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.

2.. Methods and Materials

2.1.. Dataset

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. [1]

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.

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 SVMs and hard negatives mining strategy are used to create a better method based in positives and negative examples.

2.2.2. Detection evaluation

Normally, the detection method are evaluated in precision-recall curves taking into account false and true positives and negatives.For this method, a similar approach will be tested, but taking into account the bounding boxes coordinates and the overlapping of the predicted boxes and the ground-truth ones, if the overlapping is over 50%, the prediction is consider correct.

3.. Results and Discussions

3.1.. Detection Strategy

As it can be seen in figure 5, the Average precision for the detection method was 0.029, this was unexpected as the average of the positive samples (Figure 1), was a good representation of the faces.Some of the steps that could have lead to this result are the excessive hard negatives mining with images that could have faces under the 80x80 sizes, the usage of the ground-truth coordinates for the positive samples instead of using the crops included in the subset, the resizing of the positive samples or the incorrect input of images into the detection method.For the final result explanation, 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 observe in figure 6 through 8, were some faces are detected but also an incredible amount many random objects are detected, therefore the method still lacks better training.

Nevertheless, to improve the results, many strategies could overcome the current method limitations: the usage of images without faces instead of the ones with faces lower than the 80x80 size could improve the model and the hard negative mining, parameter optimization, taking into account color,more scales and using other orientation features to train the model.The false positives follow strange patterns that could be avoided taking into account more information than the orientation and scale.

4.. Conclusions

The detection of human faces is no easy tasks, many modifications will be needed to improve the recall of the method. It is also worth noting that the detection of faces (even large ones) has many variables that could lead to a very general model as there are color and light variations in faces, the orientation, background and obstruction are also factors to take into account.

References

[1] 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.

Images

Detection Strategy

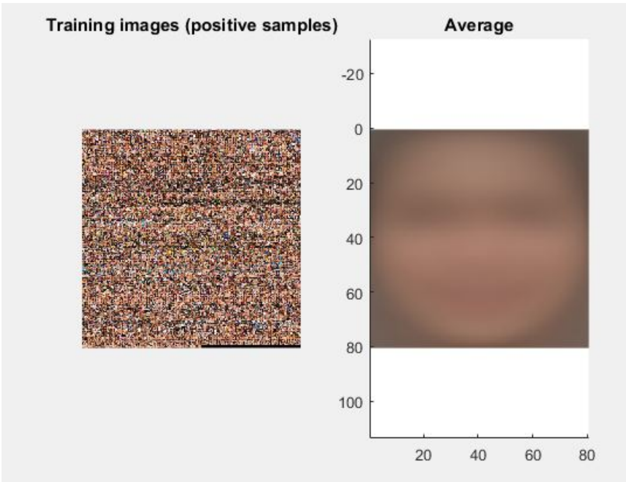

Figure 1. Positive samples set and its corresponding average

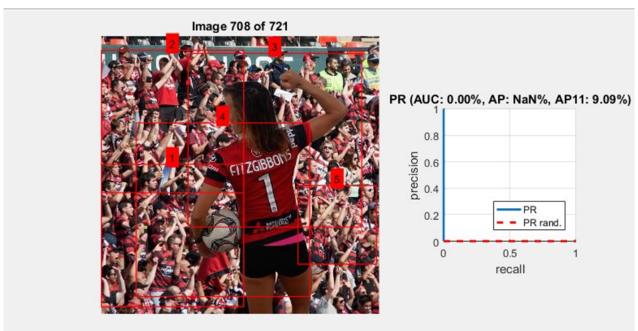

Figure 2. Negative mining process

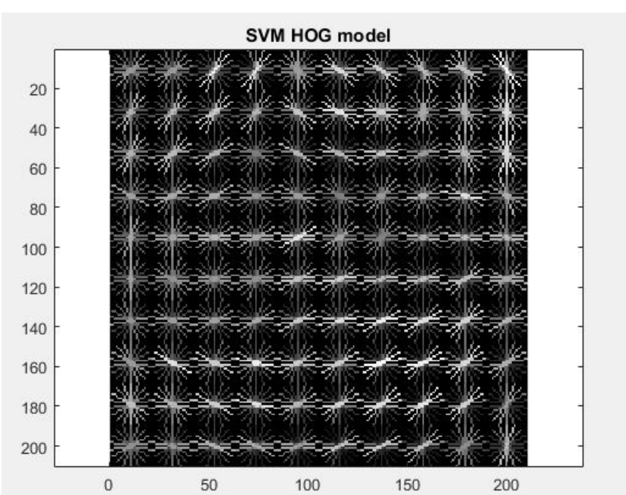

Figure 3. First SVM HOG model

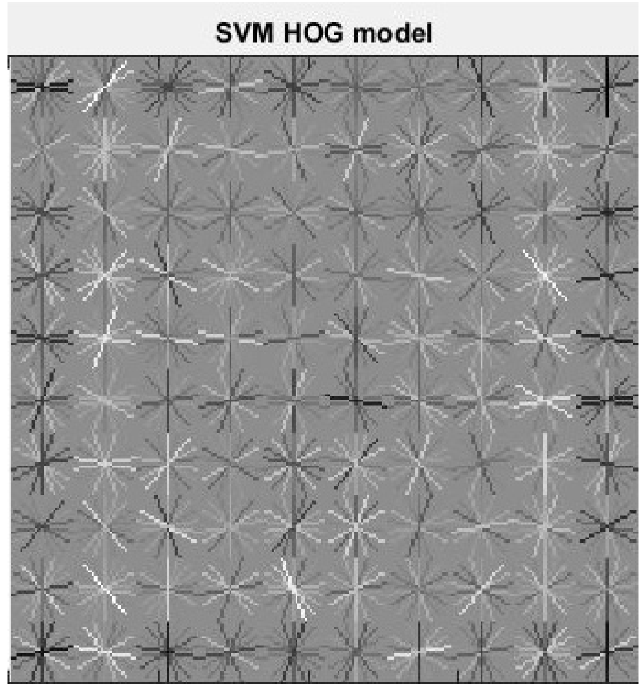

Figure 4. Third SVM HOG model

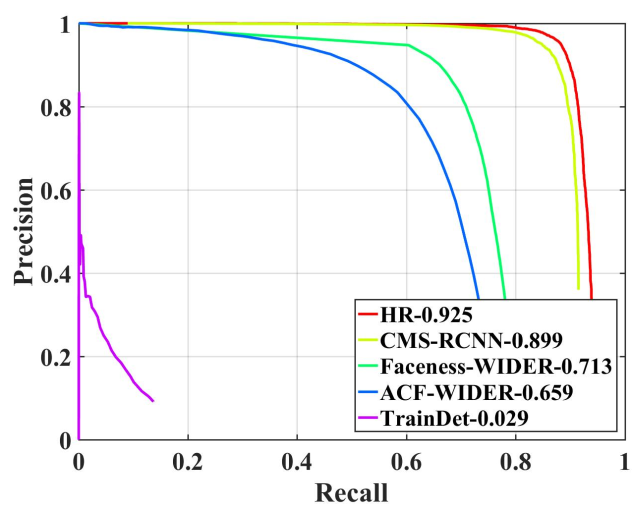

Figure 5. Precision-Recall curve for the method TrainDet

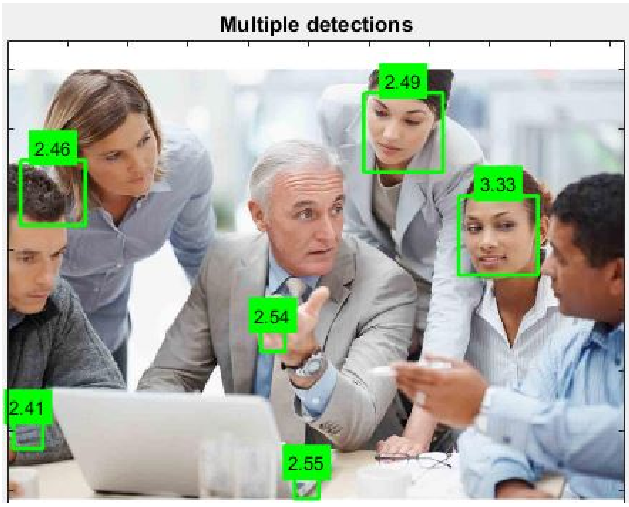

Figure 6. Example of predicted faces

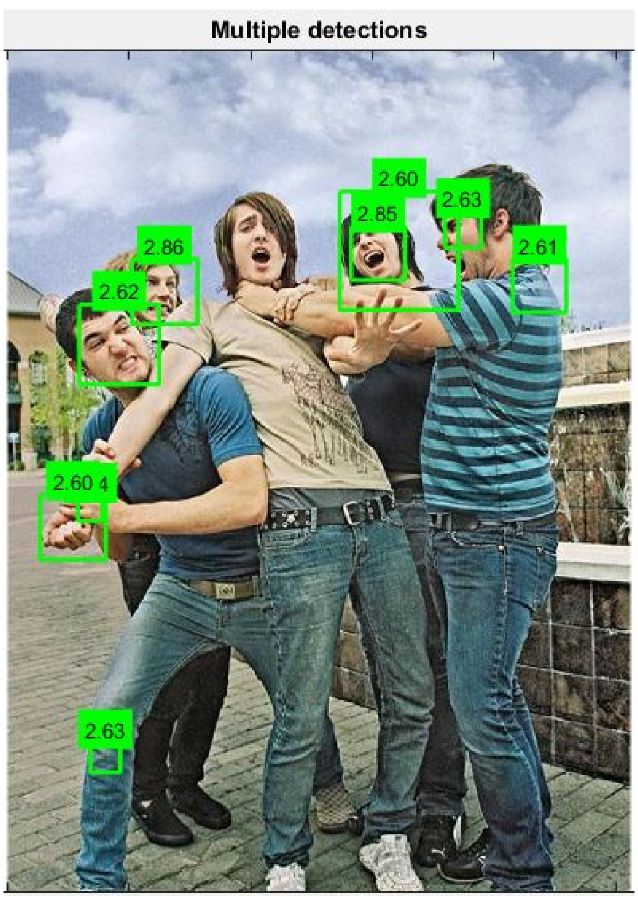

Figure 7. Example of predicted faces

Figure 8. Example of predicted faces