Traffic Sign Classifier Adaption
by Semi-supervised Co-training

Matthias Hillebrand[1], Ulrich Kreßel[1], Christian Wöhler[2], and Franz Kummert[3]

[1] Daimler AG, Group Research and Advanced Engineering, 89081 Ulm, Germany
{matthias.hillebrand,ulrich.kressel}@daimler.com
[2] Image Analysis Group, TU Dortmund, 44221 Dortmund, Germany
christian.woehler@tu-dortmund.de
[3] Applied Informatics Group, Bielefeld University, 33615 Bielefeld, Germany
franz@techfak.uni-bielefeld.de

Abstract. The recognition of traffic signs in many state-of-the-art driver assistance systems is performed by statistical pattern classification methods. Traffic signs in European countries share many similarities but also vary in colour, size, and depicted symbols, making it hard to obtain one general classifier with good performance in all countries. Training separate classifiers for each country requires huge amounts of labelled training data. A well-trained classifier for one country can be adapted to other countries by semi-supervised learning methods to perform reasonably well with relatively low requirements regarding labelled training data. Self-training classifiers adapting themselves to unknown domains always risk that the adaption will become ineffective or even fail completely due to the occurrence of incorrectly labelled samples. To assure that self-training classifiers adapt themself correctly, advanced multi-classifier training methods like co-training are applied.

Keywords: self-training, semi-supervised, co-training.

1 Introduction

The automatic recognition of traffic signs by a driver assistance system, a mature but still contemporary field of research, commonly utilises intensity-based classifiers for processing visual sensor information. In classical supervised learning-based approaches, a classifier has to be trained for each country separately, even if the traffic signs in European countries reveal only minor variations regarding colour, font, font size and depicted symbols.

Such a supervised approach is inefficient because it leads to high labelling costs for human annotators, while unlabelled data can often be acquired with justifiable expenditure. Since there are not too many variations between the different countries, an existing classifier for one country can be extended using informative data samples from other countries. An automatic class assignment is desirable to further reduce the labelling costs. In literature, this approach is known as semi-supervised learning.

Individual semi-supervised classifiers rely on their own capabilities to predict labels correctly and learn these predictions to extend their knowledge or rather their classification competence. But this approach risks the adapted classifier not improving in performance or even becoming ineffective due to the self-teaching of incorrect labelled samples. Combining two or more classifiers with different architectures may help to represent a more comprehensive variability of the domain to be learned and stabilise of the learning process.

This paper proposes an iterative co-training process where the most informative samples from a given pool of unlabelled traffic sign images are automatically selected and then classified by two classifiers which predict labels for each other. Additional knowledge of traffic sign appearance in countries unknown to the classifiers as well as noise distributions in rotation, camera angle, etc., will be obtained from a set of synthetically generated samples.

The evaluation starts with self-teaching a classifier for Germany, beginning with a relatively small initial training set. The adaption of this self-taught German classifier on Italian traffic signs without any intervention by a human expert is evaluated further as well as an adaption of a fully supervised trained German classifier. The classifiers distinguish between 14 different classes of traffic signs including speed limits, no-passing signs, and the corresponding termination signs. In the following sections, the proposed method is elaborated in detail.

2   Related Work and Applied Methods

The traffic sign recognition survey by Fu and Huang [5] provides an introduction and a brief overview of the broad field of existing approaches. A recent and comprehensive publication on automatic traffic sign recognition and classification by Lindner [10] describes a complete processing chain including the recording of training data, the detection and tracking of traffic signs, the training and evaluation of classifiers and finally the application of heuristics to bring the classification results into accordance with other traffic rules in order to construct a speed limit driver assistance system. Today, traffic sign recognition systems are available as special equipment for some car models from renowned manufacturers, but none of these systems has self-learning capabilities.

Semi-supervised learning techniques [14],[3],[13] reduce the amount of required labelled samples when training new traffic sign classifiers or adapt existing classifiers to other countries as described by Hillebrand et. al [7]. Combining two or more classifiers to obtain better classification results or to further reduce the data acquisition and labelling efforts with the help of active learning methods where applicable can be achieved by Blum and Mitchell's co-training approach [1] or by means of ensemble learning methods [11]. Recently, Cui et al. [4] analysed the application of co-training methods in semi-supervised two-class traffic sign learning scenarios.

Like Cui et al., we co-train two different classifiers, a polynomial classifier (PC) [12] with a fully quadratic structure of the feature vectors and a support vector regression (SVR) [2]. The polynomial classifier is robust against partially

mislabelled data and therefore capable of producing good classification results even if the semi-supervised learning process associates a certain fraction of samples with incorrect class labels. In contrast to this, the support vector regression is more sensitive towards mislabelled samples and just produces an exceeding amount of support vectors without achieving generalisation capability for the classification task, but is better in representing special cases not considered by a PC model function. The n-class classification capabilities are realised by n one-against-all epsilon-SVRs with radial basis functions.

Co-training relies on the assumption that two different classifiers can represent a more comprehensive variability of the feature space due to their different model hypotheses. The semi-supervised training is an iterative process. After initialising the classifiers with a certain amount of correctly labelled samples both classifiers predict labels for a certain amount of selected samples from the large set of available unlabelled samples during each iteration. It is preferable to select those samples to be classified which are the most informative ones or rather let the classifier make a lunge towards ground truth performance. On the other hand, only samples the classifier is capable of predicting a confident label for can be selected. The confidence bands method [6],[7] is used to obtain a balanced selection of samples. Confidence bands are curves enclosing a model function estimated by a regression analysis. The bands represent the areas where the true model is expected to reside with a certain probability, commonly 95%. The extent of the bands in different areas of the data space gives an idea of how well the estimated function fits the data. Each iteration finishes by adding the samples classified by classifier A along with their predicted labels to the training set of classifier B and vice versa.

Real images of traffic signs depict a wide spectrum of variations, e.g., different sizes, rotations and camera angles, translations due to inaccurate detection results, soiling, partial occlusions, or different lighting and weather conditions. Lighting conditions are normalised by a pre-processing algorithm. The other most common variations (size, rotation, camera angle, translation) are represented by virtual traffic sign samples which are generated from one ideal depiction of each traffic sign by the method described by Hoessler et al. [8], but the less frequent variations are unrepresented. In principle, an infinite amount of such virtual traffic sign samples are available. The virtual samples are utilised to support the semi-supervised learning process and give the classifiers a basic training for unknown domains.

3   Experimental Evaluation

3.1   Experimental Setup

Applying the pre-processing steps described in detail in [7] results in intensity images with a size of 17×17 pixels and adjusted lighting conditions displaying traffic sign and garbage cut-outs. The automatic traffic sign vs. garbage sample classification leads to a rate of about 5% garbage samples in the training set. This complicates a lasting semi-supervised classifier training.

Fig. 1. Visualised PCA and NMF basis vectors with size of 17×17 pixels. Left: PCA basis vectors, known as *eigenfaces* in face recognition applications; black pixels are negative values, white pixels are positive values. Right: NMF basis vectors; grey pixels are zero values, white pixels are positive values.

The intensity values are used as a feature vector of dimension 289 reduced to 23 by a Principal Component Analysis preserving about 80% of the image information. In general, co-training assumes that each sample is represented using two different feature sets that provide complementary information. To get a second view on the data for the second classifier, a non-negative matrix factorisation (NMF) [9] with the same feature dimension and a comparable mean reconstruction error on the data samples is computed. The most significant difference between the two matrix factorisation methods is that the NMF results in basis images and sample encodings with only elements equal to or greater than zero (see Fig. 1). In all experiments, the PC classifier will be trained with the NMF-reduced feature vectors and the SVR will be trained with the PCA-reduced feature vectors.

The evaluation applies the self-teaching classifiers to different learning scenarios. The performance evaluation starts with self-teaching three classifiers for German traffic signs, a PC classifier, a SVR classifier and a co-training classifier consisting of both individual classifiers. The learning process starts with a relatively small initial training set which contains 10% of the whole set. These samples are taken randomly from the whole set, but are distributed equally over all classes. The whole German data set consists of 14 classes (see Fig. 2) containing 1000 samples per class (14000 samples in total). A training set of the same composition and the same size is available from Italy. Finally, one set of virtual training samples, again with 14 classes and 14000 samples in total, is generated for each country.

The second part of the evaluation compares the performance of the three classifiers described above to a German-Italian adaption scenario: The self-trained co-training classifier from the first evaluation as well as a classifier trained with the same training set but with ground truth labels will be adapted to recognise Italian traffic signs by self-teaching samples from the Italian training set iteratively.

All evaluations consisting of random initial training sets have been repeated five times. The resulting classification rates given in the tables are mean values. The standard deviation is always included as well, except for classifiers with initial training sets consisting of all virtual samples.

The classifier performances are compared by computing correct classification rates on independent test sets. Like the training sets, the test sets consist of samples from the same 14 classes with each class represented by 250 samples (3500 samples in total). The test sets have been recorded independently of the
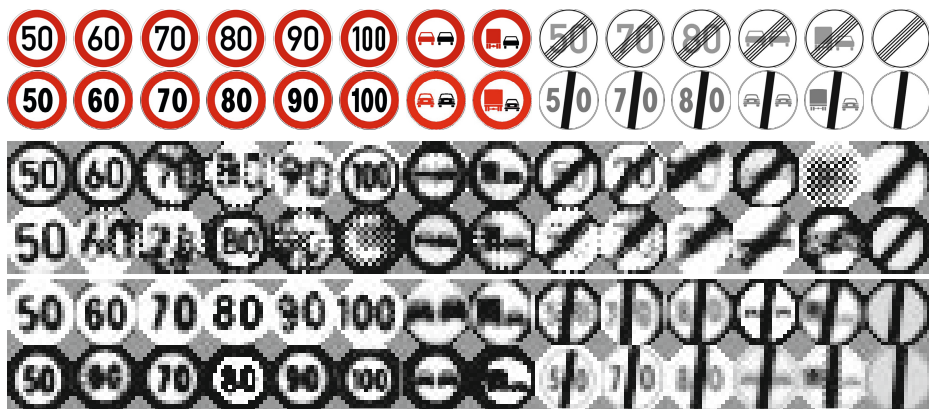
Fig. 2. Traffic sign images. Rows 1–2: Ideal depictions of 14 German and Italian signs. Rows 3–4: Two German real training samples from each class. Rows 5–6: Two Italian virtual training samples from each class.

training sets with different cameras. The same pre-processing routines have been applied to all samples from all data sets.

The data sets are not noise-free and contain a considerable amount (between 5% and 10%) of garbage samples and bad quality samples, e.g., with inaccurate cut-outs. An example of this is depicted in Fig. 2 for some German training samples. Classifiers often obtain correct classification rates above 90%, so the large number of bad quality samples in the test sets would not allow meaningful comparisons between different classifiers. For this reason, all samples not classifiable by a human expert were removed from the test sets before the 3500 samples per country were selected at random.

3.2   Reference Classifiers

The evaluation begins with the determination of reference classifiers to benchmark the performances of all following self-trained classifiers. In general, the Italian reference classifiers perform about five percentage points below the German classifiers as shown in Table 1. All classifiers trained with real samples always show the best performance. The classifiers trained with real and virtual samples perform on the same level or a little bit below. The classifiers trained only with virtual samples perform several percentage points worse. This result shows that there is no need to use virtual training samples when a sufficient amount of labelled real training samples are available.

The co-training classifiers, which combine polynomial classifier and support vector regression, perform comparable to the SVR classifiers alone. Only the PC classifiers perform somewhat worse than the other. Further investigations show that the comparatively poor performance is at least in part a consequence of the 5–10% amount of garbage samples in the training sets.

Table 1. Correct classification rates of all reference classifiers, determined on country-specific test sets. Three different German (rows 1–3) resp. Italian (rows 4–6) training sets contain only real samples, real and virtual samples or only virtual samples.

classifier	real samples	real & virtual samples	virtual samples
DE: PC	0.98	0.96	0.86
DE: SVR	0.99	0.99	0.90
DE: PC & SVR	0.99	0.98	0.89
IT: PC	0.93	0.91	0.81
IT: SVR	0.94	0.94	0.83
IT: PC & SVR	0.94	0.93	0.83

3.3   Self-teaching German Traffic Signs

This evaluation is about self-teaching a German classifier with a very small amount of labelled real training samples or even no labelled real samples at all. All three classifiers presented in Table 2, PC, SVR and PC & SVR co-training, are trained with 10% real training samples followed by self-teaching the other 90%, with 10% real and all virtual training samples followed by self-teaching the other 90% real samples, and finally with only virtual initial training samples, followed by self-teaching all 100% real samples. The classification rates are determined on the German test set.

As expected, all classifiers show their best performances for the second scenario, followed by the third scenario. The classifier from the first scenario with only 10% real initial samples (these 10% are the only samples that are definitely correctly labelled) are still close to the other classifiers' performances (except the SVR). When inspecting the standard deviations of the classification rates, the co-training classifiers which show fewer correct classifications than the PC classifiers also result in a slightly lower standard deviation.

Table 2. Correct classification rates of German self-teaching classifiers. The initial training sets contain 10% real samples, all virtual and 10% real samples or only virtual samples. The classifiers always self-taught the rest of the real German samples.

initial training set	10% real		virtual & 10% real		only virtual
classifier	class. rate	std. dev.	class. rate	std. dev.	class. rate
PC	0.94	0.017	0.95	0.015	0.94
SVR	0.81	0.041	0.91	0.033	0.91
PC & SVR	0.93	0.011	0.95	0.012	0.93

3.4   Adapting German Classifiers to Italy

This evaluation is about adapting existing classifiers trained on the complete set of German real traffic signs to Italy. The classification rates are determined on the Italian test set. As a first experiment, all three classifiers (PC, SVR and PC & SVR co-training) obtain 10% real training samples from Italy with correct labels to their training sets, followed by self-teaching the other 90%. In the second experiment, 10% real and all virtual training samples from Italy are mixed into their training sets, followed by self-teaching the other 90% real samples. Finally the classifiers are equipped only with the complete set of Italian virtual training samples, followed by self-teaching all 100% real samples.

As Table 3 shows, the best classification rates are obtained by the PC classifiers, again. But the co-training classifiers also perform reasonably well with a slightly lower deviation in their results, again. As expected after the previous evaluations, the SVRs alone perform worst of all. When using the self-trained German classifier which incorporates about 9% incorrectly labelled samples as a basic training, the performance decreases by some percentage points.

Table 3. Correct classification rates of German classifiers adapted to Italy. Rows 1–3: The classifiers have been trained initially with all German real samples. Row 4: The classifier has been semi-supervised co-trained with German training samples (contains incorrectly classified samples). Initial training sets also contain Italian samples: 10% real samples, all virtual and 10% real samples or only all virtual samples. The classifiers always self-taught the rest of the real Italian samples.

initial training set	10% real		virtual & 10% real		only virtual
classifier	class. rate	std. dev.	class. rate	std. dev.	class. rate
PC	0.89	0.032	0.90	0.028	0.81
SVR	0.83	0.038	0.85	0.040	0.79
PC & SVR	0.88	0.022	0.88	0.019	0.81
PC & SVR (DE-st)	0.81	0.056	0.82	0.051	0.77

4   Summary and Conclusions

Adapting a German traffic sign classifier to Italy by semi-supervised methods shows a reasonable performance. Although not discussed in this work, an adaptation to other European countries like Austria, Spain or France yields comparable results.

The main conclusion of this work is that classifiers with the best performances should not always be preferred for self-training tasks. In the majority of cases, the co-trained classifier does not show the best mean classification rates but the lowest standard deviation of the performance measure. This means that using this kind of classifier for a self-training task, the resulting classifier may not be the best, but supplies a reasonable performance with a considerably higher confidence.

Adding virtual training samples to the initial training sets always increases traffic sign classifier performances or stabilises the results. Due to the theoretically

infinite availability of these samples, it appears reasonable to always include them in the training sets.

All classifier training runs were processed until 100% of the training samples were labelled by the classifiers and added to the training sets. A result not revealed by the tables in the evaluation section is that many classifiers show their best performance after self-teaching a smaller amount of samples, sometimes even less than 50%. The performance of some classifiers eventually decreased slightly due to an increasing amount of incorrectly labelled samples. An early stopping criterion is therefore being developed.

References

1. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, COLT 1998, pp. 92–100. ACM, New York (1998)
2. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
3. Chapelle, O., Schölkopf, B., Zien, A. (eds.): Semi-Supervised Learning. Adaptive Computation and Machine Learning. The MIT Press (2006)
4. Cui, T., Grumpe, A., Hillebrand, M., Kreßel, U., Kummert, F., Wöhler, C.: Analytically tractable sample-specific confidence measures for semi-supervised learning. In: Proc. 21st Workshop Computational Intelligence, pp. 171–186 (2011)
5. Fu, M.Y., Huang, Y.S.: A survey of traffic sign recognition. In: 2010 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), pp. 119–124 (2010)
6. Hillebrand, M., Wöhler, C., Krüger, L., Kreßel, U., Kummert, F.: Self-learning with confidence bands. In: Proc. 20th Workshop Computational Intelligence, pp. 302–313 (2010)
7. Hillebrand, M., Wöhler, C., Kreßel, U., Kummert, F.: Semi-supervised Training Set Adaption to Unknown Countries for Traffic Sign Classifiers. In: Schwenker, F., Trentin, E. (eds.) PSL 2011. LNCS, vol. 7081, pp. 120–127. Springer, Heidelberg (2012)
8. Hoessler, H., Wöhler, C., Lindner, F., Kreßel, U.: Classifier training based on synthetically generated samples. In: The 5th International Conference on Computer Vision Systems (2007)
9. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Letters to Nature 401(1), 788–791 (1999)
10. Lindner, F.: Adaptive Traffic Sign Recognition. Ph.D. thesis, Bielefeld University (2012)
11. Rokach, L.: Pattern Classification using Ensemble Methods. Series in Machine Perception and Artificial Intelligence, vol. 75. World Scientific (2010)
12. Schürmann, J.: Pattern Classification: A Unified View of Statistical and Neural Approaches. John Wiley & Sons, Inc. (1996)
13. Xu, Z., King, I., Lyu, M.R.: More Than Semi-supervised Learning. Lambert Academic Publishing (2010)
14. Zhu, X., Goldberg, A.B.: Introduction to Semi-Supervised Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers (2009)