Traffic Sign Detection in Colour Images

Marvin Klaus[1], Daniela Schacherer[2]
[1] *Heidelberg University, M.Sc. Applied Computer Science*
[2] *Heidelberg University, M.Sc. Applied Computer Science*

Problem Definition

Our project compares two state-of-the art approaches towards the detection of traffic signs in color images. The first approach uses a linear classifier, more precisely a support vector machine (SVM), based on Histogram of oriented gradients (HOG) features. The second approach employs a recurrent neural network (RNN). Our training set was made available by the University of Bochum in the context of a traffic sign detection competition in 2013 [**Houben-IJCNN-2013**]. It consists of 900 training images which we split into 600 images for training and used the remaining 300 images for testing.



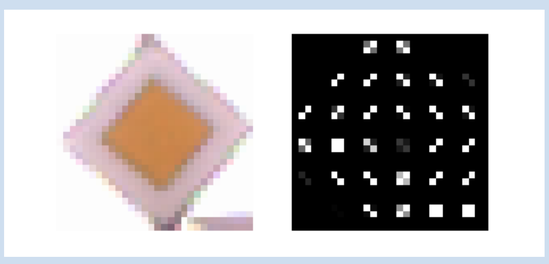
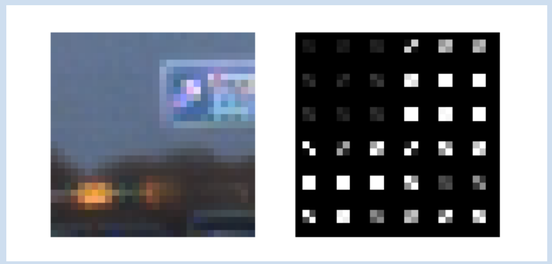
Approach 1: SVM based on HOG features

Data Preprocessing

- Preparation of positive image patches (30×30 pixels) showing traffic signs.
- Preparation of negative image patches (30×30 pixels) showing something that is not a sign.

Feature extraction

- Extraction of HOG features from the positive and negative image patches using a HOG implementation from scikit [**scikithog**].

Training

- Fit the SVM to the prepared training data.
- Hard negative mining: apply the trained SVM to some of the training images (at different scales). Each falsely detected patch is taken and explicitly added as a negative example to the training set.
- Fit the SVM using the enlarged training data.

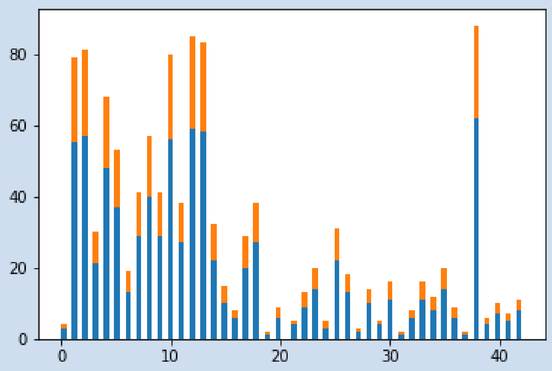
Detection of traffic signs in an image

- Slide a window across the image at different scales and obtain a prediction for every image section from the SVM.

Approach 2: RCNN

Data Preprocessing

- Create negative image patches by taking samples from traffic scenes without signs
- Create train and test dataloader with following partition



Network

- Loss function: Cross Entropy Loss
- Optimizer: Adam

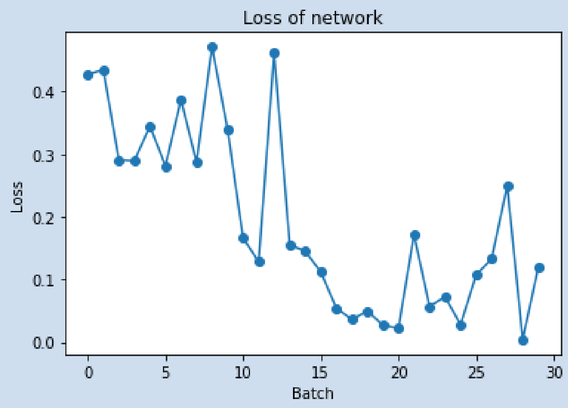
Results SVM

Performance Measure: Each predicted bounding box P is compared against every ground-truth G using the Jaccard similarity (intersection over union):

$$S(P, G) = \frac{|P \cap G|}{|P \cup G|}$$

When $S \geq 0.6$, the ground truth sign is considered as detected. Finally we record the fraction of detected ground truth signs in all images.

Results RNN



- Selective Search to find interesting regions
  - parameter: scale=300 and sigma=0.8
- Run network with every region to get signs
- Same measurement as in SVM approach to find positive results