ROSBENCH: A Simulation-Based Benchmark for Sensor Quality and Environmental Conditions Robustness in AV Perception

Matej Halinkovic, Miroslav Kunovsky, Marek Galinski
Slovak University of Technology, Ilkovičova 2, 842 16 Bratislava, Slovakia
matej.halinkovic@stuba.sk

2025 International Symposium ELMAR | 979-8-3315-9679-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/ELMAR66948.2025.11194028

***Abstract*—The perception capabilities of autonomous vehicles (AVs) rely on high-quality sensor data to accurately interpret the environment. Among the key sensing modalities, LiDAR and RGB cameras offer distinct advantages. LiDAR provides precise depth estimation and object localization, while cameras capture rich visual details. However, their effectiveness depends on factors such as resolution, measurement accuracy, and environmental conditions, making a systematic comparison essential for optimizing AV perception. Sensor fusion, which integrates multiple sensing modalities, can improve robustness by mitigating the limitations of individual sensors. The growing reliance on simulation-based research has accelerated AV development, with platforms like CARLA providing scalable, cost-effective environments to evaluate sensor performance under controlled yet diverse conditions, including adverse weather and complex traffic scenarios. In this work, we propose a comprehensive and robust simulated benchmark ROSBENCH for evaluating the robustness of Perception and Prediction systems. This benchmark can be used as a consistent reference point for validating the impact of environmental and sensor conditions on vision algorithms and is also easily configurable and thus easily adaptable.**

***Keywords*—autonomous vehicles; computer vision; perception and prediction; multimodality**

I. Introduction

Autonomous vehicles (AVs) rely on high-quality sensor data for accurate perception and decision-making. As Liu et al. [1] review, existing AV datasets vary significantly in sensor modalities, annotation styles, and environmental conditions. While LiDAR provides precise depth information and RGB cameras offer rich visuals, leading datasets like KITTI [2], NuScenes [3], and Waymo [4] differ in how they report sensor quality.

A key limitation in the field is the lack of a benchmark for evaluating not only sensor input quality but also the performance of deep learning models under varying conditions. In practice, researchers often degrade high-quality data post-hoc via downscaling, compression, or noise injection to simulate lower-quality inputs. These ad hoc modifications introduce artefacts that CNNs may overfit to, impairing generalization [5], [6]. Moreover, current AI models, typically evaluated on pristine datasets, may not generalize well to real-world scenarios with fluctuating sensor and environmental conditions.

To address this, we propose a framework that utilizes the CARLA simulator [7] to generate sensor data at predefined quality levels. By replicating realistic sensor setups and environmental conditions, our method avoids artificial degradation, ensuring reproducibility and fine-grained control over factors like resolution, weather, and lighting.

Our dataset supports not only perception tasks but the full Perception and Prediction pipeline, offering diverse environmental variations and a foundation for rigorous model evaluation under real-world-like conditions.

Our key contributions are:

- A configurable benchmark for evaluating the impact of sensor quality in AV perception.
- Realistic simulation of various environmental conditions using CARLA.
- Simultaneous multi-quality data capture without post-hoc degradation.
- Support for full Perception and Prediction pipelines

The benchmark data [1] and code [2] are publicly available.

II. Related work

To date, no existing dataset provides sensor data at controlled, varying quality levels. Standard AV benchmarks such as KITTI [2], NuScenes [3], and Waymo [4] offer high-quality sensor streams widely used for tasks like object detection. However, studies like Dodge and Karam [5] show that even minor artefacts (e.g., blur, noise) can significantly degrade CNN performance. Hjaltén [6] similarly highlights that lossy compression and downscaling lead to blur and blocking artefacts that reduce classification accuracy. These findings emphasize that artificial degradation does not reliably mimic real-world sensor imperfections and may cause models to focus on artefacts rather than robust object features. This is especially important when considering simple solutions such as those proposed by Masarykova et al. [8] which do not have the learning capacity to overcome and filter out said artefacts.

Simulation environments such as CARLA provide an alternative by enabling controlled, artefact-free generation of sensor data under diverse environmental and operational settings. This ensures reproducibility and eliminates the need for post-hoc degradation. It can also help with determining the sensoric equipment needed for sufficient readability of scenes

[1] https://tinyurl.com/bdtxstwv

[2] https://github.com/mathali/ROSBENCH.git

979-8-3315-9679-8/25/$31.00 ©2025 IEEE

and infrastructure, which is an important consideration [9]. The cost of sensors is a prohibitive factor when it comes to the adoption of advanced autonomous capabilities; as such, being able to determine the minimum viable sensor suite for accomplishing a certain task can be highly beneficial.

NuScenes stands out as a comprehensive multimodal dataset with 1,000 urban driving scenes captured under varied weather and traffic conditions. It includes six 1600×900 cameras for full 360° coverage, a 32-beam LiDAR at 20Hz, five radar sensors, and synchronized annotations and calibrations. Its standardized format supports temporal tasks like detection and motion prediction.

By aligning with NuScenes' structure, our benchmark ensures compatibility with existing models while extending capabilities to quality-controlled data generation. This compatibility enhances usability and underscores the importance of standardized tools for evaluating sensor performance.

In sum, while existing datasets provide high-quality data, they lack consistent methods for assessing the impact of sensor quality. Our work leverages CARLA's simulation to fill this gap, offering a tool for generating multi-quality sensor data without introducing artefacts.

III. Proposed benchmark

Building upon the structured sensor configurations of the NuScenes dataset, our work presents a novel framework designed to generate synthetic sensor data with configurable quality parameters. Our framework leverages the realistic simulation environment provided by the CARLA simulator to produce data at the desired quality levels directly without relying on artificial post hoc degradation.

Our approach replicates the sensor placements from NuScenes, particularly for cameras and LiDAR sensors. However, the flexibility of the CARLA simulator affords us several key advantages. First, we can position multiple virtual cameras at identical physical locations and orientations. Each of these cameras can be configured with varying imaging parameters, such as resolution, f-stop and ISO. This design enables the simultaneous capture of the same scene at multiple quality levels.

Furthermore, the use of CARLA allows us to exploit its fully controllable environment. We can systematically vary environmental conditions, including weather phenomena (e.g., rain, fog, and varying light conditions), traffic scenarios and road types to create a diverse and representative dataset. We include 500 scenes, each lasting 20 seconds, generated from different CARLA maps and under a broad spectrum of environmental conditions. This extensive variability is critical for evaluating the robustness of object detection models under real-world conditions and for assessing how sensor data quality influences model performance.

A. Generation Process

The simulation script used to control the generation process is robust, considers various environmental aspects and is fully configurable and easily adaptable. All scenes are set up in the following manner:

a) Spawn the ego vehicle at a predefined location: The location can be either manually specified or randomly selected from the list of all available spawn points on the currently loaded CARLA map.

b) Populate the environment with other actors, such as cars, trucks, motorcycles, buses, and pedestrians: The number and types of actors can be configured; in the current setup, vehicles are spawned at random positions using available spawn points. This introduces variability across scenes and enhances environmental diversity for each simulation run.

c) Attach sensors to the ego vehicle: Sensors are configured and loaded dynamically from external JSON configuration files, allowing flexible setup of sensor types, positions, and parameters. The default positioning of cameras, for all quality levels, adheres to the Nuscenes positioning as shown in Figure 1, which allows for easy validation of existing solutions.

d) Start autopilot for all actors: This initiates the movement of all spawned actors, including the ego vehicle, enabling dynamic and realistic traffic interactions during the simulation.

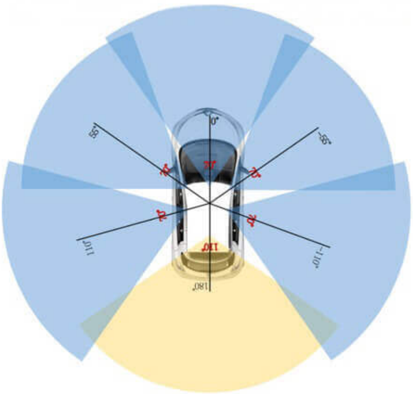

Figure 1. Default camera positioning adheres to Nuscenes [3].

Scene setup can be adjusted by modifying *scenes.json* and types and quality of sensors can simply be adjusted in *sensor_config.json*. This lets a user easily adapt the benchmark to their specific needs with ease.

All dataset components are structured according to the NuScenes data format [3], ensuring compatibility with existing tools, benchmarks, and model architectures developed for NuScenes. This includes:

- Sensor calibration files (calibrated_sensor.json)
- Ego poses and timestamps (ego_pose.json, sample_data.json)
- Object annotations with 3D bounding boxes (sample_annotation.json)
- Scene metadata and logs (scene.json, log.json)
- Sensor configuration and mappings (sensor.json, sample.json)

Camera images are stored as JPEG files and follow the NuScenes directory layout, organized by sensor name and timestamp. LiDAR point clouds are provided in binary .bin format and linked via *sample_data* entries. Data from multiple quality levels is captured simultaneously, with each quality level treated as a distinct sensor in the metadata.

CARLA provides ID tracking of actors across time, allowing for bounding box tracking across time. However, only dynamic actors, those manually spawned during simulation setup, can be reliably accessed and tracked. Static actors, such as parked cars or objects embedded in the map like traffic lights, are not directly exposed through the same interface and do not provide actor IDs or bounding boxes that can be tracked over time. Thus, to be able to facilitate the inclusion of static actors while maintaining the robustness of provided annotations, we devised the following prediction-based bounding box matching Algorithm 1, which supplements and matches unmatched bounding boxes by filtering existing boxes based on their last known positions and velocities.

Algorithm 1 Prediction-Based Bounding Box Matching

```
Data: A set of candidate tracks $candidates$, a detected bounding box $det$ with centroid $det.centroid$, prediction radius $R$.
Result: Matched track $matched$ and $match\_method$.
$matched \leftarrow$ None
if $matched =$ None and $candidates \neq \emptyset$ then
    $best\_dist \leftarrow +\infty$
    $best\_track \leftarrow$ None
    foreach $t$ in $candidates$ do
        $velocity \leftarrow t.last\_centroid - t.prev\_centroid$
        $predicted \leftarrow t.last\_centroid + velocity$
        $d \leftarrow \|det.centroid - predicted\|$
        if $d < best\_dist$ then
            $best\_dist \leftarrow d$   $best\_track \leftarrow t$
        end
    end
    if $best\_dist < R$ then
        $matched \leftarrow best\_track$
        $match\_method \leftarrow$ "prediction"
    end
end
if $matched =$ None then
    // Optional fallback if no prediction
       match
    Try order-based fallback  if still no match then
        Create a new track and set $matched$
    end
end
if $matched \neq$ None then
    Update the state of $matched$
end
```

B. Dataset Statistics

Our dataset includes sensors configured with multiple quality levels. For the cameras, we support three quality settings: Low, Mid and High with adjustable resolutions and intrinsic parameters (f-stop, ISO, shutter speed, etc.). The predetermined quality settings of RGB sensors can be seen in Table I.

TABLE I. Camera Sensor Specifications. The field of view (FOV) for the back-facing camera is set to 110°, while the FOV for all other cameras is 70°.

Quality Level	Resolution	f-stop	ISO	Shutter Speed	Bloom / Lens Flare
High	1920×1080	8.0	100	500 ms	– / –
Mid	1280×720	4.0	200	200 ms	0.3 / 0.1
Low	854×480	4.0	800	100 ms	0.4 / 0.2

For LiDAR, three quality levels are defined by parameters such as the number of beams, range, points per second, and field-of-view (FOV) settings. In addition, the capturing frequency of sensors is set to 10 fps for the cameras and 10 Hz rotation frequency for LiDAR. The predefined LiDAR quality level settings can be seen in Table II.

TABLE II. LiDAR Sensor Specifications

Quality Level	Beams	Points/sec	Atten. Rate	Dropoff	Noise
High	64	1,300,000	0.003	(0.25, 0.7, 0.3)	0.01
Mid	32	300,000	0.004	(0.35, 0.8, 0.4)	0.01
Low	16	150,000	0.005	(0.45, 0.8, 0.5)	0.02

As a result, each 20-second scene produces approximately 200 samples, and with a total of 500 scenes, this yields around 100,000 samples per quality level for each camera. Considering six cameras and three quality levels, the dataset comprises approximately 1.8 million image samples in total. A comparable sampling strategy is applied to the LiDAR sensor, generating 100,000 samples per sensor configuration amounting to 300,000 samples overall. In each scene, we provide annotations for 7 classes - car, truck, motorcycle, pedestrian, bicycle, bus, and traffic light. In total, we provide annotations for 1.3 million unique bounding boxes (Figure 3) and 13,375 unique actor instances distributed as shown in Figure 2.

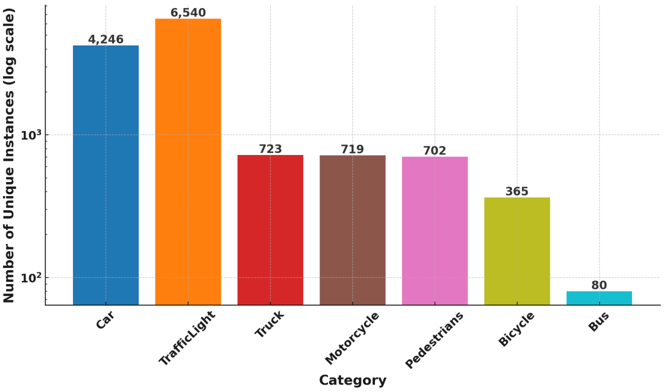

Figure 2. Unique instances per category (log scale) provided by ROSBENCH

To ensure diversity and robustness under varying environmental conditions, the dataset incorporates a comprehensive range of weather and lighting scenarios across four distinct urban maps: Town01, Town03, Town04, and Town10HD. Each map includes data collected under five predefined weather presets: ClearNoon, ClearSunset, ClearNight, HardRainNoon, and HardRainNight. These settings encompass a broad spectrum

TABLE III. Feature Comparison of ROSBENCH with Major AV Datasets (✓= Supported, ✗= Not Supported)

Dataset	RGB	360° RGB	LiDAR	Sensor Quality Variants	Weather	Lighting	Tracking / Prediction
ROSBENCH	✓	✓	✓	✓	✓	✓	✓
nuScenes [3]	✓	✓	✓	✗	✓	✓	✓
KITTI [2]	✓	✗	✓	✗	✗	✗	✓
Waymo [4]	✓	✓	✓	✗	✓	✓	✓
Argoverse [10]	✓	✓	✓	✗	✓	✓	✓
V2X-Sim [11]	✓	✓	✓	✗	✗	✗	✓

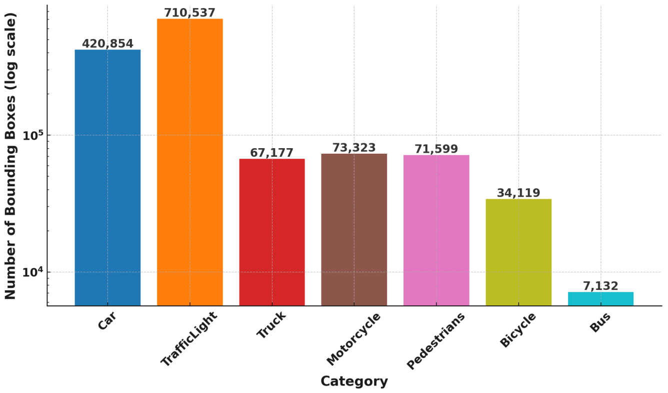

Figure 3. Bounding boxes per category (log scale) provided by ROSBENCH

of illumination levels and visibility challenges, from high-contrast daylight to low-light nighttime conditions and heavy rainfall. For each weather condition within each town, 25 scenes were recorded, resulting in a well-balanced distribution of environmental variations. This diversity, combined with the extensive volume of data, provides a robust foundation for training, validating, and testing Perception and Prediction models capable of generalizing across a wide range of real-world scenarios.

IV. Conclusion

ROSBENCH introduces a configurable and reproducible platform for evaluating autonomous vehicle perception and prediction systems. Unlike existing real-world datasets, which rely on fixed sensor setups and uncontrollable environmental conditions, ROSBENCH enables systematic manipulation of sensor parameters (e.g., resolution, noise) and scene attributes (e.g., weather, lighting). This supports rigorous benchmarking of model robustness under a wide range of scenarios.

As shown in Table III, ROSBENCH uniquely supports sensor quality variation, full 360° RGB and LiDAR coverage, and diverse environmental conditions, distinguishing it from datasets like KITTI, Waymo, and nuScenes. These features enable targeted studies on sensor degradation, adverse conditions, and fusion strategies.

Furthermore, the availability of dense ground truth and consistent object IDs over time makes ROSBENCH suitable for multi-object tracking and trajectory prediction tasks. Its fully synthetic nature also makes it ideal for sim-to-real research via domain randomization and adaptation.

Acknowledgment

This project was Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00014 and by APVV-23-0519 project “Legal and Technical Challenges of Smart Mobility to Increase Road Traffic Safety”, and by the internal grant scheme of the Slovak University of Technology to support Young Researchers.

The authors acknowledge the use of generative AI tools for language editing and grammar refinement during the preparation of this manuscript. These tools were used to improve clarity and readability, but all content, ideas, and interpretations are solely those of the authors.

References

[1] M. Liu, E. Yurtsever, E. Fossaert, X. Zhou, W. Zimmer, Y. Cui, B. Zagar, and A. Knoll, “A survey on autonomous driving datasets: Statistics, annotation quality, and a future outlook,” *IEEE Transactions on Intelligent Vehicles*, pp. 1–29, 2024.

[2] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361, 2012.

[3] 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,” 2019.

[4] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Vasudevan, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasanthakumar, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, S. Zhao, S. Cheng, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, “Scalability in perception for autonomous driving: Waymo open dataset,” 2019.

[5] S. Dodge and L. Karam, “Understanding how image quality affects deep neural networks,” 2016.

[6] A. Hjälten, “How image downscaling and jpeg compression affects image classification performance: An experimental study,” tech. rep., Umeå University, Faculty of Science, 2019. URN: urn:nbn:se:umu:diva-163308.

[7] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” in *Proceedings of The 1st Annual Conference on Robot Learning*, pp. 1–16, 2017.

[8] N. Masarykova, M. Galinski, and P. Truchly, “Single-filter cnn for vehicle recognition,” in *2024 International Symposium ELMAR*, pp. 5–8, IEEE, 2024.

[9] P. Lehoczkỳ, M. Janeba, M. Galinski, and L. Šoltés, “Testing the readiness of slovak road infrastructure for the deployment of intelligent transportation,” *Strojnícky časopis-Journal of Mechanical Engineering*, vol. 72, no. 2, pp. 103–112, 2022.

[10] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, *et al.*, “Argoverse: 3d tracking and forecasting with rich maps,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8748–8757, 2019.

[11] Y. Li, D. Ma, Z. An, Z. Wang, Y. Zhong, S. Chen, and C. Feng, “V2x-sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving,” *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10914–10921, 2022.