The German Traffic Sign Recognition Benchmark: A multi-class classification competition

Johannes Stallkamp, Marc Schlipsing, Jan Salmen
Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum, Germany
{johannes.stallkamp, marc.schlipsing, jan.salmen}@ini.rub.de

Christian Igel
Department of Computer Science
University of Copenhagen
2100 Copenhagen, Denmark
igel@diku.dk

***Abstract*— The "German Traffic Sign Recognition Benchmark" is a multi-category classification competition held at IJCNN 2011. Automatic recognition of traffic signs is required in advanced driver assistance systems and constitutes a challenging real-world computer vision and pattern recognition problem. A comprehensive, lifelike dataset of more than 50,000 traffic sign images has been collected. It reflects the strong variations in visual appearance of signs due to distance, illumination, weather conditions, partial occlusions, and rotations. The images are complemented by several precomputed feature sets to allow for applying machine learning algorithms without background knowledge in image processing. The dataset comprises 43 classes with unbalanced class frequencies. Participants have to classify two test sets of more than 12,500 images each. Here, the results on the first of these sets, which was used in the first evaluation stage of the two-fold challenge, are reported. The methods employed by the participants who achieved the best results are briefly described and compared to human traffic sign recognition performance and baseline results.**

I. Introduction

Recognition of traffic signs is a challenging real-world problem of high industrial relevance. Although commercial systems have reached the market and several studies on this topic have been published, systematic unbiased comparisons of approaches are missing and comprehensive benchmark datasets are not freely available. Sign recognition is a multi-category classification problem with unbalanced class frequencies. Traffic signs show a wide range of variations between classes in terms of color, shape, and the presence of pictograms or text. However, there exist subsets of classes (e.g., speed limit signs) that are very similar to each other. The classifier has to cope with large variations in visual appearances due to illumination changes, partial occlusions, rotations, weather conditions, scaling, etc.

Traffic signs are designed to be easily detected and recognized by human drivers. Accordingly, humans are capable of recognizing the large variety of existing road signs with close to 100 % correctness. This does not only apply to real-world driving, which provides both context and multiple views of a single traffic sign, but also to the recognition from single, cut-out images.

We present the *German Traffic Sign Recognition Benchmark (GTSRB)*, a large, lifelike dataset of more than 50,000 traffic sign images in 43 classes. We describe the design and analysis of the IJCNN 2011 competition of the same name that was built upon this dataset. We conducted experiments to determine human traffic sign recognition performance and compare them to the competition results. The competition is held in two stages, and the first stage has just finished at the time of this document's writing. We asked the participants who achieved the best results so far to provide brief descriptions of their methods, which are presented together with the classification accuracies.

The paper is organized as follows: Sec. II presents related work. Sec. III provides details about the benchmark dataset. Sec. IV addresses the competition protocol. Finally, the competition results are reported and the so far best methods are described in Sec. V before the conclusions in Sec. VI.

II. Related work

Several approaches to traffic sign recogntion have been published. In [2], an integrated system for speed limit detection, tracking, and recognition is presented. The classifier is trained using 4,000 samples of 23 classes, with samples per class ranging from 30 to 600. The individual performance of the classification component is evaluated on a training set of 1,700 traffic sign images with a correct classification rate of 94 %.

Moutarde et al. present a system for recognition of European and U. S. speed limit signs based on single digit recognition [3] using a neural network. Unfortunately, they do not provide individual classification results. The overall sytem including detection and tracking achieves a performance of of 89 % for U. S. and 90 % for European speed limits, respectively, on 281 traffic signs.

Broggi et al. [4] use several neural networks to classify different traffic signs. Shape and color information from the detection stage is used to select the appropriate neural network. Only qualitative results are provided.

In [5], a number-based speed limit classifier is trained on 2,880 images. It achieves a correct classification rate of $92.4\,\%$ on 1,233 images. However, it is not clear whether images of the same traffic sign instance are shared between sets.

Various approaches are compared on a dataset containing 1,300 preprocessed examples from 6 classes (5 speed limits and 1 noise class) in [6]. The best classification performance observed was $97\,\%$.

In [7], a classification performance of $95.5\,\%$ is achieved using support vector machines. The database comprises an

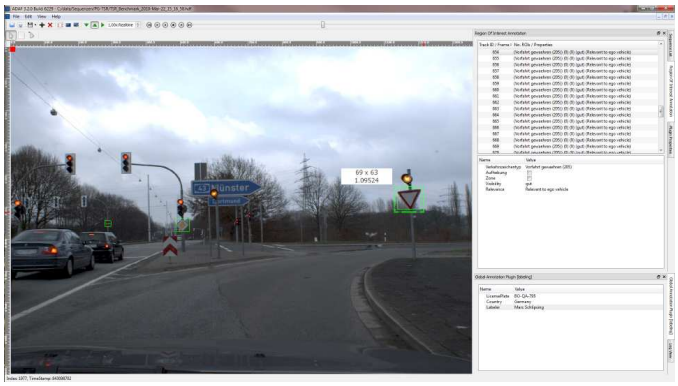


Fig. 1. Screenshot of the annotation

Fig. 2. A single traffic sign track

impressive number of ∼36,000 Spanish traffic sign samples of 193 sign classes. However, it is not clear whether the training and test sets can be assumed to be independent, as the random split only took care of maintaining the distribution of traffic sign classes (see Sec. III). To our knowledge, this database is not publicly available.

III. Dataset

A. Data collection

The dataset was created from approx. 10 hours of video that was recorded while driving on different road types in Germany during daytime. The sequences were recorded in March, October and November 2010. For data collection, a *Prosilica GC1380CH* camera was used with automatic exposure control and a frame rate of 25 fps. The camera images have a resolution of 1360×1024 pixels. The video sequences are stored in raw *Bayer*-pattern format, but extracted traffic sign images are converted to *RGB* color images [8].

Data collection and manual annotation was performed using *NISYS Advanced Development and Analysis Framework*[1] (see Fig. 1).

We will use the term *traffic sign instance* to refer to a physical real-world traffic sign in order to discriminate against *traffic sign images* which are captured when passing the traffic sign by car. The sequence of images originating from one traffic sign instance will be referred to as *track*. Each instance is unique. In other words, the dataset only contains a single track for each physical traffic sign.

From approx. 133,000 labelled traffic sign images of 2,416 traffic sign instances in 70 classes, the GTSRB dataset was compiled according to following criteria:

1) Discard tracks with less than 30 images.
2) Discard classes with less than 9 tracks.
3) For the remaining tracks: If the track contains more than 30 images, equidistantly sample 30 images.

Step 3 was performed for two reasons. First of all, the number of traffic sign images per track was very different as it strongly depends on the velocity with which the car passed the sign. Since subsequent images of a slowly passed traffic sign are very similar to each other, these images do not contribute to the diversity of the dataset. On the contrary, it causes an undesired imbalance of dependent images. Secondly, in spite of the first point, the visual appearance of a traffic sign does vary over time. Far away traffic signs result in low resolution while closer ones are prone to motion blur. The illumination may change, and the motion of the car affects the perspective with respect to occlusions. Fig. 2 provides an example. Selecting a fixed number of images per traffic sign increases the diversity of the dataset and also avoids an imbalance by strongly varying numbers of nearly identical images.

The selection procedure outlined above reduced the number of images to approx. 50,000 images of the 43 classes that are shown in Fig. 3. The relative class frequencies of the classes are shown in Fig. 4.

The set contains images of more than 1,700 traffic sign instances. The size of the traffic signs varies between 15×15 and 222×193 pixels. The images contain 10 % margin (at least 5 pixels) around the traffic sign to allow for the usage of edge detectors. The original size and location of the ROI of the traffic sign is preserved in the provided annotations. The images are not necessarily squared.

For the purpose of the competition, the dataset was split into three subsets. Set I was published as training data, Set II as test data for the online competition. Both sets may be used as training data for the final competition which will be performed on Set III (unpublished until then). Set I contains approx. 50 %, sets II and III approx. 25 % of the images each. The split was performed randomly, class-wise, and on track level, to make sure that 1) the class distribution is preserved and 2) all images of one traffic sign instance are assigned to the same set. Each of the test sets is consecutively numbered and shuffled to prevent deduction of class membership from other images of the same track. In contrast, the training set preserves the temporal structure of the images, which could be exploited by approaches capable of using privileged information [9].

[1] `http://www.nisys.de`

Fig. 3. Traffic sign classes

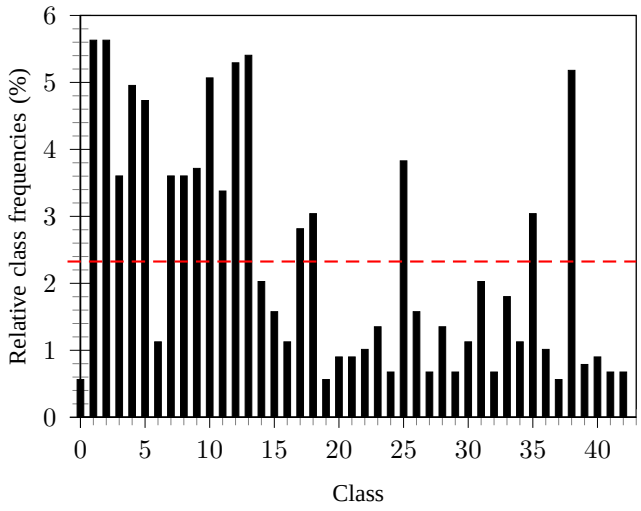


Fig. 4. Relative class frequencies in the dataset

B. Pre-calculated features

To allow scientists without a background in image processing to participate, all three sets are provided with pre-calculated feature sets. The following features are included:

1) HOG features: Three sets of differently configured HOG features (histograms of oriented gradients) [10] are provided. To compute them, the images were scaled to a size of 40×40 pixel and converted to grayscale. The sets contain feature vectors of length 1568, 1568, and 2916 respectively.

2) Haar-like features: This feature set was intended to allow participants to apply feature selection methods if desired. Just like for HOG features, images were rescaled to 40×40 and converted to grayscale. We computed 5 different types in different sizes for a total of 11,584 features per image.

3) Color histograms: This set of features was provided to complement the gradient-based feature sets with color information. It contains a global histogram of the hue values in HSV color space, resulting in 256 features per image.

IV. Competition

The competition uses the dataset presented in Sec. III. It consists of two evaluation phases. This paper focuses on the first one that was performed in the run-up to IJCNN 2011. This evaluation used Set I for training and Set II for testing.

A. Competition protocol

Participants had to classify individual images of the test set. The performance was evaluated based on the 0/1 loss.

The training set was published seven weeks before the first evaluation. This initial evaluation was designed as an online competition. At the beginning of the evaluation, the test set was provided to the participants. Results were uploaded as CSV file to the competition website[2] for evaluation. The number of submissions was (initially) not limited (see Sec. IV-C for details), to allow participating teams to submit results for different approaches.

Since the test set contains images, participants were theoretically able to manually annotate the samples with the correct class ID. Although restricted to only 3 days, the short time frame of the evaluation phase could not *guarantee* that cheating would not occur. Therefore, a second evaluation with fresh data will be held as live competition at IJCNN 2011.

To allow more thorough training of the classifiers, the class IDs for the test set have been published after the online competition. Furthermore, this mitigates any advantages a team may achieve for the final competition by investing the efforts of manual annotation.

B. Submission website

The website allows participants to upload their result files and get immediate feedback about their performance. During the online competition, results were instantly published in a public leaderboard.

After the submission deadline, some result analysis features were activated. The participants could get a more detailed

[2] `http://benchmark.ini.rub.de`

insight into their results by investigating the confusion matrix and the list of misclassified images for each of their own submissions.

We intend to introduce a second leaderboard based on the final test set after the final competition. This ranking will then be permanently open for submissions. Users will get immediate feedback about their performance after upload, but the results will not automatically be publicly visible. In order to publish results, users have to provide publication details about their approach.

C. Flaws in challenge protocol

As far as the online competition is concerned, the missing submission limit turned out to be problematic. A few participants started flooding the leaderboard with results. For some submissions, the method description did not even allow for discrimination of the methods (either because it was too cryptic or because it was the same name for all submissions only extended with running numbers). We assume the major difference between such submissions to be parameter adjustments. However, optimization w. r. t. the test set causes overfitting and biases the results. In order to protect the other teams from this misbehavior, we had to introduce a submission limit during the online competition. To avoid (or at least mitigate) penalizing teams with only a couple of submissions, we set the limit to ten submissions. This allowed most teams to submit at least one more final result. For future competitions, we would set a limit of three to five submissions and would perhaps not show the exact ranking during the submission phase.

V. Results

The competition attracted more than 20 teams from all around the world. A wide range of state-of-the-art machine learning methods was employed, including (but not limited to) several kinds of neural networks, support vector machines, linear discriminant analysis, subspace analysis, ensemble classifiers, slow feature analysis, kd-trees, and random forests. We present the results of the four best-performing teams in addition to results of baseline algorithms and an experiment to determine human traffic sign recognition performance. The results that are reported in this section are summarized in Tab. I. This table is limited to the top four teams and their characteristic methods. Details about these methods can be found in Sec. V-C. Our results are shown with team name *INI-RTCV*. The complete result table is available at the competition website.

A. Baseline

We report three kinds of baseline results: Linear discriminant analysis (LDA) on HOG features, k-nearest neighbor (k-NN) on HOG features and human performance. The LDA is based on the implementation in the Shark Machine Learning Library[3] [11]. Nearest neighbor results were computed on all HOG feature sets for 1-NN and 3-NN using l_2-distance.

[3]`http://shark-project.sourceforge.net`

TABLE I

Result overview. ID denotes the submission ID to identify the result in the leaderboard at the competition website.

CCR (%)	Team	Method	ID
98.98	IDSIA	cnn_hog3	197
98.97	sermanet	EBLearn 2LConvNet ms 108 feats	178
99.81	INI-RTCV	Human Performance	199
97.88	VISICS	IKSVM + PHOG + HOG2	183
97.35	VISICS	SRC + LDAs I/HOG1/HOG2	184
96.87	noob	HOG + LDA + VQ	84
	...	...	
96.32	INI-RTCV	HOG features (Set 2) + LDA	2
94.73	INI-RTCV	HOG features (Set 3) + LDA	3
94.51	INI-RTCV	HOG features (Set 1) + LDA	1
	...	...	
73.89	INI-RTCV	HOG 1 + 3-NN	7
73.82	INI-RTCV	HOG 3 + 3-NN	9
73.82	INI-RTCV	HOG 3 + 1-NN	6
73.65	INI-RTCV	HOG 1 + 1-NN	4
72.81	INI-RTCV	HOG 2 + 1-NN	5
72.81	INI-RTCV	HOG 2 + 3-NN	8

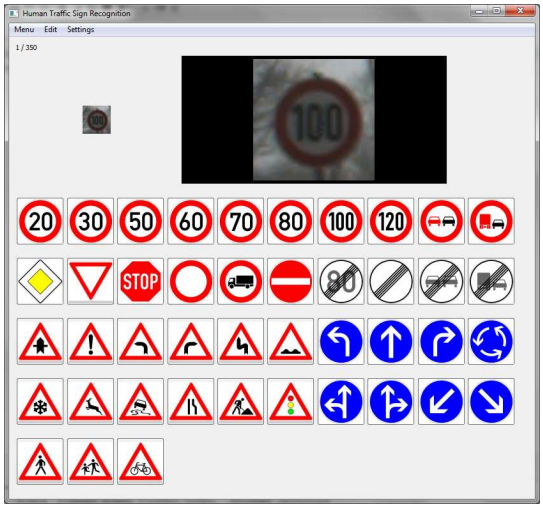


Fig. 5. Test application to determine human performance

B. Human performance

To determine the human traffic sign recognition performance on isolated images, the test set was presented in chunks of 350 randomly chosen images to 36 test persons. Over all subjects, each image was presented exactly once for classification. Each image was presented in two resolutions (see Fig. 5) — the original resolution of the image and scaled to a height of 190 pixels to improve readability of small images. The black border around the scaled image was chosen to improve contrast perception for dark and low-contrast samples. The test person assigned a class ID by clicking the corresponding button.

C. Top-ranking methods

This subsection provides an overview of the best-performing methods in the competition. The method descriptions are authored by the participants themselves. They are ordered according to their ranking in the competition.

1) Team IDSIA: Team *IDSIA* consists of Dan Ciresan, Ueli Meier, Jonathan Masci and Jürgen Schmidhuber from IDSIA, USI, SUPSI, Switzerland[4].

a) Committee of CNN and MLP: Our approach uses a flexible, high-performance GPU implementation of a convolutional neural network (CNN). We improve the performance of a single CNN by forming a committee that also includes a multilayer perceptron (MLP) trained on the provided features.

The architecture of a CNN is characterized by many building blocks set by trial and error, but also constrained by the data. In most studies a fixed, handcrafted architecture is used to perform the experiments. With respect to other implementations of similar neural network architectures on GPUs [12], [13] that are hard-coded to satisfy the hardware constraints of the GPUs, our implementation [14] is flexible and fully on-line (i.e. weight updates after each image). As subsampling layers we use max-pooling layers which are crucial for invariant object recognition. CNNs with a max-pooling layer consistently outperform conventional nets [15].

All CNNs have seven hidden layers. The output layer has 43 neurons, one for each class.

We select the ROI of the original images and resize it to 48×48 pixels. The contrast of each image is normalized independently. We try different contrast normalization methods. The best one proved to be histogram equalization.

We use a system with a Core i7-920 (2.66GHz), 12 GB DDR3 and four GTX 580 graphics cards. The implemented CNN has a plain feed-forward architecture trained by on-line gradient descent. We split the provided training set in training and validation sets and train various architectures. The best architecture is then trained on all images from the training set. Weights are initialized from a uniformly random distribution. Each neuron's activation function is a scaled hyperbolic tangent.

After having trained all the individual CNNs and MLPs, we form various committees. The MLPs have 1 hidden layer with 200 hidden units and are trained in batch mode using second order information. Individual MLPs perform worse than CNNs. Being trained on features, however, they offer an additional source of information and might correctly classify images misclassified by the CNN. Since both CNNs and MLPs produce output class probabilities, we can easily average the corresponding neuron's outputs. This averaging results in a slight performance boost, and allows us to obtain the best result with a committee of a CNN and an MLP trained on HOG features (HOG_03).

More details concerning this approach can be found in [16].

[4]{dan, ueli, jonathan, juergen}@idsia.ch

2) Team sermanet: Team *sermanet* consists of Pierre Sermanet and Yann LeCun from Courant Institute of Mathematical Sciences at New York University, United States[5].

a) Convolutional Neural Networks: Convolutional Networks (ConvNets) [17] are a biologically-inspired architecture that can learn invariant features. While traditional vision methods use hand-crafted features such as HOG, ConvNets actually learn each feature extraction stage. Features can therefore be optimized for a given task and learned without prior knowledge for any new modality where our lack of intuition makes it difficult to engineer good features. Multiple stages of features extraction provide hierarchical and robust representations to a multi-layer classifier. Each stage is composed of convolutions, non-linearities and subsampling. Non-linearities used in traditional ConvNets are the $tanh()$ sigmoid function. However more sophisticated non-linearities such as the rectified sigmoid and the subtractive and divisive local normalizations are used here, enforcing competition between neighboring features (both spatially and feature-wise). Outputs taken from multiple stages can also be combined to enrich features fed to the classifier with a multi-scale component. We use the C++ open-source implementation of ConvNets called EBLearn[6] [18]. This architecture was trained by full supervision of the (colored) traffic sign dataset (using 32×32 raw images) and reached 98.97 % accuracy during the first phase of the competition. It is interesting to note that superior networks have since then been obtained without the use of color information (fully described in [19]).

3) Team VISICS: Team *VISICS* consists of Radu Timofte and Luc van Gool from ESAT-PSI-VISICS/IBBT at the Katholieke Universiteit Leuven, Belgium[7].

a) IK-SVM based method: The method employs a fast Intersection Kernel Support Vector Machine (IK-SVM) [20] over concatenated HOG features. We used computed pyramidal HOG features over resized 28×28 pixels patches using the same settings used in [20] for handwritten digits classification. These were concatenated with the HOG Set 2, as provided by GTSRB, giving a $2172 + 1568$ dimensional feature space. We trained 43 one-against all models (one for each class) and the classification decision was taken by picking the class corresponding to the best estimated probability in the models' outputs. While running the classifiers over the testing data is relatively fast, in order of minutes, the time spent for training is big, over 15 hours. More details about choices made and the overall systems are to be found in [21].

b) l_1-minimization based method: This is a sparse representation-based classification (SRC) inspired by the increasingly popular field of compressed sensing (CS). The testing query samples are assumed to be recovered (with a very low error) as a linear combination of the sufficiently large set of training samples. Furthermore, the combination weights corresponding to the training samples from the same

[5]{sermanet,yann}@cs.nyu.edu

[6]http://eblearn.sf.net

[7]{Radu.Timofte, Luc.VanGool}@esat.kuleuven.be

class as the query sample to recover tend to be large (in l_1-norm sense). In an ideal case the remaining weights are zero. This is a sparse linear combination, with about $\frac{1}{C}$ nonzeroes, where C is the number of classes. We are interested in this sparse vector of weights which we can obtain by solving a l_1-minimization problem formulated as in [22]. We use the Homotopy solver [23] stopped after reaching a sparse support of less than 20 nonzeroes. In our challenge entries we do not use the *cross-and-bouquet* model which deals explicitly with noise, heavy corruption, occlusion in the query sample. As basic features we use HOG Sets 1 and 2 (as provided by GTSRB), and the raw grayscale pixel values (I). The features are projected using the obtained direction vectors by applying Linear Discriminant Analysis (LDA) method. Thus, we work on low, 42-dimensional spaces and benefit from the discriminant power of LDA based on the training labels. For each type of features we separately compute LDA projection matrices. The final used representation for the top scoring SRC method is a concatenation of the LDA projections of each type of features (I, HOG Set 1 and HOG Set 2). The concatenated features were normalized by l_2-norm. The running time was about two hours on a single core. More details about choices made and the overall systems are to be found in [21].

4) Team noob: Team *noob* consists of Nhat Vo[8], Subhash Challa[9] and Bill Moran[10] from University of Melbourne, Australia, and Duc Vo[11] from NICTA, Australia.

a) Discriminant Analysis on HOG features and Vector Quantization: The proposed idea is based on histograms of oriented gradients (HOG), linear discriminant analysis (LDA), and vector quantization (VQ). HOG is used to capture local object appearance and shape within traffic sign images, followed by LDA. To further improve the recognition rate and recognition speed, we apply VQ on projected samples to remove outliers or bad samples in training set. A recognition rate of 96.87 % was obtained in the competition. This whole algorithm called HOG+LDA+VQ can briefly be summarized as follows:

- HOG feature vectors are extracted from training images. We use precalculated HOG2 features provided with GTSRB dataset.
- LDA is then performed on these HOG2 features to find discriminative projections. All training HOG features will be projected on this projection to form discriminative projected features.
- VQ by k-means algorithm is performed on projected features of each class to find some representatives or codebooks which are used as templates in recognition stage. By doing this, we can remove outliers or bad training data, highly speed up recognition time and improve the performance.

More details concerning this approach can be found in [24].

[8]n.vo@pgrad.unimelb.edu.au
[9]subhash.challa@nicta.com.au
[10]b.moran@ee.unimelb.edu.au
[11]dvo@nicta.com.au

D. Result analysis

As can be seen in Tab. I, the best performing teams achieved a very high recognition accuracy which is comparable to humans. To gain a deeper insight into the results, the traffic sign classes are grouped into subsets of similar signs according to Fig. 6. The individual results per team and subset are listed in Tab. II. Since both the LDA and the k-NN approaches produced very similar results for the different HOG feature sets, only the best result each is considered. Fig. 7 shows the confusion matrices for the different approaches. The classes are ordered by subsets as defined in Fig. 6a to 6f, from left-to-right and top-to-bottom respectively. The grey lines separate the subsets.

Notably, all solutions — both human and machine — share one similarity, although to a different extent. A clustering of errors in the top-left corner, that is, in the *speed limit* subset, can be observed. Low resolution and motion blur impede the discrimination of the different numbers.

Considering the *other prohibitory* signs (s. Fig. 6b), it is noticeable that the error is generally smaller than for the speed limit signs, although this subset contains two very similar signs as well (*no overtaking* for cars and trucks). However, in case of misclassification, they were usually confused within subsets (a) and (b).

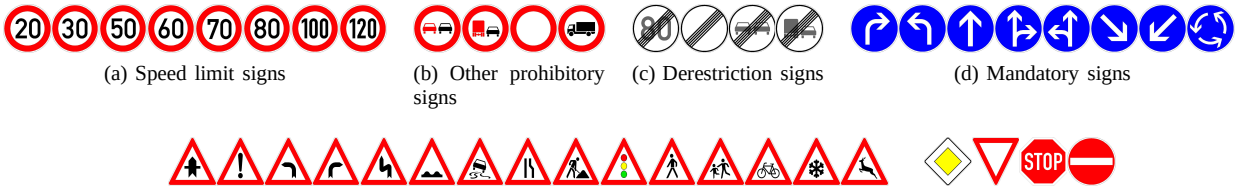
The *derestriction* signs cause little problems. The largest errors are provided by the 3-NN classifier which mostly confuses the derestriction signs among each other.

The blue *mandatory* signs are nearly perfectly recognized by humans. The machine-learned classifiers perform worse. The errors concentrate mostly on the sign classes *roundabout*, *pass on right*, and *pass on left*. The latter two are generally mounted close to the ground which makes them easily accessible. Their readability is often impaired by stickers or spray paint. The fact that they are mostly mistaken for speed limits can be attributed to the use of HOG features — which do not contain any color information — in most algorithmic approaches. Color features were only used by Team *sermanet* which reduces the confusion of the blue mandatory signs with speed limits to a minimum.

For the *danger* signs, a similar observation can be made. The focus on edge features allows classifiers to discriminate the triangular signs from other subsets, but leads to confusion within this group of traffic signs. Obviously, the provided HOG features capture the general sign shape well, but are not discriminative enough to distinguish the different pictograms. The group of human test subjects outperforms most algorithmic approaches. Only the convolutional neural networks achieve a comparable performance.

Finally, the *unique* signs are nearly perfectly classified. Since they are very different in their general shape, even the nearest neighbor approach — which generally only provided moderate accuracy — achieves a very small error rate.

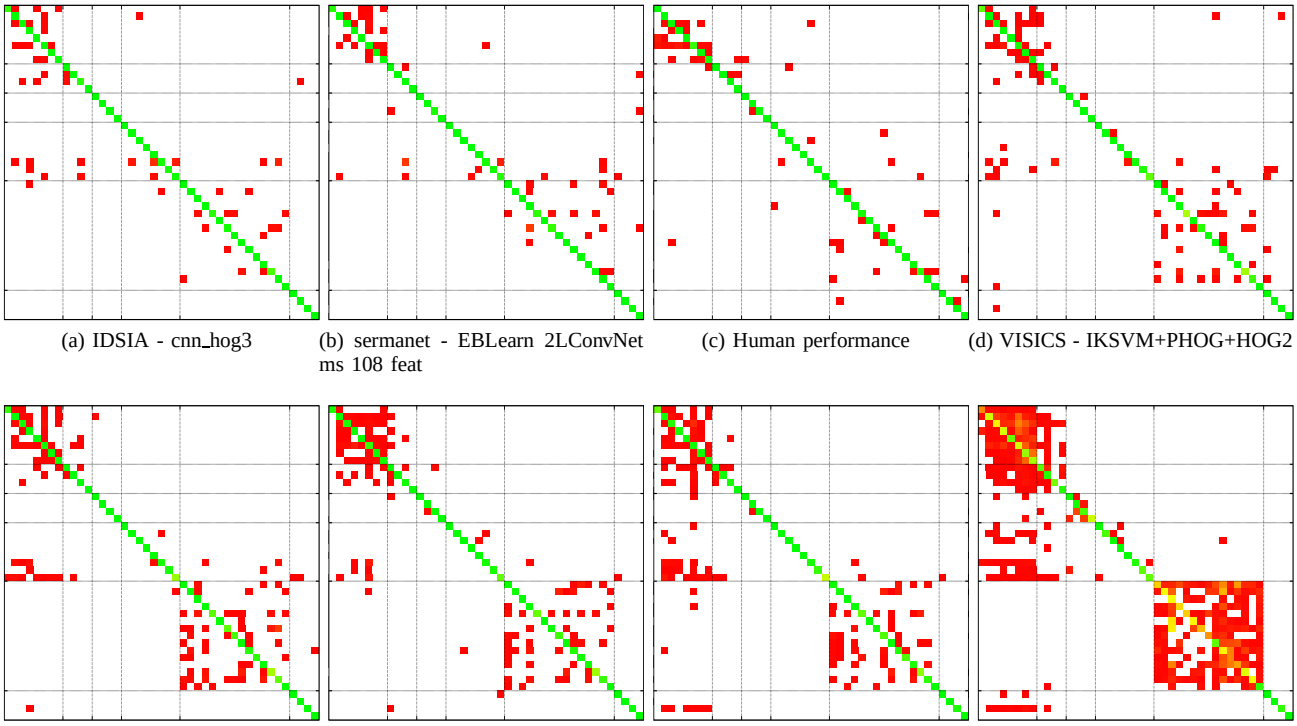
In many cases, the human errors are much more scattered than the algorithmic results. Except for the speed limit subset, most errors visible in the confusion matrices are caused by single misclassifications. These errors can be partially



(a) Speed limit signs (b) Other prohibitory signs (c) Derestriction signs (d) Mandatory signs

(e) Danger signs (f) Unique signs

Fig. 6. Subsets of similar traffic signs



(a) IDSIA - cnn_hog3 (b) sermanet - EBLearn 2LConvNet ms 108 feat (c) Human performance (d) VISICS - IKSVM+PHOG+HOG2

(e) VISICS - SRC+LDAs I/HOG1/2 (f) noob - HOG+LDA+VQ (g) INI-RTCV - HOG2 + LDA (h) INI-RTCV - HOG1 + 3-NN

Fig. 7. Confusion matrices. The grid lines separate the traffic sign subsets defined in Fig. 6. Values in [0,1]; White denotes zero, (0,1] is colored red to yellow to green.

TABLE II

INDIVIDUAL RESULTS FOR SUBSETS OF TRAFFIC SIGNS. BOLD TYPE DENOTES THE BEST RESULT(S) PER SUBSET.

	Speed limits	Other prohibitions	Derestriction	Mandatory	Danger	Unique
cnn_hog3	**99.14**	99.57	**100.00**	97.89	98.83	**100.00**
EBLearn 2LConvNet	98.87	**99.80**	99.00	97.78	98.72	**100.00**
Human Performance	97.39	99.59	99.67	**99.72**	**99.04**	99.90
IKSVM + PHOG + HOG2	97.91	99.25	99.67	96.78	96.17	99.95
SRC + LDAs I/HOG1/HOG2	97.63	99.46	**100.00**	96.05	94.54	99.95
HOG + LDA + VQ	95.73	98.50	99.33	96.72	95.39	99.90
HOG 2 + LDA	95.76	97.28	99.33	95.00	95.00	99.35
HOG 1 + 3-NN	61.39	87.28	87.00	93.39	53.83	98.76

explained by the design of the test application, which directly advanced to the next image after one of the buttons was clicked. Unintended mouse movements and double-clicks could, therefore, easily cause accidental misclassifications. This case was reported by some of the test persons.

VI. Conclusions

We presented the design and analysis of the "German Traffic Sign Recognition Benchmark" dataset and competition. The results of the competition show that state-of-the-art machine learning algorithms perform very well in the challenging task of traffic sign recognition. The participants achieved a very high performance of up to 98.98% correct recognition rate which is comparable to human performance on this dataset. Some of the human error originated from the design of the test application. For the final competition, we are confident that human performance can be "improved" by a few changes to this application to prevent pure accidental misclassifications.

We are looking forward to the final competition at IJCNN 2011 which completes GTSRB competition. This session will use the currently unpublished Set III. After the final session, the complete dataset will be published. We intend to install a new, permanent leaderboard on the competition website which allows for submissions of new results and comparison of new approaches. As many participants relied on the provided HOG features, we are curious to see whether different features can improve the recognition performance. For the future, we plan to add more benchmark tasks and data to the competition website. In particular, we consider to provide a benchmark data set for the detection of traffic signs in full camera images.

VII. Acknowledgements

We thank Lukas Caup, Sebastian Houben, Stefan Tenbült, and Marc Tschentscher for their labelling support, Bastian Petzka for creating the competition website, NISYS GmbH for supplying the data collection and annotation software, and all others that contributed to this competition.

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 *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2005, pp. 255–260.

[2] F. Moutarde, A. Bargeton, A. Herbin, and A. Chanussot, "Robust on-vehicle real-time visual detection of american and european speed limit signs with a modular traffic signs recognition system," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2007, pp. 1122–1126.

[3] A. Broggi, P. Cerri, P. Medici, P. P. Porta, and G. Ghisio, "Real time road signs recognition," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2007, pp. 981–986.

[4] C. G. Keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. Baratoff, "Real-time recognition of U.S. speed signs," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2008, pp. 518–523.

[5] A. S. Muhammad, N. Lavesson, P. Davidsson, and M. Nilsson, "Analysis of speed sign classification algorithms using shape based segmentation of binary images," in *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, 2009, pp. 1220–1227.

[6] S. Maldonado Bascón, J. Acevedo Rodríguez, S. Lafuente Arroyo, A. Caballero, and F. López-Ferreras, "An optimization on pictogram identification for the road-sign recognition task using SVMs," *Computer Vision and Image Understanding*, vol. 114, no. 3, pp. 373–383, 2010.

[7] H. S. Malvar, L. He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *Proceedings of the IEEE International Conference on Speech, Acoustics, and Signal Processing*, 2004, pp. 485–488.

[8] V. Vapnik and A. Vashist, "A new learning paradigm: Learning using privileged information," *Neural Networks*, vol. 22, no. 5-6, pp. 544 – 557, 2009.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[10] C. Igel, T. Glasmachers, and V. Heidrich-Meisner, "Shark," *Journal of Machine Learning Research*, vol. 9, pp. 993–996, 2008.

[11] K. Chellapilla, S. Puri, and P. Simard, "High performance convolutional neural networks for document processing," in *International Workshop on Frontiers in Handwriting Recognition*, 2006.

[12] R. Uetz and S. Behnke, "Large-scale object recognition with CUDA-accelerated hierarchical neural networks," in *IEEE International Conference on Intelligent Computing and Intelligent Systems*, 2009.

[13] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep big simple neural nets for handwritten digit recogntion," *Neural Computation*, vol. 22, no. 12, pp. 3207–3220, 2010.

[14] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *International Conference on Artificial Neural Networks*, 2010.

[15] D. C. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *submitted to International Joint Conference on Neural Networks*, 2011.

[16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, November 1998.

[17] P. Sermanet, K. Kavukcuoglu, and Y. LeCun, "Eblearn: Open-source energy-based learning in c++," in *Proceedings of the International Conference on Tools with Artificial Intelligence (ICTAI'09)*. IEEE, 2009.

[18] P. Sermanet and Y. LeCun, "Convolutional neural networks applied to traffic sign recognition," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN'11)*. IEEE, 2011.

[19] S. Maji and J. Malik, "Fast and accurate digit classification," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-159, Nov 2009.

[20] R. Timofte and L. V. Gool, "Fast approaches to large-scale classification," in *submitted to International Joint Conference on Neural Networks*, 2011.

[21] A. Yang, A. Ganesh, Y. Ma, and S. Sastry, "Fast l_1-minimization algorithms and an application in robust face recognition: A review," in *International Conference on Image Processing*, 2010.

[22] M. S. Asif, "Primal dual pursuit: A homotopy based algorithm for the dantzig selector," Georgia Institute of Technology, 2008.

[23] N. Vo, D. Vo, S. Challa, and B. Moran, "Discriminant analysis on histogram of oriented gradients and vector quantization for traffic sign recognition," in *submitted to International Joint Conference on Neural Networks*, 2011.