20^{th} *Computer Vision Winter Workshop*
Paul Wohlhart, Vincent Lepetit (eds.)
Seggau, Austria, February 9-11, 2015

Domain-specific adaptations for region proposals

Domen Tabernik, Rok Mandeljc, Danijel Skočaj and Matej Kristan
Faculty of Computer and Information Science, University of Ljubljana, Slovenia
domen.tabernik@fri.uni-lj.si

Abstract. *In this work we propose a novel approach towards the detection of all traffic sign boards. We propose to employ state-of-the-art region proposals as the first step to reduce the initial search space and provide a way to use a strong classifier for a fine-grade classification. We evaluate multiple region proposals on the domain of traffic sign detection and further propose various domain-specific adaptations to improve their performance. We show that edge-boxes with domain-specific learning and re-scoring based on trained shape information are able to significantly outperform remaining methods on German Traffic Sign Database. Furthermore, we show they achieve higher rate of recall with high-quality regions at the lower number of regions than the remaining methods.*

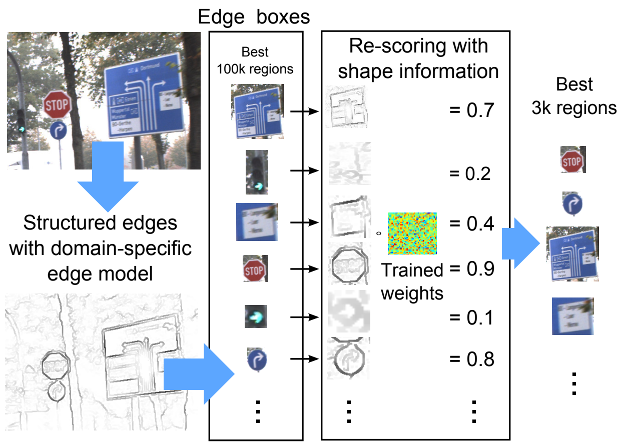

Figure 1: Overview of proposed domain-specific adaptations of edge-boxes using trained structured edges and re-scoring with shape information.

1. Introduction

The problem of detection and recognition of traffic signs has been extensively researched within the field of computer vision [18, 24, 10, 9, 17], with many proposed solutions already being deployed in real-world applications. Such applications are designed mostly for automotive safety and autonomous vehicles, and the main requirements is an excellent detection of only approximately 30 to 50 important traffic sign categories. Out of more than 400 categories, current approaches focus mostly on speed limit signs, stop and yield signs, pedestrian crossing signs and various prohibitory and mandatory signs, while information signs and direction signs are ignored.

Detection of all signs may not be crucial for automotive applications, however, they are important in road maintenance services [21], 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 2). Extending the detection to the remaining signs is desired, as it would eliminate the tedious work of manual verification. Additionally, remaining signs may also be used in current applications of autonomous vehicles to augment the navigation in case of poor GPS signal. Our work is focused on providing a way towards the detection of all traffic signs by utilizing a fast and general regions proposal algorithms. However, due to the lack of a comprehensive database with such traffic signs, we currently focus only on 40 basic categories contained in the existing datasets.

Specific combination of colors and mathematically well-defined shapes makes traffic signs stand out from the background. Several approaches utilize this information by manually hand-crafting detectors to special colors and shapes [10], and fine tuned the algorithms to them. Hand-crafted features rely on simple techniques, such as color thresholding [19, 15], Hough transform [18] and template matching [15, 14], making them fairly efficient. A downside of hand-crafted approaches are difficulty to scale to potentially very large number of cate-

Figure 2: Examples of traffic signs required for the process of road maintenance.

gories and lack of robustness to real-world changes, where traffic signs are frequently occluded by shadows, trees, vehicles, people or other road signs.

Recent methods improved robustness by relying on machine learning. Liang et al. [14] use SVM to focus on important colors for three main classes from GTSDB [9] dataset and then apply template matching to find the specific shapes. They further use HOG features and SVM with RBF kernel to classify objects. However, they still rely on hand-crafted templates to find interesting regions. This makes extension to the remaining traffic signs difficult as they are of various shapes, sizes and colors. Other state-of-the-art methods avoid hand-crafted features and utilize HOG features to achieve best results [9]. Mathias et al. [17] use integral channel features by Dollar et al. [5] for quick detection and further analyses different discriminative learning approaches of HOG features to refine the object classification. Similarly, Wang et al. [23] find coarse locations in the first stage with LDA classifier and improves accuracy in the second stage using SVM. HOG feature are used in both stages, however, low-resolution features are applied in the first stage and high-resolution in the second. While all approaches with HOG features produce state-of-the-art results [9] they cannot be easily extended to large number of traffic sign categories without creating separate models for each category.

One approach to detect the remaining traffic signs would be to focus on the distinctive color distribution separating all road traffic signs of various shapes and colors from the remaining background object. Following the inspiration of bottom-up visual attention inspired by biological systems various methods used salient region detection to reduce the initial search space to interesting regions [24, 12, 15]. Different approaches were employed to focus on the specific color distribution of traffic signs, ranging from simple thresholding of color values in color-opponent channels [15], to computing saliency map by clustering the color space with Gaussian Mixture Model and calculating per-pixel value distances [12], or to utilizing Phase Spectrum of Quaternion Fourier Transform (PQFT) with additional motion features [13].

Recently, in the field of object detection an increasing interest has been shown in development of new methods that find regions with fully enclosed visual objects [8, 2, 22, 25, 4]. Powerful, but slow, object classification algorithms, such as convolutional neural networks [11], cannot be used in exhaustive search using sliding windows. Instead, they employ pre-processing step to find region proposals, i.e., a small set of regions that may contain objects, and perform classification only on them. Novel approaches where developed with some still relying on sliding windows but using quick computation of objectness measure using single [25, 4] or multiple cues [1], while others utilized hierarchical clustering of segmented regions [22] to generate windows. Their design makes them interesting for limiting the search space in traffic sign detection. As they are class-agnostic they should be able to detect road traffic signs of various sizes, shapes and colors included in various traffic signs. Moreover, they are designed for efficiency and can be employed only once for all categories, therefore, amortizing the computational cost over all categories.

1.1. Our approach and contributions

In this paper we propose to use the region proposal methods to move towards the detection of all road-based traffic signs, including information and various direction signs. We propose a multi-stage approach with region proposals in the first level of cascade to significantly reduce the search space and allow a more powerful but slower classifier to be later used for the fine classification. This paper represents a preliminary work towards that goal and focuses on region proposal part of the cascade. We analyze various region proposal and evaluate how successfully they can be applied to the specific domain of traffic sign detection. Multiple state-of-the-art region proposals are evaluated: Objecness measure [1], a selective search [22], BING [4] and edge-boxes [25].

Furthermore, since none of the evaluated region proposals is able to produce results good enough to enable the whole pipeline to compete with the state-of-the-art traffic sign detectors, we present domain-specific adaptations as our main contribution of this work. Out of multiple domain-specific adaptations evaluated, we propose to use a cascade with domain-specific learning of edge-boxes and additional re-scoring based on learning of shape information with linear SVM (see, Figure 1). We show that domain-specific adaptation improves both the accuracy and the quality of region proposals for traffic signs. Although this method is applied to traffic sign detection, it does not use hand-crafted features that limit the method to this specific domain and may be easily applied to various other domains.

The paper is structured as follows: in Section 2 state-of-the-art region proposal algorithms are described, followed by proposed domain-specific adaptations in Section 3 . In Section 4 both state-of-the-art region proposals and domain-specific adaptations are evaluated. Conclusions are drawn in Section 5.

2. Region proposal algorithms

This section provides an overview of various state-of-the-art region proposals. For a comprehensive overview of region proposals see Houben et al. [9].

2.1. Window objectness

This early region proposal algorithm was proposed by Alexe et al. [2]. The algorithm is based on a fast evaluation of sliding windows to quickly reduce the search space of potential objects. Windows are evaluated using an objectness measure that integrates multiple weak cues. It utilizes saliency cue computed from the residual spectra of the FFT, additionally modified to bias larger windows and applied to multiple scales. The second cue, color contrast, measures the dissimilarity of the window compared to its immediate surrounding. The measure utilizes color histogram of LAB channels and computes Chi-squared distance between the window and its surrounding. The third cue captures edge density and computes the share of the edges found at the borders compared to ones at the window's center. Canny detector is used to detect the edges. The last cue measures closed boundary characteristics of the object by using superpixel straddling. Since superpixels will over-segment the object there will be a small probability that window containing the object will break the superpixel. All four measures are complementary to each other and are best integrated using Naive Bayes model.

2.2. Selective search

The approach proposed by Uijlings et al. [22] clusters individual pixels into object hypotheses using hierarchical grouping. Bottom-up approach enables objects to group from smaller regions up to bigger regions as they are being grouped together higher in the layers. This captures objects at different scales without sliding windows. Due to hierarchical segmentation, the approach favors objects with homogeneous regions. This may be well suited for traffic signs where homogeneous regions with one or two main colors are normally present in the center.

Hierarchical clustering uses segmentation by Felzenszwalb and Huttenlocher [7] to obtain initial regions and merging of two regions is performed when they are the most similar. Similarity between them is computed from four complementary measures. First measure is defined as a sum of differences between their normalized color histograms, where color histogram is created from three quantized channels. The second measure utilizes texture information by histogramming quantized edge orientations for each channel individually. The third measure computes similarity based on region sizes to encourage the merging of small regions as early in the hierarchy as possible. The last measure checks how well the two regions fit each other in order to avoid regions with holes and irregular shapes. All four measures can be efficiently propagated through the hierarchy to enable fast computation.

In [22] different strategies of combining all four measures are considered. Out of eight different color channels considered (HSV, LAB, RGB, normalized RGB, intensity and individual color channels) HSV channels performed the best. Out of different possible ways to combine similarity measure, combining all four also performed the best. We consider only HSV and all four similarity measures combined in our evaluation. Authors also evaluated combining multiple strategies together, using different color channels, combination of similarities and parameters for segmentation. However, combining multiple strategies can take more than 10-times longer as each strategy has to be run individually, thus taking significantly longer to compute. In our evaluation we consider only one strategy as our database already contains high-resolution images that take more time to process.

2.3. BING

Authors of BING [4] propose to capture objectness using the 64D norm (i.e. magnitudes) of the gradients feature. The method is based on the finding that stand-alone objects with well-defined borders and centers have a clear correlation in normed gradient space, particularly when objects are resized to small fixed sizes. The method proposes to normalize the size of object to multiple quantized sizes and collect a feature vector containing 8x8 norm of the gradients. Linear SVM is further used to find the set of weights that capture windows with fully enclosed object. In the second stage of learning, a linear SVM is used to calibrate the scores from different window sizes and to suppress the sizes that have low probability of containing an object.

The method is applied to densely sampled windows using sliding window approach and handle hundreds of thousands of windows with an efficient computation of feature vector using binarized normed gradients.

2.4. Edge boxes

Region proposals by Zitnick et al. [25] are based on a realization that a window with one fully enclosed object does not have many strong edges at the borders. In particular, strong edges rarely intersect with the borders as this would be an indication that window breaks the main outline of the object. The method computes a score based on this premise by first grouping edges into small segments, ensuring that group's sum of orientation differences does not exceed $\pi/2$. The score of the window is then computed as a weighted sum of magnitudes of segments that intersect with the borders of the window. The magnitudes are weighted based on how much of a segment lies outside of the window. Additionally, magnitudes within the window center are ignored as only border edges have been shown to be important. The authors propose an efficient way to compute this score using integral images.

The edges used by this method are extracted using structured edges by Dollar et al. [6]. This can reduce the presence of noisy edges as structured edges can be learned from the object borders.

3. Domain-specific adaptations

The domain to which we are applying the region proposals is very specific. The colors of traffic signs are designed to be very distinctive and the signs contain many homogeneous regions that are designed to pop out from the background. While the shape of traffic signs can vary, they are still designed around a small set of shapes, such as triangles, squares or circles, that fairly well separate traffic signs from the background objects. In this section we detail several proposed adaptations of region proposals that can exploit the specific characteristics of our target domain.

3.1. Saliency-based region proposals

We evaluate two region proposals generated from salient regions. Salient regions can be often present in traffic signs, particularly in the center of the sign, where homogeneous regions with single color are prominent. We consider region proposals generated by two salient region detectors: MSER [16] and WaDe key-point detector [20].

3.2. Selective search with SEED superpixels

In the selective search [22], the size of the smallest region detected is determined by the size of the initial segmentation segments. Since many traffic signs are only 20-30 pixels wide, the size of initial segments is even more important for this domain than for generic objects. We consider replacing the segmentation with the SEED superpixels [3] to obtain finer initial regions.

3.3. Domain-specific learning of BING

Many region proposals rely on a learning stage that is normally performed on a generic set of classes. We propose to utilize region proposal learning procedure to capture the visual characteristic of our target domain.

In BING [4] learning is performed on gradient features that are resized to specific sizes and aspect ratios. Particularly, window normalization is important in our domain as it will normalize traffic signs of different sizes, such as different information and directions board, to a specific size. Moreover, learning will capture rectangular, circular or triangular shape structures predominantly present in all traffic signs.

3.4. Domain-specific learning of edge-boxes

We propose two improvements to the traffic sign proposals for edge-boxes [25]. First, we can implement an adaptation of edge-boxes in a similar fashion as with BING [4] and use its own learning procedure to capture domain specific characteristics. Secondly, we propose to run region proposal algorithm in a cascade, with edge-boxes providing a bigger set of regions in the first stage, and using re-scoring with the shape information to further reduce the set of interesting regions. In the first stage of the cascade edge-boxes is set to return 10 - 20% of best region proposal. Results show (see, Figure 3) that at this

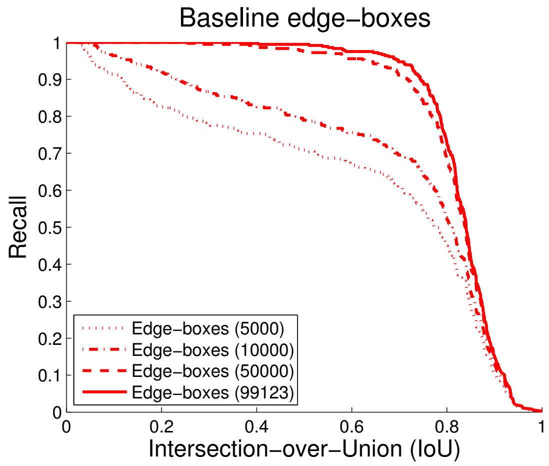

Figure 3: Performance improvements on edge-boxes when increasing the number of regions to 100.000.

range the positive samples are covered with fairly good windows. Note that due to high resolution images in our dataset and a small size of target objects, sliding window generates 250k to 500k regions, thus first stage with edge-boxes reduces a set of interesting regions to 50k - 100k regions.

3.4.1 Learning

Although edge-boxes do not perform any explicit learning, they rely on structured edges by Dollar et al. [6] that are normally trained on generic object boundaries. We train structured edges on traffic signs boundaries and allow the method to focus on borders around specific color distributions. A similar information is usually captured in various hand-crafted traffic sign detectors, however, those methods have edge detectors tuned to specific color-opponent channels [15, 13]. Instead of tuning to specific color-opponents we allow structured edges to learn which color channels are the most appropriate to find the borders of traffic signs. We trained structured edges on first 100 images from GTSDB and have manually segmented their boundaries to provide groundtruth for structured edges.

3.4.2 Re-scoring with shape information

We propose to use shape information for the re-scoring. By default, trained structured edges capture shape information fairly well. However, this information is not fully utilized in edge-boxes as the method focuses only on edges around the borders that lead out of the window, while ignoring the central edges that carry shape information.

We also perform normalization of window to specific size as windows with uniform size are invariant to changes in sizes, aspect ratios and also small degrees of rotation. Invariance to the aspect ratio is important in our domain with various rectangular boards, such as directions signs, city limits signs or information signs, which appear in multiple sizes and aspect ratios. We use simplified norming of windows by simply resizing them to specific size.

We propose the following procedure to capture shape information. Region proposals are resized to a smaller size, specifically we use 40x40 pixels that can capture enough shape information. Next, we obtain edges for each region and create feature vector from them. We can reuse domain-specific structured edges from edge-boxes and avoid additional computational cost. Feature vector is created directly from structured edges using both edge magnitudes and orientations, thus producing 3200 dimensional vector. Finally, linear SVM is trained to separate between traffic sign and non-traffic sign regions. Due to linear implementation of SVM, classification can be implemented as a dot-product between feature vector and a vector of weights.

4. Evaluation

We evaluate region proposal methods on German Traffic Sign Database [9], which contains 600 testing and 300 training images taken from vehicle mounted camera in city and countryside settings. All images have 1360x800 pixels and depict 43 different annotated traffic signs. All algorithms were tested on the testing set, while the training set was used only by the methods that require domain-specific adaptations. Baseline methods that require learning are trained on a generic dataset.

The standard evaluation of region proposals focuses on several metrics: recall, which represent the ratio of positive samples detected, the number of all regions returned by the method and the intersection-over-union (IoU) of the detected regions. The last measure is important since it captures the quality of the region proposals. Regions that cover object with only low IoU will introduce an error that propagates onwards. To capture both performance and quality of the region proposals we measure (a) recall versus IoU and (b) recall versus the number of regions proposed.

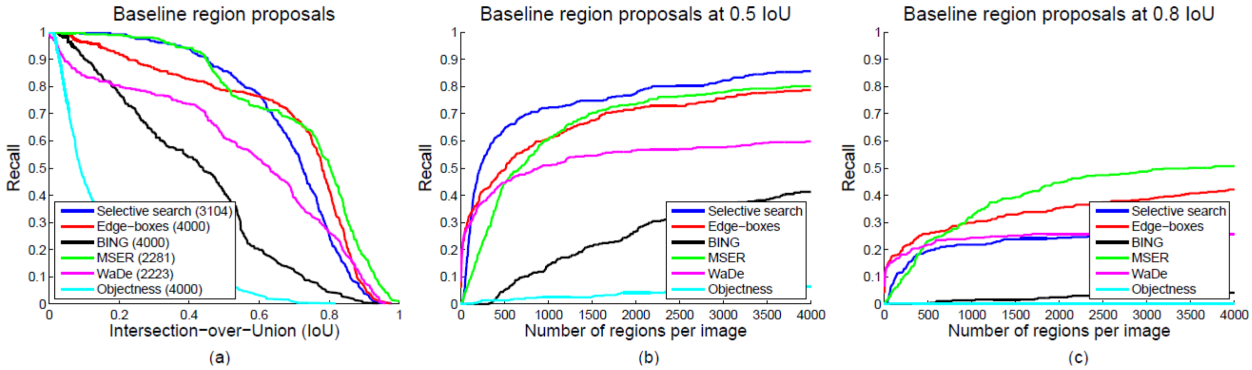

Figure 4: Results of evaluating baseline region proposals on GTSDB [9] dataset with recall over various Intersection-over-Union overlaps in (a) and recall over various number of regions in (b) for 0.5 IoU and in (c) for 0.8 IoU overlap. Note, values next to legend names in (a) are the number of regions used.

We compute the recall versus IoU measure by sorting the detection based on best IoU and thresholding IoU at various points. To ensure valid comparison between different methods the number of regions have to be fixed. For some algorithms this may not be easily achievable, however, where this is possible we set the parameters accordingly. In practice, the selective search variants produced at most approximately 3000 to 4000 regions. We adjusted the parameters of remaining methods to closely match this number. Note that standard region proposal evaluations consider only 1000 regions, however, they typically evaluate 4-times smaller images. Image samples in other evaluations also contain larger objects that are mostly present in the foreground, while GTSDB contains many small objects that are often barely visible. We account for this discrepancy by taking more than 3000 regions.

We compute recall versus the number of proposed regions by sorting the region proposals based on their score and limiting the number of regions. Note that this measure is not fully appropriate for MSER and WaDe detectors as they do not return any score. We measure recall versus the number of regions at two IoU values. One at 0.5 based on PASCAL overlap criteria and another at 0.8 to capture high-quality regions.

4.1. Baseline region proposals

Results of four baseline state-of-the-art region proposals and two salient region detectors are shown in Figure 4. Three methods, namely selective search [22], edge-boxes [25] and MSER [16], performed similarly. MSER covers most positive samples at low and high quality regions, while selective search is competitive at mid-quality regions and edge-boxes are competitive at high-quality regions. The selective search appears to perform best only at IoU of approximately 0.5 where it outperforms both MSER and edge-boxes. On the other hand, MSER performs the best at higher-quality regions, which are more important for successful classification.

More than half of the traffic signs are still not covered by any of the high-quality region proposals. This makes region proposals difficult to compete with the state-of-the-art traffic sign detectors that achieve 98 to 100% detection rate on this database [17, 23]. Both MSER and selective search have a difficulty at competing as they achieve 99% coverage at only 0.2 IoU. However, edge-boxes can achieve better coverage when enough regions are generated as 98 - 100% of samples can be covered with high-quality regions when 100k regions are generated (see, Figure 3). With MSER and selective search this is not possible as they both generate a fixed number of windows depending either on salient regions found or depending on the number and quality of initial segmentation segments.

Note that in our evaluation the objectness measure performed the worst, mainly due to a poor set of initial windows. The method constructs a dense set of initial windows, however, the implementation [1] has difficulties generating smaller windows that cover smaller traffic signs, therefore, we excluded this method from further evaluation.

4.1.1 Finer resolution

We additionally evaluate proposals at finer resolution by doubling the size of each input image. Finer reso-

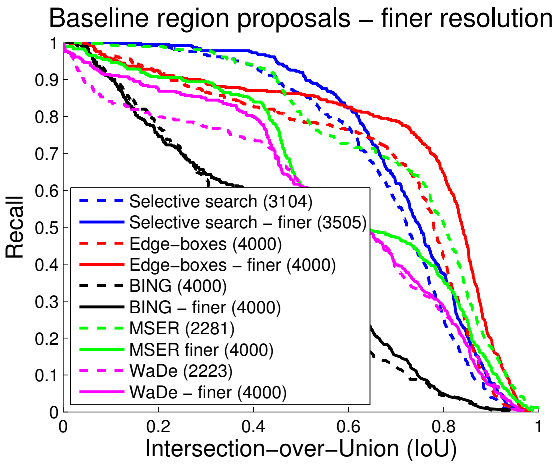

Figure 5: Comparing baseline region proposals at double the resolution. Note, values next to legend names are the number of regions used.

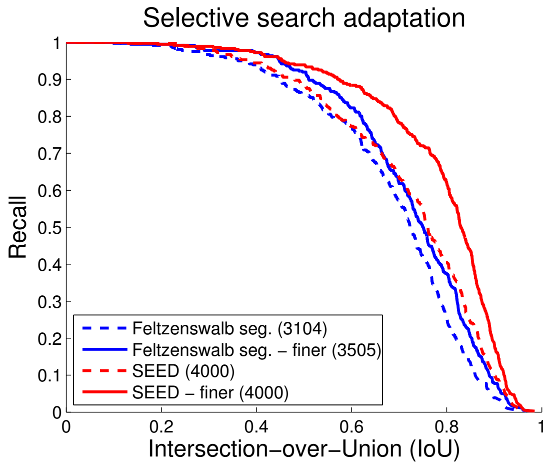

Figure 6: Results of evaluating selective search [22] using SEED [3] superpixels instead of Felzenszwalb and Huttenlocher [7] segmentation. Note, values next to legend names in are the number of regions used.

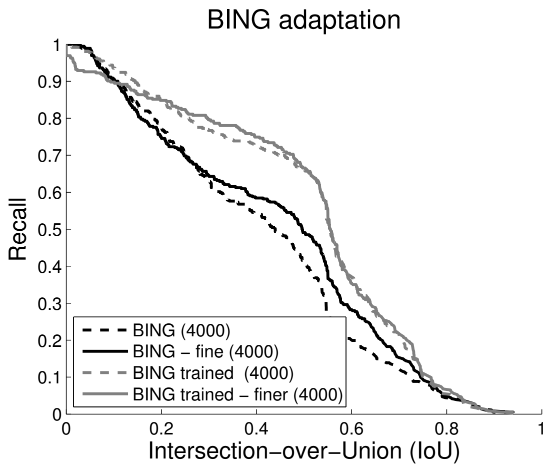

Figure 7: Results of evaluating domain-specific adaptation of BING [4] with gradient features trained on traffic signs. Note, values next to legend names are the number of regions used.

lution is better suited for our domain due to relatively small size of traffic signs. The results are shown in Figure 5, where improvements in almost all proposals can be observed. High resolution images improve edge-boxes the most, while the performance of MSER actually worsens. This happens due to a higher number of salient regions being returned, but as only the first 4000 of regions are selected to fairly evaluate all algorithms, some correct regions are discarded.

Result at the finer resolution also need to be normalized with the additional computational cost. Selective search is already slow at normal resolution, therefore, using finer resolution makes it even less useful. Higher resolution has little computational cost for BING, however, this method has the worst recall. The highest benefit is observed in edge-boxes, where multi-scale edge detection can be replaced with a single finer scale at little computational cost and a significant improvement in the performance.

4.2. SEED superpixels in selective search

We further evaluate replacing segmentation in selective search with SEED [3] superpixels to generate a higher number of regions. The results can be seen in Figure 6. A finer control over the size of initial segmentation when using SEED superpixels generates windows that capture smaller regions and improves overall performance. The performance is improved even further with finer resolution, matching the performance of edge-boxes. However, this improvement comes at a higher computational cost compared to edge-boxes, thus making selective search less attractive.

4.3. BING adaptation

Next, we evaluate the effect of domain-specific adaptation of BING with the results reported in Figure 7. The graph shows improved performance when training BING features on traffic signs over all IoU, with the highest improvement observed at the low quality regions. Despite improving the overall performance, the results are still significantly worse than in selective search or edge-boxes. The reason for

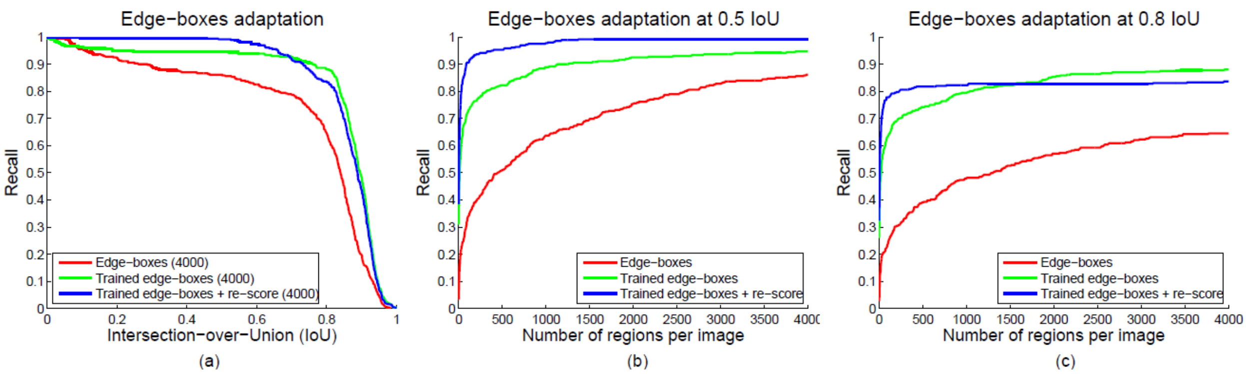

Figure 8: Results of evaluating domain-specific adaptation of edge-boxes [25] using domain-specific learning and re-scoring with shape information. Recall versus Intersection-over-Union overlaps is in (a) and recall versus number of regions in (b) for 0.5 IoU and in (c) for 0.8 IoU overlap. Note, values next to legend names in (a) are the number of regions used.

poor performance is a low resolution gradient feature that cannot sufficiently capture enough details in our domain.

4.4. Edge-boxes adaptation

With the final experiment we evaluate the effects of domain specific-adaptation applied to edge-boxes [25] as proposed in Section 3.4. The results can be observed in Figure 8. Both our proposed adaptations have proven to significantly boost the performance of region proposals, achieving recall of 0.99 at 0.5 IoU overlap and 0.90 at 0.8 IoU overlap. Learning structured edges alone is already able to capture 30% more traffic signs compared to generic structured edges. Moreover, all 90% of traffic signs are covered with high-quality regions with IoU over 0.8. Adding re-scoring with shape information further improves the region proposal, as almost 100% of traffic signs can be covered with 0.5 IoU overlap.

Additionally, an excellent performance can be achieved at a small number of windows, as can be observed in Figure 8. At both 0.5 and 0.8 IoU overlap the recall quickly converges to 0.9, requiring only between 1000 and 2000 region proposals to achieve this score.

5. Conclusion

In this paper we explored multiple region proposals in the context of traffic sign detection. We proposed to use region proposals as a first step in detection of all traffic signs to reduce the initial search space to a promising set of regions. Multiple state-of-the-art region proposals were evaluated: Objecness measure [1], selective search [22], BING [4] and edge-boxes [25]. To further increase the performance we proposed additional improvements in a form of domain-specific adaptation. Multiple adaptations where evaluated: two salient region detectors, MSER [16] and WaDe [20], replacing segmentation in selective search [22] with SEED superpixels [3], learning BING [4] feature on traffic signs and proposing domain-specific learning of edge-boxes [25] with re-scoring. The latter proved to be the most effective. We performed learning of edge-boxes by training structured edges on traffic signs, while for re-scoring we captured shape information with the magnitudes and orientations of structured edges and used linear SVM to learn the specific shape information. We showed that proposed method captures 99% of traffic signs on GTSDB [9], with 90% of objects covered with a high-quality regions. Furthermore, our proposed approach does not use hand-crafted features and is general enough to be applied to other domains as well.

In future, we will further extend the cascade using re-scoring based on trained color information. We will also evaluate the whole pipeline and explore the effects of region quality on various classifiers. We are also planing on assembling a new dataset containing traffic signs with additional categories, including direction signs, information signs and various road marking signs.

References

[1] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In *CVPR*, pages 73–80, 2010.

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

[2] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2189–202, Nov. 2012.

[3] M. V. D. Bergh, X. Boix, and G. Roig. SEEDS: Superpixels extracted via energy-driven sampling. *European Conference on Computer Vision*, pages 1–19, 2012.

[4] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr. BING: Binarized Normed Gradients for Objectness Estimation at 300fps. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3286–3293, June 2014.

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

[6] P. Dollár and C. Zitnick. Structured forests for fast edge detection. *International Conference on Computer Vision*, 2013.

[7] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2):167–181, Sept. 2004.

[8] J. Hosang, R. Benenson, and B. Schiele. How good are detection proposals, really? *British Machine Vision Conference*, pages 1–25, June 2014.

[9] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel. Detection of traffic signs in real-world images: The German traffic sign detection benchmark. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Aug. 2013.

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

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012.

[12] A. C. Le Ngo, L.-M. Ang, K. P. Seng, and G. Qiu. Colour-based bottom-up saliency for traffic sign detection. *2010 International Conference on Computer Applications and Industrial Electronics*, (Iccaie):453–457, Dec. 2010.

[13] C. Li, W. Song, L. Xiao, Y. Hu, and X. Pan. Salient traffic sign video detection based on hypercomplex frequency domain. *Proceedings of the 33rd Chinese Control Conference*, pages 7379–7382, July 2014.

[14] M. Liang, M. Yuan, X. Hu, J. Li, and H. Liu. Traffic sign detection by ROI extraction and histogram features-based recognition. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Aug. 2013.

[15] K. Lim, K. Seng, and L. Ang. Intra color-shape classification for traffic sign recognition. *Computer Symposium (ICS), 2010*, 2010.

[16] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, Sept. 2004.

[17] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool. Traffic sign recognition How far are we from the solution? *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Aug. 2013.

[18] F. Moutarde, A. Bargeton, A. Herbin, and L. 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.

[19] C. F. Paulo and P. L. Correia. Automatic Detection and Classification of Traffic Signs. *Eighth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '07)*, pages 11–11, June 2007.

[20] S. Salti, A. Lanza, and L. Di Stefano. Keypoints from symmetries by wave propagation. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2898–2905, June 2013.

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

[22] J. Uijlings and K. van de Sande. Selective search for object recognition. *International Journal of Computer Vision*, 2013.

[23] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang. A robust, coarse-to-fine traffic sign detection method. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–5, Aug. 2013.

[24] W.-j. Won, M. Lee, and J.-w. Son. Implementation of road traffic signs detection based on saliency map model. *2008 IEEE Intelligent Vehicles Symposium*, pages 542–547, June 2008.

[25] C. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. *European Conference on Computer Vision*, 2014.