PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Acta Informatica Pragensia
2022, Volume 11, Issue 3, pp. 458–466
https://doi.org/10.18267/j.aip.195

Miscellanea Open Access

Use of Intelligent Navigation and Crowd Collaboration for Automated Collection of Data on Transport Infrastructure

Tomáš Tvrzský

Telematix Software, a.s., Branická 66, 147 00 Prague, Czech Republic

Corresponding author: Tomáš Tvrzský (tvrzsky@telematix.cz)

Abstract

The article briefly presents the main results of an applied research project to the professional public. The project output is a solution that enables the recognition of selected types of traffic signs using artificial intelligence for image recognition. This computationally intensive process is implemented in mobile phones. In order to achieve the involvement of the general public in the collection of data on transport infrastructure, the entire solution is part of navigation for mobile phones and supported by two functions that motivate users to collect data, i.e., scan the area in front of the vehicle with the phone's camera. The first function is the projection of the route into the real environment (the so-called augmented reality mode), and the second function is the possibility of video recording the drive. The video recording is cryptographically signed to ensure authenticity in administrative or judicial proceedings, e.g., when proving the course and circumstances of a traffic accident. The collection of data on transport infrastructure is completely anonymous in compliance with applicable laws. The data about recognized traffic signs will not only serve the navigation provider to improve the user experience but the processed data will also be exported to community-created world maps (project OpenStreetMap).

Keywords

Navigation; Camera; Video recording; Intelligent transportation systems; Crowdworking; Cryptography; Real-time image processing; Traffic sign recognition.

Citation: Tvrzský, T. (2022). Use of Intelligent Navigation and Crowd Collaboration for Automated Collection of Data on Transport Infrastructure. *Acta Informatica Pragensia*, 11(3), 458–466. https://doi.org/10.18267/j.aip.195

1 Introduction

The aim of the article is to briefly present a solution developed as part of an applied research project that was supported by the Technology Agency of the Czech Republic. The main goal of the project was to develop a software module that would be integrated into an existing car navigation solution of the main investigator of the project, Telematix Software company. This module enables the recognition of selected types of traffic signs using artificial intelligence approaches for the task of image recognition. This computationally intensive process is implemented and executed in mobile phones equipped with processors with artificial intelligence support, including image recognition.

The implemented solution has some specifics by which we want to motivate navigation users to use their camera and anonymously collect data about transport infrastructure. The results of traffic sign recognition will be integrated not only into the Telematix Software map data, but also into OpenStreetMap [1] – a collaborative project that creates a free editable geographic database of the world.

The structure of the article corresponds to the aim, which is to briefly present the main parts of the proposed and implemented solution to the professional public. Section 2 presents the starting points of the problem context and the research approaches applied in our research. Section 3 briefly presents the designed and implemented solution. The conclusion section summarizes the main benefits and further development opportunities.

2 Background

The proposed solution uses the principles of crowdworking, where many people work on a certain task (Jäger et al., 2019). In this case, it is the task of automated mapping of transport infrastructure. Crowdworking is typically implemented in a narrowly focused community working on a selected task that can be parallelized. Compared to regular work, there is no substantial social support between workers (who often do not even know each other), but the motivation for work performance tends to be different aspects such as financial remuneration or other benefits (Ihl et al., 2020).

In order to use the principle of crowdworking in our case, it is necessary to figure out how to properly motivate people for such an activity. In addition to altruistic motivation – helping to improve publicly available and globally used map materials – such work must also be motivated by other benefits. Therefore, Dynavix car navigation[2], which has tens of thousands of users and is available for Apple iOS and Google Android mobile device operating systems, was chosen as the default platform for the implementation of the traffic sign recognition module. These users are offered new features:

- Augmented reality navigation, which requires the use of a mobile phone camera. The obtained image data can then be processed further while ensuring the full anonymity of the user.
- Recording video footage of events in front of the vehicle (dashcam function) and its authentication using cryptography.

The use of real-time video analysis is not a new task from the research point of view. There are a number of context-oriented solutions, e.g., fire and smoke detection (Saponara et al., 2020), monitoring engineering structures as a part of monitoring transport infrastructure facilities (Gura et al., 2021), automatic road environment classification (Tang & Breckon, 2011). The area dealing with traffic sign recognition has been developing rapidly, especially in the last 15 years. New solutions dealing with traffic sign recognition using various methods have been and are still being proposed (e.g., Ruiz & Serrat, 2022; Khnissi et al.,

[1] See, https://www.openstreetmap.org

[2] See, https://www.dynavix.com

2022). The vast majority of the proposed solutions are so-called ex-post, where data processing takes place on powerful computing machines.

However, the specific conditions in which we decided to implement it are new – when providing a navigation service, the mobile phone directly recognizes traffic signs in real time from the video, which ensures anonymity and secure data processing from the navigation user perspective (i.e., no image materials are sent to the server). Although conceptual proposals for a technical solution existed[3], the practical implementation itself is rarely mentioned in the professional literature. In practice, we found only one navigation system [2022] that recognizes speed limits from traffic signs in real time, according to our search of navigation applications for mobile phones – this is Sygic navigation[4]. However, this task is simpler than the solution presented here. Moreover, no such solution existed at the time of project preparation [2018]. Based on the aforesaid, the presented solution can still be considered original.

3 Brief Description of the Proposed Solution

3.1 Module for recognizing selected traffic signs in a mobile phone

TensorFlow (2022) was used as the main tool. It is an end-to-end open source platform for machine learning. Deep learning methods were used for traffic sign recognition, which enabled simultaneous sign detection and classification. A general pre-trained classification model was used, and its recognition capabilities were modified to recognize Czech traffic signs. The model is able to recognize multiple traffic signs, their class and location in the source image.

In accordance with the project proposal, the model focused on the recognition of traffic signs that may be important for navigation or for route calculation:

- main road and its end,
- give right of way,
- stop give way,
- maximum permitted speed and its end,
- ban on the entry of vehicles weighing more than,
- ban on the entry of vehicles with a width over,
- ban on the entry of vehicles with a height above,
- ban on the entry of all vehicles in both directions,
- ban on the entry of all vehicles,
- one-way traffic.

Artificial intelligence learning requires an extensive dataset of traffic signs, which would be based on videos (or photos) from the driver's point of view that capture regular traffic in the Czech environment. Because no such dataset was freely available, the researchers decided to create one. The dataset was produced in Prague traffic and contains photos in full resolution and colour[5]. The photos were taken with several cameras and under different lighting conditions. A wide variety of captured photos improves the classification model learning. Each traffic sign had to be manually labelled and metadata (e.g., type of traffic sign) were added to each traffic photo.

[3] At the time of the project proposal, there were at least 7-year-old conceptual proposals (e.g., Ling & Seng, 2011). However, for example, the Scopus citation database contains only five records for the keywords ["traffic sign recognition" AND mobile AND phone] searched anywhere in the title, abstract or keywords.

[4] See, https://www.sygic.com

[5]A dataset containing over 10,000 training photos was created between 2020 and 2022.

The photos and metadata from the previous step needed to be normalized. Normalization was done using Python scripts and libraries. During this preparation, individual photos and their metadata were recoded to the given maximum dimensions – 1280x720 pixels in landscape mode and 720x1280 pixels in portrait mode.

Furthermore, during this step, it was determined which classes (types of traffic signs) the learning algorithm should take into account. The entire input dataset was divided according to the selected criteria into three subsets:

- training set – contains photos and metadata on which the learning model is trained, i.e., on which the algorithm creates a model of the found relationships between the data;
- test set – serves the learning algorithm for evaluating the relationships found, which were created by training; thanks to the evaluation of the test set directly during training, the algorithm can correct its learning and thus avoid, for example, overlearning the model (i.e., a situation where the learned model would work very well on training data, but fail on real, as yet unknown data);
- validation set – serves for additional evaluation of the quality of the finished learned model.

When dividing the input data into these three subsets, it is necessary not to share any data between the individual subsets (if possible) because otherwise, there is a bias in the validation step or a low-quality model that would not perform well in the localization and classification of traffic signs in real photos. The prepared data from the previous step are used as input for the learning algorithm. The TensorFlow framework was used to train the model. This framework was controlled using Python scripts.

The learning process is iterative and takes place in individual steps. The learning algorithm evaluates the quality of the current model using the so-called loss function. The value of this function expresses the model's error rate – both the localization error (searching for the place in the photo where the traffic sign should be located) and the classification error (determining the type of sign found).

The longer the learning process runs, the more iterations will take place, and a better model can be created. The loss function starts with high values (sharply greater than 1) and gradually approaches lower values (about 0.05), where it oscillates. With further iterations, the learned model stops improving. After performing a sufficient number of training iterations (in our case, approximately 40,000 iterations) and the loss function decreasing (to a value oscillating between 0.01 and 0.09), the learning was terminated. The final recognition model (so-called inference graph) was generated from the saved last learning intermediate state.

The resulting recognition model was used to verify its quality. Again, Python scripts and libraries provided by the TensorFlow framework and a prepared input subset of validation data can be used. For each input photo and its metadata, the model performs traffic sign recognition – the model can generate many labels (i.e., potentially identify multiple traffic signs from a single photo) containing the exact location and size of the region of interest, along with the classified class (i.e., the type of traffic sign), and determines the confidence value. The determination confidence is a value from the interval [0;1], where 0 means the lowest confidence and 1 means the highest confidence (100%). In our case, we removed all traffic sign designations that had a confidence lower than 0.6.

Subsequently, the traffic sign recognition module was integrated into the Dynavix navigation application so that the module was able to use the instruction set and hardware support of mobile phones that contain dedicated processors for image processing. The NNAPI library [6] has been modified for hardware acceleration of traffic sign recognition. Thus, Dynavix navigation can itself check the available hardware

[6] See, https://www.tensorflow.org/lite/performance/nnapi/

resources of the mobile phone and delegate the recognition process to the Graphics Processing Unit, Digital Signal Processor or other specialized hardware for processing neural network operations.

As part of the first tests, the processor load was tested in image recognition for different classes of mobile phones. From our first results, it can be said that phones with a performance at the level of 150,000 AnTuTu points (today's lower class) do not have fundamental performance problems with traffic sign recognition.

3.2 Module for projection of calculated route into real environment

The aim of the development of the augmented reality module using the projection of the calculated route into the real environment is to motivate the end users of the system to use the Dynavix navigation software in a way that allows the processing of the image obtained from the mobile phone camera (i.e., for the mobile phone to "have a view" of the scene in front of the vehicle)[7].

The unique function of displaying the calculated route for navigating to the real scene was used for this purpose. So far, only Sygic navigation provides such a function. The results of the implementation are not yet entirely satisfactory but suitable. The current state is shown in Figure 1.

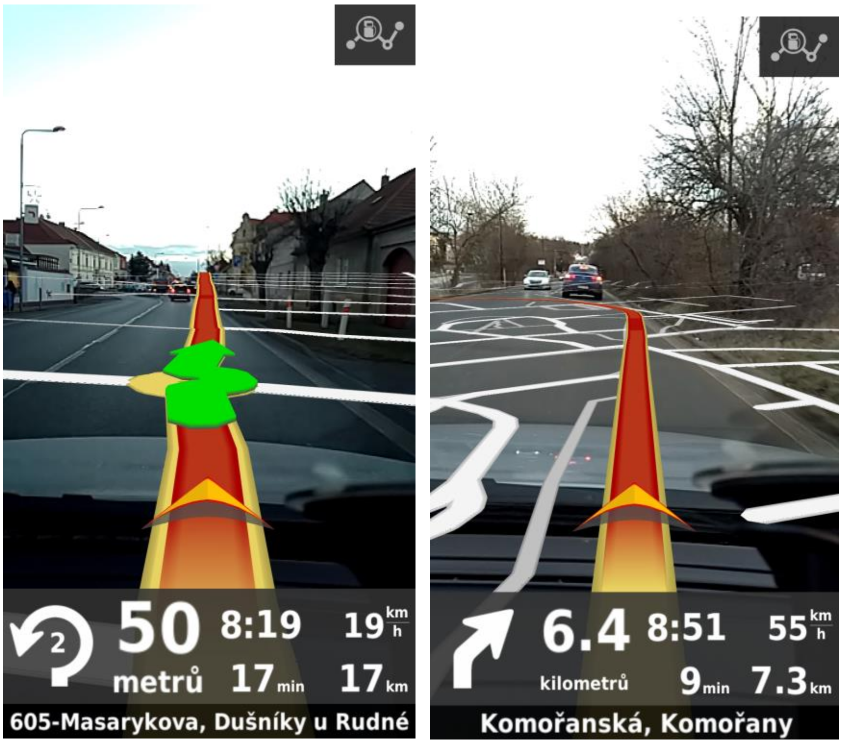


Figure 1. *Illustrative sample from a test drive.*

3.3 Module for authenticated video recording of driving and events around the vehicle

The goal of the development of the module for authenticated video recording of driving and events around the vehicle is, similar to the module for route projection into augmented reality, to offer end users a function that will motivate them to use the mobile phone camera while driving. In principle, the unique function of authentication of the video file is also intended to serve this purpose, which will prove that no one has subsequently manipulated the footage from the on-board camera in the event of a hearing in an

[7] In the future, using another camera on the car board can also be considered if the data from this camera are available to third parties – this can happen with the expansion of the Android Auto and Carplay systems.

administrative or judicial proceeding (e.g., when demonstrating the course and circumstances of a traffic accident).

The video file is authenticated using additional data and metadata that can be easily obtained from mobile phone sensors (e.g., position based on World Geodetic System, or WGS, date, time, etc.). At the same time, a single-purpose web application was developed that, after uploading a video file, verifies that the video was captured with Dynavix software and has not been edited.

Several settings were designed for recording with the phone's camera to adapt the recording well to the mobile phone's capabilities. During development, the WebM recording format (WebM, 2022) was used for the output video, guaranteeing good compatibility across different platforms. The WebM format belongs to the group of Extensible Binary Meta Language video formats, which is a generalized file format for any kind of data, aiming to be a binary equivalent to Extensible Markup Language or XML (libEBML, 2022). It provides a basic framework for storing data in XML-like tags. This format contains the "signature" tag, where the signature of the video file is stored. The video is signed on an ongoing basis in the phone's memory using the Elliptic Curve Digital Signature Algorithm or ECDSA on the "P-256" curve, "SHA-256" hash. The VP8 video codec (RFC 6386, 2011) with subtitles in Web Video Text Tracks format (W3C WebVTT, 2019) was used for video recording. Data from the Global Positioning System (position based on WGS, time in Coordinated Universal Time, or UTC) and machine-readable data for possible further processing in JavaScript Object Notation format are stored in the captions.

3.4 Server part with statistical processing of image recognition results and database sharing with other users

The recognition results (typically very data-efficient, e.g., type of traffic sign, the position where the recognition took place, the direction in which the vehicle was travelling and where the mobile phone was located) were processed in the central server part of the system, where they are statistically compared with results obtained from other users of the system. Based on these results and according to the settings of the recognition reliability requirements, the server part will then evaluate whether the performed traffic sign recognition can be considered validated and provide signs for further use. The server part is built on a user interface of our map server. An example of working with detected traffic signs in the map is shown in the Appendix. When working with the obtained data, it is possible to use the following parameters:

- setting the period for which recognized signs should be displayed;
- selecting the type of sign to display;
- selection of the probability range of traffic sign recognition (e.g., only signs for which the recognition probability was higher than 75% can be displayed);
- selecting the road class on which the sign was recognized;
- filter for determining the distance to the nearest intersection from the point of traffic sign recognition;
- the possibility of displaying the vehicle speed vector for each case of traffic sign recognition.

The mentioned parameters can be easily adjusted in the web interface and the filtered data can be displayed on the map. Several support functions have also been implemented for the statistical processing of data on recognized traffic signs in a map context:

- map-matching of events with the assignment to the correct side of the road (use of the velocity vector);
- options for filtering data assigned to the map;
- scenarios were developed for several typical use cases (e.g., taking into account the location of the nearest intersection, taking into account the time of day, taking into account the weather).

4 Conclusion

The presented solution answers the issues associated with collaborative projects such as OpenStreetMap. The main problem is the varied quality of geodata in individual regions. This is due to the fact that the individual volunteers who manage and expand the map coverage do not have uniform approaches and tools for this work. The quality of map data is very important for car navigation. Many errors when navigating using navigation applications on a mobile phone are caused by errors in the map data and, therefore, not by the navigation application itself. In this context, we would not only like to improve the user experience for users of Dynavix navigation, but we would also like to contribute to the mapping of geodata additions (mainly in the Czech Republic) in OpenStreetMap.

We hope that the results of this research project will be appreciated by many people around the world who use map data from OpenStreetMap for their navigation. We would like to export to the basic OpenStreetMap database once a year. In the coming months, we plan to further improve individual modules, based on feedback from Dynavix navigation users.

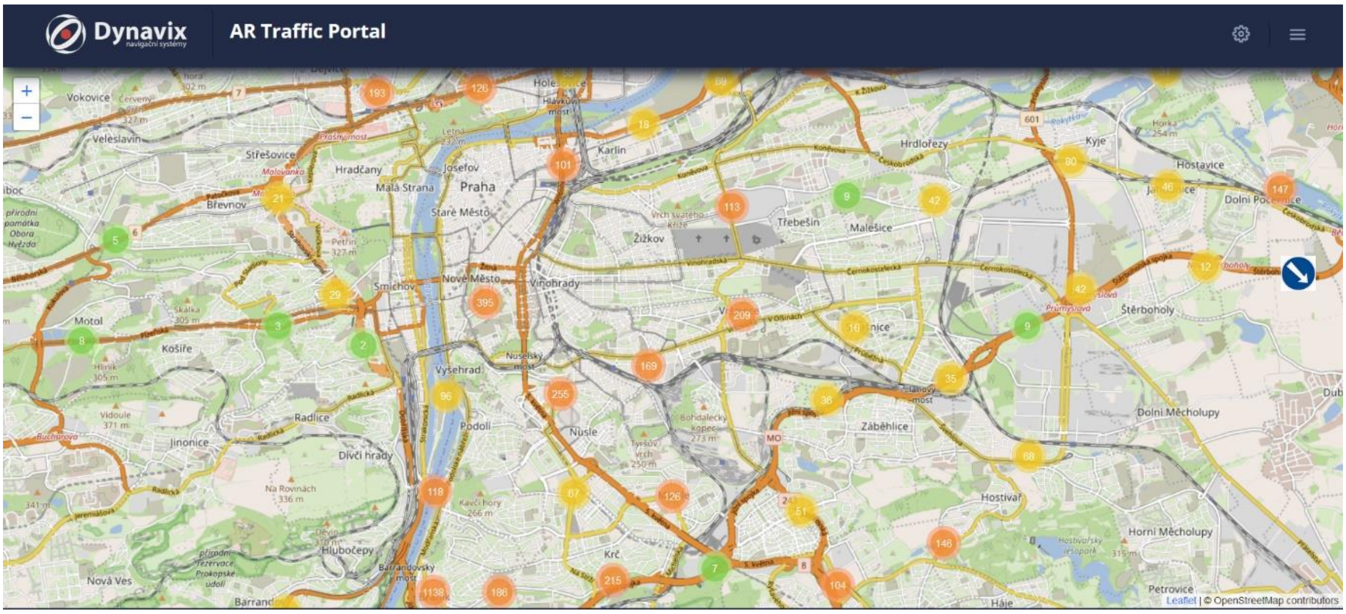
Additional Information and Declarations

Funding: The project researchers received public financial support from research grant CK01000099 administered through the Technology Agency of the Czech Republic.

Conflict of Interests: The author works in a company developing the Dynavix navigation used in the presented research project.

Author Contributions: The author confirms being the sole contributor of this work.

Appendix



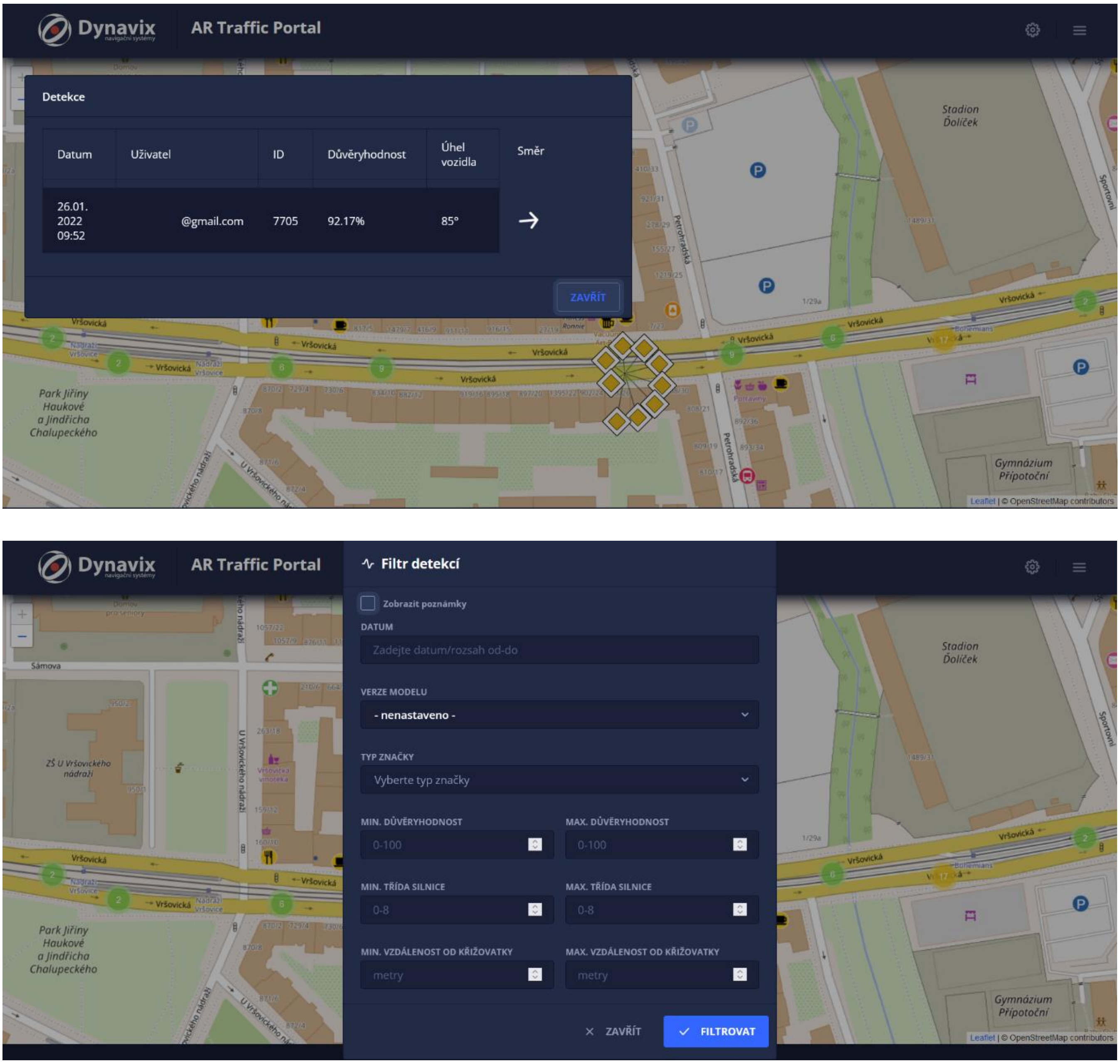


Figure A1: *Example of working with detected traffic signs in the map of the server part of the proposed solution. The first image shows the number of detected traffic signs in Prague. The second image shows detailed information about the selected traffic sign detection (detection date, ID, confidence, vehicle angle, and direction in which the traffic sign was detected). The last image shows the detection filter setting options.*

References

Gura, D., Markovskii, I., Khusht, N., Rak, I., & Pshidatok, S. (2021). A Complex for Monitoring Transport Infrastructure Facilities Based on Video Surveillance Cameras and Laser Scanners. *Transportation Research Procedia*, 54, 775–782. https://doi.org/10.1016/j.trpro.2021.02.130

Ihl, A., Strunk, K. S., & Fiedler, M. (2020). The mediated effects of social support in professional online communities on crowdworker engagement in micro-task crowdworking. *Computers in Human Behavior*, 113, Article no. 106482. https://doi.org/10.1016/j.chb.2020.106482

Jäger, G., Zilian, L. S., Hofer, C., & Füllsack, M. (2019). Crowdworking: working with or against the crowd? *Journal of Economic Interaction and Coordination*, 14(4), 761–788. https://doi.org/10.1007/s11403-019-00266-1

Khnissi, K., Jabeur, C. B., & Seddik, H. (2022). Implementation of a Compact Traffic Signs Recognition System Using a New Squeezed YOLO. *International Journal of Intelligent Transportation Systems Research*, 20(2), 466–482. https://doi.org/10.1007/s13177-022-00304-6

libEBML. (2022). EBML – Extensible Binary Markup Language. https://matroska-org.github.io/libebml/

Ling, W. H., & Seng, W. C. (2011). Traffic sign recognition model on mobile device. In *2011 IEEE Symposium on Computers & Informatics* (pp. 267-272). IEEE. https://doi.org/10.1109/isci.2011.5958925

RFC 6386. (2011). VP8 Data Format and Decoding Guide. https://datatracker.ietf.org/doc/html/rfc6386

Ruiz, I., & Serrat, J. (2022). Hierarchical Novelty Detection for Traffic Sign Recognition. *Sensors*, 22(12), 4389. https://doi.org/10.3390/s22124389

Saponara, S., Elhanashi, A., & Gagliardi, A. (2020). Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *Journal of Real-Time Image Processing*, 18, 889–900. https://doi.org/10.1007/s11554-020-01044-0

Tang, I., & Breckon, T. P. (2011). Automatic Road Environment Classification. *IEEE Transactions on Intelligent Transportation Systems*, 12(2), 476–484. https://doi.org/10.1109/tits.2010.2095499

TensorFlow. (2022). Create production-grade machine learning models with TensorFlow. https://www.tensorflow.org

W3C WebVTT. (2019). WebVTT: The Web Video Text Tracks Format. https://www.w3.org/TR/webvtt1/

WebM. (2022). WebM: an open web media project. https://www.webmproject.org

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951