The DLR Highway Traffic Dataset (DLR-HT): Longest Road User Trajectories on a German Highway

Clemens Schicktanz*, Lars Klitzke*, Kay Gimm*, Richard Lüdtke*, Karsten Liesner*, Henning Hajo Mosebach*, Fin Heuer*, Axel Wodtke*, Lennart Asbach*

***Abstract*— In recent years, several datasets containing trajectories of road users have been published, providing valuable insights for the analysis and modeling of traffic participant behavior. However, road user trajectories from highway datasets are often limited to lengths of less than 2.5 km, restricting the analysis of consecutive traffic scenarios, such as multiple lane changes. To address this gap, we introduce the DLR Highway Traffic dataset, the longest road user trajectory dataset from a German highway. This dataset contains 38,209 trajectories, along with local weather and road condition data collected over a period of 10 h at the Testbed Lower Saxony. A comparison with other publicly available datasets reveals that our dataset, with trajectories reaching up to 6,428 m in length, contains the longest trajectories from German highways, enabling the analysis of long-duration traffic scenarios. With a total of 143,371 km, our dataset is approximately three times larger than the largest existing German highway dataset, the highD dataset, which covers 44,500 km. However, it is 28 times smaller than the largest highway dataset, the I-24 MOTION dataset, which covers approximately 4,050,000 km. In contrast, our dataset stands out by including additional raw data beyond just trajectories, such as locally recorded weather data and road condition data. Furthermore, the traffic volume data, derived from the trajectory data, provide valuable insights into traffic flow. Additionally, the trajectory data are available in OpenSCENARIO format, facilitating the visualization and simulation of traffic scenarios. Overall, the dataset provides valuable resources for researchers seeking to conduct data-driven behavior modeling. It is available for non-commercial use and can be directly downloaded from `https://doi.org/10.5281/zenodo.14811064`.**

***Index Terms*— Highway Trajectory Dataset, Road User Behavior, Automated Vehicles, Traffic Research, Open Data**

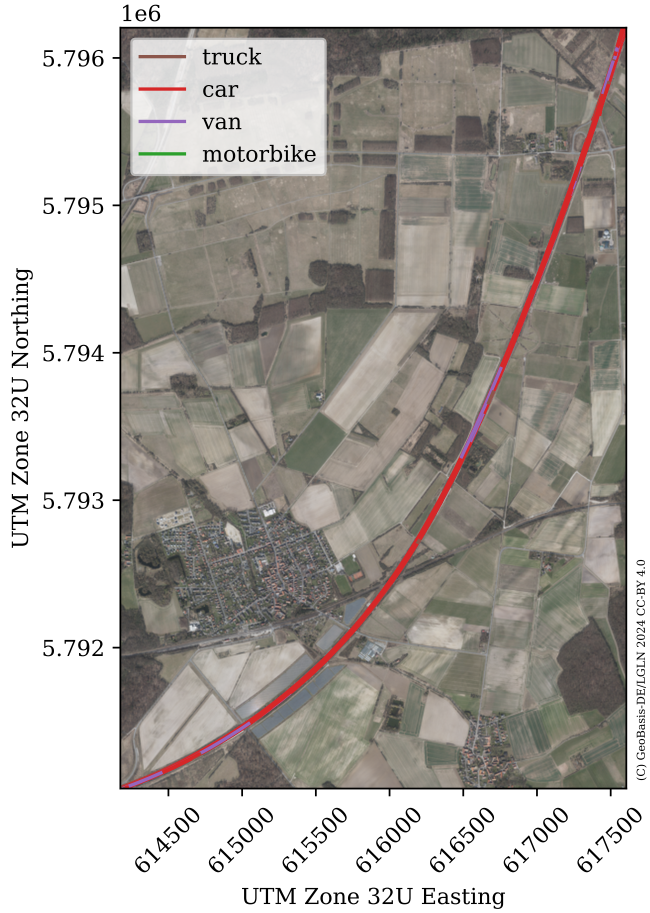

Fig. 1: Visualization of trajectories from the DLR-HT dataset over a 15-minute period.

I. Introduction

Traffic is a major challenge in today's world, causing accidents, injuries, and contributing to harmful carbon dioxide emissions that worsen climate change [1], [2]. To tackle these problems, researchers focus on making traffic systems safer and more efficient. This relies heavily on collecting and analyzing traffic data, which helps to understand how people behave on the road and to find practical solutions [3].

In recent times, the collection and sharing of traffic data, particularly datasets containing vehicle trajectories, has gained significant attention due to its usefulness in a variety of research areas [4]. Stationary cameras, which are commonly used to capture traffic data through video surveillance, have proven to be a reliable source of information. With recent technological advancements, drones have emerged in the field of traffic data collection as an alternative [5]. However, the datasets collected by drones tend to cover only small portions of the road, which can limit the ability to analyze broader traffic patterns.

To overcome these limitations, it is necessary to implement continuous data collection across larger areas. Testbeds designed for long-term operation are ideal for generating such comprehensive datasets. The *German Aerospace Center (DLR)* began its work with automated driving testbeds

* Clemens Schicktanz, Lars Klitzke, Kay Gimm, Richard Lüdtke, Karsten Liesner, Henning Hajo Mosebach, Fin Heuer, Axel Wodtke, Lennart Asbach are with the German Aerospace Center (DLR), Institute of Transportation Systems, Braunschweig, Germany, {clemens.schicktanz, lars.klitzke, kay.gimm, richard.luedtke, karsten.liesner, henning.mosebach, fin.heuer, axel.wodtke, lennart.asbach}@dlr.de.

TABLE I: Comparison of our dataset with relevant publicly available datasets ("-" indicates data not available, "*" indicates approximately).

Dataset Name (Year)	Duration [h]	Number of Trajectories	Max. Length of a Single Trajectory [km]	Total Length of all Trajectories [km]	Number of Object Classes	Min. Number of Lanes per Direction	Number of Cameras
NGSIM US-101 (2005)	0.75	9,206	0.64	5,892*	0	5	8
highD (2018)	16.5	110,500	0.42	44,500	2	2	1
exitD (2021)	16.1	69,430	0.42	27,300	6	2	1
AUTOMATUM (2021)	30	60,000	0.66	39,600*	4	2	1
HIGH-SIM (2021)	2	-	2.44	-	3	3	3
Zen Traffic Data (2023)	5	18,000	2.00	36,000*	2	2	38
I-24 MOTION (2022)	47	600,000*	6.75	4,050,000*	7	4	276
DLR-HT (2024) (ours)	*10*	*38,215*	*6.42*	*143,371*	*6*	*2*	*118*

back in 2009 [6], including the creation of the *Application Platform for Intelligent Mobility (AIM)* [7] for urban traffic and the *Testbed Lower Saxony* [8] for highway traffic. Both testbeds have been continuously expanded and improved through several projects, collecting various types of data, including video recordings, trajectory data, weather conditions, and road surface information. Recently, data from the AIM Research Intersection has been published as *DLR Urban Traffic dataset (DLR UT)* [9].

This publication introduces the *DLR Highway Traffic dataset v1.1.0* [10], a publicly available dataset recorded at the Testbed Lower Saxony on the German highway A39 near Braunschweig. This dataset provides valuable insights into real-world traffic conditions, offering a long-distance, stationary-recorded dataset of vehicle trajectories. The position data of these trajectories is shown in Fig. 1.

Such long-distance trajectory datasets allow for the analysis of changes in driving behavior, such as variations in speed, following distance, or overtaking patterns.

This publication is conceptually similar to the publication for the DLR UT dataset [9] and therefore follows a comparable structure, as outlined in the following. In Section II, we provide an overview of other publicly available traffic datasets and highlight the unique aspects of our contribution. Following that, we outline the methods used for data collection and processing in Section III. Section IV includes key statistics of the dataset such as the length of captured trajectories. In Section V, we discuss the contributions of the dataset, its limitations, and potential avenues for future work, before concluding the publication in Section VI.

II. Related Work

A. Dataset Comparison

In the field of traffic research, several publicly available datasets exist. This section outlines how our dataset differs from previously published ones and highlights its contribution to the field.

As our contribution is based on the trajectory data from our dataset, we compare it with other trajectory datasets. These datasets can be categorized based on their data collection methods: recordings from moving vehicles, recordings from flying drones, and recordings from stationary sensors mounted on infrastructure. In this review, we focus exclusively on datasets recorded by drones and stationary infrastructure, as these are more comparable to our dataset than those collected from moving vehicles.

The *Next Generation Simulation (NGSIM)* [11] dataset is the most widely used dataset and was recorded in 2005 using stationary sensors on a freeway [12]. The first large-scale trajectory dataset recorded by a drone is the *highD* [13] dataset, which has also been widely used since its publication in 2018. Parts of the highD dataset were converted to OpenSCENARIO format [14] and made publicly available [15]. However, the data is no longer accessible to the public. The *exiD* [16] dataset published data from seven locations featuring highway on- and off-ramps to enable the analysis of behavior in these specific on/off-ramp and merging traffic scenarios. The *AUTOMATUM DATA* [17] dataset provides a drone-based highway dataset for the development and validation of software for automated driving. It includes highly accurate trajectories of road users from twelve different locations in Germany. A limitation of these datasets is that the trajectories are recorded in an area of less than 1 km in length. Consequently, the behavior of road users in scenarios covering distances beyond this range cannot be analyzed.

The *HIGH-SIM* [18] dataset focuses on expanding the test area and, therefore, the maximum possible trajectory length. It offers trajectory data extracted from helicopter recordings captured by 3 cameras. Although the dataset has a relatively short temporal span of only 2 h compared to the previously mentioned datasets, it stands out due to its significantly longer trajectories, with a maximum length of 2.4 km. The *Zen Traffic Data* [19] dataset is comparable to the HIGH-SIM dataset with a temporal span of 5 hours and a spatial coverage of 2 km. In addition, it includes information on road surfaces and traffic volume such as our dataset. The *TUMTraf A9 Highway* [20] dataset contains data from the 3 km long *Providentia++* testbed near Munich, Germany. However, current releases do not yet contain trajectory data, but only raw sensor data from installed cameras and LIDAR sensors, which is why this dataset has not been included in Table I.

Table I presents a comparison of the aforementioned datasets based on various parameters. It is important to note that Table I does not represent all existing datasets. Table I builds upon Table 1 from the publication [21] about the *I-24 MOTION* dataset, as our dataset is similar to the dataset and should thus be compared with similar references. Our dataset and the I-24 MOTION dataset distinguish themselves from previously published datasets due to their maximum trajectory length of over 6 km. While the I-24 MOTION dataset, recorded in the United States, contains trajectories up to 6.75 km in length, our dataset features trajectories up to 6.42 km, making it the dataset with the longest road user trajectories on German highways.

Another difference between our and the I-24 MOTION dataset is the number of cameras used for data recording as shown in Table I. With 118 cameras, we use less than half of the cameras employed in the recording of the I-24 MOTION dataset. However, we only capture traffic on at least 2 lanes per direction, whereas the I-24 MOTION dataset records traffic from at least 4 lanes per direction.

B. Data Usage

To highlight possible uses of our dataset, the following examples from related studies are provided.

In [22] trajectory data from the highway is used to develop and evaluate a novel approach for lane-change maneuver identification and extraction using a primitive-based representation of traffic data. Although the method has shown to robustly work with trajectory data from a test vehicle, a follow-up work [23] demonstrated the application of the method to infrastructure-based traffic data collection. In [23] on-ramp scenarios are extracted to analyze the merging behavior of traffic participants. This subject is of significant concern and is the focus of extensive research due to the potential for conflict [24] and the high degree of complexity of the scenario [25], which needs to be managed by both human operators and automated vehicles. In [26], [27], [28], additional use cases are presented, utilizing similar trajectory data from urban areas for behavior analysis. These include the analysis of rare and critical traffic scenarios for the development of test scenarios for autonomous vehicles [26], [28], as well as the quantification of the impact of traffic congestion on safety and efficiency [27].

In conclusion, the examples presented underscore the dataset's potential for investigating the behavior of traffic participants, emphasizing its wide-ranging applicability in transportation research.

III. Method

The presented dataset was captured using stationary infrastructure-mounted sensors and then processed through a data processing pipeline. This section introduces both components: the technical setup for data acquisition and the data processing pipeline, and also provides information on tooling for using the data.

A. Technical Setup

The trajectory data is extracted from video recordings captured by 118 multi-sensor systems, which are mounted on 59 poles along the A39 highway, spanning from the Wolfsburg/Königslutter interchange to the Cremlingen exit. The field of view of each multi-sensor system overlaps with that of the neighboring system, as shown in Fig. 2. This configuration establishes a redundant system that effectively tracks all relevant objects on the highway, minimizing the risk of occlusion and enabling continuous monitoring of road users.

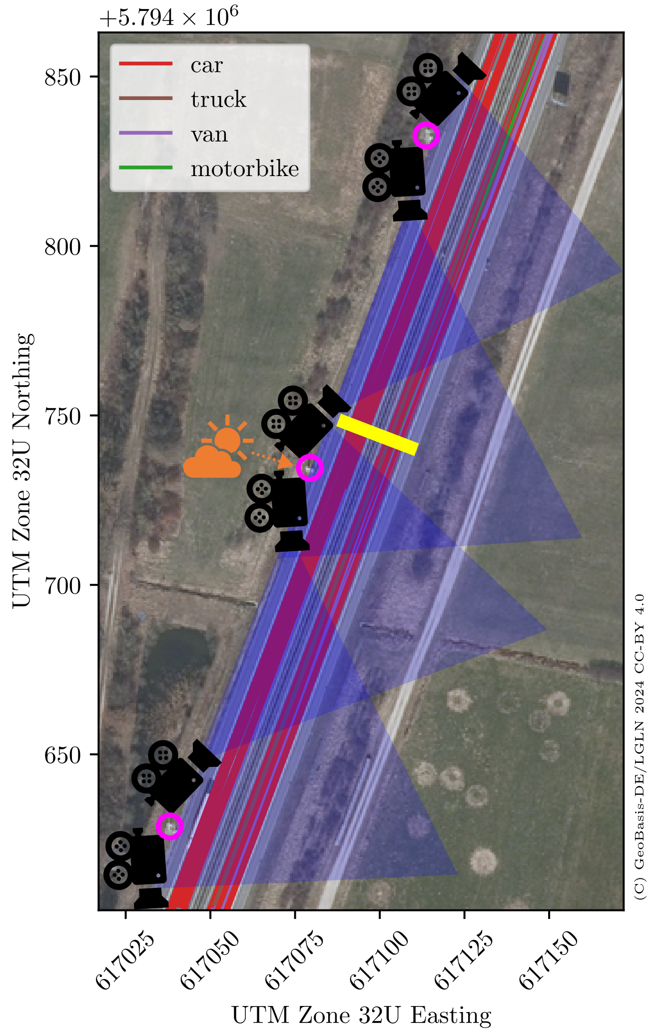

Fig. 2: Overlap of camera fields of view (blue) at Testbed Lower Saxony, with trajectory data (see legend), symbolic representations of poles (pink), cameras (black), weather station (orange), and virtual optical loops (yellow).

Each multi-sensor system consists of two *GiGEVision mvBlueCOUGAR-109b* cameras, with active infrared lighting positioned between them to enhance scene visibility, as shown in Fig. 3. Both cameras are linked to a signal processing server, which requests data from the primary camera at a frequency of 10 Hz. The primary camera subsequently triggers the secondary camera and the infrared lighting to ensure synchronized image capture from both devices.

The weather and road condition data is collected from sensors mounted on a pole at UTM coordinates

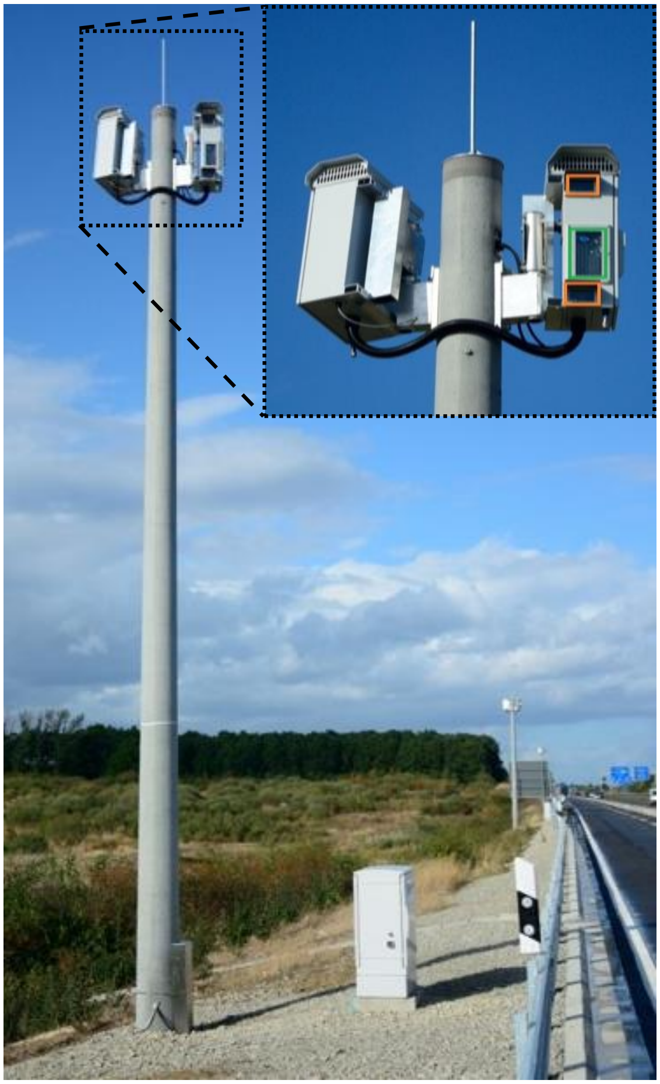

Fig. 3: Multi-sensor system mounted on pole next to the A39 highway. Cameras (orange) and infrared lighting (green) are highlighted.

32U 617078 5794734 as depicted in Fig. 2. General weather information is collected using the *Weather Transmitter WXT536* sensor. The *Present Weather and Visibility Sensor PWD22* is responsible for determining visibility. The *Remote Road State Sensor DSC211* and *Remote Road Surface Temperature Sensor DST111* sensors are employed to assess the condition of the road surface.

B. Data Processing Pipeline

The processing of camera images is performed directly on-site in a server room located next to the highway. To detect and track objects in the camera images and generate trajectory data, the same processing steps are applied as those used at the AIM Research Intersection. For a detailed description of the data processing pipeline, we refer to the publication [9], specifically *Section III.B*, which explains the processing steps in detail. In brief, detected pixel changes are used to create 3D voxels that can be described as object hypotheses. When a sufficient accumulation of object hypotheses is achieved, an object pose is detected at the corresponding location. This principle is illustrated in the images of Fig. 4.

Fig. 4: Object detection. Accumulation of voxels (blue) in the left image, which form a bounding box (turquoise) in the right image.

The resulting trajectory data is stored in a database. For the release of our dataset, the data was retrieved from the database, and post-processing steps were applied to enhance data quality. To eliminate unusable measurements, trajectories shorter than 4 s, those with a total travel distance of less than 40 m, and those with a distance of less than 40 meters between the first and last object pose, were discarded. Additionally, the median object dimensions and classification probabilities were calculated and appended to each object pose of a trajectory. Since not all objects are continuously tracked across the entire testbed, particularly under bridges where trajectories are occasionally interrupted, an algorithm was developed to link trajectories that are likely to belong to the same object. Each trajectory was predicted forward based on its last pose, and the prediction was compared with the starting points of other trajectories within the next 7 s in a radius of 300 m. If the other trajectory had a speed difference of less than 8 m/s and the orientation difference was less than 10°, the trajectory was considered a candidate for merging. The trajectory with the least deviation between its start position and the predicted position of the first trajectory was selected as the successor. The two trajectories were merged by replacing the successor's ID with that of the first trajectory. Finally, a Kalman filter was applied to smooth the entire trajectory, interpolate missing values, and derive speed and acceleration data. Interpolated values were marked in the dataset.

Weather and road condition data were not post-processed but were directly exported from the database.

Metadata was generated in a separate post-processing step from the processed trajectory data. To generate the metadata of traffic volume, virtual inductive loops, as described in [29], were used to simulate conventional inductive loops and determine the crossing time of an object at a specific position. The position of the optical loop was near the weather station, at UTM 32U Northing 5,794,749 (see Fig. 2), to facilitate the best possible comparison between traffic volume data and weather data.

Furthermore, the trajectories were converted into OpenSCENARIO XML format v1.2.0 [14] to facilitate direct simulation and visualization of the data within a simulator. The conversion was carried out using the Python library scenariogeneration [30]. In this process, trajectory data was transformed into FollowTrajectoryActions, and the actions were added and removed from the simulation using StartTrigger and StopTrigger based on their respective start and end timestamps. Moreover, the objects were assigned to an object class from the VehicleCatalog of esmini [31] that corresponds to their classification.

C. Tooling

To facilitate the use of the dataset, we provide a Python library, *Traffic Analysis and Situation Interpretation (TASI)* [32], which simplifies downloading, visualizing and analyzing the data. The library is regularly updated with new functionalities to improve the analysis of traffic data.

IV. Dataset Description

The DLR HT dataset v1.1.0 was generated using the system and data processing pipeline described in the previous section. This section provides an overview of the dataset's key features, helping potential users evaluate its relevance for their specific use case. A comprehensive description, including details on the dataset's format and structure, can be found in the accompanying documentation, available on Zenodo alongside the dataset [10].

A. Recording Area

The data were collected at the Testbed Lower Saxony on the German highway A39. The recording area spans a section of the highway measuring 6,428 m, located between the Scheppau and Cremlingen interchanges. The dataset includes the two on and two off ramps at the Scheppau interchange. However, it does not cover the section at the Cremlingen interchange.

B. Dataset Overview

The raw data of the dataset comprises trajectory data of traffic participants, along with local weather data and road condition data and was recorded on Monday, October 7, 2024, from 6:00 to 16:00 UTC+00:00 which is 8:00 to 18:00 local time (UTC+02:00). The trajectory data is indexed by object ID and timestamps, including detailed information about the center position, velocity, acceleration, dimensions, and classification of each object. The object classification is represented by probability values, indicating the likelihood of each object belonging to one of the following classes: pedestrian, bicycle, motorbike, car, van, or truck. The weather data provide information on wind, sunlight, precipitation, visibility, and more. The road condition data provide information on surface temperature, water layer thickness, and more.

In addition to the raw data, the dataset also includes metadata such as traffic volume and OpenSCENARIO files, which have been extracted from the raw data. The traffic volume data provides the number of objects per lane at a specific location on the testbed near the weather station, as depicted in Fig. 2. The OpenSCENARIO files contain all raw data ready for replay in simulation environments.

C. Data Categories

The following subsections provide a detailed description of the different data categories included in the dataset.

1) Trajectory Data: This subsection presents the trajectory data, including details on the object classes, duration, and length of the trajectories.

a) Object Classes: The dataset contains a total of 38,209 trajectories (93,697,250 object poses), with the majority being trajectories of cars (26,322). The next most common categories are trucks (7,758), vans (4,003), and motorbikes (126).

b) Duration of Trajectories: The total duration of the trajectories is 1,301 h (54 days), with an average trajectory duration of 2 min. The shortest trajectory is 4 s and the longest trajectory is 5 min 2 s.

c) Length of Trajectories: The total length of the trajectories is 146,391.856 km, with an average trajectory length of 3.831 m. Among the object classes, trucks (3,931 m), cars (3,885 m), and vans (3,359 m) have longer trajectories compared to motorbikes (1,604 m). The detailed distribution of trajectory length per object class is shown in Fig. 5.

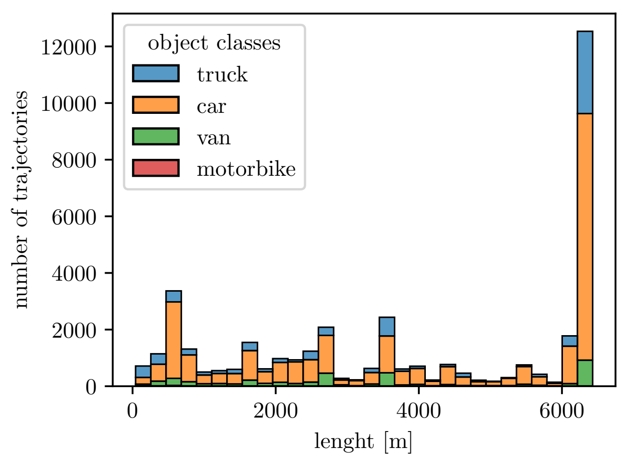

Fig. 5: Distribution of trajectory length per object class.

The variation in trajectory length and the clusters in the distribution shown in Fig. 5 can be explained by the Scheppau interchange and the bridges, where the view on the objects is interrupted. The largest cluster in the distribution, comprising 35 %, consists of trajectories that originate from objects tracked across the entire testbed, each longer than 6 km. A smaller cluster with a length of around 5,600 m likely corresponds to vehicles traveling from the south to the Scheppau exit in the north, exiting the testbed approximately

800 m before reaching the northern end. Additionally, the second-largest peak can be explained by vehicles entering the highway at Scheppau to the north and then exiting the testbed again 600 meters later. The other peaks at approximately 1,700 m, 2,700 m, and 3,500 m can be attributed to interrupted trajectories under the bridges. For instance, the distance from the entrance at Scheppau to the first bridge is 1,700 m.

2) Weather Data: Weather data was recorded every 10 s, resulting in 3,600 records per day. The dataset was recorded under very favorable weather conditions in early autumn. There was no precipitation, snow, or hail, and visibility was the whole day at the maximum sensor value of 20 km. The wind blew from various directions at speeds ranging from 0 to 5 m/s. The air temperature fluctuated between 9 and 18 °C.

3) Road Condition Data: The road condition data was recorded every 30 s, resulting in a total of 1,200 records in the dataset. Due to the favorable weather conditions, the road surface remains consistently in good condition. Since no precipitation was observed and the road surface temperature did not reach minus degrees, no influence on driving behavior is assumed. As shown in Fig. 6, the air temperature and road surface temperature were similar throughout the day, with the air temperature always slightly above the road surface temperature, which fluctuated between 9.7 and 20.1°C.

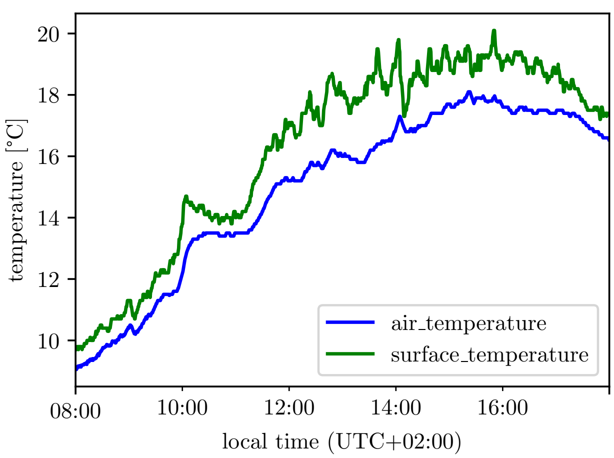

Fig. 6: Air and surface temperature data from the dataset.

4) Traffic Volume: In the dataset, trajectories at all lanes are available. The highway experiences significantly higher traffic volume in the direction from north to south, with 14,154 objects recorded by the virtual optical loops, compared to the direction from south to north, with 9,090 objects. In total, the most traffic (7,133 objects) is on the left lane from north to south. Less traffic is on the right lane (7,021 objects), followed by the right (6,167) and left (2,923) lanes from south to north. With 1,008 objects per hour, the highest traffic volume is recorded on the left lane from north to south during the hour starting at 16:00 local time. The traffic volume per lane over time is depicted in Fig. 7.

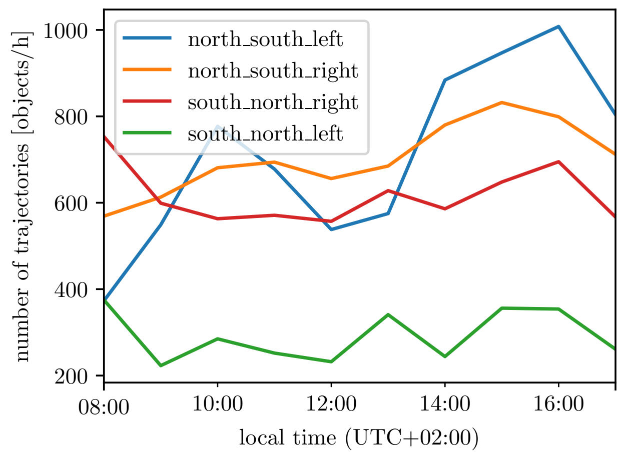

Fig. 7: Traffic volume per hour and lane.

5) OpenSCENARIO: We converted the trajectories into OpenSCENARIO files, which define FollowTrajectoryActions for the simulation of vehicle movement. Fig. 8 shows a screenshot of the replay of the first OpenSCENARIO file from our dataset in esmini [31]. In the foreground, a red van is visible in the right lane, moving from north to south. In the background, additional traffic participants are visible, captured by the system at that time. The complete trajectories of all objects simulated in this timestep are shown in orange.

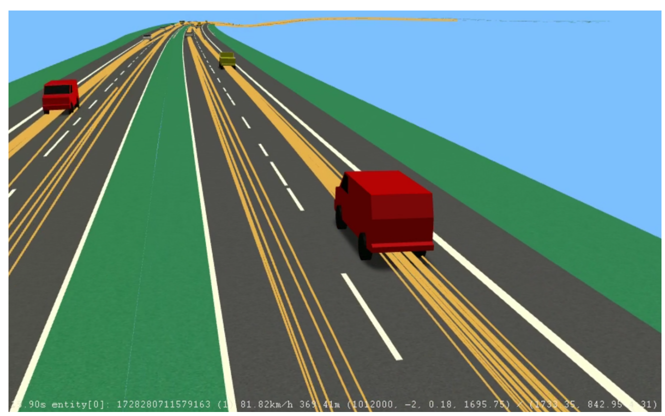

Fig. 8: Screenshot of the replay of the first OpenSCENARIO file from the dataset in esmini.

V. Discussion

In this section, we will discussion the limitations of our dataset, highlight its key contributions, and outline potential directions for future research.

A. Dataset Limitations

As shown in Table I, our dataset includes six object classes, which is the second-highest number of classes, with only one dataset having seven classes. A more detailed classification could be achieved through the use of neural networks. This would enable a more differentiated analysis of traffic participant behavior based on object class.

Furthermore, the area of the testbed for which the weather data describe the environmental conditions is limited. The measurements included in the dataset reflect the conditions at the location of the weather station and cannot be generalized to the entire testbed, which spans over 6 km in length.

For certain traffic behavior analyses, it is necessary to represent the position of traffic participants relative to the lane. A limitation of the dataset is that the high-precision digital map of the testbed in OpenDRIVE format [33] is not publicly available. However, a digital map can be created using OpenStreetMap [34] and Carla [35]. Although this map does not include all road details, such as shoulder lanes, it can still be used to determine the position of traffic participants relative to the lanes.

B. Dataset Contributions

The analysis in Table I shows that our dataset is exceptional compared to existing publicly available datasets in the maximum length of trajectories. While longer trajectories from highways are only available in the I-24 MOTION dataset, our dataset stands out in three aspects when compared to it. Firstly, our dataset includes the classification of motorcycles as objects, enabling the analysis of motorcyclist behavior. Secondly, our dataset includes data on road conditions, providing up-to-date information on factors such as water, ice, and snow layer thickness. Thirdly, our dataset includes metadata derived from the raw data. The traffic volume data facilitates the analysis of traffic efficiency, while the OpenSCENARIO data enables direct replay of the data in a compatible simulator, making it suitable for a wide range of scenarios. These include accurately replicating vehicle behavior in traffic simulations, testing advanced driver assistance systems, and conducting safety analyses. It is particularly valuable for creating realistic and repeatable tests in simulated environments, allowing for the evaluation of traffic behavior.

C. Future Work

The aforementioned metadata are just initial examples. In future work, the data will be used to analyze various traffic scenarios and model the behavior of traffic participants within these scenarios. Selected results will be used to enrich the dataset, providing a broader data foundation for researchers and facilitating studies on traffic safety and efficiency. We encourage users to share analyzed metadata to create a more comprehensive picture of traffic dynamics. The dataset's versioning allows for continuous expansion, for instance, lane changes can be identified in the trajectory data and added to the dataset. This facilitates efficient analyses of lane change behavior, without users needing to detect the events on their own.

VI. Conclusion

This publication presents the DLR HT v1.1.0 dataset [10], which contains traffic data collected at the Testbed Lower Saxony near Braunschweig, Germany, from Monday, October 7, 2024. To provide a detailed representation of traffic conditions, the dataset includes not only trajectory data but also weather and road condition data. Additionally, the dataset features metadata such as traffic volume data and data in OpenSCENARIO format, enabling a replay and simulation of the data. The provided metadata simplifies the process of working with the raw data and offers a more detailed description of traffic conditions. Overall, it can be concluded that the dataset distinguishes itself from already published datasets by being the trajectory dataset with the longest trajectories from German highways. This provides a broad data foundation for researchers to investigate human road user behavior.

References

[1] H. Naci, D. Chisholm, and T. D. Baker, "Distribution of road traffic deaths by road user group: a global comparison," *Injury Prevention*, vol. 15, no. 1, pp. 55–59, 2009. [Online]. Available: https://injuryprevention.bmj.com/content/15/1/55.short

[2] J. Lu, B. Li, H. Li, and A. Al-Barakani, "Expansion of city scale, traffic modes, traffic congestion, and air pollution," *Cities*, vol. 108, p. 102974, 2021.

[3] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2018.

[4] Y. Zheng, "Trajectory data mining," *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no. 3, pp. 1–41, 2015.

[5] E. N. Barmpounakis, E. I. Vlahogianni, and J. C. Golias, "Unmanned aerial aircraft systems for transportation engineering: Current practice and future challenges," *International Journal of Transportation Science and Technology*, vol. 5, no. 3, pp. 111–122, 2016.

[6] S. Bauer, E. Bretschneider, J. Rataj, and T. Schumann, "Aim – application platform intelligent mobility," 2009. [Online]. Available: https://elib.dlr.de/60757/

[7] S. Knake-Langhorst, K. Gimm, T. Frankiewicz, and F. Köster, "Test site aim – toolbox and enabler for applied research and development in traffic and mobility," *Transportation Research Procedia*, vol. 14, pp. 2197–2206, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352146516302381

[8] German Aerospace Center, "Test bed lower saxony: Innovation laboratory for the mobility of the future," 2024. [Online]. Available: https://www.dlr.de/en/ts/research-transfer/research-services/test-bed-lower-saxony

[9] C. Schicktanz, L. Klitzke, K. Gimm, S. Knake-Langhorst, G. Rizzo, H. H. Mosebach, and K. Liesner, "Dlr urban traffic dataset (dlr ut)," 2025. [Online]. Available: https://doi.org/10.5281/ZENODO.14773161

[10] C. Schicktanz, L. Klitzke, K. Gimm, R. Lüdtke, H. H. Mosebach, K. Liesner, F. Heuer, A. Wodtke, and L. Asbach, "Dlr highway traffic dataset (dlr ht)," 2025. [Online]. Available: https://doi.org/10.5281/zenodo.14811064

[11] U.S. Department of Transportation, Federal Highway Administration, "Next generation simulation (ngsim)," 2005. [Online]. Available: http://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm

[12] L. Li, R. Jiang, Z. He, X. M. Chen, and X. Zhou, "Trajectory data-based traffic flow studies: A revisit," *Transportation Research Part C: Emerging Technologies*, vol. 114, pp. 225–240, 2020.

[13] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2118–2125.

[14] Association for Standardization of Automation and Measuring Systems (ASAM e. V.), "Asam openscenario version 1.2.0: User guide," 2022. [Online]. Available: https://www.asam.net/standards/detail/openscenario/

[15] A. Tenbrock, A. König, T. Keutgens, and H. Weber, "The conscend dataset: Concrete scenarios from the highd dataset according to alks regulation unece r157 in openx," in *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*. IEEE, 2021, pp. 174–181.

[16] T. Moers, L. Vater, R. Krajewski, J. Bock, A. Zlocki, and L. Eckstein, "The exid dataset: A real-world trajectory dataset of highly interactive highway scenarios in germany," in *2022 IEEE Intelligent Vehicles Symposium (IV)*, 2022, pp. 958–964.

[17] P. Spannaus, P. Zechel, and K. Lenz, "Automatum data: Drone-based highway dataset for the development and validation of automated driving software for research and commercial applications," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 1372–1377.

[18] X. Shi, D. Zhao, H. Yao, X. Li, D. K. Hale, and A. Ghiasi, "Video-based trajectory extraction with deep learning for high-granularity highway simulation (high-sim)," *Communications in transportation research*, vol. 1, p. 100014, 2021.

[19] T. Seo, Y. Tago, N. Shinkai, M. Nakanishi, J. Tanabe, D. Ushirogochi, S. Kanamori, A. Abe, T. Kodama, S. Yoshimura, *et al.*, "Evaluation of large-scale complete vehicle trajectories dataset on two kilometers highway segment for one hour duration: Zen traffic data," in *2020 International Symposium on Transportation Data and Modelling*, 2020.

[20] C. Creß, W. Zimmer, L. Strand, M. Fortkord, S. Dai, V. Lakshminarasimhan, and A. Knoll, "A9-dataset: Multi-sensor infrastructure-based dataset for mobility research," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 965–970.

[21] D. Gloudemans, Y. Wang, J. Ji, G. Zachar, W. Barbour, E. Hall, M. Cebelak, L. Smith, and D. B. Work, "I-24 motion: An instrument for freeway traffic science," *Transportation Research Part C: Emerging Technologies*, vol. 155, p. 104311, 2023.

[22] L. Klitzke, C. Koch, and F. Köster, "Identification of lane-change maneuvers in real-world drivings with hidden markov model and dynamic time warping," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–7.

[23] L. Klitzke, K. Gimm, C. Koch, and F. Köster, "Extraction and analysis of highway on-ramp merging scenarios from naturalistic trajectory data," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 654–660.

[24] O. Siebinga, A. Zgonnikov, and D. A. Abbink, "Human merging behavior in a coupled driving simulator: How do we resolve conflicts?" *IEEE Open Journal of Intelligent Transportation Systems*, vol. 5, pp. 103–114, 2024.

[25] Z. e. a. Kherroubi and S. Aknine, "Review of driver behaviour modelling for highway on-ramp merging," *IET Intelligent Transport Systems*, vol. 18, no. S1, pp. 2793–2813, 2024. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/itr2.12572

[26] C. Schicktanz and K. Gimm, "Detection and analysis of corner case scenarios at a signalized urban intersection," *Accident Analysis & Prevention*, vol. 210, p. 107838, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000145752400383X

[27] C. Schicktanz, L. Klitzke, and K. Gimm, "Microscopic analysis of the impact of congestion on traffic safety and efficiency at a signalized intersection: A case study," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 2827–2834.

[28] C. Schicktanz and K. Gimm, "Detection and analysis of critical interactions in illegal u-turns at an urban signalized intersection," *Data Science for Transportation*, vol. 7, no. 1, p. 3, 2025.

[29] L. Klitzke, C. Schicktanz, K. Gimm, C. Koch, and F. Köster, "Scenario mining in the urban domain: Exploiting the topology of a road network for maneuver annotation and scenario extraction," *Authorea Preprints*, 2023.

[30] pyoscx, "scenariogeneration," https://github.com/pyoscx/scenariogeneration, 2025, version 0.14.5.

[31] esmini, "esmini," https://github.com/esmini/esmini, 2025, version 2.45.2.

[32] L. Klitzke and C. Schicktanz, "Tasi: A python library for traffic data analysis and situation interpretation," 2024. [Online]. Available: https://github.com/DLR-TS/TASI

[33] Association for Standardization of Automation and Measuring Systems (ASAM e. V.), "Asam opendrive version 1.4.0," 2015. [Online]. Available: https://www.asam.net/standards/detail/opendrive/

[34] J. Bennett, *OpenStreetMap*. Packt Publishing Ltd, 2010.

[35] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16. [Online]. Available: https://carla.readthedocs.io/en/0.9.11/tuto_G_openstreetmap/