DETECTION AND RECOGNITION OF TRAFFIC SIGNS FROM DATA COLLECTED BY THE MOBILE MAPPING SYSTEM

H. Gezgin [1*], R.M. Alkan [2]

[1] ITU, Graduate Program of Geographical Information Technologies, 34469 Maslak Istanbul, Türkiye - turkhi@itu.edu.tr
[2] ITU, Civil Engineering Faculty, Dept. of Geomatics Engineering, 34469 Maslak Istanbul, Türkiye - alkanr@itu.edu.tr

KEY WORDS: Object Detection & Recognition, Traffic Sign, Deep Learning, Mobile Mapping.

ABSTRACT:

Autonomous vehicles and high-resolution maps are key elements of future transport systems. Detection and recognition of traffic signs is an important element for the safe driving of autonomous vehicles and the development of high-resolution maps. In this study, it is aimed to accurately detect and identify traffic signs based on the data collected by the mobile mapping system in order to ensure the safe movement of autonomous vehicles in traffic. A low-cost method is proposed with the ResNet-50 model for an autonomous vehicle to automatically detect and recognise traffic signs while moving on the road. As a result of the model training, 0.99 accuracy and 0.016 loss were obtained. The success of the method was first observed on images randomly selected from the dataset. Then, a real-time test was performed on a low-cost webcam. The tests showed that the handled method detects and identifies the traffic sign quickly and accurately

1. INTRODUCTION

The correct management of the traffic sign inventory is an important task in ensuring the safety and efficiency of traffic flow. Most of the time this task is performed manually. Traffic signs are captured using a camera mounted on the vehicle. Manual positioning and sign recognition, checking for consistency with the existing database is performed offline by a human. However, such manual work can be extremely time consuming when applied to thousands of kilometres of roads. Automating this process would therefore significantly reduce the amount of manual work and improve safety through faster detection of damaged or missing traffic signs (Balali and Golparvar-Fard, 2016). The purpose of traffic signs is to be easily recognisable by pedestrians and drivers, to warn and guide them both day and night. The uniqueness of traffic signs and the fact that they are designed to have distinguishable features such as simple shapes and uniform colours pose a challenge to their detection and recognition. However, the development of a robust real-time traffic sign detection and recognition system is also challenging due to delays in testing time. Recent advances in the field of deep learning have shown promising results in the detection and recognition of general objects. Previous work has utilised deep learning approaches for traffic sign detection and recognition to some extent (Zhu et al., 2016a); but their evaluation has focused on a very limited subset of traffic sign categories (Zhu et al., 2016b). Various techniques proposed by the computer vision community are reviewed and compared in detail with their advantages and disadvantages. A comprehensive survey on vision-based traffic sign detection and recognition systems is provided (Wali et al., 2019). According to this study, detection in traffic sign detection and recognition systems consists of finding the traffic sign bounding box, while traffic sign recognition involves classification by giving a label to the image. The real-world traffic environment is generally complex. Another problem in traffic sign detection is that the complex environment or objects prevent traffic signs from being seen. To solve this problem, researchers have used image segmentation (Kamal et al., 2020). Such methods have proven to aid traffic sign detection in complex environments. However, it is difficult to use image segmentation in practical applications due to its high computational cost. Common traffic sign recognition methods found in the literature are: colour-based, such as (RGB), (CIELab) and (HIS) in different colour spaces (Bello-Cerezo et al., 2016); shape-based, such as Hough Transform (HT) and Distance Transform (DT); texture-based, such as Local Binary Patterns (LBP) (Gonzalez et al., 2016) and hybrid. With these methods, a feature vector is extracted from the image at a lower computational cost. The class labelling of the feature vector is then obtained using a classifier such as Support Vector Machine (SVM) or Deep Learning based (DL) methods. Among the Deep Learning based methods, Convolutional Neural Networks (CNN) have been widely adopted given their high performance in both detecting and recognising traffic signs in images and point clouds (Balado et al., 2020). In colour-based approaches (Xu et al., 2019), captured images are segmented into subsets of connected pixels that share similar colour properties. Traffic signs are then extracted by colour thresholding segmentation based on intelligent data processing. Colour space selection is important in the detection phase, so captured images are usually transformed into a specific colour space where the signs are more prominent (Luo et al., 2018). A traffic sign recognition method using a kernel-based extreme learning machine classifier with deep perceptual features is proposed (Zeng et al., 2017). Although traditional traffic sign detection methods based on machine learning have achieved better results than methods based on colour and shapes, these machine learning methods need to extract HOG features or Harr features. This makes the computational process more complex. The other most common method for classification is the use of artificial neural networks (ANNs). This method has gained increasing popularity in recent years due to advances in general purpose computing in graphics processing units (GPGPU) technologies (Satılmış et al., 2019). It is also popular due to its robustness, better adaptability to changes, flexibility and high accuracy (Ugolotti et al., 2012). It is explained (Carrasco et al., 2012) that ANN-based methods have some limitations, such as their slowness and instability in ANN training due to a very large step.

A lightweight and accurate ConvNet with a sliding window detector is proposed to detect traffic signs in GTSDB (Aghdam et al., 2016). In this study, it is stated that it is possible to locate traffic signs by processing 37.72 high resolution images per second in multi-scale by using sliding window application.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-183-2024

Speed is also an important consideration when detecting traffic signs. When an autonomous car is operating under real-time conditions, the vehicle needs to be able to accurately detect and classify traffic signs as soon as possible to minimise the risk of danger. The deep learning based approach proposed in this study is different from previous related works. Unlike traditional approaches involving manually prepared features and machine learning (Greenhalgh and Mirmehdi, 2012; Zaklouta and Stanciulescu, 2012) full feature learning with end-to-end learning is proposed. Unlike other deep learning-based traffic sign detection methods, the method in this study, which is based on Mask R-CNN, uses the region recommendation network and deep networks based on the ResNet-50 (He et al., 2016) architecture instead of using a separate method to generate region recommendations (Tabernik and Skocaj, 2019).

In this paper, the issue of learning and detecting multiple traffic sign categories for road-based traffic sign inventory management is addressed. As a main contribution, a deep learning based system is proposed to train a large number of traffic sign categories using convolutional neural networks on a dataset generated from images collected by a mobile mapping system. This system is based on Resnet50, a state-of-the-art algorithm that has shown great accuracy and speed in object detection.

2. EXPERIMENTS

In this section, we describe the procedures performed on the ResNet50 model for training and evaluating traffic sign images from a mobile mapping system to obtain the most successful result. ResNet50, a 50-layer deep convolutional neural network built specifically for image recognition applications, has produced superior performance in a number of important image recognition tests, including ImageNet. One reason for using ResNet is that it makes it possible to train up to 1000 layers, resulting in good performance. The design of ResNet50 (Shabbir et al., 2021) is built around the concept of residual learning, a method for solving the problem of vanishing gradients in deep neural networks. The images obtained by mobile mapping are preprocessed and then reshaped into a matrix according to the ResNet50 architecture.

2.1 Mobile Mapping System

Over the last few decades, the evolution of mobile mapping systems (MMSs) has been met with increasing interest. These systems are widely used to deliver valuable assets in a variety of applications. Underlying this growing popularity is the wide availability of low-cost sensors, improvements in computational resources, maturity of mapping algorithms, and the need for digital maps with geographic information systems (GIS) data. Many MMSs have gained the ability to combine hybrid sensors to complement each other and provide a more informative, robust and stable solution. Mobile mapping technology has undergone significant development in the last few decades with algorithmic advances in photogrammetry, computer vision and robotics (Wang et al., 2019). In addition, increased processing power, data collection speed and storage capacity have further facilitated the process. Current studies aim to use direct referencing and multisensor systems to minimise human intervention during data collection and processing. Mobile Mapping Systems (MMS) are divided into two categories: traditional vehicle-based systems and non-traditional lightweight/portable systems. Vehicle-based systems consist of vehicle-mounted sensor sets. Such systems achieve the highest accuracy compared to other mobile mapping systems. Vehicle-mounted systems are used for 3D city modelling, road asset management and condition assessment applications. A standard MMS includes sensors such as Global Navigation Satellite Systems (GNSS), Inertial Measurement Unit (IMU), odometers (DMI). In the mobile mapping system used, panoramic camera, tunnel camera, GNSS sensors, IMU, Odometer and LiDAR are mounted on the vehicle as shown in Figure 1.

Figure 1. MMS measurement system on a car.

2.2 Resnet-50

ResNet is a deep neural network introduced to solve the degradation problem in deep neural networks. Degradation refers to the phenomenon that as the depth of a neural network increases, its performance on the training set starts to decrease (Simonyan and Zisserman, 2015). The Residual Network, or ResNet for short, is a deep neural network architecture that has revolutionised the field of computer vision. Introduced in 2015 by researchers at Microsoft Research, ResNet made a breakthrough by enabling the training of extremely deep neural networks with up to hundreds of layers, which were previously impossible to train due to the vanishing gradient problem. One of the main advantages of the ResNet model over traditional deep neural networks is its ability to alleviate the vanishing gradient problem. As neural networks get deeper, it becomes increasingly difficult to train them due to the vanishing gradient problem, where the gradient of the loss function shrinks exponentially as it propagates back through the layers. In addition to the vanishing gradient problem, ResNets offer other advantages over traditional deep neural networks. They are computationally efficient, require fewer parameters and operations, and are easier to optimise, leading to faster convergence and better results. ResNet-50 is one of the variants of ResNet (He et al., 2016), a convolutional neural network. ResNet-50 has over 23 million trainable parameters, which is much smaller than existing architectures despite having 50 layers. Figure 2 shows the Resnet-50 architecture. Resnet links have emphasised the hopping approach. Architecturally, if any layer is detrimental to the performance of the model in a flat network, it is skipped due to the presence of jump links. A small change has been made in ResNet-50 from older models, whereas before that shortcut links skipped two layers, in ResNet-50 and later models they skip three layers.

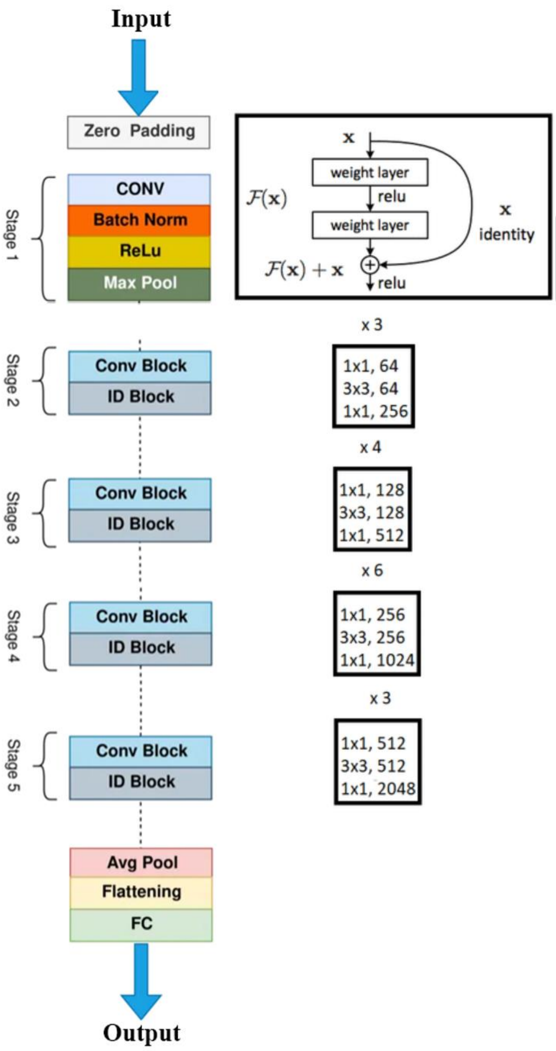


Figure 2. Resnet-50 Model architecture.

2.3 Dataset

Various machine learning and deep learning algorithms have been proposed on the BelgianTS and German Traffic Sign Detection Benchmark (GTSRB) open datasets, which are widely used in traffic sign detection and identification studies. In reality, these two datasets cannot adequately represent complex traffic sign detection scenarios. The reason for this is that their image resolution and image size are low and they do not reflect realistic road conditions. In this study, the images obtained from the mobile mapping system were used as the data set as shown in Figure 3. The size of the images in the data set was reduced to 32x32 pixels. It can be seen in Figure 4 that as the camera gets closer to the sign, the image resolution gradually increases and the angle changes. Therefore, when selecting the validation set, the images need to be blended to ensure that the validation set is meaningful. The dataset contains 73,139 traffic sign samples and consists of 43 different classes. 20% of the dataset is reserved for testing.

Figure 3. MMS vehicle.

Figure 4. Traffic sign captured from the image received from the MMS vehicle.

2.4 Settings

For networking and model training, we use TensorFlow Keras, a deep learning library built in Python programming language, which provides suitable conditions for defining and training all types of deep learning models. Specific libraries such as torch, torchvision, sklearn, matplotlib, seaborn, os, pandas, etc. are used to build the network for traffic sign detection and identification. Simulation experiments are performed on a NVIDIA GeForce GTX 1660Ti GPU. The batch size and epoch are set to 32 and 50, respectively. While Adam is selected as optimizer, Softmax is used as activation function. Some tweaks are made to improve the model performance. The changes affect the required computation time, convergence speed and the utilisation of processing units. CNN architectures perform better with large data sets. The performance of deep learning models can be improved by increasing the existing data instead of collecting new data. For this reason, data augmentation was applied to increase the size of the data set in this study. The basic principle of data augmentation is to create additional training data by

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-183-2024 | © Author(s) 2024. CC BY 4.0 License.

applying some deformation operations to the existing data (Salamon and Bello, 2017). There are many data augmentation techniques such as rotating the image at different angles, rotating horizontally and vertically, adding noise and colour manipulation to the image.

Softmax is usually used in the classification layer of the pre-trained deep model. In this research, Softmax is used as the activation function in classification. The Softmax function ensures that the elements of a vector are in the range [0, 1] and sum to 1. This property allows each element of the output vector to represent the probability of belonging to a class. The softmax formula is given in Equation 1.

$$softmax(x)i = \frac{e^{xi}}{\sum_{n=1}^{N} e^{x_n}} \; for\, j = 1 \ldots . N \qquad (1)$$

where x is the vector of raw outputs from the neural network. The i-th entry in the softmax output vector softmax(x) can be thought of as the predicted probability of the test input belonging to class i. The value of e ≈ 2.718

2.5 Results and Analyses

Accuracy is one of the most used metrics to measure the success of a deep learning model. Accuracy is known as the ratio of the number of correct predictions to the total number of predictions made and could be calculated with the following equations:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

where TP = count of examples for which the actual label is true and the model made a correct prediction
TN = number of examples for which the actual label is false and the model made a correct prediction
FP = count of samples for which the actual label is false but the model made an incorrect prediction
FN = count of examples for which the actual label is true but the model made an incorrect prediction

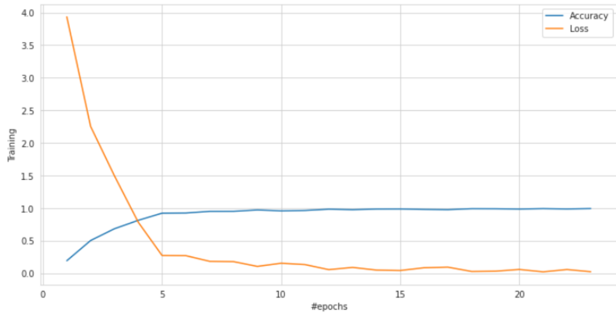
The accuracy of the proposed model and its comparison with other CNN algorithms are calculated and given Table 1.

CNN	Accuracy (-)
AlexNet	0.894
GoogLeNet	0.898
VGG16	0.905
VGG19	0.903
ResNet-50	0.901
Inceptionv2	0.903
Inceptionv3	0.901
Inceptionv4	0.89
SENet	0.875
Our Method	**0.996**

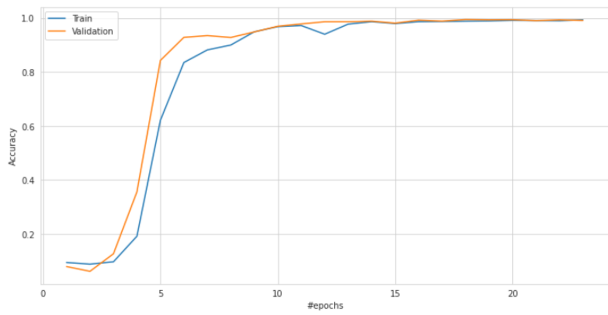
Table 1. Classification accuracy (in percent) comparison of CNN-Based models).

In this study, it is observed that the proposed model outperforms the others in terms of accuracy.

The change in Accuracy and Loss values of the network trained for traffic sign detection is shown in Figure 5. As a result of 50 epochs taken during training, Loss: 0.03172951191663742, Accuracy: 0.9916598200798035.



(a) Accuracy and Loss values obtained as a result of model training.



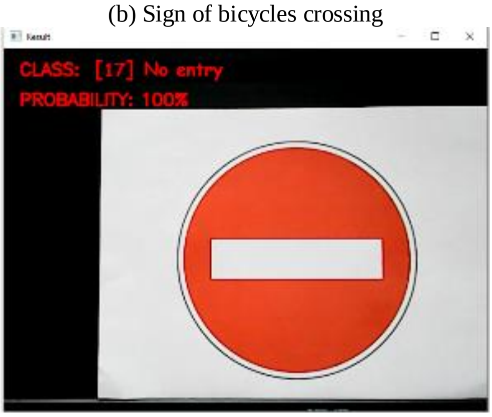
(b) Train and Validation values obtained as a result of model training.

Figure 5. Training result learning curves of the proposed method

After training the proposed model, it was first tested in a digital environment on samples from the dataset. After observing that the results were successful, a test environment was set up to reflect the real environment. With a simple webcam, real-time tests were performed on samples generated in real traffic sign sizes. Figure 6 shows examples of some traffic sign classes from the tests. As can be seen, the developed model correctly classifies the traffic signs by processing the images obtained from the webcam. As a result of the tests of the proposed model, both the class IDs and the name of the sign are written in red on the upper left side of the result screen for the traffic signs No passing, Bicycles crossing, No entry, Stop, Speed limit 50 km/h and Turn left ahead. The results of the selected examples proved that the model predicts the class of the traffic sign quickly and accurately with high accuracy for traffic signs with different colours and shapes.

(a) Sign of no passing

(b) Sign of bicycles crossing



(c) Sign of no entry

(d) Stop sign

(e) Sign of speed limit 50 km/h

(f) Sign of turn left ahead

Figure 6. Tests with real size traffic signs of some classes and classification results.

3. CONCLUSIONS

In this paper, ResNet50 is used to extract features of images, detect and identify traffic signs. To reduce the computation time, ResNet50 is used as it is a pre-trained convolutional neural network with 50 layer depths. The detector proved to be very fast and has very high results with an average accuracy of over 95% for multiple images. The developed model quickly detected and correctly classified 43 different classes of traffic signs.

The ResNet model has a complex architecture and requires excess computational resources for training. Its deeper architecture may overfit small datasets or require more time for training on large datasets.

In future studies, the model can be trained with a dataset containing more images. By improving the quality of the cameras, much better results can be obtained. The effect of image rotation and image enhancement on model performance can also be investigated.

REFERENCES

Aghdam, H. H., Heravi, E. J., & Puig, D., 2016: A practical approach for detection and classification of traffic signs using convolutional neural networks. *Robotics and autonomous systems*, 84, 97-112.

Balado, J., González, E., Arias, P., & Castro, D., 2020: Novel approach to automatic traffic sign inventory based on mobile mapping system data and deep learning. *Remote Sensing*, 12(3):442. https://doi.org/10.3390/rs12030442

Balali, V., & Golparvar-Fard, M., 2016: Evaluation of multiclass traffic sign detection and classification methods for U.S. roadway asset inventory management. *Journal of Computing in Civil Engineering*,30(2),04015022. https://doi.org/10.1061/(ASCE)CP.1943-5487.000049

Bello-Cerezo, R., Bianconi, F., Fernández, A., González, E., & Di Maria, F., 2016: Experimental comparison of color spaces for material classification. *Journal of Electronic Imaging*, 25(6):061406. https://doi.org/10.1117/1.JEI.25.6.061406

Carrasco, J. P., de la Escalera, A., & Armingol, J. M., 2012: Recognition stage for a speed supervisor based on road sign detection. *Sensors*, 12(9), 12153-12168.

Gonzalez, E., Bianconi, F., & Fernandez, A., 2016: An investigation on the use of local multi-resolution patterns for image classification. *Information Sciences*, 361-362, 1-13. https://doi.org/10.1016/j.ins.2016.04.044

Greenhalgh, J., & Mirmehdi, M., 2012: Real-time detection and recognition of road traffic signs. *IEEE Transactions on Intelligent Transportation Systems*, 13(4), 1498-1506. https://doi.org/10.1109/TITS.2012.2208909

He, K., Zhang, X., Ren, S., & Sun, J., 2016: Deep residual learning for image recognition. *In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

Kamal, U., Tonmoy, T. I., Das, S., & Hasan, M. K., 2020: Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint. *IEEE Transactions on Intelligent Transportation Systems*, 21(4), 1467-1479. https://doi.org/10.1109/TITS.2019.2911727.

Luo, H., Yang, Y., Tong, B., Wu, F., & Fan, B., 2017: Traffic sign recognition using a multi-task convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 19(4), 1100-1111.

Salamon, J., & Bello, J. P., 2017: Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3), 279-283. https://doi.org/10.1109/LSP.2017.2657381.

Satılmış, Y., Tufan, F., Şara, M., Karslı, M., Eken, S., & Sayar, A., 2019: CNN based traffic sign recognition for mini autonomous vehicles. In Information Systems Architecture and Technology: *Proceedings of 39th International Conference on Information Systems Architecture and Technology–ISAT* 2018: Part II (pp. 85-94). Springer International Publishing.

Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A. and Dar, S.H., 2021: Satellite and scene image classification based on transfer learning and fine tuning of ResNet50. *Mathematical Problems in Engineering*, Article ID 5843816 1-18. https://doi.org/10.1155/2021/5843816

Simonyan, K. and Zisserman, A., 2015: Very deep convolutional networks for large-scale ımage recognition. *In 3rd International Conference on Learning Representations (ICLR 2015)*, 1–14. Computational and Biological Learning Society.

Tabernik, D., & Skočaj, D., 2019: Deep learning for large-scale traffic-sign detection and recognition. *IEEE Transactions on Intelligent Transportation Systems*, 21(4), 1427-1440.

Ugolotti, R., Nashed, Y. S., & Cagnoni, S., 2012: Real-time GPU based road sign detection and classification. *In Parallel Problem Solving from Nature-PPSN XII: 12th International Conference*, Taormina, Italy, September 1-5, 2012, Proceedings, Part I 12 (pp. 153-162). Springer Berlin Heidelberg.

Xu, X., Jin, J., Zhang, S., Zhang, L., Pu, S., & Chen, Z., 2019: Smart data driven traffic sign detection method based on adaptive color threshold and shape symmetry. *Future generation computer systems*, 94, 381-391.

Wang, Y., Chen, Q., Zhu, Q., Liu, L., Li, C., & Zheng, D., 2019: A survey of mobile laser scanning applications and key techniques over urban areas. *Remote Sensing*, 11(13), 1540.

Wali, S. B., Abdullah, M. A., Hannan, M. A., Hussain, A., Samad, S. A., Ker, P. J., & Mansor, M. B., 2019: Vision-based traffic sign detection and recognition systems: Current trends and challenges.*Sensors*,19(9),2093. https://doi.org/10.3390/s19092093

Zaklouta, F., & Stanciulescu, B., 2012: Real-time traffic-sign recognition using tree classifiers. *IEEE Transactions on Intelligent Transportation Systems*, 13(4), 1507-1514. https://doi.org/10.1109/TITS.2012.2225618.

Zeng, Y., Xu, X., Shen, D., Fang, Y., & Xiao, Z., 2017: Traffic sign recognition using kernel extreme learning machines with deep perceptual features. *IEEE Transactions on Intelligent Transportation Systems*, 18(6), 1647-1653. https://doi.org/10.1109/TITS.2016.2614916.

Zhu, Y., Zhang, C., Zhou, D., Wang, X., Bai, X., & Liu, W., 2016: Traffic sign detection and recognition using fully convolutional network guided proposals. *Neurocomputing*, 214, 758-766. https://doi.org/10.1016/j.neucom.2016.07.009

Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S., 2016: Traffic-sign detection and classification in the wild. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2110-2118. https://doi.org/10.1109/CVPR.2016.232.

APPENDIX

The data set used in this study will be also utilized in the PhD thesis and another possible future studies. For that reason, the data set is not shared as an open access.