Quality of region proposals in traffic sign detection and recognition

Domen Tabernik, Rok Mandeljc, Danijel Skočaj

Faculty of Computer and Information Science, University of Ljubljana
domen.tabernik,rok.mandeljc,danijel.skocaj@*fri.uni-lj.si*

Abstract

In this paper we evaluate a two stage detector for traffic sign detection proposed in our previous work which addresses the detection of all traffic signs by reducing the search space with region proposals in the first stage. This work evaluates the whole pipeline with several different feature types in the second stage and explores the effect of region proposals quality, i.e. an overlap with the object, on the second stage classification. We asses the quality of regions using three proposed metrics and show that on German Traffic Sign Detection Database the required region overlap for optimal performance is around 0.7 and 0.8 or higher, while at lower overlaps classifiers becomes too sensitive.

1 Introduction

Within the field of computer vision, traffic signs detection and recognition has become an extensively researched problem [12, 19, 8, 7, 11]. The detection of 30 to 50 common traffic signs, such as speed limit signs, stop and yield signs, pedestrian crossing signs has been given significant attention due to many practical uses in automotive safety and autonomous vehicles applications. On the other hand the detection of the remaining signs, such as information signs and direction signs, has not developed as fast due to limited applicability. Nevertheless, extending the detection to the remaining signs would have applicational use, particularly, in road maintenance services [15] where an important task is verification of presence or absence of all road-based traffic signs, including verification of various information signs, special road marking signs and various direction signs (see, Figure 1). Detection of all traffic signs would eliminate manual verification in such tasks, while it could also be useful in current applications of autonomous vehicles to augment the navigation when GPS signal is poor.

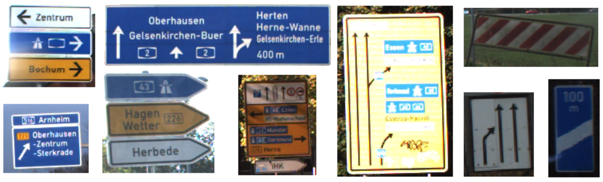

Figure 1: Examples of traffic signs with high degree of shape and color variability.

Existing approaches for traffic sign detection and recognition range from methods that focus on hand-crafted features, such as color thresholding [14, 10], Hough transform [12] and template matching [10, 9], to methods that utilize general features and strong classifiers to achieve state-of-the-art results [7]. Among later, Integral Channel Features [3] have proven useful for quick detection while HOG [2] features can be used to refine the object classification [11, 18]. While such approaches produce state-of-the-art results they may not be scalable for the detection of remaining signs as many specialized features would be required in hand-crafted features or many separate models from general features and classifiers would be needed to cover 300 or more remaining categories.

In our previous work [16] we proposed to address this problem with a two stage pipeline using region proposal algorithms in the first stage and object classification in the second stage. Utilized region proposals have been originally developed for generic object detection to find regions that fully enclose any visual object [1, 17, 20]. This approach is independent of specific class characteristics and is suitable for detection of remaining traffic sign objects. Since region proposals return a smaller set of regions, a more powerful classifiers can be then used in the second stage. In [16] we explored different region proposal algorithms suitable for traffic sign domain and further proposed domain-specific adaptations for edge-boxes [20], to improve the overall accuracy of traffic sign detection. On German Traffic Sign Detection Database [7] our proposed solution generated regions that covered 90% of groundtruth objects with high-quality regions and 99% of objects with lower-quality regions. However, only the first stage of the proposed pipeline was evaluated and no affects on the performance of the second stage classifiers were explored.

In this work we evaluate the full pipeline proposed in [16]. In particular, we focus on the problem of region proposals quality and their effect on the final classification performance. Since regions will not always perfectly overlap with the object there remains an open question of how does the region quality (i.e. their overlap with an object) influence the performance of the classifier. In [16] a definition of low and high quality regions were set arbitrarily to 0.5 and 0.8 overlap respectfully. In this work we now measure the required quality of regions based on combined second stage classifier and show that regions with overlap of at least 0.7 or 0.8 are required to compete with state-of-the-art detectors. Furthermore, we evaluate the full pipeline on GTSDB [7] with classification stage composing feature extraction and classification. We evaluate several feature types and several color-spaces and use the best one in region quality evaluation.

This paper is structured as follows: in Section 2 our detection pipeline with the features and classifier is described, in Section 3 the evaluation and the results are presented and in Section 4 we conclude with the discussion.

2 Detection pipeline

Our detection pipeline comprises of two stages: (i) in the first stage a smaller set of regions is extracted using domain-specific region proposals and (ii) in the second stage a classification of the region is performed.

2.1 Domain-specific adaptations of edge-boxes

In this section we present a brief description of the region proposal algorithm used, while we refer the reader to [16] for a detailed description.

We utilize a state-of-the-art algorithm edge-boxes [20] and further preform a domain-specific adaptations with two improvements as proposed in [16]. First, structured edges [4] are learnt on traffic sign domain, and second, region proposals are run in a cascade, where edge-boxes provide an initial set of regions (around 100 000 regions) which are then re-scored with a shape information to further reduce the search space. The shape is extracted as trained structured-edge features, resized to uniform size of 40x40 pixels and classified with the linear SVM classifier to separate between traffic sign and non-traffic sign regions.

2.2 Object classification

We utilize classification process composed of feature extraction and feature classification. We use and evaluate several different types of features which have proven useful in both general object detection tasks as well as in traffic sign detection and recognition. All features are extracted from patches of 40x40 pixels in size and are resized using bi-linear interpolation. Additionally, we evaluate the effect of using several different color-space representations: *RGB, HSV, LAB, LUV, and YCbCr.*

Feature classification is performed using one-vs-all linear Support Vector Machine. In particular, we use the linear SVM from the LIBLINEAR [5] package, with the default L2- regularized L2-loss dual SVC solver, and without bias (B = 0).

Histogram of Oriented Gradients (HOG)

We use the VLFeat1 implementation of the 31-dimensional UoCTTI HOG variant [6], which includes both signed and unsigned gradient information. The number of orientations is left at the default value of 9. The cell size was set to 5, which is the same value as the one used in GTSRB to pre-compute the HOG Sets 1 and 2. The resulting feature vector has 1984 elements for single, and 5952 elements for three-channel images.

Feature	Category	Color-space				
		rgb	hsv	lab	luv	ycbcr
HOG	danger	0.760	0.787	0.795	**0.804**	0.754
	mandatory	0.758	0.746	0.781	0.779	**0.784**
	prohibitory	0.974	0.973	0.976	**0.983**	0.977
LBP	danger	0.846	0.856	0.856	0.830	0.819
	mandatory	**0.862**	0.823	0.823	0.840	0.842
	prohibitory	**0.987**	0.986	0.986	0.986	0.984
ICF	danger	0.710	0.682	0.716	**0.751**	0.789
	mandatory	0.773	0.733	0.778	**0.762**	0.792
	prohibitory	0.969	0.920	0.966	**0.946**	0.972

Table 1: Reported averaged precision (i.e. area under precision-recall curve) for different feature types and different color-spaces. Max among various color-spaces for specific feature type and category are shown in bold.

Local Binary Patterns (LBP)

We obtain the histograms of LBP [13] using the VLFeat implementation of the algorithm. The input 40 × 40 patches are divided into 5 × 5 cells, and a histogram of 58 quantized patterns is computed for each cell. The obtained histograms are unrolled into a single feature vector, which has 3712 and 11 136 elements for single and three-channel images, respectively.

Integral Feature Channels (ICF)

Integral Channel Features [3] consist of six gradient orientation channels, a single gradient magnitude channel, and one or three color channels. When applied to input 40 × 40 patches with shrink factor 2 (the rest of parameters are left at default), the resulting feature vectors are 3200-dimensional (single-channel) or 4000-dimensional (three-channel).

3 Evaluation and results

We evaluate our detector on German Traffic Sign Detection Database [7]. Following standard protocol for this database the detector was trained separately for three super-class categories: *prohibitive* signs, *mandatory* signs and *danger* signs. Learning of first stage, namely domain-specific adaptation of edge-boxes, was performed the same as in [16]. The second stage was trained on output of the first, with best 2000 candidate regions retained for each training image. The remaining regions were split into a positive and negative sample set, using regions with groudtruth overlap of 0.8 or more for positives and groundtruth overlap of 0.4 or less for negatives. Remaining regions were discarded. Additionally, each positive sample was duplicated three times with random jittering, while in negative samples only 100 samples per image were retained for training. Altogether, around 60 000 training images were used to train all three categories.

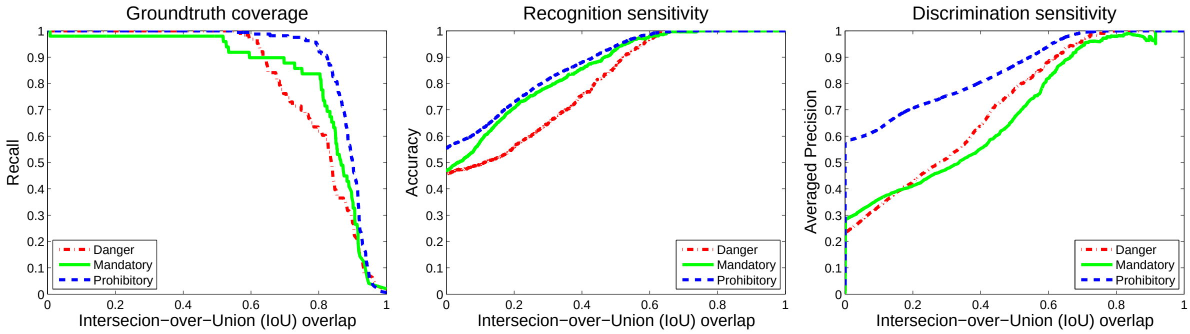

Figure 2: Region quality assessment with three metrics evaluated over varying overlap: (a) recall for groundtruth coverage, (b) accuracy for recognition sensitivity and (c) averaged precision for discrimination sensitivity. Bases on last two graphs the optimal overlap is between 0.7 and 0.8 or higher.

3.1 Detector performance

We evaluate the whole pipeline using precision-recall metrics, where we report averaged precision (AP) for each category. Results with different types of features and color-spaces are shown in Table 1. Among different feature types LBP has consistently performs the best on all three categories, with AP of 0.987 for *prohibitive* signs, 0.862 for *mandatory* sings and 0.846 for *danger* signs. Compared to other feature types, they perform similarly for *prohibitive* signs with only 1-2 % worse performance, while classifiers for *mandatory* and *danger* signs perform by around 5-10 % worse. Among different colors-spaces there is no consistent winner. LBP features perform best with RGB color-space, HOG performs best with LUV while ICF performs best with YCBCR.

3.2 Region quality assessment

We further evaluate the quality of region proposals with respect to the used classifier to asses the impact of region proposals on the second stage classifier. We define a quality of a region as an intersection-over-union (IoU) overlap with the groundtruth object, where highest quality regions have 100% overlap with the object.

We propose to asses the region quality effect on classifier with three important metrics. All three metrics are computed with respect to the IoU overlap to produce the metrics-vs-overlap graph and each overlap value/point in this graph is considered as a threshold and only detections with the overlap above this threshold are used to calculate the metric value at that point. The following metrics are used:

Recall metric: captures the percentage of groundtruth objects that are covered with the region proposals. A similar metric is reported in [16].

***Accuracy metric*:** captures recognition sensitivity of a classifiers to correctly identify object region when overlap changes. This metric will reveal how low can the region overlap with the object be before the classifier losses its ability to recognize it.

***Averaged precision metric*:** captures discrimination sensitivity of a classifier to correctly discriminate between background regions and object regions. We measure this metric by combining thresholded detections at each overlap point with all background detections i.e. with detections that do not overlap with any groundtruth object at all. Combined detections are sorted by classifier score and averaged precision is calculated as area under the precision-recall curve.

We evaluate region quality only with LBP feature and RGB color-space classifier. All three metrics are reported in Figure 2.

For *prohibitive* category (blue dashed line), the metrics reveal that all groundtruth objects are covered very well, with recall dropping by only 10% even at high quality regions with overlap of 0.8. The sensitivity of classifier to low quality regions also appears fairly small, as both accuracy and averaged precision can be maintained at 100% rate for overlap of 0.6 for former and for overlap of 0.7 for latter.

For *danger* and *mandatory* categories all three metrics reveal the reason for poor performance. Focusing on *mandatory* category (green full line), the averaged precision drops quickly, between overlaps of 0.8 and 0.9. The drop is not significant but it points to poor ability to discriminate between mandatory sings and background signs. First graph in Figure 2 also reveals a quick drop of recall by 10% at overlap of 0.5. However, it remains at around 80% until overlap of 0.8, which accounts for slightly increased overall averaged precision of the whole classifier compared to the *danger* signs. On the other hand with category of *danger* signs (red doted line) the classifier does not have any problems with recognition and background discrimination, as both metrics can be maintained at 100% rate for at least overlap of 0.8 for averaged precision and 0.6 for accuracy. The problem is revealed in the recall metric, which shows a good detection rate only for overlaps of 0.6 or less, while it drops to 60% at overlaps of 0.8.

4 Conclusion

In this work we have presented a full pipeline evaluation of traffic sign detection and recognition algorithm as proposed in [16]. The pipeline consist of two stages: (a) region proposal algorithm with domain-specific adaptation of edge-boxes and (b) object classification with feature extraction and classification. We evaluated several different feature types for the second stage, while also varying several color-spaces. LBP [13] feature with RGB color-space performed the best in this case. Furthermore, we focused on assessment of the quality of region proposals and their effect on the second stage classification performance. We assessed region quality using three different metrics with respect to the region quality (i.e. overlap) to capture: (a) coverage of the region proposals (recall metric), (b) sensitivity to recognition when overlap changes (accuracy metrics) and (c) sensitivity to discriminative ability of background regions (averaged precision). Based on the presented metrics we have shown that the classifier starts losing performance with overlaps of 0.6 or less and that overlaps of around 0.7 and 0.8 or higher are needed to achieve the best results. This confirms the values of 0.5 for low and 0.8 for high quality regions defined that were originally arbitrarily set in [16].

In our future work we will explore two possible paths for improving performance of our detector. We will consider introducing several improvements into region proposals to increase the recall rate at higher quality regions, while we will also consider using stronger second stage classifier, such as convolutional neural networks, which could be less sensitive to lower quality regions.

Acknowledgments. This work was supported in part by ARRS research program P2-0214 and ARRS research project L2-6765.

References

[1] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari. Measuring the objectness of image windows. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2189–202, November 2012.

[2] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. In *Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[3] P Dollár, R Appel, and W Kienzle. Crosstalk cascades for frame-rate pedestrian detection. In *European Conference on Computer Vision*, pages 1–14, 2012.

[4] P Dollár and CL Zitnick. Structured forests for fast edge detection. In *International Conference on Computer Vision*, pages 1841–1848, 2013.

[5] Re Fan, Kw Chang, and Cj Hsieh. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning*, 9(2008):1871–1874, 2008.

[6] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.

[7] Sebastian Houben, Johannes Stallkamp, Jan Salmen, Marc Schlipsing, and Christian Igel. Detection of traffic signs in real-world images: The German traffic sign detection benchmark. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. Ieee, August 2013.

[8] Jesmin F. Khan, Sharif M. a. Bhuiyan, and Reza R. Adhami. Image Segmentation and Shape Analysis for Road-Sign Detection. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):83–96, March 2011.

[9] Ming Liang, Mingyi Yuan, Xiaolin Hu, Jianmin Li, and Huaping Liu. Traffic sign detection by ROI extraction and histogram features-based recognition. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. Ieee, August 2013.

[10] KH Lim, KP Seng, and LM Ang. Intra color-shape classification for traffic sign recognition. In *International Computer Symposium*, pages 642–647, 2010.

[11] Markus Mathias, Radu Timofte, Rodrigo Benenson, and Luc Van Gool. Traffic sign recognition - How far are we from the solution? In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. Ieee, August 2013.

[12] Fabien Moutarde, Alexandre Bargeton, Anne Herbin, and Lowik Chanussot. Robust on-vehicle real-time visual detection of American and European speed limit signs, with a modular Traffic Signs Recognition system. *IEEE Intelligent Vehicles Symposium*, 51(33):1122–1126, 2007.

[13] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[14] Carlos Filipe Paulo and Paulo Lobato Correia. Automatic Detection and Classification of Traffic Signs. In *Eighth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '07)*, pages 11–11. Ieee, June 2007.

[15] S Segvic and K Brkic. A computer vision assisted geoinformation inventory for traffic infrastructure. In *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 66–73, 2010.

[16] Domen Tabernik, Rok Mandeljc, Danijel Skočaj, and Matej Kristan. Domain-specific adaptations for region proposals. In *Proceedings of the 20th Computer Vision Winter Workshop*, 2015.

[17] JRR Uijlings and KEA van de Sande. Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154–171, 2013.

[18] Gangyi Wang, Guanghui Ren, Zhilu Wu, Yaqin Zhao, and Lihui Jiang. A robust, coarse-to-fine traffic sign detection method. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–5. Ieee, August 2013.

[19] Woong-jae Won, Minho Lee, and Joon-woo Son. Implementation of road traffic signs detection based on saliency map model. In *2008 IEEE Intelligent Vehicles Symposium*, pages 542–547. Ieee, June 2008.

[20] CL Zitnick and P Dollár. Edge boxes: Locating object proposals from edges. In *European Conference on Computer Vision*, pages 391–405, 2014.