object detection, such as face detection, is a growing field where speed and accuracy are key goals
- Approach was the first to achieve real-time face detection
- Detector can run at 15 frames per second on on a conventional 700 MHz Intel Pentium III
- Detection rates are competitive with some of the best methods to date
- A key feature that distinguishes this approach from previous ones is that it doesn't need auxiliary information like pixel colour, and image differences in video sequences to achieve high frame rates.
- Instead, it does so by working with only a single grey scale image. This auxiliary information can be incorporated to achieve even higher frame rates.

Key notions of the Viola-Jones Algorithm:

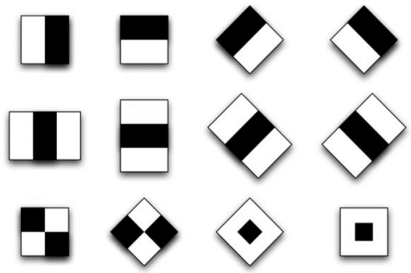
- Haar-like Features - simple rectangular features
- Integral Image - Calculating the Integral Image summed area table necessary for quick calculation by less number of references
- AdaBoost Learning Algorithm - creates a small set of only the best features to create more efficient classifiers
- Cascade Filter - discards negative windows early to focus more computational time on possible positive windows

Work Distribution:

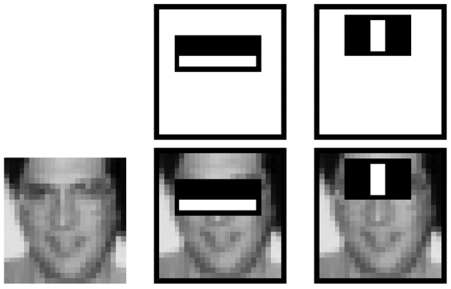
- Chunduru Sai Pranav: AdaBoost Implementation, Cascade Classifier, Documentation
- Supreeth Reddy Gangam: AdaBoost Implementation, Cascade Classifier, Results
- Pusapati Romaharshan: ViolaJones Algorithm(including haar-like and integral image), Weak Classifier Implementation, Results
- Pothugunta Venkat: ViolaJones Algorithm(including haar-like and integral image), Weak Classifier Implementation, Documentation

Haar-like Features:

- Simple features that provide just above random accuracy
- Each feature is a difference calculation of the white area minus the dark area
    - Numerous Haar-like features available to choose from as shown below.



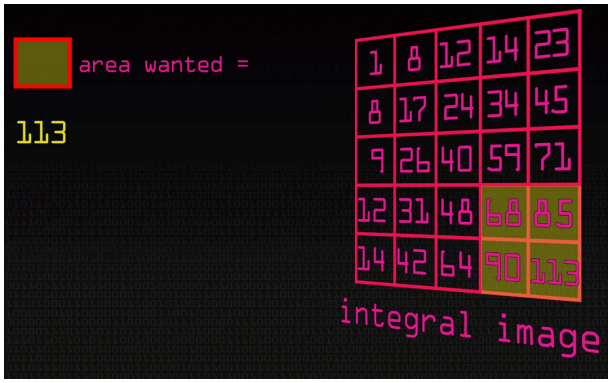
- The two features are shown in the top row and then overlaid on a typical training face in the bottom row.
- The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks.
- The feature capitalises on the observation that the eye region is often darker than the cheeks.
- The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.



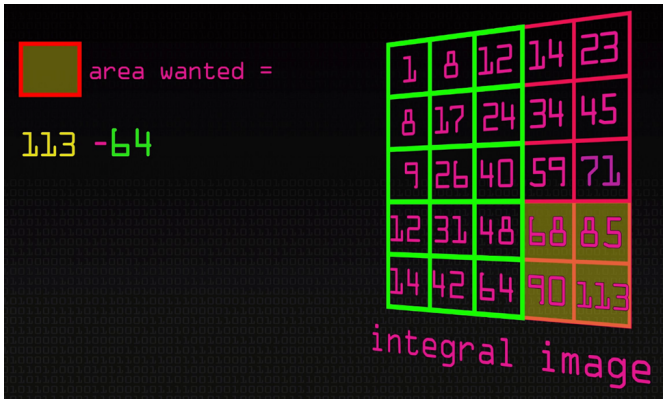
Integral Image:

•Summed area table in which every block is the summation of the previous blocks.
•Point of origin is the upper left corner, meaning previous blocks are defined as those above and to the left.

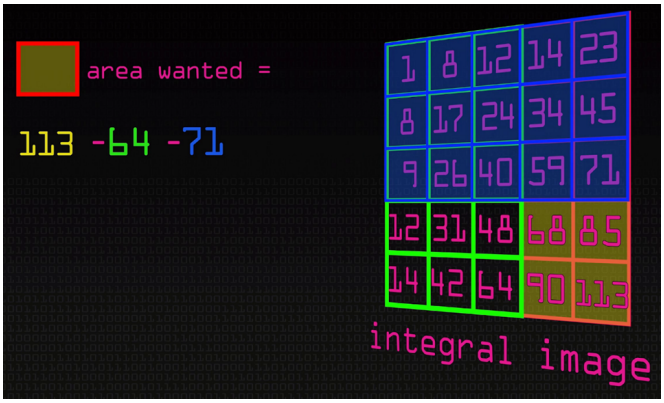
Example: Suppose we want to calculate the value of the highlighted area shown below.



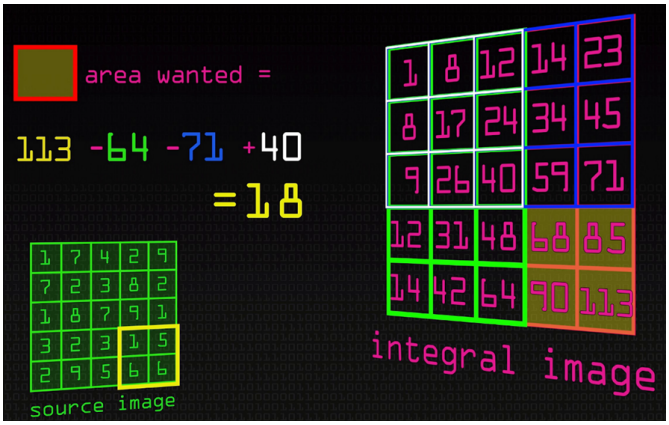
Step 1:



Step 2:



Finally,



As haar-like features are rectangular, we can evaluate them in constant time using the integral image concept as follows:

- Two-rectangle (Edge) Features - 6 memory lookup
- Three-rectangle (Line) Features - 8 memory lookups
- Four-rectangle Features - 9 memory lookups

The integral image we obtained as part of the results:

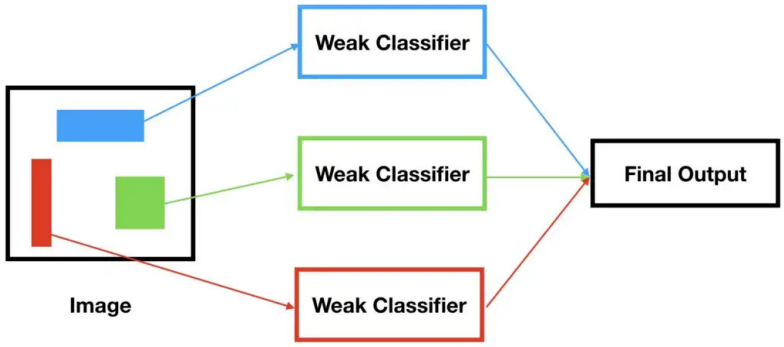
AdaBoost Learning Algorithm:

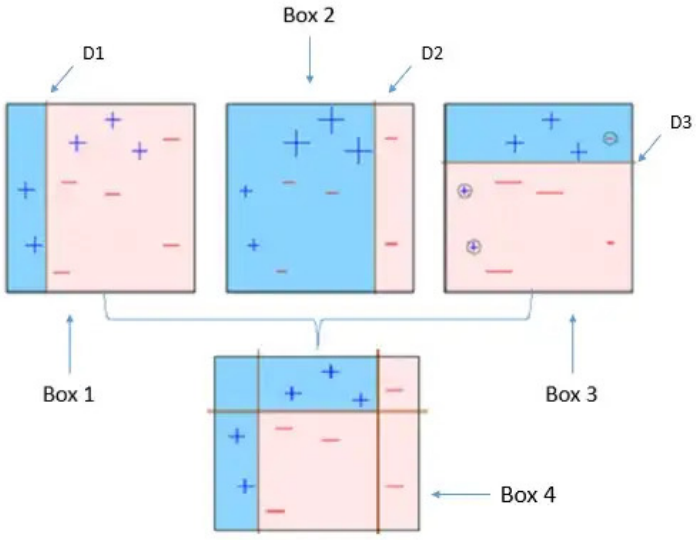
- Learning algorithm used to train the classifier and select the best subset of features
- Aggressive approach that will disregard the majority of features
- The weak learning algorithm is designed to select the single rectangle feature which best separates the positive and negative examples.

- The weak classifier is as follows:

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases}$$

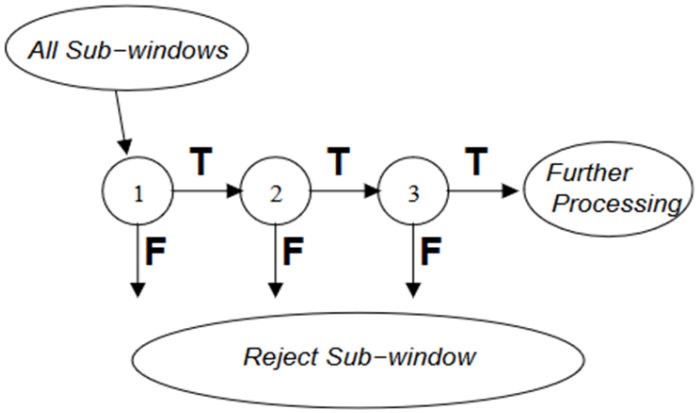
- Each weak classifier h_j consists of:
  - Feature f_j
  - Threshold Θ_j
  - Polarity p_j





Cascade Filter:

- Strong features are formed into a binary tree
- Positive matches are sent along to the next feature
- Negative matches are rejected and exit evaluations
- Reduces the amount of computation time spent on false windows
- Threshold values may be adjusted to tune accuracy
- Lower thresholds yield higher detection rates and more false positives



Dataset Used:

The data is described at http://cbcl.mit.edu/software-datasets/FaceData2.html, and I downloaded from www.ai.mit.edu/courses/6.899/lectures/faces.tar.gz and compiled into pickle files.

Each image is 19x19 and greyscale.
Training Data(training.pkl): 2429 face images, 4548 non-face images.
Testing Data(test.pkl): 472 faces, 23573 non-face images.

Training:

We trained the viola jones classifier with 10, 50, 200 features each and saved them in 10.pkl, 50.pkl, 200.pkl files respectively. We then trained attentional cascades with different layers and saved them in cascade.pkl, cascade1.pkl, cascade2.pkl files.

```
- cascade.pkl
  - An Attentional Cascade of classifiers looking at 1 feature, 2 features, 5 features, 10 features, and 50 features.
```

```
- cascade1.pkl
  - An Attentional Cascade of classifiers looking at 1 feature, 2
features, and 10 features.
- cascade2.pkl
  - An Attentional Cascade of classifiers looking at 1 feature, 10
features, 50 features, and 100 features.
```

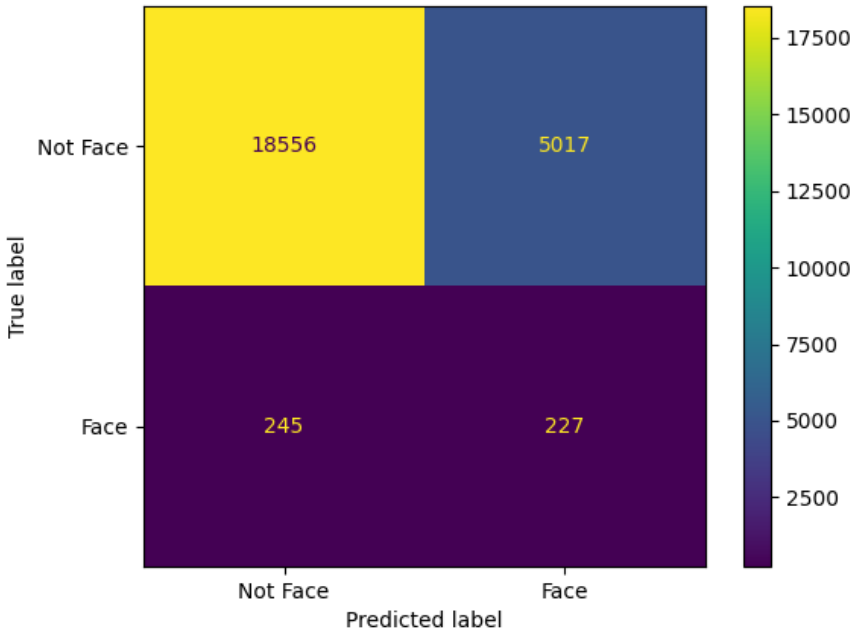
Testing:

We tested our classifiers with test.pkl

Our Results:

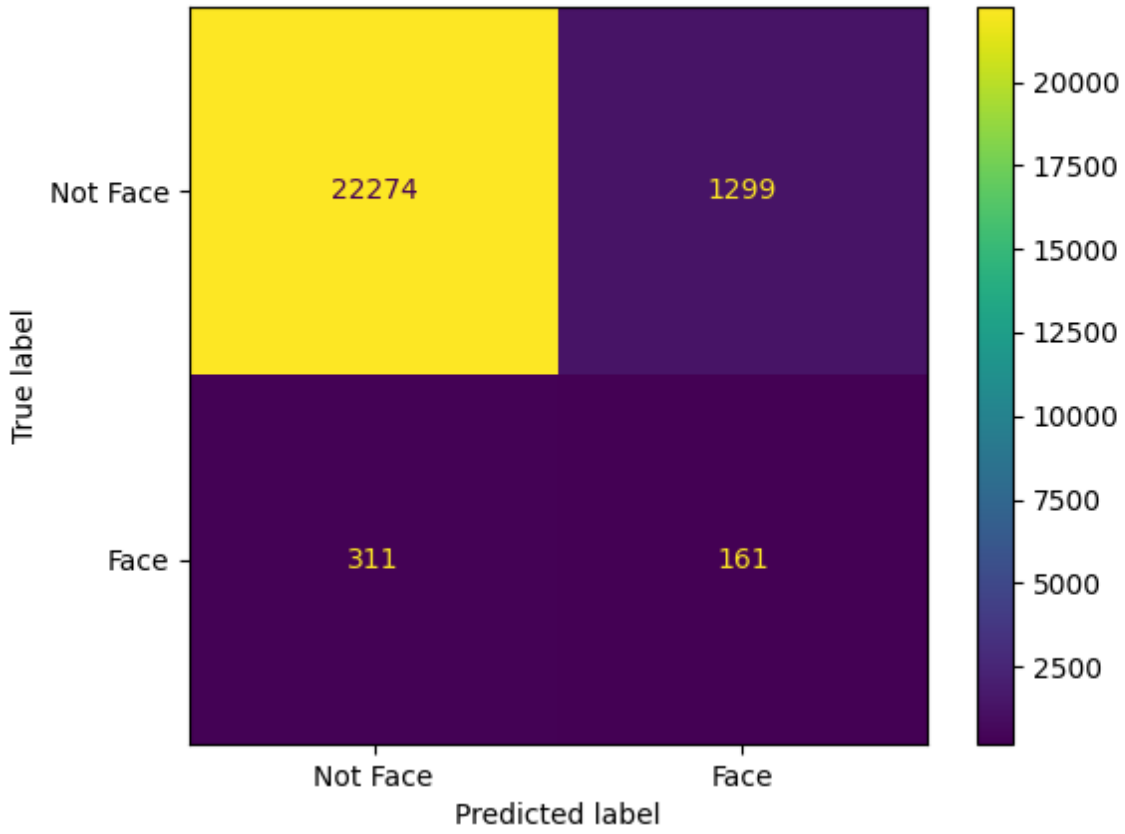
For 10.pkl

False Positive Rate: 5017/23573 (0.212828)
False Negative Rate: 245/472 (0.519068)
Accuracy: 18783/24045 (0.781160)
Average Classification Time: 0.000617



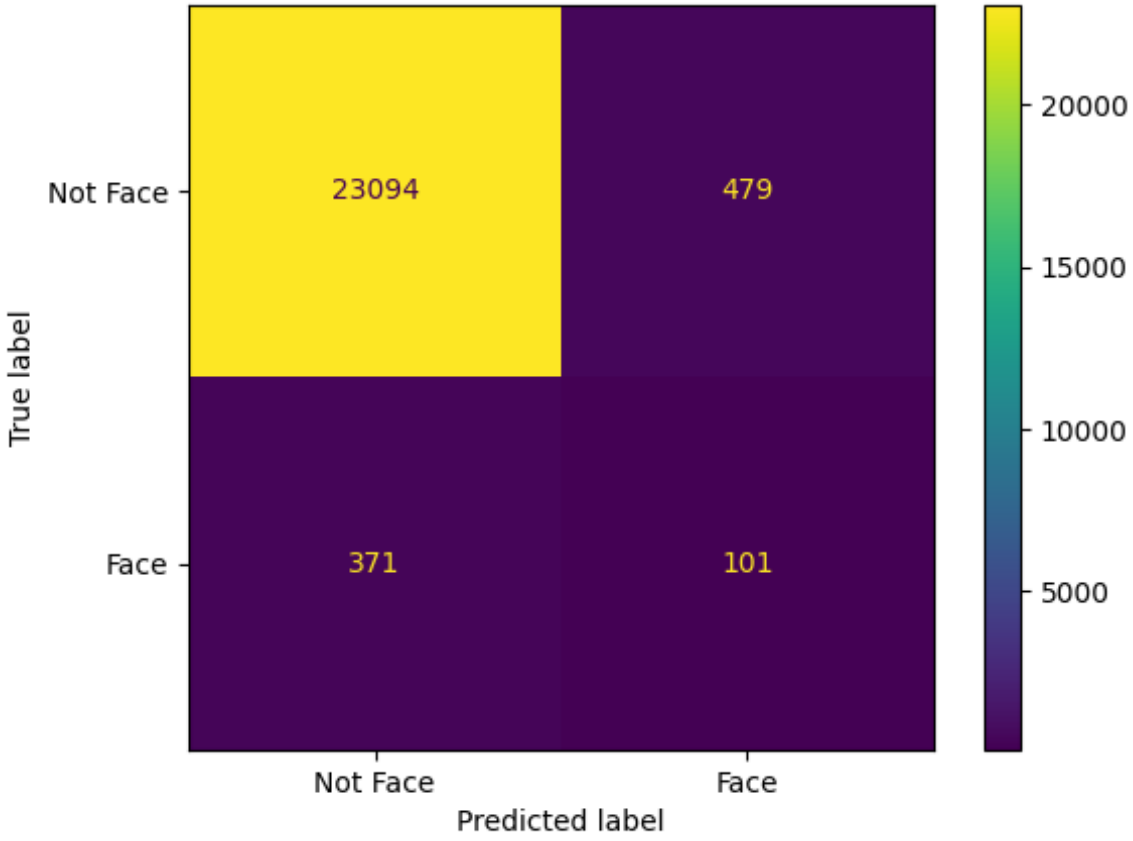
For 50.pkl

False Positive Rate: 1299/23573 (0.055105)
False Negative Rate: 311/472 (0.658898)
Accuracy: 22435/24045 (0.933042)
Average Classification Time: 0.000747



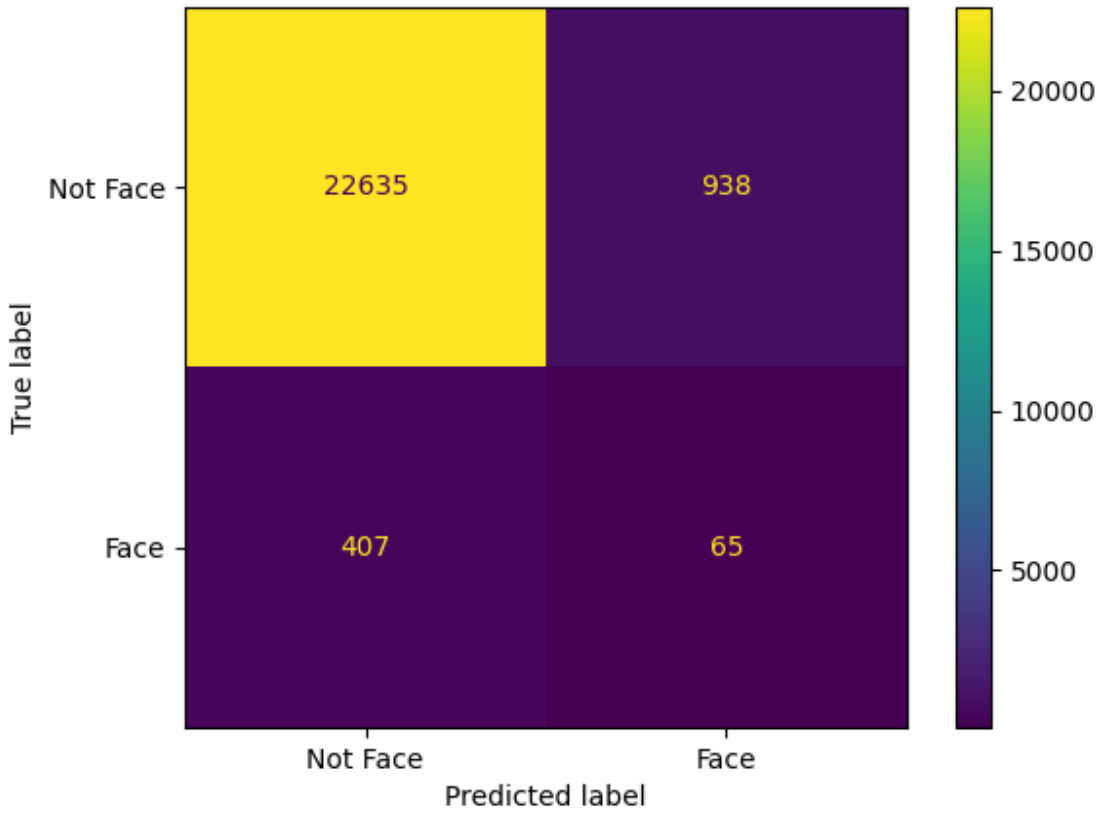
For 200.pkl

False Positive Rate: 479/23573 (0.020320)
False Negative Rate: 371/472 (0.786017)
Accuracy: 23195/24045 (0.964650)
Average Classification Time: 0.001135



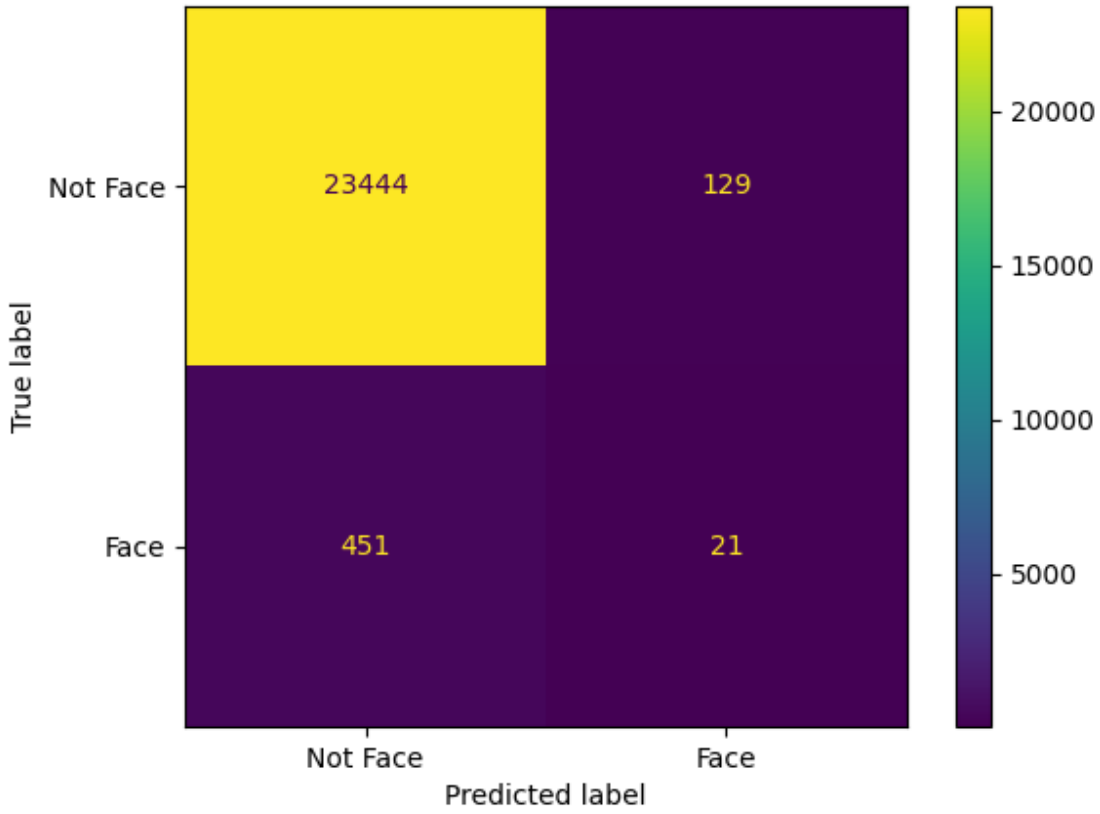
For cascade1.pkl

False Positive Rate: 938/23573 (0.039791)
False Negative Rate: 407/472 (0.862288)
Accuracy: 22700/24045 (0.944063)
Average Classification Time: 0.000837



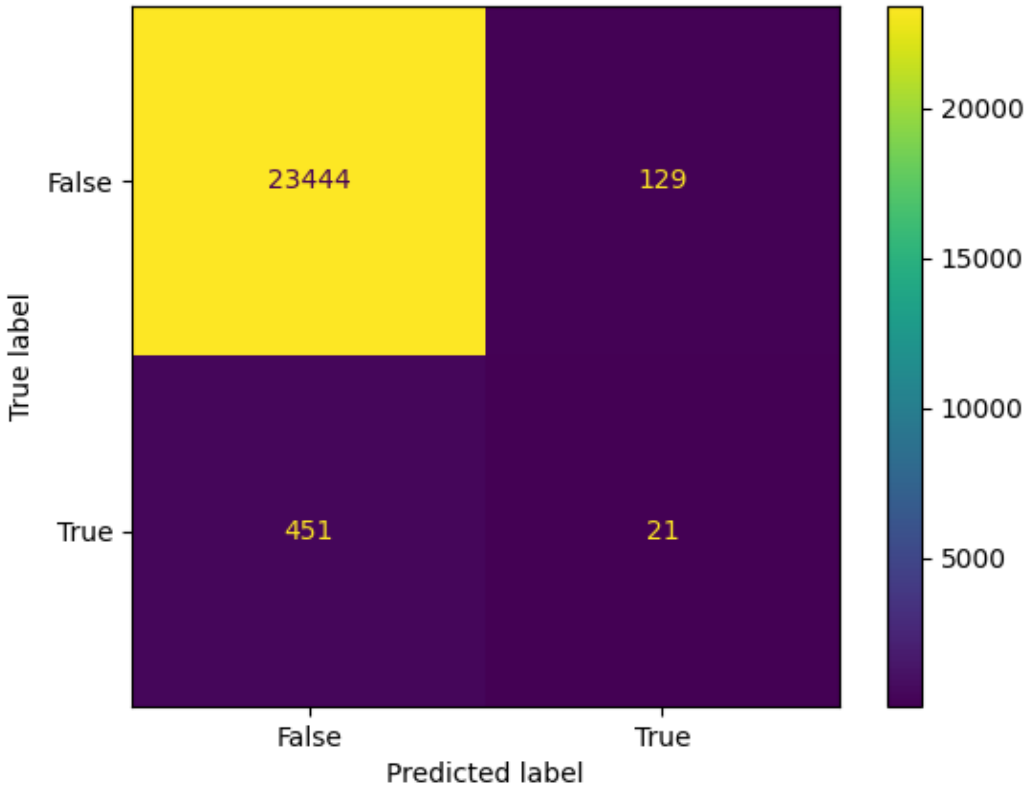
For cascade.pkl

False Positive Rate: 129/23573 (0.005472)
False Negative Rate: 451/472 (0.955508)
Accuracy: 23465/24045 (0.975879)
Average Classification Time: 0.000355

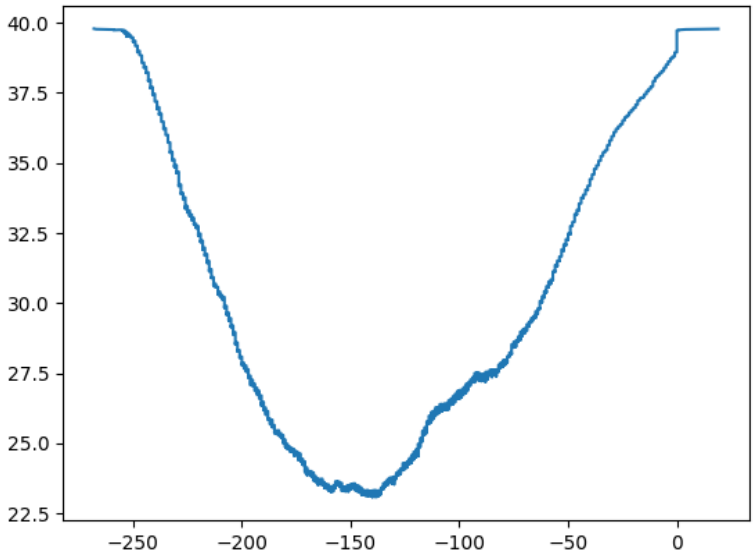


For cascade2.pkl

False Positive Rate: 203/23573 (0.008612)
False Negative Rate: 409/472 (0.866525)
Accuracy: 23433/24045 (0.974548)
Average Classification Time: 0.000777



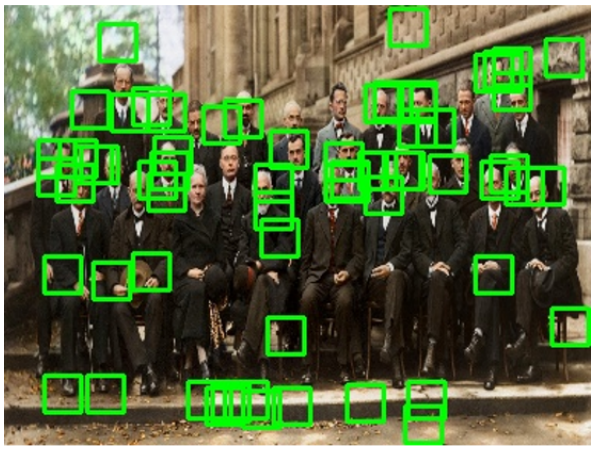
Threshold vs Errors:



Face Detection:-

Cascade1.pkl:- 1, 2, and 3 features

Cascade2.pkl:- 1, 10, 50, and 100 features

Cascade.pkl :- 1, 2, 5, 10, and 50 features