



OPEN

DATA DESCRIPTOR

Camera and telemetry data from Slovak roads in various light and weather conditions

Marek Galinski¹✉, Tomáš Milesich², Matej Janeba¹, Ján Danko², Peter Lehoczky¹,
Luboš Magdolen², Lukáš Šoltés¹ & Adam Tomčala¹

In this paper, we describe the dataset captured with our proprietary data capture solution mounted on top of a Land Rover Defender vehicle. The captured data are the real data of drives on various Slovak roads. The total dataset consist of almost 33 hours of driving with a automotive grade FPD Link camera with 30 fps and with additional sensors such as high-precision GNSS sensor and modem towards mobile data connectivity LTE and 5G. There are various road types and weather and conditions. Road types are 2-lane highway, 3-lane highway, urban environment, 1st class road (main, national), 2nd class road (main, regional), 3rd class road (local). Weather and light conditions are direct sunlight, normal daylight, rain/fog and dawn/twilight. All frames have captured its position, velocity, heading and connectivity information.

Background & Summary

The dataset AIL Albus, presented in this paper, was recorded as part of a research project funded by the Ministry of Transport of the Slovak Republic. Researchers from the Automotive Innovation Lab (AIL) at STU tested the readiness of Slovak road infrastructure for the deployment of intelligent transportation systems¹. In line with open-science principles and acknowledging that projects funded by public money should have publicly available outcomes, we have prepared the collected data for public release to facilitate further research. This paper describes how the data were collected, the sensors used, the organization of the dataset, and how it can be processed and interpreted.

Among the key aspects of intelligent and autonomous vehicles, environment perception and network connectivity are paramount. The ability to accurately detect relevant objects around a vehicle—such as other cars, cyclists, pedestrians, and traffic signs—is crucial not only for future autonomous vehicles but also for various intelligent driver assistance systems. Technologies like intelligent speed assistants or lane-keeping assistants, which rely on reliable detection of traffic signs or road lanes, are already required by law in certain regions, such as the EU. On the other hand, vehicle-to-everything (V2X) communication technology, which enables wireless connectivity between vehicles and other road participants or infrastructure, is envisioned to play a pivotal role by facilitating the timely and reliable exchange of critical information.

In the field of computer vision, object detection and classification are well-established research areas where deep neural networks have been applied with significant success. Numerous datasets have been published for training models in this domain. The German Traffic Sign Recognition Benchmark (GTSRB)² and the German Traffic Sign Detection Benchmark (GTSDB)³ are two of the earlier, robust, and widely used datasets by the research community. Similar, often smaller, datasets focused on specific countries' road and traffic signs have also been released, such as the MASTIF dataset captured on Croatian roads⁴. While the traffic signs in our dataset are not manually and perfectly labeled, it offers continuous video footage, capturing multiple frames of each visible road sign from various distances. This provides opportunities for tasks like continuous detection and tracking of the same road sign across consecutive frames, enhancing temporal analysis and model robustness.

In the domain of connected vehicles, stringent requirements exist, particularly concerning the reliability and low latency of communication between various traffic participants, to support crucial safety features. While specific technologies have been standardized, it is essential to evaluate them in realistic scenarios. Numerous experiments

¹Slovak University of Technology, Faculty of Informatics and Information Technologies, Bratislava, 842 16, Slovakia.

²Slovak University of Technology, Faculty of Mechanical Engineering, Bratislava, 812 31, Slovakia. ✉e-mail: marek.galinski@stuba.sk



Fig. 1 Land Rover Defender equipped with sensors for data collection.

have been conducted for this purpose, both in specialized test beds and in regular day-to-day traffic (e.g.⁵). Such tests are critical to assessing the practical capabilities of emerging technologies like 5G and communication paradigms such as Vehicle-to-Everything (V2X), and how well they meet the expected communication requirements. Moreover, they provide valuable insights into the current state of technologies deployed in real-world conditions. For instance, the authors of⁶ propose a model for evaluating mobile network performance in video-based tele-operated driving scenarios, validated through their own measurements. Similarly, our AIL research team has utilized the dataset presented in this paper to assess mobile network performance in Slovakia for various potential use cases, as documented in⁷. Specifically, this assessment was conducted with respect to the V2X communication requirements defined by 3GPP⁸ for V2X applications across different levels of driving automation⁹.

Additionally, in the context of traffic flow and vehicle dynamics modeling, large-scale datasets like the Next Generation Simulation (NGSIM) project¹⁰ provide multi-vehicle trajectory data for broad traffic analysis. In contrast, our dataset, limited to a single vehicle, is less suited for large-scale modeling. However, the use of a highly precise GNSS system with up to 14 mm horizontal accuracy offers value for some studies, such as analyzing individual vehicle behavior.

In this paper, we contribute a dataset that can support various research efforts in intelligent transportation systems, with multiple potential use cases. The dataset consists of continuous camera recordings from a single vehicle, coupled with mobile network telemetry data. Each measurement includes highly accurate geographical positioning data from GNSS (with up to 14 mm horizontal accuracy). These data can be used to model and evaluate mobile network performance or assess GNSS positioning in different environments, such as under bridges or in mountainous terrain, where visual data from the camera can help contextualize the signal environment.

Regarding visual data, we also provide synthetically labeled road lanes and traffic signs using pre-trained neural network models, which can be utilized directly. Since the data were collected by driving through selected road sections multiple times and under specified weather conditions, the dataset allows for comparisons across different weather conditions. This opens up possibilities for studies related to the comparison of road sign visibility, detection of weather conditions, or analysis of road surfaces.

Furthermore, the dataset captures the current state of various technologies in Slovakia, where all measurements were conducted. For example, researchers can evaluate road sign quality or assess the readiness of the mobile network infrastructure (at the time of measurement) to support future connected mobility solutions. This dataset might also serve as a benchmark for assessing the preparedness of road infrastructure and communication systems in Slovakia for intelligent transportation applications.

Methods

For all drives, Land Rover Defender (shown in the Fig. 1) equipped with an AIL System Solution rooftop box, which is a custom-built proprietary solution of the Automotive Innovation Lab research laboratory at STU has been used. This evaluation system provides variability in sensors setup. For this dataset, following data were collected:

- Timestamp (in format YYYY-MM-DD HH:MM:SS)
- Vehicle position (Latitude, Longitude; measured in Degrees)
- Vehicle speed (measured in cm/s)
- Altitude of the vehicle (meters above sea level; measured in cm)
- Vehicle direction according to the compass (0–360°)
- Data connection speed (Downlink, Uplink; measured in kbps)
- Response time of access to the mobile network (ping to the ISP network, ping to the SIX.sk network; measured in milliseconds)
- Mobile network access technology (2G, 3G, 4G, 5G-NR; String format)
- Accuracy of position measurement using GNSS (measured in millimeters)
- Number of potentially visible GNSS satellites based on time and position
- Camera images (JPEG format, Full HD resolution, 30 fps)

Time alignment method. To align data from sensors with different sample rates into one dataset file data buffers for sensors where used. In the method, sensor with highest sampling rate acts as primary data source to which every auxiliary data source is aligned to. The method is based on assumption that in the time between two consecutive samples acquired in primary data source there will be no two consecutive samples acquired by any other data source, because highest sampling rate has primary data source. Samples acquired from auxiliary data source is updated in separate buffers and then aligned and stored into file together with acquired data from primary data source. The method was chosen because the acquired data is aligned into table. In the dataset, camera is considered as primary data source. In the Pseudo-code 1, data recording process is described. The pseudo-code describes two processes. First process is executed as a separate process per each auxiliary data source. Second process is executed only per primary data source as described before.

```
## Definition of variables

buffers = [] # list of all buffers for auxiliary data sources
observation = [] # represents single observation as list of records(rows)
frames_folder # folder where camera frames are saved

## Executes in each Auxiliary data source 'A' independent processes

while capture == True:
    sample_A = auxiliary_datasource.getData()
    if sample_A != previous_sample_A:
        previous_sample_A = sample_A
        buffer_A.push(sample_A)

## Executes in Primary data source 'P' (Camera) as main process

record_number = 1
while capture == True:
    camera_frame = primary_datasource.getData()
    record = []

    if sample_P != previous_sample_P:
        previous_sample_P = sample_P
        camera_frame_name = str(record_number) + camera_frame.format
        record.append(actual_time)
        record.append(camera_frame_name)
        frames_folder.save(file=camera_frame, name=camera_frame_name)
        for buff in buffers:
            record.append(buff.pop())

    observation.append(record)

observation.save_as_csv()
```

Pseudo-code 1. Pseudo-code for data recording process.

Sensor setup. In the following subsections we describe the setup of visual sensor, GNSS sensor and setup of the mobile network modem. Overview of the vehicle sensor setup can be seen in the Fig. 2.

Front camera. For recording of camera frames we used D3RCM-OV10640-953 rugged camera module by DesignCoreTM with 1.3 MegaPixel and FPD-LinkTM III interface. This camera suits automotive and industrial production needs. The lens Sunny 4083 provides 52 degree horizontal field of view (FOV) with aperture of 1.6 and 6.14 mm focal length. The resolution of each frame is 1280×1080 , meaning the frames are in 4:3 ratio. Full specification of the camera can be found in the datasheet¹¹.

GNSS Sensor - U-blox C102-F9R. The GNSS development kit from the company U-blox C102—F9R is a device for determining high-precision position. The module has the option of connecting a dual-band external antenna, which together with Real-Time Kinematics (RTK) over the Internet Protocol using NTRIPv3 can ensure the fixed accuracy of the position - determining the position of the antenna with a maximum deviation of 14 mm in terms of horizontal accuracy. At the same time, the device is connected to all four commercially available satellite positioning systems (GPS, Galileo, GLONASS and BeiDou). The device supports communication via USB 2.0, I2C, SPI and UART buses. In the solution, data is obtained via USB interface in UBX message format. From

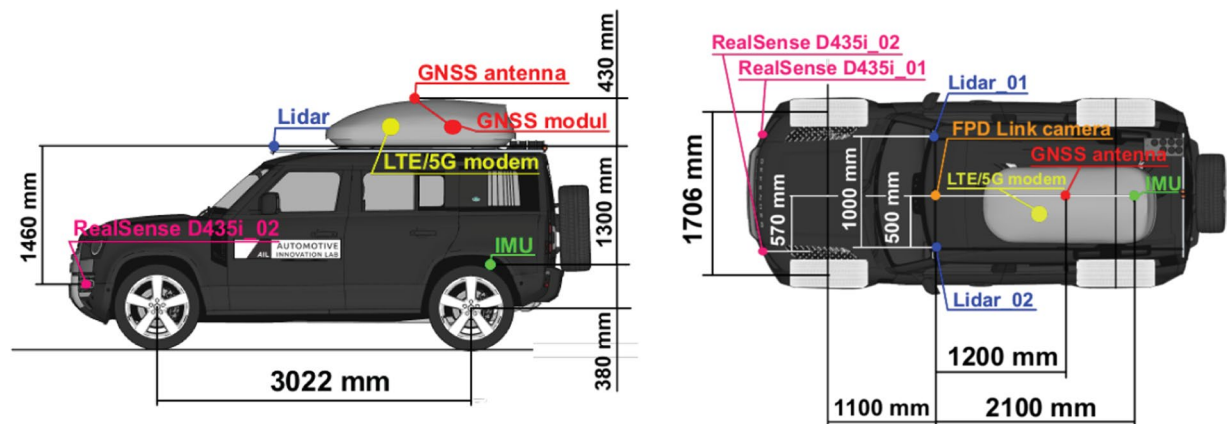


Fig. 2 Physical sensor layout with important dimensions.

the messages, position together with accuracy, ground speed, ground course and potentially visible satellites has been parsed and then recorded.

Mobile network modem - Teltonika RUTX50. A network modem and router that provides connection to 2G, 3G, 4G (LTE) and 5G NSA mobile network. The device has a modem that supports all transmission bands of mobile operators used in the territory of the Slovak Republic. It supports transfer rates up to 1 Gbps, ensuring connectivity and transfer rate measurements are not bottlenecked by the device. Information about radio access network type is obtained via opened socket on the device, in form of text response to terminal command. Information about downlink and uplink speeds has been obtained using Iperf4 program, and information about network latency towards mobile network edge router and towards cloud is obtained using Ping program.

Data Records

The dataset described in this paper can be downloaded from the website of Automotive Innovation Lab at STU (link to the dataset: <https://data.aail.sk/dataset-aail-albus/>).

The raw data is initially divided into folders according to the specific road categories where the measurements were conducted. The primary folder division is structured as follows: `d1` (highway D1), `d2` (highway D2), `c1` (1st class road), `c2` (2nd class road), `c3` (3rd class road), and `city` (urban environment). For each type of road were performed measurements under diverse light and weather conditions. The subsequent categorization is presented as follows: `direct` (direct sunlight), `normal` (normal daylight), `rain` (rain or fog), and `twilight` (dawn or twilight). Multiple drives were conducted on each road category, encompassing diverse lighting and weather conditions, with the primary objective of capturing lane detection on the roads as well as road sign detection. Our measurements were made on the 14th, 18th, 19th, 21st, 24th, 25th of November 2022 and 4th, 5th, 6th, 7th, 19th, 22nd of January 2023. The total size of the provided data is 1.7 TB. The dataset consists of 96 separate drives, with four drives for each weather condition and road category. Altogether, the dataset contains 3 558 705 samples.

Road categories. For the purpose of creating this dataset, various Slovak roads were identified for data collection. The roads were selected based on their type and environment to encompass a wide variety of road types and settings, including rural areas, hills, and urban environments.

According to the Slovak Act on Roads (Act No. 135/1961), roads in Slovakia are classified into highways, roads (including first, second, and third-class roads), local roads (comprising streets and parking areas within municipalities), and purpose-built roads. The classification primarily reflects the roads' significance within the overall network. It is noteworthy that the legislation also assigns maintenance responsibilities for each road type to different organizational bodies, often indirectly impacting the quality of road surfaces and road markings. For this reason, all road types (excluding purpose-built roads, which are relatively uncommon) were included in our dataset.

The whole road selection process utilized stratified sampling approach. As mentioned earlier, all road types (except purpose-built roads) were used for the primary stratification. In the region of interest, specific road sections from each type were chosen to represent diverse environmental conditions. For instance, the selected second-class road includes a mountain pass and traverses forested, sparsely populated areas, a typical scenario for many regions in Slovakia. In contrast, the selected first-class road crosses predominantly flat fields and small towns. To capture urban environments, we also included road sections passing through a major city, which features multiple road types, ranging from highways to small streets and parking areas.

The final selection of road sections where data was collected is presented in Fig. 3. The corresponding list of roads is as follows:

- **City - Bratislava (Green)** ± 13 km long route across urban roads in Bratislava, containing part of inner-city highway. Different types of urban roads (including local roads according to the law) were part of this route with various levels of traffic.
- **Highway - D1 (Pink)** ± 12 km long route on a 3-lane highway that is considered the busiest in the country. No tunnels or complicated surroundings is present here.

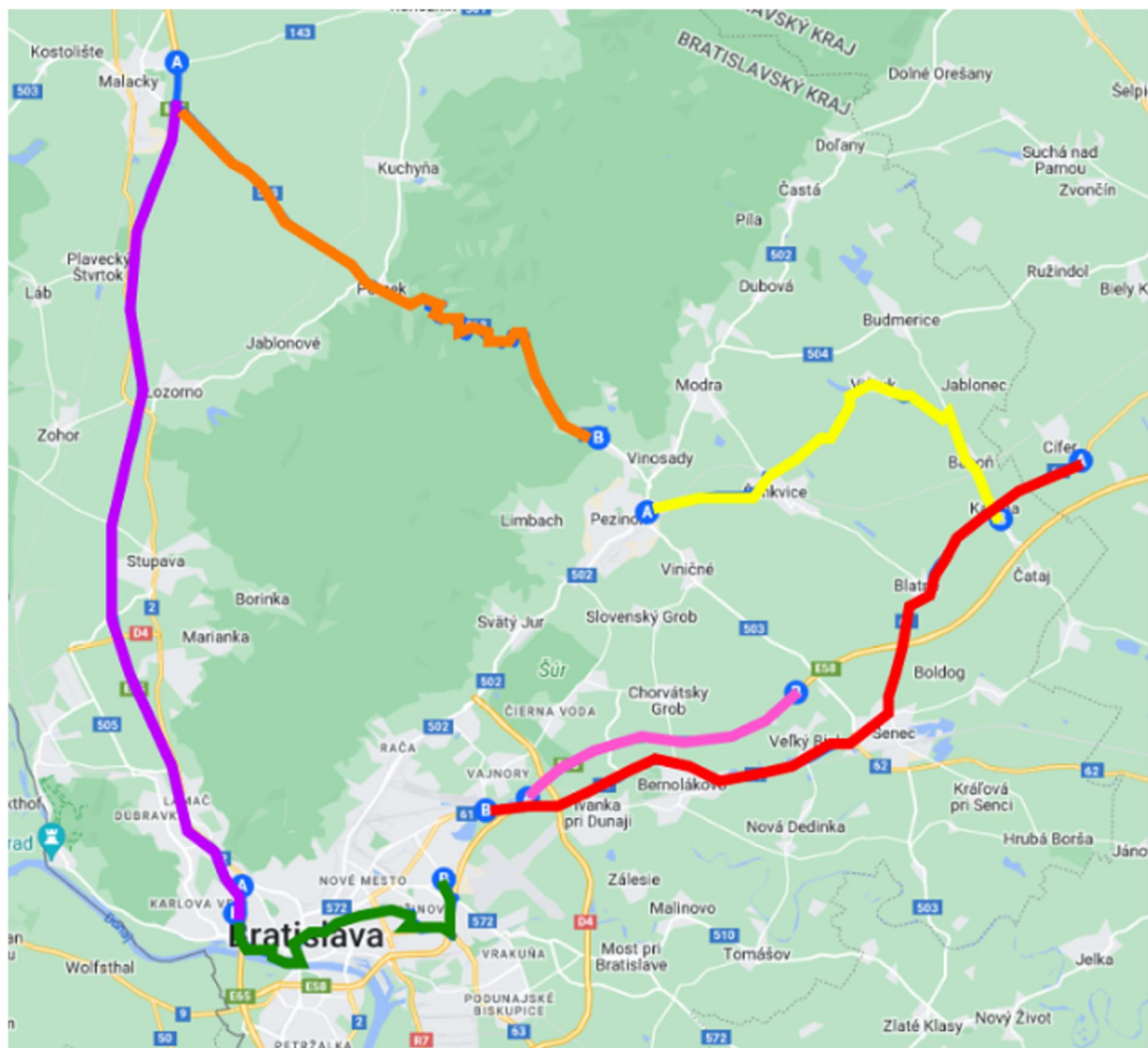


Fig. 3 Map showing roads where data was collected.

- **Highway - D2 (Purple)** ± 32 km long route on a 2-lane highway that contains tunnel Sitina with length of 1415 m.
- **1st class - I/61 (Red)** ± 30 km long route on the 1st class road in Bratislava region. Contains both rural and urban environment as well as intersection with highway and several roundabouts.
- **2nd class - II/503 (Orange)** ± 25 km long route on a 2nd class road, that is most complicated in this dataset, as it goes through hills in a forest, meaning that mobile data connection is poor and GNSS accuracy is impaired.
- **3rd class - various (Yellow)** ± 19 km long route across several 3rd class roads which are meant for local rural transportation. These roads are mostly empty or with very small amount of traffic. Also these roads are the least maintained, which results in worse technical quality.

Weather and light conditions. Each selected road segment was driven multiple times under varying weather and light conditions, with each drive labeled according to a single condition category. For simplicity, these conditions were grouped into four categories: direct sunlight, normal daylight, rain or fog, and dawn or twilight. While all experiments were classified within these specific categories, this does not exclude the occurrence of other conditions during parts of the sample, as such conditions can change unexpectedly. Thus, the classification reflects the predominant conditions observed in each drive. The following methodology was applied when categorizing the weather and light conditions:

- **Direct sunlight** No atmospheric precipitation was present, and the sky was either clear or with low cloud cover. Sunlight was direct, causing objects within the camera's field of view to cast distinct, sharp shadows. The direction of the sunlight was not constrained, with an effort made to capture diverse lighting conditions

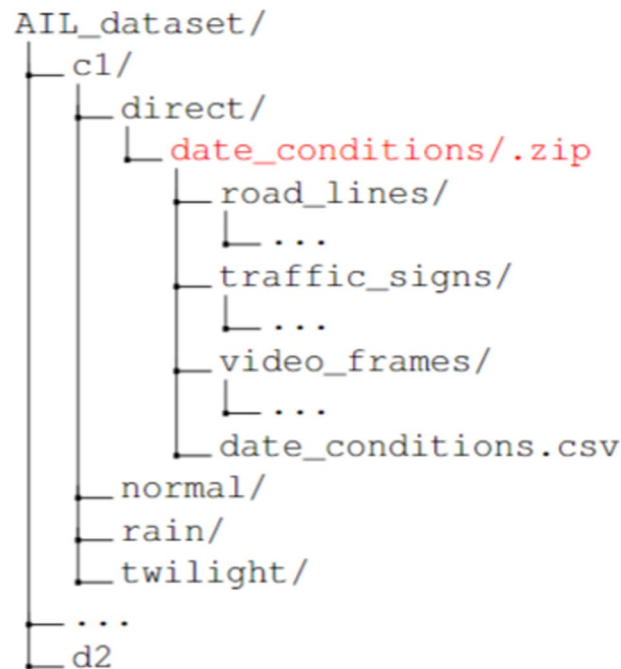


Fig. 4 File structure of the dataset.

- during the drives. Situations where the evening sun shines directly into the camera are also included. Due to the nature of the road sections, these conditions varied based on the vehicle's current position and orientation.
- **Normal daylight** No atmospheric precipitation was present, and the sky was fully overcast or had high cloud cover. Sunlight was diffused, resulting in minimal or no shadows being cast by objects within the camera's field of view.
 - **Rain or fog** Atmospheric precipitation (rain or sleet) or fog was present, resulting in wet road surfaces and condensation on the front camera lens. Cloud cover and sunlight intensity during rainfall were not taken into account, nor was a quantitative criterion for rain intensity defined.
 - **Dawn or twilight** No atmospheric precipitation was present, and cloud cover was not a determining factor (any cloud cover was possible). The drives were conducted either between the onset of daylight and sunrise or between sunset and complete darkness. During these periods, lighting conditions changed rapidly in terms of light intensity, and portions of the samples may be classified as night.

Data description. All sensor readings, the position of traffic signs and lanes, as well as video frames from each drive, are organized within a folder structure based on the type of road, and weather & lighting conditions, as shown in Fig. 4. The data from the corresponding drive are zipped in one file in the format `date_condition.zip`, where the date and conditions are placeholders for the start date of the drive (in format `day-month_hour-minute`) and the weather condition & light condition, respectively.

Data collected from U-blox C102-F9R and Teltonika RUTX50 devices during individual measurements are stored in respective folders within the file `date_condition.csv`, which includes the following attributes:

- **time** - timestamp that composed of hours, minutes, seconds, and milliseconds,
- **frame_name** - name of video frame in `.jpg` format,
- **lat** - latitude, in degrees
- **lon** - longitude, in degrees
- **hmsl** - height above mean sea level in `mm`,
- **ggs** - ground speed, in `cm/s`
- **gc** - ground course (compass), in degrees
- **hacc** - horizontal accuracy in `mm`,
- **satellite** - number of potentially visible GNSS satellites
- **ping_edge** - ping to the ISP network, in `ms`
- **ping_six** - ping to the SIX.sk network, in `ms`
- **iperf_down** - download speed in `kbits/sec`,
- **iperf_up** - upload speed in `kbits/sec`,
- **netmode** - mobile network access technology, (2G, 3G, 4G, 5G (NSA))

A part of all measurements are also files that provide the results of the detection of road lanes and road-signs on individual images from the relevant drive, which were achieved using appropriate machine learning

Road Category	Weather Condition	Number Of Samples	Accuracy of position [mm]	Number of potentially visible GNSS satellites	Ping to the Edge [ms]	Ping to the Cloud [ms]	Downlink [kbps]	Uplink [kbps]
c1	direct	205433	29.3 (156.64)	49.0 (3.35)	71.64 (219.18)	75.41 (219.06)	39474.7 (25394.73)	14316.82 (8250.51)
	normal	206844	25.01 (151.04)	35.0 (21.07)	59.35 (127.62)	63.51 (127.68)	45345.7 (28717.26)	15648.31 (8329.55)
	rain	200128	49.98 (551.52)	48.0 (1.95)	78.08 (232.84)	82.37 (231.95)	42454.92 (27073.54)	15073.54 (8008.16)
	twilight	214982	37.95 (388.2)	53.0 (2.38)	63.7 (149.3)	68.93 (150.42)	39051.47 (25186.47)	15026.25 (8520.1)
c2	direct	201806	275.8 (1291.19)	38.0 (21.55)	85.12 (108.17)	87.55 (105.67)	36178.02 (31527.6)	12057.44 (8379.89)
	normal	179724	316.22 (971.74)	48.0 (2.45)	73.38 (108.7)	76.73 (107.84)	37866.69 (30016.62)	12456.41 (9056.55)
	rain	191970	297.65 (2298.79)	45.0 (1.86)	73.46 (108.98)	76.71 (106.1)	40834.95 (33274.63)	13328.54 (10769.03)
	twilight	176767	568.37 (3912.11)	47.0 (1.68)	75.05 (113.58)	79.55 (113.65)	41830.07 (33599.8)	15409.45 (12180.67)
c3	direct	146371	114.9 (539.0)	50.0 (3.2)	52.54 (72.16)	56.72 (70.5)	49341.6 (30088.63)	17523.31 (7894.37)
	normal	144624	156.23 (555.93)	47.0 (1.6)	57.14 (77.29)	61.04 (76.42)	47355.32 (29617.4)	18093.47 (8636.18)
	rain	141059	100.76 (348.36)	35.0 (20.63)	50.5 (66.98)	54.78 (66.84)	49979.09 (32396.09)	18241.35 (8341.45)
	twilight	148151	84.95 (433.95)	51.0 (1.51)	54.02 (97.23)	57.94 (96.56)	53227.62 (34648.63)	19026.62 (8586.52)
city	direct	171785	119.25 (662.12)	48.0 (1.21)	33.95 (30.06)	39.43 (28.16)	146112.08 (64036.14)	28063.94 (9962.78)
	normal	175263	65.81 (345.64)	47.0 (1.24)	40.43 (122.06)	44.31 (107.45)	184341.41 (59434.0)	42626.04 (15490.77)
	rain	159669	106.01 (549.89)	51.0 (1.85)	32.71 (19.99)	38.05 (19.23)	179129.48 (60961.33)	37900.68 (13028.33)
	twilight	176089	84.75 (604.77)	44.0 (15.72)	33.16 (21.81)	39.0 (21.8)	181545.62 (63239.65)	36554.45 (16095.45)
d1	direct	46338	31.27 (97.64)	48.0 (3.78)	99.09 (381.76)	102.75 (381.7)	60208.3 (33784.49)	11048.29 (3801.59)
	normal	47990	15.34 (7.71)	44.0 (2.59)	55.58 (94.06)	84.03 (265.68)	61281.32 (33735.11)	10899.85 (3603.2)
	rain	44704	17.46 (49.19)	50.0 (1.34)	47.87 (103.78)	52.68 (102.6)	72176.93 (32336.74)	16682.09 (8569.16)
	twilight	49094	15.41 (8.27)	52.0 (1.37)	129.39 (566.87)	134.16 (565.84)	58298.97 (32747.0)	10013.7 (5001.93)
d2	direct	125991	105.1 (3994.13)	47.0 (1.24)	65.06 (226.14)	69.99 (226.43)	64716.32 (64035.47)	17790.07 (11811.95)
	normal	122124	1120.09 (6610.63)	47.0 (2.39)	85.73 (256.15)	106.27 (363.63)	66462.1 (63750.1)	16330.28 (11283.27)
	rain	121524	691.94 (7115.93)	47.0 (3.45)	48.52 (79.94)	74.72 (310.74)	71034.07 (66098.06)	19491.3 (13110.1)
	twilight	118709	632.14 (5282.98)	49.0 (3.74)	67.71 (273.69)	56.28 (90.67)	63026.59 (62035.88)	20519.94 (13202.57)

Table 1. Mean values and standard deviation (in parenthesis), aggregated by weather condition and road category.

techniques. In the Table 1, mean values and standard deviations for numeric data can be seen. In the table, observations has been aggregated by weather conditions per road category. Number of sampels per each weather condition and road category has been provided in the table as well. The Fig. 4 shows the accessible dataset structure.

Road lane detection. The files related to lane detection are in the `road_lines` folder of the respective measurement. We provide two file formats that contain information about the road lanes in a particular video frame.

The first file type, named `worker_number.log`, contains data in JSON format, where `number` is a placeholder indicating the thread in which the file was created. The first key's label is `frame` and encapsulates the name of video frame. The second key's label is `detections` and contains a three-dimensional array where the second dimension represents the particular road lane. The third dimension represents the points of that road lane on the corresponding frame.

The second type of file is in `.csv` format and contains attribute `lines`, which represents the number of detected lanes. For each detected lane, we provide the attributes `lnumber_points` and `lnumber_length`, which represent the number of detected points of the corresponding road lane on the image, and the length of the road lane in pixels, respectively. The placeholder `number` represents particular road lane starting from 0.

Roadsigns detection. Files containing data regarding the detection of traffic signs are within the `road_signs` folder. We offer detection for various categories of road signs, including danger, priority road, mandatory, yield, prohibitory, stop, and other. Just as in the case of lane detection, we furnish files in both JSON and CSV formats.

The JSON files for traffic sign detection share identical names, placeholder, and keys with those used in lane detection. The `frame` key designates the name of the respective frame. The `detections` key comprises a two-dimensional array, with the second dimension containing details about the detected road signs. This specifically encompasses the road sign's category, probability score, and the coordinates of the road sign within the frame.

The `.csv` format files contain the following attributes: `frame` (name of video frame), `sign_class` (traffic sign category), `sign_confidence` (probability score), `sign_x1`, `sign_y1`, `sign_x2`, and `sign_y2` (x and y coordinates of the detected traffic sign).

Video frames anonymization. For the purpose of publishing the dataset, we have anonymized individual video frames from every measurement. Specifically, our focus was on anonymizing the license plates of vehicles and anonymizing individuals in the video frames. The successful achievement of this objective was facilitated by leveraging the capabilities of YOLOv4 together with easyocr, enabling the precise identification of vehicles and

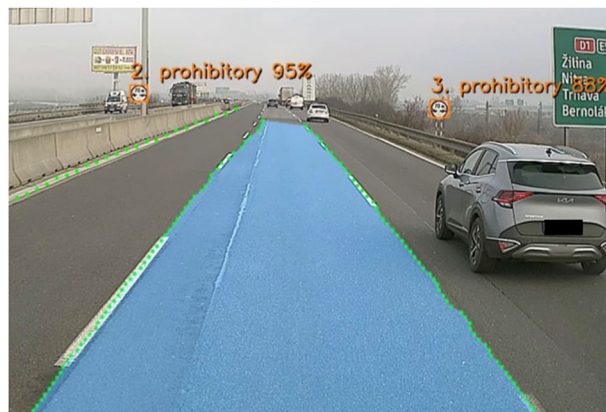


Fig. 5 Lane and road signs detection example.

pedestrians in the images. To anonymize the license plates of vehicles detected by YOLOv4, we utilized a license plate recognition model available in the GitHub repository (link to GitHub project: https://github.com/spmallick/mallick_cascades/tree/master/haarcascades/). To detect only a part of the vehicle's license plate, we employed the easyocr tool, which is specifically designed for text detection in images. After identification of frames that need to be anonymized, parts of images that could lead to concerns about privacy have been blurred manually to ensure that private data have been removed. In addition, to the best knowledge of authors, no pictures of protected sites have been identified in the dataset.

Technical Validation

Basic experiments and evaluations has been done using data from this dataset - i.e. preliminary lanes and road-signs detection.

Preliminary lanes and roadsigs detection. For the purpose of the original project, we did the lane detections and road signs detection on top of the original data and the results of both of these detections are present in the provided dataset. However, this detection has not been evaluated against other models and approaches in terms of performance or reliability.

For detection of the lanes we used existing “Ultra-Fast-Lane-Detection-V2”¹² model, that was created for the purpose of lane detection in difficult conditions while maintaining sufficient detection speed. It is therefore able to provide relevant results even in cases where the driving lanes are not fully visible, whether due to poor lighting conditions, coverage by other vehicles or poor visibility caused by poor road maintenance.

The model YOLOv4 based on neural network was used to search for traffic signs on individual video frames. The advantage of this model is its high detection speed, which can run in the real time without any problems. For road signs classification, a variant of the YOLOv4 tiny L3 model was specifically applied using the Darknet framework (link to GitHub project: <https://github.com/AlexeyAB/darknet/>). The given model was trained on an annotated dataset based on the well-known German Traffic Sign Recognition Benchmark (GTSRB) dataset, where the original 43 types of traffic signs were mapped to 7 categories². An illustrative frame from both lane and roadsigs detection is depicted in Fig. 5.

Usage Notes

This dataset can be used for free for any non-commercial purpose - i.e. further research, student projects, etc. In case of doubt about the right to use, do not hesitate to contact the authors. When using this dataset for your projects and research, please cite this paper.

The license for this dataset is Creative Commons (CC BY) that lets others remix, adapt, and build upon our work, and although their new works must also acknowledge us, they don't have to license their derivative works on the same terms.

The recorded videos are anonymized using a program that has obfuscated license plates and individuals on the roads. However, it's possible that not all videos have been anonymized, and some video frames might still raise privacy concerns. If you have any issues regarding non-anonymized video frames or potential privacy breaches, please don't hesitate to contact our team.

Code availability

No custom code was used specifically for this paper. All data were collected using the AIL System Solution rooftop box, a device protected by an industrial utility model registered with the Industrial Property Office of the Slovak Republic. The post-processing steps were tailored for the outcomes of the research project conducted for the Ministry of Transport of the Slovak Republic. Details regarding the use of open-source libraries can be found in the paper.

Received: 1 March 2024; Accepted: 11 December 2024;

Published online: 30 December 2024

References

1. Lehoczký, P., Janeba, M., Galinski, M. & Šoltés, L. Testing the readiness of slovak road infrastructure for the deployment of intelligent transportation. *Strojnícky časopis - Journal of Mechanical Engineering* **72**, 103–112 (2022).
2. Stallkamp, J., Schlipsing, M., Salmen, J. & Igel, C. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks* **32**, 323–332, <https://doi.org/10.1016/j.neunet.2012.02.016> (2012).
3. Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M. & Igel, C. Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In *International Joint Conference on Neural Networks* 1288 (2013).
4. Šegvić, S., Brkić, K., Kalafatić, Z. & Pinz, A. Exploiting temporal and spatial constraints in traffic sign detection from a moving vehicle. *Machine vision and applications* **25**, 649–665 (2014).
5. Monroy, I. T. *et al.* Testing facilities for end-to-end test of vertical applications enabled by 5g networks: Eindhoven 5g brainport testbed. In *2018 20th International Conference on Transparent Optical Networks (ICTON)*, 1–5 (IEEE, 2018).
6. Perez, P., Ruiz, J., Benito, I. & Lopez, R. A parametric quality model to evaluate the performance of tele-operated driving services over 5g networks. *Multimedia Tools and Applications* **81**, 12287–12303 (2022).
7. Janeba, M. *et al.* Evaluating slovak roads for vehicle-to-everything communication 3gpp requirements. In *Proceedings of the XXI Polish Control Conference*, 420–430 (Springer, 2023).
8. 3GPP. Service requirements for enhanced v2x scenarios. *Technical Specification (TS) 22.186, 3rd Generation Partnership Project (3GPP)* (2019).
9. Shadrin, S. S. & Ivanova, A. A. Analytical review of standard sae j3016 taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles with latest updates. *Avtomobil'. Doroga. Infrastruktura*. **10** (2019).
10. Kovvali, V. G., Alexiadis, V. & Zhang, L. Video-based vehicle trajectory data collection. Tech. Rep. (2007).
11. D3 Engineering. Designcore d3rcm-ov10640-953 datasheet Accessed on August 22, 2023 (2021).
12. Qin, Z., Zhang, P. & Li, X. Ultra fast deep lane detection with hybrid anchor driven ordinal classification. *IEEE transactions on pattern analysis and machine intelligence* (2022).

Acknowledgements

This article was written thanks to the generous support under the Operational Program Integrated Infrastructure for the project: “Support of research activities of Excellence laboratories STU in Bratislava”, Project no. 313021BXZ1, co-financed by the European Regional Development Fund. The research was also supported by the APVV-19-0401 and the KEGA 025STU-4/2022 projects. This research was also supported by Ministry of Transport of Slovak Republic.

Author contributions

M.G. was leader of this project and main author of the system to collect data. T.M. provided valuable technical insights on how to collect these data. M.J. is author of the software that synchronises data in the final log and was part of most of the experiment. J.D. provided valuable insight on construction and power system of the vehicle. P.L. was part of most of the experiments and did the preliminary data processing. L.M. is the head of mechanical engineering lab and provided the vehicle itself. L.S. provided valuable insights on splitting sensors from vehicle, where possible. Also took care of administrative project management. A.T. did all the anonymization and sorted data into the final downloadable form.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024