Image classification system based on Deep Learning applied to the recognition of traffic signs for intelligent robotic vehicle navigation purposes

Diego Renan Bruno., and Fernando Santos Osório., Member, IEEE - *University of São Paulo*

Abstract - **This paper presents a system for classifying images based on Deep Learning and applied in the recognition of traffic signals aiming to increase road safety increased road safety using autonomous and semi-autonomous intelligent robotic vehicles. This Advanced Driver Assistance System (ADAS) is a system created to automate vehicles, but also to help the human drivers to increase safety and the respect of traffic rules while driving the car. The system must be able to classify several different traffic signs (e.g. maximum speed allowed, stop, slow down, turn ahead, pedestrian), thus helping to make navigation within the local traffic rules. The obtained results are promising and very satisfactory, where we get 97.24% of test accuracy in a well known traffic sign benchmark dataset (INI - German Traffic Sign Benchmark).**

Keywords - Deep Learning; Tensorflow; Traffic Sign; Robotics Vehicles; ADAS.

I. INTRODUCTION

The structured road scenarios and urban street environments needs general and proper traffic rules, including varied information for drivers, so that a vehicle can safely navigate on it, whether it is in an autonomous or semi-autonomous mode, or even being conducted by a human. The most used signs are: traffic signs (speed control, stop and direction), traffic lights (danger, attention, proceed, traffic semaphore) and traffic lanes (regulation of flows and space). Traffic signs are used to assist the driver in the task of driving, however, also informing about the local traffic rules [1].

These signs are of great importance not only for the driver of a semi-autonomous/intelligent vehicle but also for a vehicle that travels autonomously and shares the same rules with other cars and humans. In this work, a Deep Learning based image classification method was applied for the recognition of traffic signs in benefit of road safety applied to intelligent robotic vehicles.

The research area of detection and classification of traffic signs has grown a lot in the last decade, mainly with the works that use machine learning techniques. This increase is closely linked to the evolution of robotized vehicles, where Advanced Driver Assistance Systems (ADAS) are applied to assist in the task of driving safely (conducted by a human driver) and also for autonomous vehicles. Safe navigation in traffic environments also depends on this information provided by means of traffic sign plates, traffic lights and traffic lanes. In this work it was proposed the use of Deep Learning, adopting the TensorFlow tool [2], to classify the traffic signals.

II. RELATED WORKS

Traffic sign recognition have been studied for a long time, but the in the 80's and 90's, most of the works suffered from problems with sign detection, illumination, noise and partial occlusion, and most of them obtained relatively poor performances of 70% to 80% of accuracy considering small datasets. The improvement in the development of machine learning techniques and the available datasets allowed to obtain better results more recently.

In a paper by Yaong Yu et. Al. [1], a system was developed based on an algorithm capable of classifying traffic signals for the benefit of navigation of robotic vehicles. The system uses a mobile laser scanning (MLS) sensor.

In a work done by Timofte and Zimmermann [3], a system was developed that aims at the detection of signaling signs in favor of the recognition of traffic rules. In this work, a method of plate recognition based on 3D image analysis was developed using a technique called Minimum Description Length principle (MDL). This was one of the first works using 3D images in favor of the analysis of traffic signs [3].

In a paper by Zhou and Deng [4], a system based on LIDAR (Light Detection And Ranging) and recognition algorithms for the analysis of images of signaling plates was developed. The LIDAR sensing technology is similar to ultrasonic sensing, however, using a light beam as a signal. The system aims to analyze 3D images in order to achieve greater robustness in the detection of plates. Through 3D point-cloud data (colors and point agglutination) the signaling of the board is analyzed. The algorithm that identifies the characteristics of each detected board is based on a Support Vector Machine (MVS), however, being applied as a classifier [4].

In a work developed by Soilán [5], each traffic signal is automatically recognized using the Light Detection And Ranging (LIDAR) technique in conjunction with algorithms based on semantic analysis applied as classifiers [5]. The identification of the plates and their signs are done via geometric classifiers.

In a paper by Stallkamp et. Al. [14] it is possible to visualize a general idea about the benchmark adopted [6] and that it was also applied to compare various machine learning techniques to classify traffic signals.

For the recognition of traffic signals, in this work we used TensorFlow, an open source software library for Machine Intelligence and Deep Learning implementation. This recognition system presented very good results and were compared with other techniques used for this same purpose. A traffic sign benchmark dataset from Germany containing 21 classes and 21,000 images was used for the tests. This dataset is available for download at [5]. The tests performed were compared with the works of other authors who also used this dataset.

The remainder of the text of this paper is organized as follows: In section 2 the theoretical background is presented and the deep learning system is presented (TensorFlow) applied in the recognition of traffic signals. In section 3 the algorithm for plate detection is presented. Section 4 presents the tests performed in the recognition system. Finally, in section 5 the results of this work are presented.

III. THEORITICAL BACKGROUND AND PROPOSED MODEL

A. Deep Learning: TensorFlow Tool

The open source software library for machine learning TensorFlow [2] is widely used today for image recognition applications. It has being applied from the recognition of images found in the Internet (ImageNet Competition), up to the recognition of objects in real world environment, used to assist visually impaired persons. TensorFlow was developed for machine learning and research into Deep Artificial Neural Networks (ANNs). In the system developed in this work, the images are passed to a Deep Learning model, adopting a DCNN Net (Deep Convolutional Neural Network / ConvNet) implemented using TensorFlow. The TensorFlow is utilized to classify the traffic signs previously detected by the attention/segmentation module of a computer vision system.

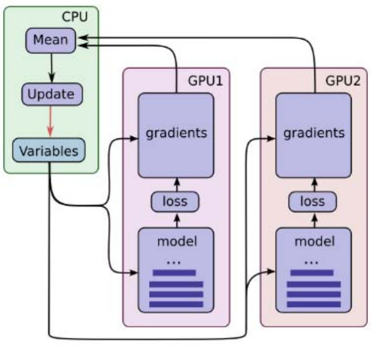


Figure 1: Configuration of the hardware architecture used by the Tensorflow in image recognition [2]

TensorFlow can work with GPUs (Graphics Processing Units) in parallel or using only CPUs (Central Processing Units). Each GPU can calculate the gradients for different sets of images [2], speeding-up the training of large datasets. This makes possible to do ANN training with better load distribution and optimization of the dataset for training. The final trained network usually can be used to classify new patterns running it in real-time applications.

The structure applied by Google in the construction of their Convolutional Neural Network (CNN) in its Deep Learning Tensorflow based applications, as implemented in the famous GoogLeNet, uses a lot of hidden layers, which is what defines this type of "deep" networks. The number of layers can be optimized according to the problem in question. For this, there is also an interactive simulator that can help in understanding the operation of the Tensorflow in its classification tasks [7].

In Figure 2, an example of an optimized network in this simulator [7] could be observed, which was constructed with 1 input layer and 3 hidden layers, containing [7x8x8x4] neurons respectively [15] [16]. This model was applied for a classification of the two spirals problem samples (a classical problem in neural learning). More information about Tensorflow can be found in [2][7].

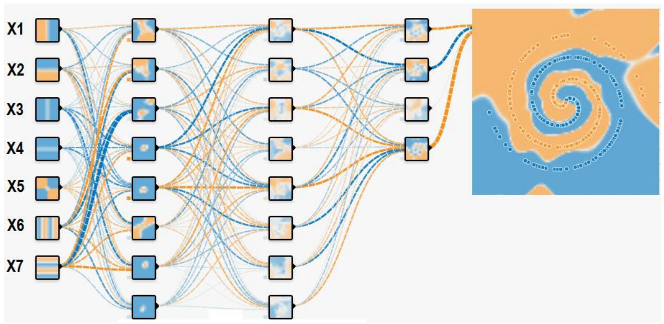


Figure 2: Example of a neural network architecture used by the Tensorflow [2]

B. Methodology Training and Testing

The training used in this work was applied in the upper layer of a CNN, based on the Inception's Network [2] [15] [16], 70% of the dataset was used to make the network training (adapt the network parameters). The remaining 30% is applied only to evaluate the network performance in the test steps [16]. In order to do that, the images were chosen randomly to generate these two sets (training/test). The accuracy of the system was evaluated through the test images (30% test set), consequently, this data set is not contained in the training data (validation accuracy).

If the training does not show good results later in the classification task (test set), it means that the network is working on particular characteristics in the set of training images, probably suffering of overtraining/overfit, and this network usually is not very useful for the classification [2].

This problem can be seen in traffic signs where the background characteristics (vegetation, lane, cars) become relevant in the training and recognition. This should be avoided, using bigger training datasets, and also using a good detection method to extract the background from the interest image (better segmentation).

C. Dataset

In figure 3 are shown 12 classes of the German traffic signals dataset (INI) [6], that are used in the neural classifier. In total, 21 classes containing a total of 21,000 images are available and were used. The main objective is related to the recognition of the traffic signs of regulation and warning. These signs should strengthen the rules for road safety.

Figure 3: Examples of traffic signs from the database used [6]

In order to make the evaluations of the proposed machine learning method comparable with other methods, public data provided by Institut Für the Neuroinformatik were used [6]. The available data set contains images between 15x15 and 250x250 pixels resolution and are in Portable Pixmap Format (PPM) format. The images were converted to JPG (Joint Photographic Experts Group) format, configured with no loss in image data and quality. TensorFlow works with images from this last image format.

IV. ALGORITHM FOR DETECTION OF TRAFFIC SIGNS

The algorithm for detection of traffic plates uses the technique of Slide Window. Given an input image the algorithm slides a template on it, thus generating several clippings. These snippets are then sent to the Deep Learning classification system. This algorithm is quite simple, it was used only to validate the classification system using other images that are not in the original image dataset (knowledge base) of cropped images [6]. To test this algorithm it was used a database of images also made available by the Institut Für the Neuroinformatik [6].

Figure 4 shows the operation of the algorithm. The template (rectangle) is slide over all the image until its end. At this point, it is not yet possible to predict if a traffic sign has already been correctly detected or if there are still other traffic signs to be detected. In the example of Figure 4, only the speed limiting plate has been detected and then passed to the DCNN to be classified.

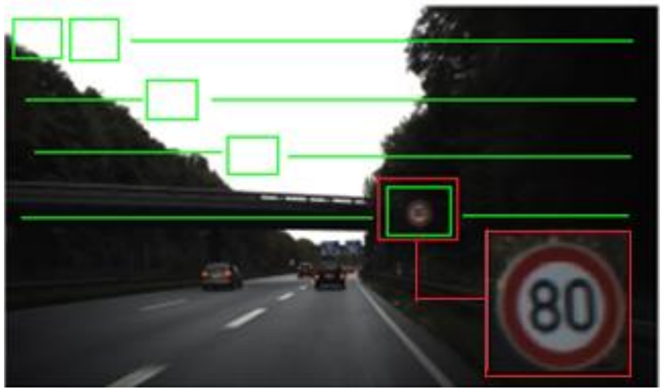


Figure 4: Slide window algorithm execution

The algorithm must traverse the mask over all the image, segmenting the fragments and sending it to the classifier. Obviously fragments other than those of interest will be classified as not being sign plates. Only the fragments of interest, (traffic signs), will present a good result in the classification. The template traverses the image by lines and columns (i, j) until the scan is finished. Remember that this algorithm was not applied in the original set of test images (30% test dataset - already segmented), but only in other images captured through a visual attention system that captures scenes with traffic signs in real environments. These images do not undergo any kind of pre-processing before being processed by the slide window algorithm.

V. EXPERIMENT AND RESULT

A. Tests performed for segmented and tagged images

In this stage, 10 training iterations of the DCNN training system TensorFlow were performed. The database (with labeled images) used was 70% (training) and 30% (tests). Remember that the images are divided randomly again for each training iteration, thus defining which images will be in the training and testing set. For this reason 10 iterations of the classification system were performed.

Each DCNN training has a time cost of 1 hour and 30 minutes. The adopted hardware was the following (table 1) - For all tests performed no GPU was used, so training can be speed-up significantly.

Hardware used in the tests	
CPU	Core i7 - 5500U
CPU speed	2.4 GHZ
RAM memory	8 GB
Operational system	Ubuntu 16.04 LTS (64 bits)
GPU	Not used

Table 1: Hardware used in the tests

All the tests performed had a recognition rate higher than 94%. This result indicates that DCNN Tensorflow is very efficient, mainly because it is working with a big dataset (many sign examples), including low resolution images and various environmental situations at the time of image capture (occlusion, blur, lack of illumination). Test results can be viewed on table 2.

Set of tests performed Recognition efficiency (%)	
Test 1	95.4
Test 2	94.7
Test 3	97.10
Test 4	99.8
Test 5	97.0
Test 6	95.9
Test 7	98.3
Test 8	97.5
Test 9	98.5
Test 10	98.2
Average results	97.24

Table 2: Recognition efficiency (% of Accuracy in Test Set)

In the next section will be presented other tests that are directed to the recognition of traffic signs in unlabeled images.

B. Tests performed for no tagged images

In this section the tests performed with unlabeled images and unsegmented traffic signs are presented. For these tests a dataset was used that contains images that have not yet been processed to identify the region of interest. The regions of interest, in these captured images by the vision system, must represent the zone where is the traffic sign. As the focus of this work is not on traffic sign detection/segmentation but on classification, the slide window algorithm was used for this task (this algorithm was explained in the previous section).

Figure 5: Example of detection of two traffic signs "STOP" [6]

Considering the Figure 5, it can be observed that two warning plates ("STOP") are present in the scene. The slide window algorithm after segmenting these two regions of interest (representing the traffic signs) must send them to the DCNN Tensorflow for classification.

To test this system of automatic detection and classification it was used a set of 300 real world environment images. Not all of these images contain traffic signs, and some other may contain more than one sign (at most 3).

The DCNN Tensorflow system presented good classification results in the unlabeled images. The system obtained a recognition rate greater than 75% for all classified images. For unlabeled images set, this lower result was caused because the slide window algorithm is not always able to detect and segment the traffic sign completely. Another problem is also related to the severe variations of conditions in the environment (rain, illumination, blur).

C. Testing for Images with Occlusion

In some cases the slide window algorithm can detect and segment the traffic signs in the wrong way. The obtained result can be cropped images (Table 3). However, this problem can also happen in some natural situations due to a real object occlusion. As for example, trees in front of the traffic sign, highway vegetation, deteriorated, or even, another object that hidden the sign (car, truck, motorcycle). These situations can create problems to detect the traffic sign.

However, Deep Learning works very well in situations where some information is not available for classifying an image. This is very evident in the process of classifying traffic signs with occlusion. Some tests were also performed for these type of images (Table 3). The recognition system always displays the first 5 traffic signs that can represent the detected image. The results appear in a decreasing way, always giving the greater percentage to the best estimate. In order to consider a good classification, the first place must have a score of at least 50% greater than the second placed.

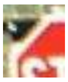
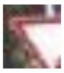
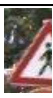
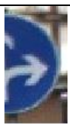
Traffic Sign	Type of traffic signal	Accuracy in classification (%)
	STOP	99.1
	STOP	99.6
	STOP	98.3
	STOP	96.2
	PREFERENCE	96.2
	PREFERENCE	97.6
	PEDESTRIAN	88.5
	PEDESTRIAN	89.2
	FOLLOW IN FRONT OR RIGHT	94.1
	FOLLOW IN FRONT OR RIGHT	88.2

Table 3: Classification of traffic sign with occlusion

The image occlusion cause a major problem for speed limit traffic signs. For this case, the information of only one region of the traffic sign is not enough, so a larger region is necessary to allow a good classification. This problem is not only found in artificial systems, but is also a problem for humans. When we cannot observe a complete picture, we can only make an estimate of which object it would be. However, even intelligent humans sometimes do not achieve good results.

In Figure 6, a situation can be observed where two traffic signs have zero at the end (Figure 6 (a) and (b)). If only the end zero is detected (Figure 6 (b)) then it is impossible to correctly classify whether it would be a 70km or 80km sign (or even another zero-ended traffic sign).

Figure 6: Severe problem of occlusion of traffic signs [6] (a) 80km (b) 70km and (c) problem of occlusion

By means of the graph of Figure 7 it is possible to observe the problem that maximum velocity traffic signs generate in the classification results. The reason for this lower efficiency in classification is related to the problem of occlusion. Since in these cases of speed limit it is very difficult to recognize numbers when there is an occlusion/cut (Figure 6).

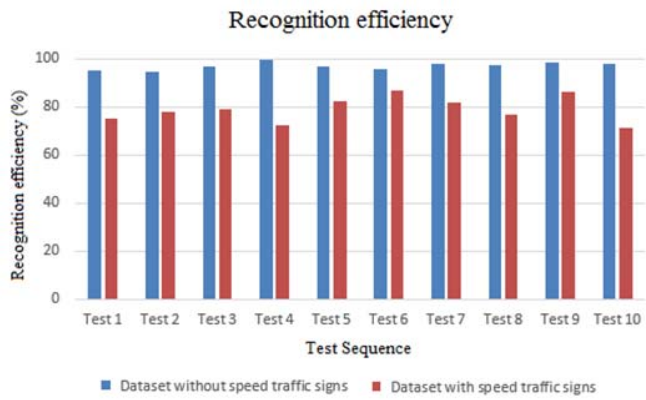


Figure 7: Comparison of classification tests

VI. CONCLUSION AND FUTURE WORK

We present the application of a Deep Learning method based on a Deep Convolutional Neural Networks (DCNN) using TensorFlow for the recognition of traffic signs. TensorFlow DCNN presented good results for this application involving a dataset widely used by researchers in the ADAS area and traffic sign recognition [6]. One of the main positive features of this tool is its ability to work with low resolution images and also with occlusions. This is very important in real environments, since intelligent robotic vehicles will not always have all the necessary/perfect information, and should also work in not very favorable environmental conditions.

Other authors [8,9,10,11,12,13] also used the Institut Für Neuroinformatik dataset [6]. The results presented in this article can be compared directly to them considering Table 4. The presented results in this paper are among the top 5 five. On the other hand, one interesting contribution presented in this paper is related to the performance of the system using other images not previously segmented or annotated, and the problems related to the occlusion in speed limit signs.

In the work of Ciresan et al [10], Stallkamp et al. [8], Zaklouta et al., [9] and Stallkamp et al., [12] CNNs were used for the classification of traffic signs. CNN results were compared with human visual recognition. In the test performed in this work, the CNN won.

In the work of Gecer et al [13] a system based on COSFIRE (Combination of Shifted Filter Responses) was applied. This filter combines the responses of a selection of Gaussian filters. The system also presented results that surpass human visual recognition [13].

TEAM	METHOD	TOTAL	SUBSET All signs
[10] - Committee of CNNs	Committee of CNNs	99.46%	99.46%
[13] - COSFIRE	Color-blob-based COSFIRE filters for object recogn	98.97%	98.97%
[8] - INI-RTCV	Human Performance	98.84%	98.84%
[11] - sermanet	Multi-Scale CNNs	98.31%	98.31%
[9] - CAOR	Random Forests	96.14%	96.14%
[12] - INI-RTCV	LDA on HOG 2	95.68%	95.68%
[12] - INI-RTCV	LDA on HOG 1	93.18%	93.18%
[12] - INI-RTCV	LDA on HOG 3	92.34%	92.34%

Table 4: Results that were submitted for dataset used [6]

The results presented in these studies were of great scientific contribution to the area of classification of traffic signs, however, in none of these papers the problem was completely solved, since there are still problems of detection and segmentation of traffic signs. This problem of visual attention is of great importance, therefore, it must help (and speed-up) in the process of acquisition of images for the machine learning and classification system. In the current research in this domain, 2D detection have been losing space for new 3D detection methods. The latter presents less vulnerability to physical interference from the environment (e.g. illumination). Briefly, the methods of 3D computer vision are of greater robustness.

A future work proposed in this article is linked with these methods of 3D sign plate detection and segmentation. For this, a visual attention system with 2D and 3D data fusion will be used. For the detection of traffic signals a 3D vision system will be used, eliminating some of the major problems encountered in 2D detection/segmentation systems. The 2D data will only be used for image processing and traffic sign recognition, based on the previously segmented sign plate.

ACKNOWLEDGMENT

We thank CAPES for the financial support to Diego Renan Bruno of graduate program the University of São Paulo at São Carlos.

REFERENCES

[1] Y. YU, J. LI, C. WEN, H. GUAN, H. LUO AND C. WANG, "Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data," ISPRS journal of photogrammetry and remote sensing, vol. 113, p. 106–123, 2016.

[2] TENSORFLOW. An open-source software library for Machine Intelligence :< https://www.tensorflow.org/get_started/>.

[3] TIMOFTE, R.; ZIMMERMANN, K.; GOOL, L. V. Multi-view traffic sign detection, recognition, and 3D localisation (2014). Available in: <http://www.vision.ee.ethz.ch/~timofter/traffic_signs/>. Access in: 10/04/2017.

[4] ZHOU, L.; DENG, Z. LIDAR and vision-based real-time traffic sign detection and recognition algorithm for intelligent vehicle. 17th International IEEE Conference on Intelligent Transportation Systems (2014).

[5] SOILÁN, M.; RIVEIRO, B.; MARTÍNEZ-SÁNCHEZ, J.; ARIAS, P. Traffic sign detection in MLS acquired point clouds for geometric and image-based semantic inventory - ISPRS Journal of Photogrammetry and Remote Sensing (2016) p. 92–101.

[6] INSTITUT FÜR NEUROINFORMATIK (INI) - German Traffic Sign Benchmak..Available:<http://benchmark.ini.rub.de/?section=gtsrb&subsection=dataset>. Access in: 10/04/2017.

[7] TENSORFLOW. Tinker With a Neural Network Right Here in Your Browser. Don't Worry, You Can't Break It. We Promise. <http://playground.tensorflow.org/>

[8] STALLKAMP, J.; SCHLIPSING, M.; SALMEN, J.; IGEL, C. Human Performance, INI-RTCV , Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, August 2012, Neural Networks (32), pp. 323-332

[9] ZAKLOUTA, F.; STANCIULESCU, B. Random Forests, CAOR , Traffic sign classification using K-d trees and Random Forests , Traffic sign classification using K-d trees and Random Forests .Hamdoun, August 2011, International Joint Conference on Neural Networks (IJCNN) 2011

[10] CIRESAN, D.; MEIER, U.; MASCI, J.; SCHMIDHUBER, J. Committee of CNNs, IDSIA , Multi-column deep neural network for traffic sign classification, Multi-column deep neural network for traffic sign classification, August 2012, Neural Networks (32), pp. 333-338

[11] SERMANET, P.; LECUN, Y. Multi-Scale CNNs, sermanet , Traffic sign recognition with multi-scale Convolutional Networks, Traffic sign recognition with multi-scale Convolutional Networks, August 2011, International Joint Conference on Neural Networks (IJCNN) 2011

[12] STALLKAMP, J.; SCHLIPSING, M.; SALMEN, J.; IGEL, C. LDA on HOG 1, INI-RTCV, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. August 2012, Neural Networks (32), pp. 323-332

[13] GECER,B.; AZZOPARDI, G.; PETKOV, N.; Color-blob-based COSFIRE filters for object recogn, COSFIRE, Color-blob-based COSFIRE filters for Object Recognition, Color-blob-based COSFIRE filters for Object Recognition, 2017, Image and Vision Computing(57), pp. 165-174

[14] STALLKAMP, J.; SCHLIPSING, M.; SALMEN, J.; IGEL, C. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, 2011

[15] SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; AND . WOJNA, Z.; "Rethinking the Inception Architecture for Computer Vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 2818-2826

[16] DONAHUE, J.; JIA, Y.; VINYALS, O.; HOFFMAN, J.; ZHANG, N.; TZENG, E.; DARRELL, T. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. Available< https://arxiv.org/pdf/1310.1531v1.pdf> Access in: 28/04/2017.