Neural Networks 32 (2012) 323–332

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2012 Special Issue

Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition

J. Stallkamp[a,*], M. Schlipsing[a], J. Salmen[a], C. Igel[b]

[a] Institut für Neuroinformatik, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany

[b] Department of Computer Science, University of Copenhagen, Universitetsparken 1, 2100 Copenhagen, Denmark

ARTICLE INFO

Keywords:
Traffic sign recognition
Machine learning
Convolutional neural networks
Benchmarking

ABSTRACT

Traffic signs are characterized by a wide variability in their visual appearance in real-world environments. For example, changes of illumination, varying weather conditions and partial occlusions impact the perception of road signs. In practice, a large number of different sign classes needs to be recognized with very high accuracy. Traffic signs have been designed to be easily readable for humans, who perform very well at this task. For computer systems, however, classifying traffic signs still seems to pose a challenging pattern recognition problem. Both image processing and machine learning algorithms are continuously refined to improve on this task. But little systematic comparison of such systems exist. What is the status quo? Do today's algorithms reach human performance? For assessing the performance of state-of-the-art machine learning algorithms, we present a publicly available traffic sign dataset with more than 50,000 images of German road signs in 43 classes. The data was considered in the second stage of the German Traffic Sign Recognition Benchmark held at IJCNN 2011. The results of this competition are reported and the best-performing algorithms are briefly described. Convolutional neural networks (CNNs) showed particularly high classification accuracies in the competition. We measured the performance of human subjects on the same data—and the CNNs outperformed the human test persons.

1. Introduction

Traffic sign recognition is a multi-category classification problem with unbalanced class frequencies. It is a challenging real-world computer vision problem of high practical relevance, which has been a research topic for several decades. Many studies have been published on this subject and multiple systems, which often restrict themselves to a subset of relevant signs, are already commercially available in new high- and mid-range vehicles. Nevertheless, there has been little systematic unbiased comparison of approaches and comprehensive benchmark datasets are not publicly available.

Road signs are designed to be easily detected and recognized by human drivers. They follow clear design principles using color, shape, icons and text. These allow for a wide range of variations between classes. Signs with the same general meaning, such as the various speed limits, have a common general appearance, leading to subsets of traffic signs that are very similar to each other. Illumination changes, partial occlusions, rotations, and weather conditions further increase the range of variations in visual appearance a classifier has to cope with.

Humans are capable of recognizing the large variety of existing road signs in most situations with near-perfect accuracy. This does not only apply to real-world driving, where rich context information and multiple views of a single traffic sign are available, but also to the recognition from individual, clipped images.

In this paper, we compare the traffic sign recognition performance of humans to that of state-of-the-art machine learning algorithms. These results were generated in the context of the second stage of the *German Traffic Sign Recognition Benchmark* (GTSRB) held at IJCNN 2011. We present the extended GTSRB dataset with 51,840 images of German road signs in 43 classes. A website with a public leaderboard was set up and will be permanently available for submission of new results. Details about the competition design and analysis of the results of the first stage are described by Stallkamp, Schlipsing, Salmen, and Igel (2011).

The paper is organized as follows: Section 2 presents related work. Section 3 provides details about the benchmark dataset. Section 4 explains how the human traffic sign recognition performance is determined, whereas the benchmarked machine learning algorithms are presented in Section 5. The evaluation procedure is described in Section 6, together with the associated public leaderboard. Benchmarking results are reported and discussed in Section 7, before conclusions are drawn in Section 8.

* Corresponding author. Tel.: +49 234 3225566; fax: +49 234 3214210.
E-mail addresses: johannes.stallkamp@ini.rub.de (J. Stallkamp), marc.schlipsing@ini.rub.de (M. Schlipsing), jan.salmen@ini.rub.de (J. Salmen), igel@diku.dk (C. Igel).

0893-6080/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2012.02.016

2. Related work

It is difficult to compare the published work on traffic sign recognition. Studies are based on different data and either consider the complete task chain of detection, classification and tracking or focus on the classification part only. Some articles concentrate on subclasses of signs, for example on speed limit signs and digit recognition.

Bahlmann, Zhu, Ramesh, Pellkofer, and Koehler (2005) present a holistic system covering all three processing steps. The classifier itself is claimed to operate with a correct classification rate of 94% on images from 23 classes. Training was conducted on 4000 traffic sign images featuring an unbalanced class frequency of 30–600 examples. The individual performance of the classification component is evaluated on a test set of 1700 samples.

Moutarde, Bargeton, Herbin, and Chanussot (2007) present a system for recognition of European and US speed limit signs. Their approach is based on single digit recognition using a neural network. Including detection and tracking, the proposed system obtains a performance of 89% for US and 90% for European speed limits, respectively, on 281 traffic signs. Individual classification results are not provided.

Another traffic sign detection framework is presented by Ruta, Li, and Liu (2010). The overall system including detection and classification of 48 different signs achieves a performance of 85.3% while obtaining classification error rates below 9%.

Broggi, Cerri, Medici, Porta, and Ghisio (2007) apply multiple neural networks to classify different traffic signs. In order to choose the appropriate network, shape and color information from the detection stage is used. The authors only provide qualitative classification results.

In the work by Keller, Sprunk, Bahlmann, Giebel, and Baratoff (2008), a number-based speed limit classifier is trained on 2880 images. It achieves a correct classification rate of 92.4% on 1233 images. However, it is not clear whether images of the same traffic sign instance are shared between sets.

Gao, Podladchikova, Shaposhnikov, Hong, and Shevtsova (2006) propose a system based on color features inspired by human vision. They report recognition rates up to 95% on 98 British traffic sign images.

Various approaches are compared on a dataset containing 1300 preprocessed examples from 6 classes (5 speed limits and 1 noise class) by Muhammad, Lavesson, Davidsson, and Nilsson (2009). The best classification performance observed was 97%.

In the study by Maldonado Bascón, Acevedo Rodríguez, Lafuente Arroyo, Caballero, and López-Ferreras (2010), a classification performance of 95.5% is achieved using support vector machines. The database comprises ∼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 Section 3). To our knowledge, this database is not publicly available.

Obviously, the results reported above are not comparable, as all systems are evaluated on proprietary data, most of which is not publicly available. Therefore, we present a freely available, extensive traffic sign data set to allow unbiased comparison of traffic sign recognition approaches.

3. Dataset

This section describes our publicly available benchmark dataset. We explain the process of data collection and the provided data representation.

3.1. Data collection

The dataset was created from approx. 10 h of video that were 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 GC 1380CH* camera was used with automatic exposure control and a frame rate of 25 fps. The camera images, from which the traffic sign images are extracted, have a resolution of 1360 × 1024 pixels. The video sequences are stored in a raw *Bayer*-pattern format (Bayer, 1975).

Data collection, annotation and image extraction was performed using the *NISYS Advanced Development and Analysis Framework (ADAF)*,[1] an easily extensible, module-based software system (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 a *track*. Each instance is unique. In other words, the dataset only contains a single track for each physical traffic sign.

3.2. Data organization

From 144,769 labelled traffic sign images of 2416 traffic sign instances in 70 classes, the GTSRB dataset was compiled according to the 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 car passes different traffic sign instances with different velocities, depending on sign position and the overall traffic situation. In the recording, this leads to different numbers of traffic sign images per track (approximately 5–250 images per track). Consecutive images of a traffic sign that was passed with low velocity are very similar to each other. They do not contribute to the diversity of the dataset. On the contrary, they cause an undesired imbalance of dependent images. Since the different velocities are not uniformly distributed over all traffic sign types, this would strongly favour image classes that are present in low-speed traffic (*Stop*, *Yield-right-of-way*, low speed limits).

Secondly, the question arises why to keep multiple images per track at all. Although consecutive images in long tracks are nearly identical, the visual appearance of a traffic sign can vary significantly over the complete track, as can be seen in Fig. 2. Traffic signs at high distance 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 and background. Selecting a fixed number of images per traffic sign both increases the diversity of the dataset in terms of the variations mentioned above and avoids an undesired imbalance caused by large numbers of nearly identical images.

The selection procedure outlined above reduced the number to 51,840 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 1700 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 traffic sign within the image (region of interest, ROI) is preserved in the provided annotations. The images are not necessarily square. Fig. 5 shows the distribution of traffic sign sizes, taking into account the larger of both dimensions of the traffic sign ROI.

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