Assignment 5: Face Detection

Mario Döbler

Institute for Signal Processing and System Theory
University of Stuttgart

1 Introduction

In this assignment we look into a face detector that uses HOG features to train a linear SVM and a multi-scale sliding window detector that classifies image patches at different scales.

Even though we won't get state-of-the-art performance, as with deep learning approaches, the performance will still be decent and prediction is much faster. Successes of face detection can be traced back to influential work, such as [RBK98] and [VJ01]. In this assignment you will implement the face detector from [DT05].

Pipeline:

- Feature extraction using Histogram of Oriented Gradients (HOG) for positive samples (face crops) and negative samples (randomly sampled patches from non-face scenes)
- Train a linear SVM
- Multi-scale sliding window detector for individually classifying patches at different scales
- Non-maximum suppression to remove duplicate/overlapping detections

2 Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients (HOG) is/used to be a popular feature descriptor for object detection. We won't present the details of HOG, but if you're interested you can either look at the original paper *Histograms of Oriented Gradients for Human Detection, by Dalal and Triggs* or have a look at the documentation from scikit-image `https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_hog.html`.

An example visualization of HOG features is illustrated in Figure 1.

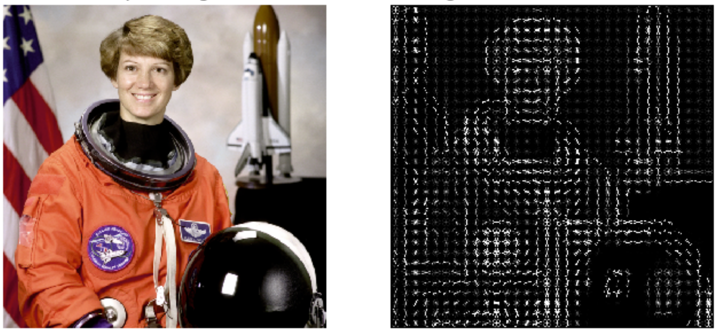

Figure 1: Visualization of HOG features (8 orientations and 16×16 pixels per cell were used)

3 Data Collection

The choice of training data is crucial for this task. While an object recognition system is typically trained and tested on a single database (as in the Pascal VOC challenge), face recognition papers have traditionally been trained on heterogeneous datasets. We will use three databases: (1) face crops as positive training samples, (2) non-face scenes for negative training data, and (3) test scenes with ground truth face locations.

For face crops, we use the UTKFace dataset (`https://susanqq.github.io/UTKFace/`). It provides 20k+ aligned and cropped faces. Data augmentation is a viable option since faces in an overall scene are not necessarily aligned in the same way (e.g. they might be slightly rotated). You're also welcome to try out different datasets and report your results.

The SUN2012 dataset (`https://groups.csail.mit.edu/vision/SUN/`) contains 16k+ images from various scenes. Some scenes might contain faces, but it's unlikely that we will randomly extract a patch that contains a "perfect" face. In case you have too much time to spare, feel free to remove them.

And last but not least, the FDDB: Face Detection Data Set and Benchmark (`http://vis-www.cs.umass.edu/fddb/`) is used as our test dataset.

4 Tasks

1. When using a linear classifier, what should hold for our features? How can SVMs efficiently perform non-linear classification?

2. SVMs are one of the most robust classifiers, how?

3. Implement the extraction of positive and negative features. For the neg-

ative feature extraction it's recommended to randomly sample patches at different image scales.

4. Train a linear SVM (use LinearSVC from sklearn) using the extracted features.

5. Implement a sliding window, multi-scale detector and run it on the provided test data. In case that this seems too complicated to you, you can start with a single-scale detector and extend it to a multi-scale detector.

Optional tasks:

- Hard negative mining is the idea to run the trained detector on the training data and use the false positives as additional negative samples for another training iteration. Multiple iterations of training, predicting, and including new negative samples are possible.

- Data augmentation for face images.

- Consider using additional/other datasets for extracting positive and negative features.

References

[DT05] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[RBK98] Henry A Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, 20(1):23–38, 1998.

[VJ01] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.