HOG

Federico Ulloa
Universidad De Los Andes
Cra. 1 18a-12, Bogotá, Colombia
f.ulloa10@uniandes.edu.co

Esteban Vargas
Universidad De Los Andes
Cra. 1 18a-12, Bogotá, Colombia
e.vargas11@uniandes.edu.co

Abstract

Histograms of Oriented Gradients (HOG) is a popular feature used in detection problems. It consists in computing the oriented gradients of each pixel in a window, compute the histogram and then concatenate all the histograms of an image. This can also be done in pyramidal way (multi-scale HOG), calculating HOG in different dimensions of the same image. In this laboratory, we used the multi-scale HOG features to train an SVM for face detection. For this, a subset of the WIDER FACE: A Face Detection Benchmark was used to train and evaluate the algorithm through a precision-recall curve. We obtained very poor results with our method, and we concluded that our descriptor was too simple, that a single SVM was not enough, and that further exploration with HOG hyperparameters is needed.

1. Introduction

The Histograms of Oriented Gradients (HOG) is a very popular descriptor in computer vision. It is a simple concept, which consists in: first, divide a chosen window into cells, compute the histogram of oriented gradients in each cell, and finally putting all these histograms together to make up a single feature vector.

This feature vector can be used to perform detection of objects in images. In this case, a multi-scale HOG is used to perform face detection. Multi-scale HOG consists in computing the HOG feature in not just the original size of the training image, but to re-scale it to different sizes and get a group of HOG features for each image. [1] This is done in a pyramidal way, hence it is also called Pyramidal HOG, or PHOG. We have, then, that the hyperparameters for HOG are only the size of the window in which HOG will be applied in the images and, for PHOG, the different dimensions of the images in which HOG will be performed. With these extracted features from the training images, the highest responses to these features are taken from test images, which are then labeled as a detection. This feature is often then used to train a classifier depending on the application.

In this laboratory, a multi-scale HOG is used for face detection. This is a widely studied problem due to its very broad application possibilities. As it is done in detection problems, the evaluation method is a precision-recall curve, in which precision is the amount of true positives detected over the total of detections and recall is the true positives detected over all detections.

2. Methodology

2.1. Dataset

The dataset used is a small subset of the WIDER FACE: A Face Detection Benchmark. [2] The data was divided into Training images, Training Crops images and Validation images. Each of them were subdivided into 61 different categories. The images came in a wide variety in sizes and orientations, in a JPG format. The training set contained scenes that belong to each of the categories, the Train Crops contained only cropped faces of images that belong to each of the categories and, lastly, the Validation set also contained images of scenes for each category. For the initial negatives used, a random subset of the ImageNet dataset was used. [3] These images were chosen as none of them contains human faces, so they serve well as initial negatives to train the SVM.

2.2. Multi-scale HOG

The implemented methods are based on the Visual Geometry Group at Oxford object category detection tutorial. [4] First, the training images (crop images) are loaded and re-sized into 64x64, so that we get comparable features and for the algorithm to be less expensive computationally.

Then, multi-scale HOG features are extracted from these images with a window size of 8x8 and scales from 1 to 8. Figure 1 shows the average of all the cropped training images. Later, random patches from the negative images were extracted to reduce even more the chances of getting faces in the negatives. All these were then used to train an SVM classifier. Figure 2 shows the resultant HOG feature descriptor learned by the SVM, its is important to notice that

it seems to represent a face, in the outside part it have a round pattern and inside a higher response on vertical gradient which corresponds to the nose of the face. Also, 8x8 hog size seems to represent correctly the face, as have all the important parts without having to much specificity's that would lead in overfitting.

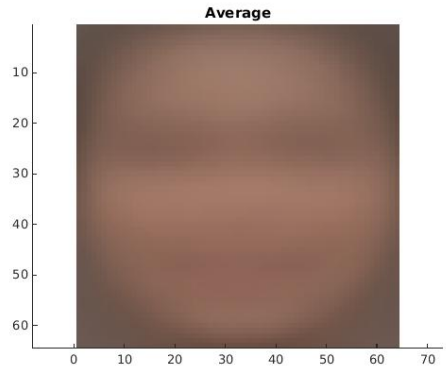

Figure 1: Average of cropped training images

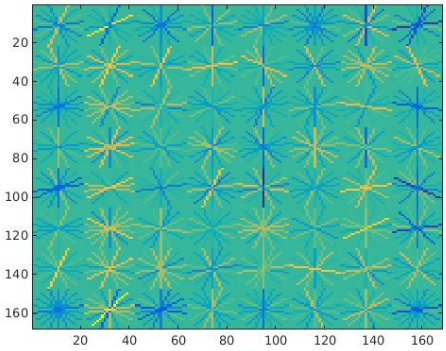

Figure 2: HOG descriptor

Next, the SVM is tested with the non-cropped training images, suppressing the non-maximums and keeping only the detections with the highest score in each image. Lastly, the classifier is evaluated with the validation set.

3. Results

After running the evaluation in the validation set, we obtained the precision-recall curve shown in Figure 3. It can be seen that we obtained 0 as a result, so clearly our method did not work.

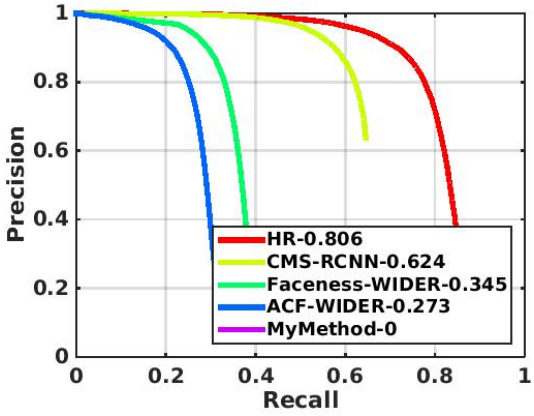

Figure 3: Precision-Recall curve

Figures 4 through 7 show some examples of the detections obtained by our method. It can be observed that detections are often in regions of the faces but very rarely it detects the actual complete face, or it is not at all over faces. Furthermore, there are a lot of false positives, and they are mainly due to rounded parts of the images, which makes sense because the descriptor learned by the svm is practically a circle with some patterns inside, so its really easy for the algorithm to get confused with watches for example, or mouths. In the other hand, the average of cropped training images looks similar to a front face which tell us that the cropped images were correctly align, but also tell us that it will be impossible for the algorithm to detect rotated faces as seen in figure 6 were the algorithm just detect correctly the front face.

With this, we can say that the descriptor built is very simple by far, as extracting the average of all the images gives a very vague notion of an actual face, as this is a very complex database and faces come in a very big variety in size, orientation, rotation, color, brightness and also often have occlusion. So it is evident that training a single SVM with all these images was not a good approach.

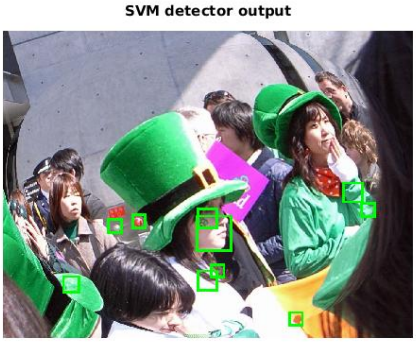

Figure 4: Detection example 1

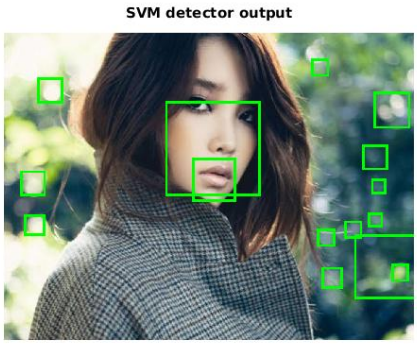

Figure 5: Detection example 2

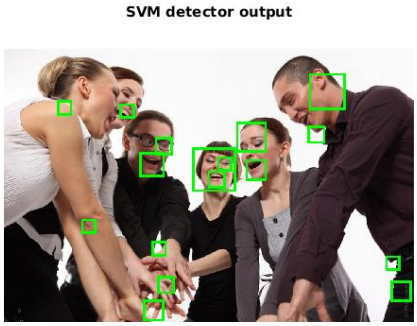

Figure 6: Detection example 3

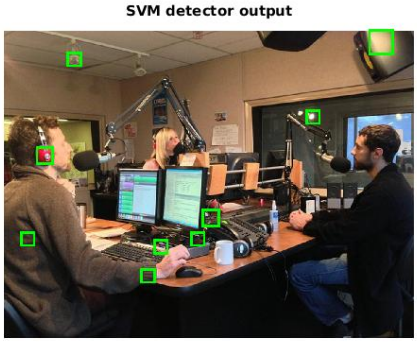

Figure 7: Detection example 4

For better results, we propose training a SVM per category of the dataset or even sub-dividing again the dataset into groups in which faces look similar among each other. This way, a classifier for each category could be trained better, including the hard negative mining to improve the training, and this way, although it would be a lot more expensive in computation, better results could be obtained.

Lastly, it is also important to note that further exploration of the HOG descriptor must be done. Meaning that different window sizes and more scales of the images should be included. This is because, as mentioned earlier, this database is very wide and challenging, so faces are present in many different scales and orientations.

4. Conclusions

Our multi-scale HOG with SVM method gave us a result of 0, meaning that it did not work. We observed that the descriptor built was way too simple to apply in this very broad database.

Patterns in the false positives were observed, being most of them in rounded parts of the image, reinforcing the fact that the descriptor was too vague.

As this database contains a big amount of faces in quite different varieties, we concluded that a single SVM is simply not enough. Hence, we propose to subdivide the images in their preexisting categories or in new categories with images containing similar faces and training a SVM for each of the categories.

Lastly, we concluded that a wider experimentation with the multi-scaled HOG hyperparameters is needed, as the ones used might have been cut short for this database.

References

[1] L. Zhang and L. Li, "Improved Pedestrian Detection Based on Extended Histogram of Oriented Gradients", Applied Mechanics and Materials, vol. 347-350, pp. 3815-3820, 2013.

[2] "WIDER FACE: A Face Detection Benchmark", Mmlab.ie.cuhk.edu.hk, 2018. [Online]. Available: http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace/.

[3] "ImageNet", Image-net.org, 2018. [Online]. Available: http://www.image-net.org/.

[4] "VGG Practical", Robots.ox.ac.uk, 2018. [Online]. Available: http://www.robots.ox.ac.uk/ vgg/practicals/category-detection/.