Face Detection: Training a Cascade Object Classifier

Giovanni Mantovani and Andrea Palmieri

April 12, 2024

1 Introduction

The aim of this report is to explore the training process of multiple face detection models utilizing the Viola-Jones cascade object detector algorithm within the MATLAB environment. Our approach involves training various models by changing the following parameters FeatureType and FalseAlarmRate. Following the training phase, we conduct a evaluation of the trained models by computing the average precision AP. Additionally, we visually represent the performance of each model through precision-recall curves.

2 Methodology

2.1 Positive images

The positive images class consisted of 6.713 face cutouts, in grayscale, measuring 36x36 pixels.

2.2 Negative images augmentation

For the augmentation of the negative class images, several techniques were applied to diversify the dataset. Initially, each original image was augmented by creating two additional versions: one flipped vertically and one flipped horizontally. Subsequently, rotation was performed 25 times on the original image, 25 times on the horizontally flipped image, and 25 times on the vertically flipped image. This rotation process introduced variations in object orientation. Additionally, random noise was added to the original image 10 times to simulate variations in pixel intensity and texture. In total, each original image resulted in 88 augmented images (1 original + 2 flips + 75 rotations + 10 noise additions). Consequently, the negative images dataset was augmented from 274 to 24,112 images.

2.3 Viola-Jones Detector Training

The Viola-Jones cascade object detector algorithm was trained using the positive and negative class images. The `trainCascadeObjectDetector` function was utilized, with the following parameters:

- **FalseAlarmRate**: 0.2, 0.3, 0.4
- **NumCascadeStages**: 20
- **NegativeSamplesFactor**: 4
- **FeatureType**: LBP, HOG
- **ObjectTrainingSize**: [36, 36]

These parameters influence various aspects of the training process and the characteristics of the resulting detector.

The `FalseAlarmRate` parameter determines the algorithm's thresholds for accepting regions as containing "faces", affecting the balance between accurate detections and false positives.

`NumCascadeStages` controls the number of cascade stages trained and was set to 20 in order to train as many stages as possible.

The `NegativeSamplesFactor` parameter determines the ratio of negative samples to positive samples used in training, impacting the robustness of the classifier against false detections.

The choice between `LBP` and `HOG` for `FeatureType` affects the type of features extracted from the images. `LBP (Local Binary Patterns)` features focus on capturing texture information by comparing pixel intensities, while `HOG (Histogram of Oriented Gradients)` features emphasize local gradient orientations to detect shape and structure within images.

Finally, `ObjectTrainingSize` specifies the dimensions of the training images, aligning them with the size of the face images in the dataset for optimal training efficiency.

3 Results

In this section, we present the results of the face detection classifiers for different combinations of 'FalseAlarmRate' and 'FeatureType'. They are summarized in Table 1, showcasing the impact of these parameters on the face detection process.

Classifier	Num Of Stages	Feature Type	False Alarm Rate	Accuracy
Classifier1	6	LBP	0.2	0.6127
Classifier2	7	LBP	0.3	0.5415
Classifier3	9	LBP	0.4	0.5075
Classifier4	7	HOG	0.2	0.6252
Classifier5	8	HOG	0.3	0.5801
Classifier6	10	HOG	0.4	0.5780

Table 1: Summary of Experiment Results

Additionally, the precision-recall curves for each classifier are shown in Figures 1 to 6. These curves provide insights into the performance of each classifier in terms of precision and recall, highlighting the trade-offs between these two metrics.

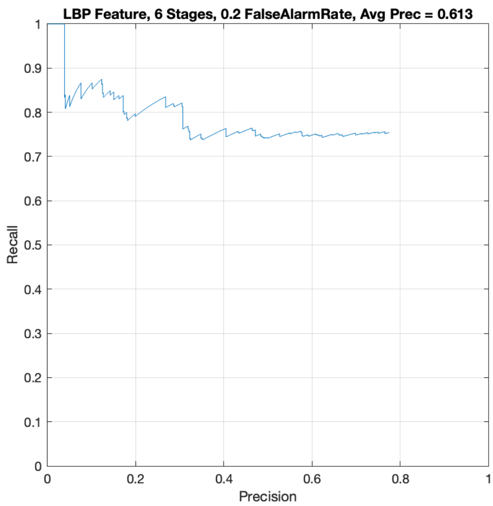

Figure 1: Precision-Recall Curve of Classifier with FalseAlarmRate set to 0.2, LBP features, and 6 stages

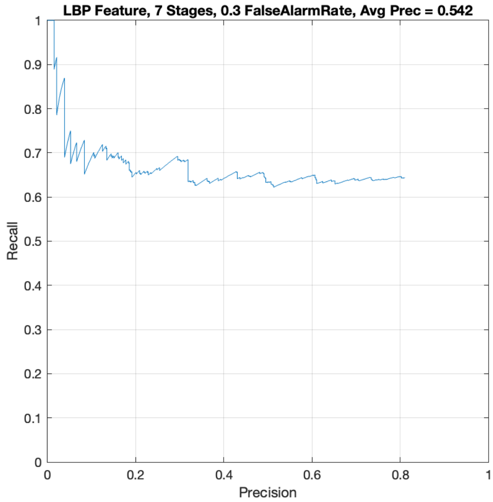

Figure 2: Precision-Recall Curve of Classifier with FalseAlarmRate set to 0.3, LBP features, and 7 stages

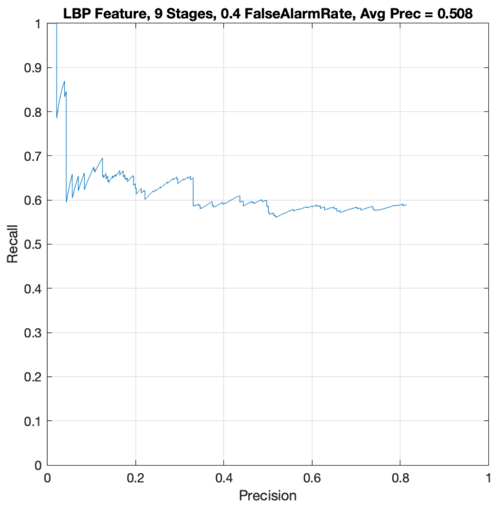

Figure 3: Precision-Recall Curve of Classifier with FalseAlarmRate set to 0.4, LBP features, and 9 stages

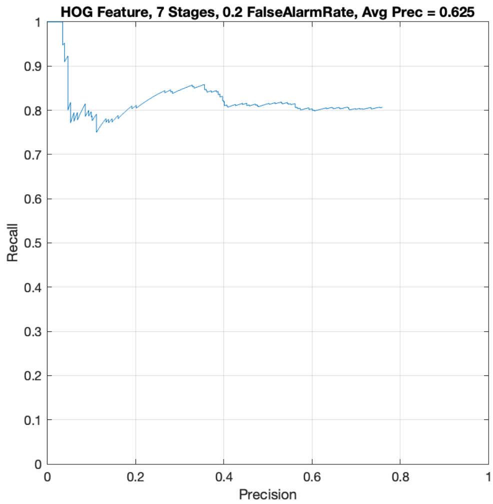

Figure 4: Precision-Recall Curve of Classifier with FalseAlarmRate set to 0.2, HOG features, and 7 stages

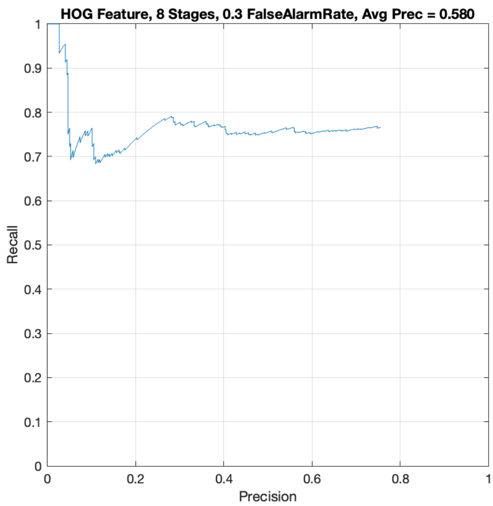

Figure 5: Precision-Recall Curve of Classifier with FalseAlarmRate set to 0.3, HOG features, and 8 stages

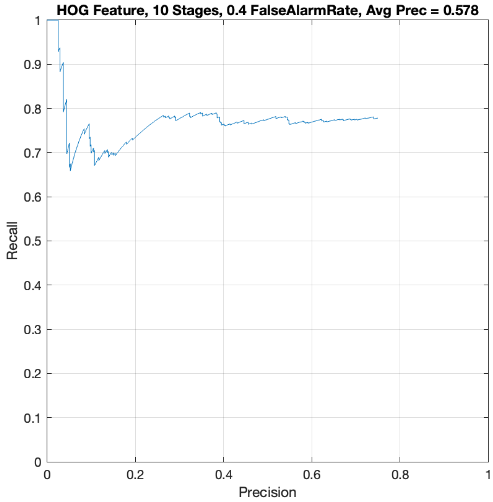

Figure 6: Precision-Recall Curve of Classifier with FalseAlarmRate set to 0.4, HOG features, and 10 stages

4 Conclusion

After conducting the training experiment on a Viola-Jones cascade object detector for face detection, several key findings have emerged.

Our results highlight a generally low performance of our classifiers, indicating that the negative images dataset lacked diversity and did not present significant challenges to the classifiers. This is shown by the low average precision scores obtained across all classifiers, with none exceeding 0.63. Moreover, the performances of classifiers using different features (HOG or LBP) were similar for a FalseAlarmRate of 0.2, whereas classifiers trained on HOG features and FalseAlarmRate of 0.3 and 0.4 achieved better average precision that their LBP counterparts. Overall, our findings highlight the importance of dataset diversity in training robust face detectors and underscore the need for more challenging negative image datasets to improve classifier performance.