Adverse Weather Benchmark Dataset for LiDAR-based 3D Object Recognition and Segmentation in Autonomous Driving

Dominik Weikert
Otto-von-Guericke-University
Magdeburg, Germany
dominik.weikert@ovgu.de

Christoph Steup
Otto-von-Guericke-University
Magdeburg, Germany
christoph.steup@ovgu.de

Sanaz Mostaghim
Otto-von-Guericke-University
Magdeburg, Germany
sanaz.mostaghim@ovgu.de

***Abstract*—Current developments in flexible mobility solutions are striving towards autonomous electric driving in the areas of public transport and logistics. The benefits of such systems are lower costs, higher availability, and greater flexibility. Object detection and segmentation techniques based on LiDAR sensors to complement camera and GPS data are essential for reliable behavior in autonomous driving. However, little research has been done to evaluate these techniques in representative adverse weather conditions such as rain, fog, or snow. Consequently, this paper presents a new dataset based on adverse weather data present in already widely used public datasets. The existing data is complemented with additional weather labels to facilitate the evaluation of object detection and segmentation in various weather conditions. To generate a baseline, a state-of-the art 3D object recognition is evaluated using the enhanced dataset. The results show a strong impact of the weather conditions on the performance of the evaluated baseline algorithm, indicating the relevance of the benchmark.**

***Index Terms*—dataset, autonomous driving, adverse weather**

I. INTRODUCTION

The proportion of autonomous electric vehicles for passenger transport on public roads will increase in the future, aiming to reduce individual traffic and minimize traffic volume. Autonomous vehicles are already capable of handling many standard traffic situations and precisely locating themselves, which is essential for safe and reliable autonomous operation. However, difficult weather conditions, communication outages, and limitations in satellite navigation can lead to reduced and inaccurate environmental perception and localization. This creates three concrete problems: absence of reliable navigation because of loss of localization reduced localization, emergency braking due to erroneously detected obstacles and complete loss of functionality, through the activation of safety mechanisms. To enable reliable and safe autonomous driving under such conditions, intelligent automated solutions are needed to assess the situations and mitigate them. An important step towards this is the development of methods capable of behaving reliably in adverse weather conditions. However, the effectiveness of such methods hinges on the availability of extensive datasets for training artificial intelligence-based methods. While several datasets are available for research purposes, some of which are outlined in Table I, they predominantly focus on daytime driving in sunny or cloudy weather. By integrating several diverse datasets, a more comprehensive weather dataset can be obtained. These datasets lay the groundwork for developing and validating methods that effectively combat or remove weather-related disturbances from sensor data. In addition to the scarcity of specific weather conditions in the individual datasets, the heterogeneity of their design and data format further complicates the task of establishing a standardized and cohesive weather-aware dataset. As such, extension of the available data with weather information obtained from weather databases, as well as manual or automatic labeling, is critical. In this paper, we propose a combined and homogenized dataset extended with weather information, offering a broader spectrum of scenarios for training and testing autonomous systems.

This work was partially funded by the German Ministry for Education and Research (BMBF) with grant number 01IS22061A.

TABLE I
DATASETS TO BE INCLUDED IN THE WEATHER DATASET

	nuScenes	**Apolloscape**	**CADCD**	**Waymo**
LiDAR	✓	✓	✓	✓
Camera	✓	✓	✓	✓
GPS	(✓)	✓	✓	
Rain	✓	✓		✓
Snow			✓	
Nighttime	✓			✓

II. RELATED WORK

While some datasets try to provide data from different environments and conditions, there is a large bias towards standard scenarios in most available datasets. Consequently, multiple datasets need to be combined to enhance the availability of special situations. Table I provides an overview of the currently planned integrations of open datasets. **NuScenes** [1] is an autonomous driving dataset containing data collected from

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00031

several driving routes in Boston and Singapore. In addition to LiDAR and camera data, full bus data and radar data are included. While there is no GPS data of the vehicles published, the GPS coordinates of the starting points as well as the route layout are available.The **ApolloScape** [2] Dataset for autonomous driving includes about 100000 image frames and 80000 LiDAR frames of urban traffic in varying conditions. In addition to the provided data collected by camera and LiDAR sensors, GPS and IMU data is provided. The **Canadian Adverse Driving Conditions Dataset (CADCD)** [3] is focused specifically on data collected in adverse weather conditions, specifically snow and winter conditions. Camera, LiDAR, GPS and IMU data is provided. The **Waymo Open Dataset** [4] is a diverse dataset providing labeled camera and LiDAR, as well as limited IMU data, mainly collected in San Francisco and Phoenix. No GPS information is provided; however, the data is already annotated with simple weather information such as ”sunny” or ”rain”.

The discussed datasets show a notable lack of weather annotation, and challenging weather situations are largely underrepresented in them. Furthermore, they lack a standardized format, which complicates the integration. These limitations highlight the need for dataset standardization, combination and expansion.

III. The proposed dataset

This work presents a combined dataset, unifying the different datasets discussed above into a single dataset. Additionally, by using the location information of each data point, weather information for the collected data can be obtained from online weather databases. Thus, weather annotations are provided to facilitate evaluation of methods on specific weather scenarios. The combined and annotated data is stored in a relational database. The relational schema, as depicted in Fig. 1, is designed to systematically organize and manage the combined dataset. The primary table in this schema is the weather table, where each entry is uniquely identified by a pair of primary keys consisting of time and location. The base weather attribute is a string describing the weather condition, based on the World Meteorological Organization (WMO) Code Table. In the case where weather annotations but not exact locations are provided, it is adapted to match the WMO. For each sensor type (e.g., LiDAR, camera), a dedicated table is established. While the current focus of this work is mainly on LiDAR-specific data, sensor data beyond LiDAR will be included in the future. The sensor tables are linked to the weather table through the time and location foreign keys, associating the sensor measurements with corresponding weather conditions. Using this relation, subsets of data containing only specific weather, times of day, and sensor data, can easily be retrieved. Finally, a table detailing the pose and transformation matrix from local to global coordinates for each sensor are provided in an additional table. This will facilitate fusing data from different sensors, such as merging all LiDAR point clouds into a single cloud, or enhancing LiDAR data with camera pixel data and vice versa.

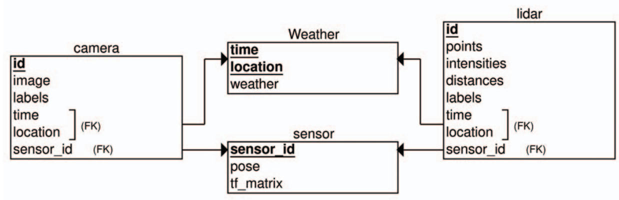

Fig. 1. Relational Schema of the Dataset

IV. Evaluation

To evaluate the relevance of the proposed dataset, a state-of-the-art 3D object recognition approach [5] is evaluated on two subsets of the Waymo Open Dataset. The first subset contains only sensor data without precipitation while the second subset includes rain.The model is trained according to the settings suggested by the original authors. On the sunny subset, the trained model reached an Intersection over Union (IoU) of 0.609. On the rainy subset, the IoU fell to 0.3167. While the two sets differed significantly in size, this is still a very significant drop in prediction quality and suggests the relevance of a dataset containing weather annotations and further benchmarks.

V. Conclusion

This paper proposed a new, unified dataset for autonomous driving in adverse weather conditions. By combining the data of multiple open datasets, and augmenting the existing data with weather annotations, the available data for AI-based autonomous systems can be increased significantly regarding challenging weather. Evaluation of a state-of-the-art approach to 3D object recognition showed a significant reduction in detection quality in rainy weather, which underscores the potential usefulness of the proposed data set. Future work on the dataset will include further datasets, more diverse sensor data, additional manual labeling regarding weather, and the incorporation of especially collected adverse weather data.

References

[1] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nuscenes: A multimodal dataset for autonomous driving,” in *CVPR*, 2020.

[2] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, “The apolloscape dataset for autonomous driving,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 1067–10 676.

[3] M. Pitropov, D. E. Garcia, J. Rebello, M. Smart, C. Wang, K. Czarnecki, and S. Waslander, “Canadian adverse driving conditions dataset,” *The International Journal of Robotics Research*, vol. 40, no. 4–5, p. 681–690, Dec. 2020. [Online]. Available: http://dx.doi.org/10.1177/0278364920979368

[4] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, “Scalability in perception for autonomous driving: Waymo open dataset,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[5] X. Lai, Y. Chen, F. Lu, J. Liu, and J. Jia, “Spherical transformer for lidar-based 3d recognition,” in *CVPR*, 2023.