21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Learning to Detect Traffic Signs: Comparative Evaluation of Synthetic and Real-world Datasets

Andreas Møgelmose
VAP, AAU & CVRR, UCSD
am@create.aau.dk

Mohan M. Trivedi
CVRR Lab, UC San Diego
mtrivedi@ucsd.edu

Thomas B. Moeslund
VAP Lab, Aalborg University
tbm@create.aau.dk

Abstract

This study compares the performance of sign detection based on synthetic training data to the performance of detection based on real-world training images. Viola-Jones detectors are created for 4 different traffic signs with both synthetic and real data, and varying numbers of training samples. The detectors are tested and compared. The result is that while others have successfully used synthetic training data in a classification context, it does not seem to be a good solution for detection. Even when the synthetic data covers a large part of the parameter space, it still performs significantly worse than real-world data.

1 Motivation

With the emergence of more advanced sensors embedded in cars, the field of Traffic Sign Recognition (TSR) has seen increasing interest over the last decade. TSR systems can be used in a number of scenarios, ranging from Driver Assistance Systems (DASs) - as described in [14] - to fully autonomous cars.

Many sign detection systems (see section 2) rely on large amounts of training data to work. Over the past two years, a few traffic sign datasets has shown up: The GTSRB dataset [12, 13], the Swedish Traffic Signs Dataset [6], and the KUL Belgium Traffic Signs Dataset. A commonality among these datasets is that they contain European Signs conforming to the Vienna Convention. Since signs differ from region to region and in many cases from country to country, an interesting proposition is to use synthetically generated training data, saving a lot of time and effort in gathering the data. Synthetic training data has not yet been widely used in the field of TSR, but is worth researching since very few datasets from outside of Europe exist. A recent survey [9] shows that research on the detection and recognition of traffic signs outside countries conforming to the Vienna Convertion on traffic signs is lacking in general. This paper investigates if using synthetic data for the detection of traffic signs is feasible.

The role and importance of high quality, representative datasets in the development of TSR systems cannot be overemphasized. Collection of such datasets is an expensive task (in time as well as effort). Issues in training, annotations in the real-world, and semi-supervised learning for object recognition is treated in [11]. Since signs have a well-defined appearance, the idea of using synthetic data emerge. The use of synthetic training in sign detection is not yet widespread, prompting this paper. Our paper is focused closely on the generation of synthetic training data for detection purposes. It is also the first of its kind dealing with US signs. In [4, 3], generation of synthetic data specifically for classification is investigated. In [10], some aspects of detecting non-US signs with synthetic data is discussed. The detection task is somewhat harder the classification due to the lack of knowledge about whether a sign is present, where it is, and what size it has.

The following section briefly covers the general workings of TSR systems, followed by a section on how we generate synthetic training data. Towards the end of the paper, the performance of synthetic training data is compared to the performance of real-world training data when used to train a simple AdaBoost cascade with Haar-like features [15].

2 TSR: General approaches

Overviews of TSR can be found in [9, 8, 2]. TSR can be split into two main stages: Sign detection and sign classification, as seen in fig. 1. Not all detection approaches require training as such, since they are using a theoretical model of the sign, based on e.g. the shape. With that said, many papers present Machine Learning (ML) based approaches. In [1], an AdaBoost Cascade similar to the one used in this paper was used, albeit on

978-4-9906441-1-6 ©2012 IAPR

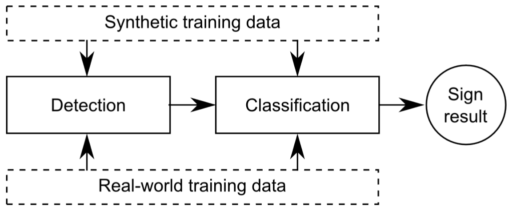


Figure 1: Flow for ML-based TSR-systems. The stages can be trained with synthetic or real-world data, and two stages does not have to be trained with the same type.

specific color channels. In [5], the image is segmented with a HSI threshold and then classifies the resulting blobs using a linear Support Vector Machine (SVM) on Distance to Bounding Box (DtB) features. DtB features are measurements of the distance between the edge of the blob and its rectangular bounding box.

3 Synthetic training data for detection

The question this paper tries to answer is: Can we substitute real-world training data with synthetic in ML based sign detection systems? The idea is to generate synthetic training images from a drawn template. Template examples can be seen in fig. 2.

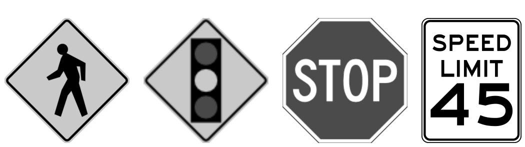
The goal is to emulate how signs of the given type might look on pictures from the real world. In order to do this, several transformations are made randomly to the template:

Hue variations emulates faded signs and color casts due to lighting of the natural scene. Done by adding to/subtracting from the hue-parameter in the HSV color space.

Lighting variations emulates shadows and variations in exposure. Done by adding to/subtracting from the value-parameter in the HSV color space.

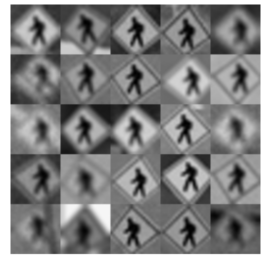
Rotations around the x-, y-, and z-axis with the origin in the center of the template. Emulates signs captured from different perspectives.

Backgrounds taken from a real image are added to the template. This emulates the various backgrounds a sign might have in real life.

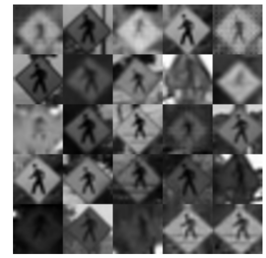


(a) Pedestrian crossing (b) Signal ahead (c) Stop (d) Speed limit

Figure 2: Examples of typical US sign templates.



(a) Synthetic training images. Template in fig. 2a



(b) Real-world training images.

(c) Synthetic training images. Template in fig. 2c

(d) Real-world training images.

Figure 3: Samples from the training image sets.

Gaussian blur is added to emulate an unfocused camera. It should be noted that Gaussian blur does not really emulate the bokeh produced by an unfocused lens, but emulating bokeh properly is not a straightforward task, and it would likely not give any notable detection benefit.

Gaussian noise to emulate sensor noise.

Occlusions are added in the form of tree branches growing in front of some signs.

Each transformation should be applied with a random parameter within some realistic boundaries. Samples of training images can be seen in fig. 3.

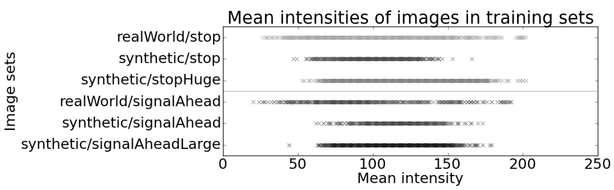
To evaluate whether the synthetic datasets cover the same variance in appearance as the real-world data, we compare the distributions in intensity- and blur-values among training sets. In fig. 4a a plot of the mean of the intensities in the training images is shown. Each point in the plot is a single image. Data for the detectors of two different signs is shown. In a few sets, the intensity span does not match, but the large 5000 image stop sign set is similar to real-world data. Another parameter is shown in fig. 4b: Blur. Blur is calculated as

$$B = \frac{1}{n}\sum_{i=0}^{n} e_i \tag{1}$$

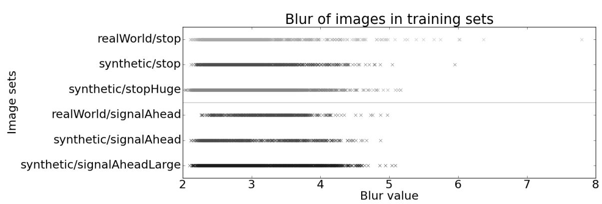
where B is the blur-value, n is the number of vertical edges in an image and e_i is the edge width of a specific edge pixel, given as the distance between the pixels

Table 1: Results of the comparative evaluations of detectors

Training type	Training images (positive/negative)	Stages	Signs to find	TP	FP	FN
Stop						
Real-world	1218/2500	20	103	76 (73.8%)	11	27
Real-world	1686/3000	20	103	75 (72.8%)	8	28
Synthetic	1218/2500	17	103	18 (17.5%)	2	85
Synthetic	5000/10000	19	103	26 (25.2%)	5	77
Synthetic	1218/2500	10	103	60 (58.3%)	1500	43
Pedestrian crossing						
Real-world	364/800	20	40	29 (72.5%)	10	11
Real-world	1044/2000	20	40	30 (75%)	2	10
Synthetic	364/800	14	40	11 (27.5%)	28	29
Speed limit 35						
Real-world	253/500	20	21	15 (71.4%)	1	6
Synthetic	253/500	7	21	5 (23.8%)	32	16
Synthetic	2000/4000	7	21	6 (28.6%)	6	15
Signal ahead						
Real-world	597/1500	20	56	42 (75%)	10	14
Real-world	859/2000	20	56	38 (67.9%)	4	18
Synthetic	597/1500	13	56	14 (25%)	117	49
Synthetic	2000/4000	13	56	16 (28.6%)	53	48



(a)



(b)

Figure 4: Distribution of two parameters in the training sets.

with the local maximum and minimum intensity around the edge pixel. The measure is described further in [7]. This shows that the blur variance is covered well by the synthetic data.

4 Comparative evaluation

To compare the synthetic training data to training data obtained from real footage, a simple Viola-Jones based detector [15] was trained for the four sign types illustrated in fig. 2. The choice of detection algorithm is not crucial, as the purpose of this paper is not to find a perfect traffic sign detector, but rather look at the relative differences between detectors trained with synthetic and real-world images. It was trained with an image size of 20x20 pixels in all cases, except for the rectangular speed limit sign, trained with 18x24 pixels.

The detectors created with various numbers of training images was tested on a set of real-world images, collected from cars in conjunction with this lab's research. The results can be seen in table 1.

With all signs, the real-world data performs significantly better than the synthetic data. Providing more training data in the synthetic case does help, but even a large increase (more than a doubling) of the training data does not make the synthetic data perform comparably to the real-word data. All detectors were trained with a target of 20 stages, but some terminated earlier due to a sufficiently good fit to the training data, and others were lowered to give better detection performance at the cost of more false positives. It is indeed possible for the synthetic detector to find more true signs, but at a huge cost in false positives, and still not as good as the real-world detector.

Even in the cases (like the stop sign detector with 5000/10000 training images) where the synthetic data spans nearly the same space as the real-world detector, the synthetic detector fails to achieve a detection rate anywhere near the real-world data.

5 Concluding remarks

We discussed a research study to assess the feasibility of using carefully synthesized training datasets for developing traffic sign detectors. In this research, output from a synthetic training generator has been used to train a stock AdaBoost cascade and its performance compared with real-world training images. The real-world training data consistently performs significantly better than the synthetic training data, even in cases where the synthetic data seems to span a similar set of appearances. This leads to the conclusion that there is simply no substitute for real-world images in the case of detection.

An ML-approach to setting the synthetic data generation parameters would be a logical place to go from here, if further study of synthetic data for detection is desired. It is also possible that the system could benefit from further transformations to the template image, such as motion blur. Other works have shown promising results in using synthetic training data for classification of signs. An interesting direction of research could be to explore hybrid (real and synthetic) datasets for TSR approaches.

References

[1] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, pages 255–260. IEEE, 2005.

[2] H. Fleyeh and M. Dougherty. Road and traffic sign detection and recognition. In *10th EWGT Meeting and 16th Mini-EURO Conference*, pages 644–653, 2005.

[3] H. Hoessler, C. Wöhler, F. Lindner, and U. Kreßel. Classifier training based on synthetically generated samples. In *Proceedings of 5th international conference on computer vision systems. Bielefeld, Germany*, 2007.

[4] H. Ishida, T. Takahashi, I. Ide, Y. Mekada, and H. Murase. Identification of degraded traffic sign symbols by a generative learning method. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 531–534. IEEE, 2006.

[5] S. Lafuente-Arroyo, S. Salcedo-Sanz, S. Maldonado-Bascón, J. A. Portilla-Figueras, and R. J. López-Sastre. A decision support system for the automatic management of keep-clear signs based on support vector machines and geographic information systems. *Expert Syst. Appl.*, 37:767–773, January 2010.

[6] F. Larsson and M. Felsberg. Using fourier descriptors and spatial models for traffic sign recognition. *Image Analysis*, pages 238–249, 2011.

[7] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. A no-reference perceptual blur metric. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages III–57–III–60 vol.3, 2002.

[8] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund. Traffic Sign Detection and Analysis: Recent Studies and Emerging Trends. In *15th IEEE International Conference on Intelligent Transportation Systems*, 2012.

[9] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund. Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey. *IEEE Intelligent Transportation Systems Transactions and Magazine*, Special Issue on MLFTSR, dec 2012.

[10] G. Overett, L. Tychsen-Smith, L. Petersson, N. Pettersson, and L. Andersson. Creating robust high-throughput traffic sign detectors using centre-surround HOG statistics. *Machine Vision and Applications*, pages 1–14, 2011.

[11] S. Sivaraman and M. M. Trivedi. A General Active Learning Framework for On-road Vehicle Recognition and Tracking. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):267–276, June 2010.

[12] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 1453–1460. IEEE, 2011.

[13] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*, (0):–, 2012.

[14] M.M. Trivedi and S.Y. Cheng. Holistic sensing and active displays for intelligent driver support systems. *Computer*, 40(5):60–68, 2007.

[15] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2001.