Lab 9: HOG Face Detection

Cristian M. Amaya
Universidad de los Andes
cm.amaya10@uniandes.edu.co

Abstract: In this laboratory, a HOG-based face detection method will be train with a subset of the WIDER FACE datasets. To crated the detection method, the basic object category detection method from the Oxford Visual Geometry Group will be used.

Keywords: Detection, face, HOG, multi-scale

1.. Introduction

Face detection is one of the first steps in many deep face methods that seek to find specific features inside faces like gestures, traits or the different key-points alignment, techniques such as Deep Face Alignment or Deep Face Recognition have plenty of industrial applications as they can be implemented to serve different purposes.

In this practice, an Multiple Scale Oriented Histograms based Detection method will be adapted to run on a subset of the modern WIDER FACE database, based in a generic object class detection method based in the VLfeat library and provided by the Oxford Visual Geometry Group. Additional, a face detector will be trained based in the Viola-Jones algorithm to compare the results the detection accuracy.

2.. Methods and Materials

2.1.. Datasets

WIDER FACE:Face detection dataset that consist of 32,203 images and 393,703 annotated faces, the images are separated in 61 event classes. The annotations are given as bounding box ground truth text files, in this practice only a small subset of test and training images will be used, the faces that will count as positives are the ones that are larger than an 80x80 pixels image. [2]

Places365: Scene recognition database that consist of 365 scene categories, the validation set will be used to provide negative samples for the face recognition methods.[3]

2.2.. Complete strategy

To be able to used the detection method, first the training images and bounding box annotations have to be processed in a specific format to be used with the method. The patches of faces will be resize to the same size, and only one face larger than 80x80 will be consider from each training image in the first attempt. In the second attempt, a subset of training face crops will be use form the given training set will be use as the positive samples. The average of the training instances can be seen in figure 1, it is a good example of the orientations of the majority of faces.

To create negative samples, the validation set of Places365 will be used to generate background samples to train the methods and hard negative mining.

2.2.1. Multi-scale HOG strategy

The detection method was created by the Oxford Visual Geometry Group, its based in the creation of a model from the extraction of Histograms of Oriented Gradients(HOG) features out of the train images. To obtained better results, multiple scales of the train images are taken into account to better detect the faces with the model, additionally SVM and a hard negatives mining strategy are used to create a better method based in positives and negative examples, as shown in figure 2.

2.2.2. Viola-Jones Algorithm

Boosting strategy for object detection, the method uses Integral Images to allow for faster feature extraction and a cascade of combining classifiers that enables to discard the background rather quickly and enables the method to focus on the face detection. It's very famous for its high detection rates for face detection and in this laboratory the Cascade Object Detector class from Matlab will be used, which implements Viola-Jones algorithm.[1]

2.2.3. Detection evaluation

Normally, the detection methods are evaluated in precision-recall curves and Average Precision (AP), as establish in the PASCAL VOC criterion.For this method, evaluation method given by the WIDER FACE database is used, taking into account the bounding boxes coordinates and the overlapping of the predicted boxes and the ground-truth ones with three given levels of difficulty.

3.. Results and Discussions

3.1.. HOG Face Detection

3.1.1. First Attempt: One face by training image

As it can be seen in figure 3, the first trained detector based in the Oxford method has an average precision (AP) of 2.9%, some of the steps that could have lead to this result are the excessive hard negatives mining with training images that could have faces under the 80x80 sizes. As the precision-recall curve shows, the recall was incredibly low, this means that the number of false positives is relatively high compare to the true positives, as it can be seen in figure 4, some of the faces are detected but also a high amount of random objects are detected, therefore the method still lacks better training.

3.1.2. Second Attempt: Training Crops

Using the training faces crops as positive examples and the Places365 for the Hard Negative Mining improve the result of the new Oxford detectors, the main parameters that was changed was the number of negative images to be taken into account and the number of iterations of the Hard Negative mining method. This changes made the detectors more precise, even though a limit was found after using 400 negative images and 5 hard negatives mining iterations, the maximum AP for the new methods was 3.7%.

Even with changing other parameters as the HOG cell size or the scales, the method was still rounding the same average precision, the result is still low but we have to take into account that face detection is not at easy task, specially when using only orientations for multi-scale detection.

3.2.. Viola-Jones Algorithm

The Viola-Jones method was much better in terms of detection accuracy and computation time, just training the detector with 100 images produce a similar result to the best HOG detector (3.6%). The best method was trained with 300 training images, 11 training stages and Haar as the feature type, the average precision for this method was 6,6%.

A better detector with this algorithm could be produced by training with all the training faces crops, but the computational time for the detector's training stages could extend to days. Nevertheless, it is still possible to achieve better results, as we can appreciate in figure 6, the recall was pretty high for this method and the precision increase by taking more positives and negative examples.

3.3.. Final Detectors

The two final detectors precision-recall curves can be observe in figure 8, the HOG detector achieve high precision measures and low recall and the Viola-Jones achieve high recall while doing poorly on the precision, perhaps taking into account both detection predictions outputs and integrating them into a single method could achieve a more accurate detector with decent precision and recall.

4.. Conclusion

Face detection is no easy task in natural images, many factors as orientation, occlusion and scale have to take into account. The basic strategy with HOG does not consider color or occlusion, although the Oxford method consider multiple scales.

In any case, both detectors are good enough considering the detection task is a complex, additionally the Viola-Jones algorithm could be train with all the training crops and achieve a higher accuracy.

References

[1] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[2] S. Yang, P. Luo, C. C. Loy, and X. Tang. Wider face: A face detection benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[3] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, and A. Oliva. Places: An image database for deep scene understanding. *arXiv preprint arXiv:1610.02055*, 2016.

Images

HOG Face Detection

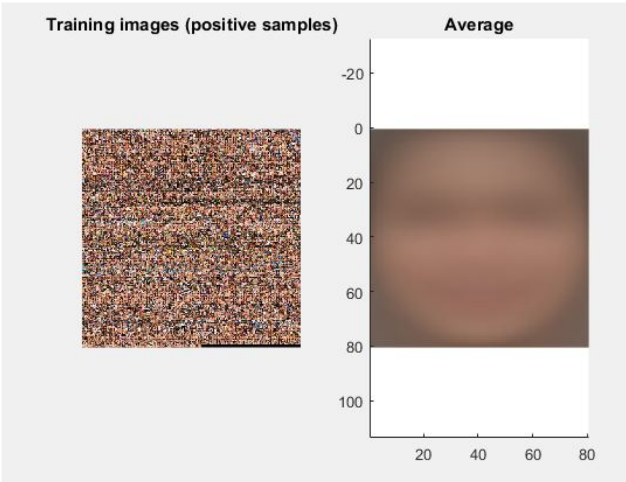

Figure 1. Positive samples set and its corresponding average

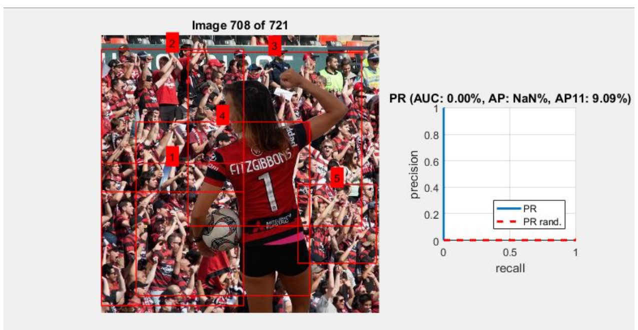

Figure 2. Negative mining process

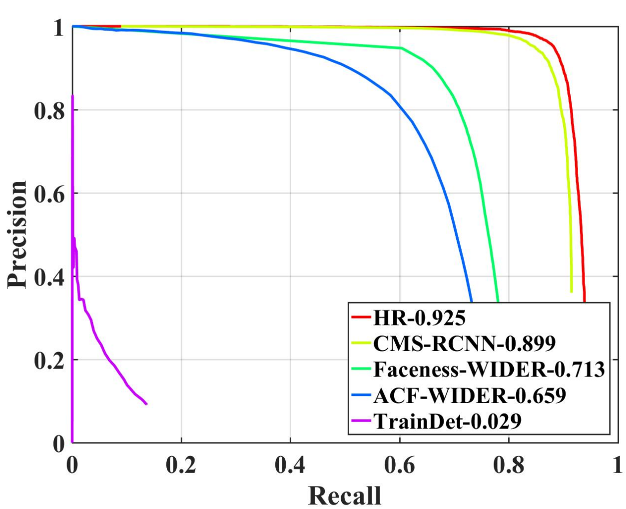

Figure 3. Precision-Recall curve for the method TrainDet

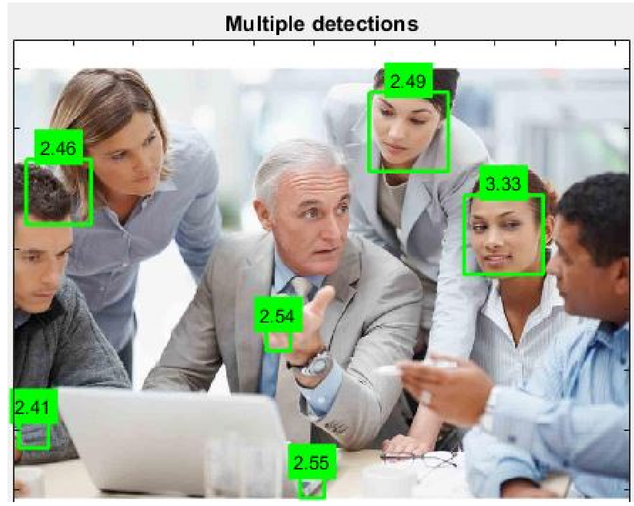

Figure 4. Example of predicted faces

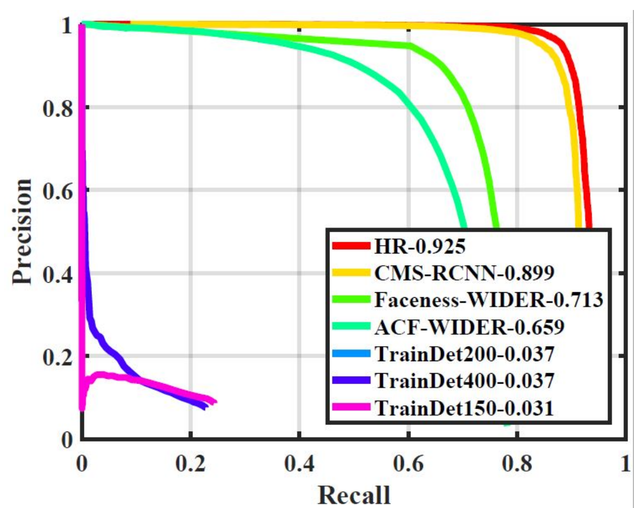

Figure 5. Precision-Recall curve for the new TrainDet methods

Viola-Jones Algorithm

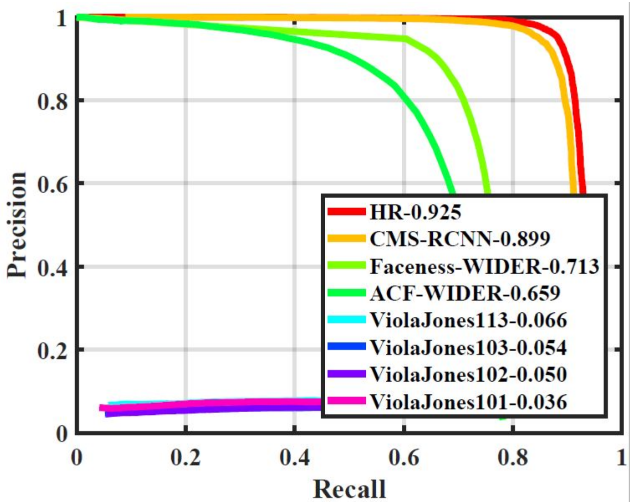

Figure 6. Precision-Recall curve for the Viola-Jones methods

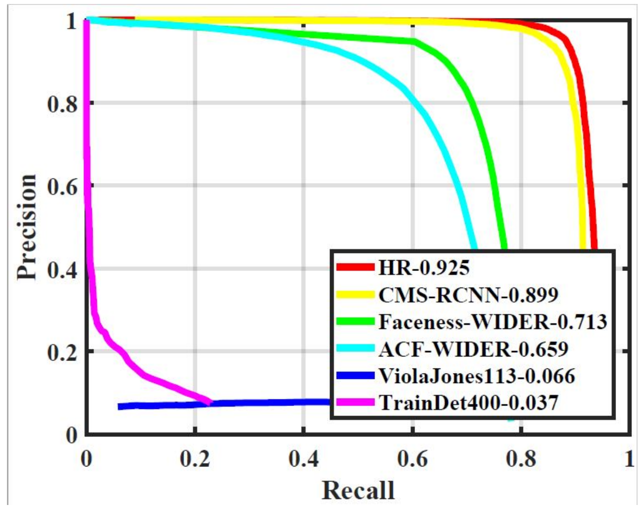

Figure 8. Precision-Recall curve for the final methods

Figure 7. Detection example

Final Detectors