POLITECNICO DI TORINO
Repository ISTITUZIONALE

Descriptor: Context-Aware Collaborative Perception in Autonomous Driving Dataset (ConVeX)

Original
Descriptor: Context-Aware Collaborative Perception in Autonomous Driving Dataset (ConVeX) / Palena, M., Selvaraj, D.C., Chiasserini, C.F., Cerquitelli, T.. - In: IEEE DATA DESCRIPTIONS. - ISSN 2995-4274. - 3:(2026), pp. 531-538. [10.1109/IEEEDATA.2026.3706437]

Availability:
This version is available at: 11583/3010047 since: 2026-06-25T06:24:44Z

Publisher:
IEEE

Published
DOI:10.1109/IEEEDATA.2026.3706437

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

16 October 2026

An IEEE Societies and Technical Councils Publication
IEEE Data Descriptions

Received 00 April, 2024; revised 11 April, 2024; accepted 00 April, 2024; Date of publication 00 May, 2024; date of current version 24 June, 2024.

Digital Object Identifier 10.1109/DD.2024.0624000

Descriptor: *Context-Aware Collaborative Perception in Autonomous Driving Dataset (ConVeX)*

Marco Palena[1], Dinesh Cyril Selvaraj[2], Carla Fabiana Chiasserini[1,2,3] (FELLOW, IEEE), Tania Cerquitelli[4]

[1]CNIT, Italy
[2]CARS@Polito, Politecnico di Torino, Italy
[3]Chalmers University of Technology, Sweden
[4]Department of Control and Computer Engineering, Politecnico di Torino, Italy

CORRESPONDING AUTHOR: Marco Palena (e-mail: marco.palena@polito.it).

ABSTRACT Collaborative perception, using vehicle-to-everything (V2X) communication to share sensor data between connected vehicles and between the latter and the network infrastructure, has emerged as a prominent solution to extend the view of single autonomous vehicles. The effectiveness of this paradigm, however, may be hindered by the presence of adverse weather conditions and changes in lighting, often affecting real-world scenarios. Thus, assessing the robustness of collaborative perception to environmental contingencies is still an open issue. Importantly, although some large-scale datasets for collaborative perception, comprising realistic and simulated data, are now publicly available, most of them lack diversity in terms of environmental conditions in the autonomous driving scenarios they collect, making it difficult for researchers to assess how such conditions may affect perception performance. We thus introduce ConVeX, an extensive multi-agent synthetic dataset for collaborative perception (CP) that reproduces different realistic driving scenarios (urban, rural, highway), road layouts, and weather and lighting conditions. Remarkably, ConVex includes multi-modal data (i.e., images from RGB (Red-Green-Blue) cameras, LiDAR (Light Detection and Ranging) points, and GNSS (Global Navigation Satellite System) coordinates) collected by different vehicles, and ground-truth annotations for object detection.

IEEE SOCIETY Vehicular Technology Society (VTS)

DATA DOI/PID <10.21227/naks-xp87>

DATA TYPE/LOCATION

INDEX TERMS Autonomous vehicles, Collaborative perception, Computer vision

BACKGROUND

Autonomous driving has garnered significant interest from academia and industry due to its potential to deliver safer and more efficient driving experiences. However, most current autonomous driving systems focus on individual vehicles, which limits their environmental perception because of restricted fields of view, thereby continuing to pose safety risks. Collaborative perception—where autonomous vehicles and road infrastructure devices cooperate through vehicle-to-everything (V2X) communication—has recently emerged as a promising approach to overcoming these limitations. Nevertheless, various factors such as adverse weather conditions and insufficient lighting may still lead to low perception accuracy, exacerbating safety concerns [1], [2]. Weather conditions like rain, snow, and fog often degrade the color and texture quality of images while introducing noise that distorts the distribution of LiDAR points. Similarly, low-light environments may cause the loss of details and contrast in images. At the same time, intense direct light from either the sun or artificial light sources may result in overexposed images and induce noise in LiDAR-collected data [3]. These data inaccuracies in environment perception can cause errors or malfunctions in the vehicle's perception system, limiting

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

the applicability of autonomous driving to only specific, highly favorable conditions [4].

Developing solutions to mitigate these issues requires access to comprehensive datasets that encompass a wide range of weather conditions and lighting variations. However, most datasets commonly used to train perception algorithms for autonomous driving lack diversity in these aspects, as they are typically biased toward daytime and clear-weather conditions [3], [5]. As a consequence, recent years have witnessed the emergence of an increasing number of datasets designed to represent a wide range of weather and lighting conditions [5]–[9]. However, due to inherent biases and the difficulty of capturing a broad and diverse range of weather and lighting scenarios in real-world settings, these datasets are often synthesized using autonomous driving simulators such as CARLA [10]. Still, these datasets typically collect data from a single sensing agent per frame, making them unsuitable for supporting research on collaborative perception. A notable exception is Adver-City [11], a multi-agent, multi-modal synthetic dataset for collaborative perception specifically designed to capture a variety of adverse environmental conditions. Furthermore, most studies fail to treat weather conditions and time of day as distinct sources of noise in the perceived data. For instance, variations in lighting are often categorized as separate instances of weather conditions, disregarding the challenges that arise when both factors occur simultaneously – such as a rainy or snowy road at night. Additionally, existing adverse-condition datasets, including synthetic ones, are not free from biases, as specific operating scenarios often occur only in a subset of the overall scenes and, in general, adverse conditions are not systematically varied across the different scenes. Most of these datasets are designed to maximize diversity by augmenting different scenes with adverse weather and lighting conditions. This makes it difficult to assess to which extent adverse conditions impact the performance of perception algorithms in relation to the scene. In contrast, we focus on a dataset that enables a rigorous assessment of the robustness of collaborative perception algorithms under adverse operating scenarios by collecting data from the same scenes across varying conditions. Table 1 summarizes these observations, highlighting the characteristics of existing datasets for collaborative perception and adverse conditions.

To fill the above gaps and overcome the limitations of existing datasets, in this work, we propose ConVeX, a large-scale, *multi-agent*, synthetic dataset designed to *support V2X collaborative perception research with a focus on adverse operating conditions*. In contrast to similar works like [11], ConVeX is designed to evaluate collaborative V2X perception robustness under systematically varied adverse weather and lighting, enabling precise ablation studies across a variety of urban. rural, and highway scenarios. Further, we leverage the advanced capabilities of SCANeR™ Studio [12], a state-of-the-art industrial autonomous driving simulation platform, to collect images and LiDAR points from multiple vehicular and roadside agents in a variety of virtual scenes, encompassing urban, rural, and highway scenarios.

Unlike existing synthetic datasets, which are typically generated using open-source simulators like CARLA, our dataset is created using a proprietary, industrial-grade driving simulator widely employed by car manufacturers. Running simulations with this tool requires specialized expertise, significant time, and considerable computing resources. Thus, making this dataset publicly available will greatly benefit the autonomous driving research community. For each scene, we collect data under a wide range of operating conditions by running multiple simulations with various combinations of weather and time of day settings. To the best of our knowledge, ConVeX is the first synthetic, multi-agent dataset for collaborative perception specifically designed to address robustness of collaborative perception under adverse operating conditions.

Furthermore, to support downstream perception tasks, such as detection and tracking, we annotate each data sample with 3D bounding boxes, in addition to labelling each data sample with contextual information such as time of day and weather conditions. Importantly, we also make the raw simulation data publicly available, to allow other researchers to extend the dataset with additional annotations.

In summary, ConVeX offers the following key features:

- It provides multi-agent sensory streams that represent each scene as perceived by various vehicles and infrastructure devices;
- It does so in diverse driving scenarios, road layouts, and operating conditions determined jointly by weather and lighting variations, providing data under different sets of conditions for each scene to minimize biases;
- It provides per-vehicle multi-modal data, integrating RGB images with LiDAR points and GNSS coordinates;
- It includes ground-truth annotations for the widely popular task of object detection and can be easily extended to support other tasks as well;
- It provides data obtained via accurate and extensive simulations that would not be otherwise available due to the proprietary nature of the simulator;
- It makes data easily accessible, eliminating the need for expertise in designing relevant scenarios or using the simulator, as well as the expense of substantial computational resources.

COLLECTION METHODS AND DESIGN

Collecting autonomous driving data in real-world environments is both time-consuming and resource-intensive. Additionally, field tests are notoriously difficult to control and replicate, often raising safety concerns for operators, especially under adverse conditions. Autonomous vehicle (AV) simulators provide an appealing alternative for data collection, allowing researchers to create specific scenarios that would be challenging or impossible to reproduce in real-world settings while offering precise control over operating

TABLE 1. Comparison of collaborative perception datasets for autonomous driving. Column "Comm. scenario" indicates the target communication paradigm, while "Adverse" and "Robustness" indicate, respectively, whether a dataset is targeted at adverse environmental conditions and whether it is designed to assess collaborative perception robustness to adverse conditions.

		Goal			Sensors					Scale			
Dataset	**Source**	**Comm. scenario**	**Adverse**	**Robustness**	**Cam**	**LiDAR**	**Depth**	**GNSS**	**IMU**	**Scenarios**	**Scenes**	**Frames**	**Agents**
V2V-Sim [13]	Sim [14]	V2V				✓				> 1	5.5K	51K	1-63
V2X-Sim [15]	Sim [10], [16]	V2X			✓	✓	✓	✓	✓	3	100	10K	2-5
OPV2V [17]	Sim [10], [16], [18]	V2V			✓	✓		✓	✓	$8 + 1$	73	11K	2-7
DAIR-V2X-C [19]	Real	V2I			✓	✓		✓	✓	1	100	39K	2
V2XSet [20]	Sim [10], [18]	V2X			✓	✓				5	55	11K	2-5
DOLPHINS [21]	Sim [10]	V2X			✓	✓				3	6	42K	3
DAIR-V2X-Seq [22]	Real	V2I			✓	✓			✓	1	95	20K	2
V2V4Real [23]	Real	V2V			✓	✓				1	67	15K	2
RACECAR [24]	Real	none			✓	✓		✓	✓	2	11	50K-130K	1-6
WHALES [25]	Sim [10]	V2X			✓	✓				> 5	n.a.	70K	8
TUMTraf-V2X [26]	Real	V2I			✓	✓		✓	✓	1	8	5K	2
RCooper [27]	Real	I2I			✓	✓				2	410	50K	2-4
V2X-Real [28]	Real	V2X			✓	✓		✓	✓	2	68	171K	4
ACDC [5]	Real	none	✓		✓					> 3	n.a.	4K	1
GTA [7]	Sim [29]	none	✓		✓					n.a	n.a	300K	1
Adver-City [11]	Sim [10]	V2X	✓		✓	✓		✓	✓	5	110	24K	5
ConVex (ours)	**Sim [12]**	**V2X**	✓	✓	✓	✓		✓		**12**	**156**	**26K**	**8-9**

conditions. In this study, we adopt this methodology by using an AV simulator to collect data from a wide range of virtual scenes, each evaluated under diverse operating conditions. Our approach enables the creation of a comprehensive dataset that accurately captures the complexity of real-world scenarios. Below, we explain our rationale for the simulator selection and the methods we used for data collection, processing, and labelling.

Tool Selection & Hardware Setup

Several open-source AV simulation tools exist, with CARLA [10], LG SVL [30], and AirSim [31] standing out as the most feature-rich and widely used. Unlike previous works relying on these academic, open-source tools, we collected data using SCANeR™ Studio [12], a professional, comprehensive simulation platform developed by AVSimulation, widely used in the automotive industry for testing and validating vehicle dynamics, ADAS (Advanced Driver Assistance Systems), and autonomous driving systems.

Differently from the aforementioned open-source simulators, SCANeR™ Studio offers a comprehensive suite of tools and models to create a highly realistic virtual world, encompassing road environments, vehicle dynamics, traffic, sensors, driving behavior, and environmental conditions. Compared to CARLA, which is primarily focused on autonomous driving research and sensor testing in urban environments, SCANeR™ Studio offers significantly more accurate models for simulating vehicle dynamics and environmental conditions. This enhanced realism is crucial for generating high-fidelity sensory data, particularly under adverse conditions.

To run our experiments, we used a workstation equipped with an 8-core Intel i7-11700K processor (3.6 GHz), 32 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU with 12 GB of RAM.

Dataset Design

Our data collection strategy is guided by the goal of capturing the same set of virtual scenes across a wide range of operating conditions. Unlike most previous works focusing on adverse-condition datasets, we treat weather (precipitation) and lighting variations (time of day) as independent factors influencing sensor performance, evaluating each scene under every possible combination of these conditions. Our approach enable researchers to assess the impact on perception algorithms of different operating conditions, individually or in combination. For instance, researchers can identify whether performance degradation stems from a specific weather or lighting variation by evaluating their algorithm on data collected from variations of the same base scene.

We set up 13 individual base scenes in SCANeR™ Studio, using a mix of pre-existing and custom-made maps. These scenes are designed to capture a wide variety of road layouts and driving conditions, in the *urban*, *rural*, and *highway* environments. For each scene, we collect sensory data from multiple agents, including connected autonomous vehicles (CAVs) and roadside units (RSUs). Table 2 provides a summary of the base scenes used to build the ConVeX dataset.

As depicted in Fig. 1, each vehicular agent is equipped with a suite of high-resolution RGB camera, LiDAR and GNSS sensors. Each color camera has a resolution of $1,600 \times 900$ pixels, with 120° horizontal FOV, 60° vertical FOV and focal length of 65 mm, recording frames at 20 fps. The simulated LiDAR sensor is mounted vertically on the roof of the vehicle, at a height of 1.8 m from the ground, and replicates the specifications of a Velodyne HDL-64E (64 channels, from -24.8° to $+2^\circ$ of vertical field of view, 20 Hz revolution frequency, 80 m of maximum range and $130K$ points per frame). Since SCANeR™ Studio does not allow for the deployment of fixed sensors within the

TABLE 2. Overview of the 13 base scenes used to create the ConVeX dataset and their distinctive features: number of frames, annotations, and agents (along with their relative average distance), as well as number of moving and static perception targets and their average number per frame.

Id	Scenario	Description	No. frames	No. annot.	No. agents		Targets		Avg no. targets per frame	Avg Agent Distance [m]
					CAVs	RSUs	Dyn.	Stat.		
U1_TJ_NTL	Urban	Unregulated T junction	756	4756	6	3	21	47	38	49.73
U4_TJ_TL		Regulated T junction	1716	8994	5	3	25	28	45	74.66
U6_4J_TL		Regulated 4-way junction	2676	18628	6	3	26	28	42	104.83
U7_4J_TL		Regulated 4-way junction	3468	14960	6	3	18	4	21	51.77
U8_RA_S		Roundabout small	1284	6545	6	3	19	6	25	39.32
U9_RA_L		Roundabout large	2316	6050	5	3	17	24	39	86.66
U10_SS_X		Straight segment, crosswalk	1776	8164	7	2	17	2	18	122.55
U11_SS_C		Straight segment, construction site	1536	4973	7	2	18	3	21	104.72
R1_2LR	Rural	2-lanes road	3180	19216	6	3	21	7	35	233.38
R3_NWR		Narrow and winding road	3408	5444	6	3	17	1	12	207.42
H4_TB	Highway	Toll booths	612	3384	7	2	19	32	40	268.62
H5_EE		Highway entering	1536	3057	7	2	15	21	20	160.54
H6_EE		Highway exiting	1824	5999	6	3	18	22	30	222.34

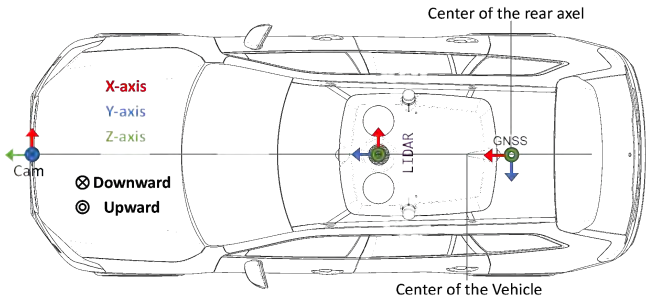

FIGURE 1. Layout of sensors for vehicular agents.

road infrastructure, we implemented RSUs as a workaround by attaching sensors to stationary pedestrian actors, which will be invisible in the simulations. The sensors of the infrastructural agents are positioned 15 m above the ground, looking diagonally downward at 35° and oriented toward key points of interest in the road infrastructure, such as intersections or toll booths. Each RSU is equipped with a color camera and a LiDAR sensor, both sharing the same global coordinates, and their specifications are identical to those of the vehicular agents. The sensor layout is identical across all vehicles and RSUs.

Regarding weather conditions, we consider four scenarios: *clear*, *cloudy*, *rainy*, and *snowy*. Regarding the time of day, we differentiate between twilight, daylight, and nighttime settings, corresponding to 6:00 a.m., 12:00 p.m., and 8:00 p.m., respectively. Fig. 2 shows a frame taken from a sample scene under all possible combinations of weather and lighting conditions.

Collection Methodology

For each base scene, we ran individual simulations collecting sensory data under all combinations of weather and time of day, thus obtaining a total of 156 scenes. After synchronizing each sensor in SCANeR™ Studio, we ran each simulation collecting data at a frequency of 20 Hz, resulting in each scene being composed of number of frames varying between 756 and 3,468. For each frame, we extract video streams from the RGB camera of the agents as well as LiDAR point clouds. In addition, we collect GNSS data from all vehicles appearing in the scene. We also instructed SCANeR™ Studio to generate a list of ground-truth bounding-box annotations for each potential detection target in every frame. To support a variety of downstream detection tasks, we included dynamic and static objects in our target list. Both types of objects are categorized into three main classes, namely, *vehicles*, *vulnerable vehicles*, and *pedestrians* for dynamic objects, and *traffic sign*, *traffic signal*, and *barrier* for static objects. Each main class is further divided into several subclasses. Fig. 3 illustrates the target classes breakdown and their occurrences in the ConVeX dataset.

Bounding boxes are derived from ground-truth annotations generated directly by the simulation environment. First, the 3D coordinates of the 8 corner points of each object's bounding box, expressed in the global reference frame of the simulation, are extracted from SCANeR™ Studio. They are then post-processed to obtain a compact representation consisting of the 3D position of the box center (x, y, z), its dimensions (w, l, h), and its 3D orientation (q_w, q_x, q_y, q_z). We then filter these annotations by removing the bounding boxes that contain no LiDAR points. Finally, we filter out those without any suitable annotations, yielding 26,088 annotated frames, each comprising the sample image and point cloud data for each agent.

VALIDATION AND QUALITY

We now present scenes statistics and the generated annotations to emphasize ConVeX data diversity and quality.

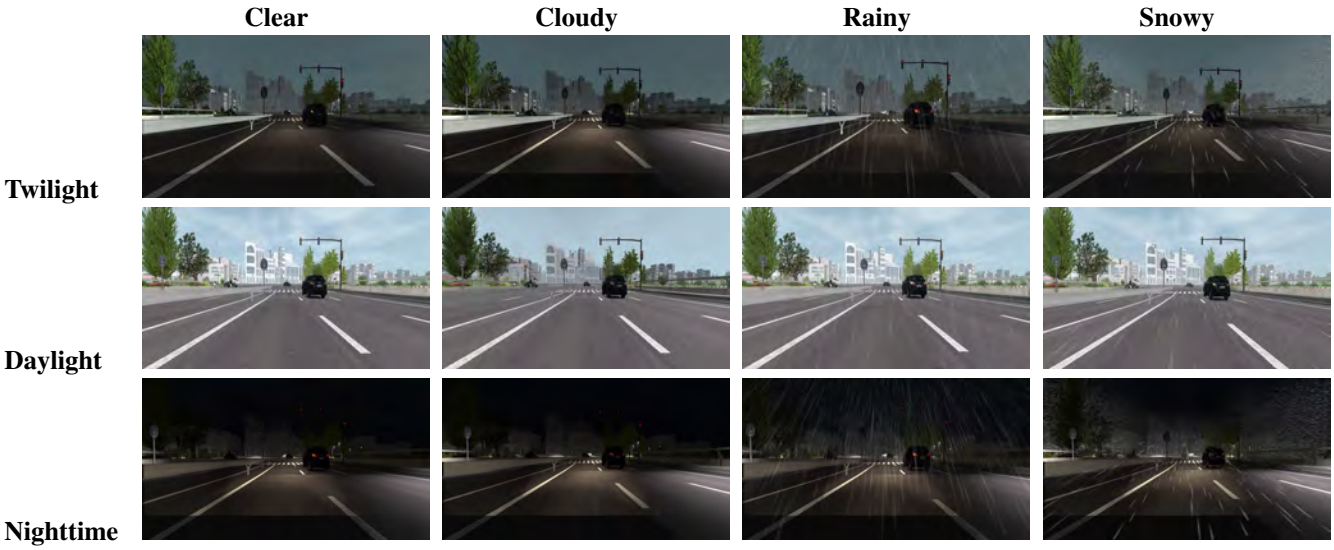

FIGURE 2. Sample scene depicted under different operating conditions.

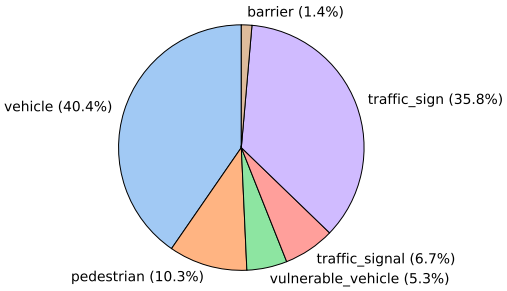

FIGURE 3. Distribution of main target classes in the ConVeX dataset.

Scenes statistics

Our scenes feature vehicles travelling at different speeds, from stationary up to a maximum of 136 km/h. Fig. 4 (a) illustrates the vehicle speed distribution in the urban, rural, and highway scenarios, where the speed values have been computed by considering only vehicles that are within 70 meters from at least another agent and by averaging, for each of such vehicles, its speed over all scenes and frames. We observe that, consistently with real-world behavior, urban scenes, primarily including vehicles approaching intersections, are characterized by lower speeds, mostly between 20 and 30 km/h. In rural and highway scenes, instead, vehicles tend to travel faster, with speeds mainly between 30 and 40 km/h in the rural scenario and higher than 90 km/h on highways.

The dataset includes scenes with varying vehicle density. Fig. 4 (b) illustrates the distribution of the minimum Euclidean distance between each pair of vehicles within a 70-meter range from at least one agent. Again, distance values are averaged across all scenes and frames. Urban scenes tend to be more crowded, with most vehicles located within 20 m from another vehicle. Highway scenes, while less dense than urban ones, still show quite high values of vehicular density, with most vehicles falling within a 20 to 30-meter range from at least one other vehicle. This is direct a consequence of the specific considered highway scenarios (i.e., vehicles approaching a toll booth or entering/exiting the highway via ramps. In contrast, rural scenes exhibit lower vehicle density, with some vehicles separated by as much as 90 m from the nearest vehicle.

Fig. 4 (c) shows the distribution of minimum longitudinal distances between CAV agents and other vehicles travelling on the same road and lane, providing a more realistic measure of inter-vehicle spacing in autonomous driving scenarios. Compared to Euclidean distances, all scenarios shift toward larger separations: urban traffic peaks around 20-–30 m, while rural and highway traffic are more spread out, with most vehicles maintaining 30—40 m headways, reflecting free-flow conditions and lane-aligned spacing.

Annotation statistics

Fig. 5 presents the occurrence of the number of annotations per frame, with most frames containing 3 to 6 bounding boxes. Fig. 6, instead, highlights the distribution of the size of the bounding boxes across different classes of target objects. The wide range of vehicle sizes underscores the the broad spectrum of real-world vehicles that is represented in our database.

Benchmark

To demonstrate how the ConVeX dataset can be used to assess the robustness of CP algorithms under adverse conditions, we provide a preliminary benchmark using CoBEVT [32], a multi-agent, multi-camera Birds Eye View (BEV) semantic segmentation model. We first split the ConVeX dataset into train/validation/test sets using a 60/20/20 ratio, yielding 15,652/5,218/5,218 frames per set. The split is stratified across scenes to ensure a representative mix of urban, rural, and highway scenarios.

We evaluate two variants of CoBEVT on the test set. The first (CoBEVT-V2X-Sim) is the CoBEVT model trained on

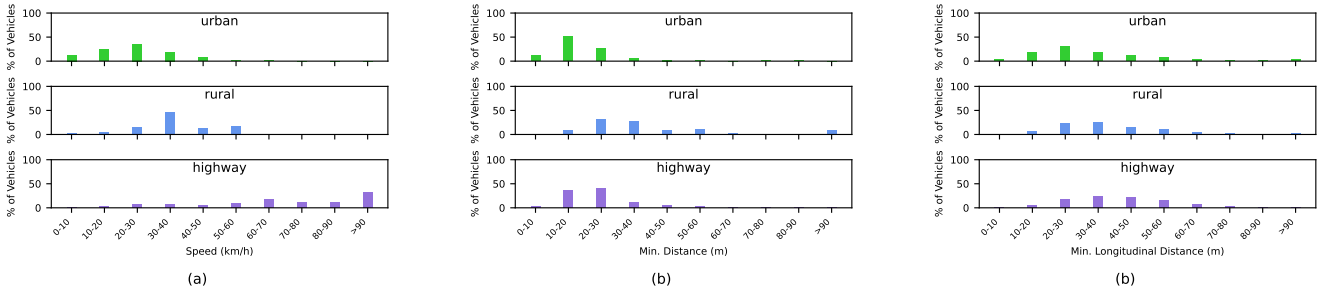

FIGURE 4. Distribution of average vehicle speed (a), minimum inter-vehicle distance (b), and minimum longitudinal distance between each CAV agent and the other vehicles on the same lane (c) for each scenario, computed considering only those vehicles that are within a 70-meter range from at least one agent.

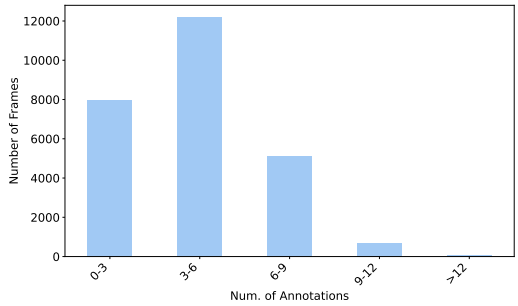

FIGURE 5. Occurrence of the number of annotations per frame.

the V2X-Sim dataset (consisting mostly of clear weather and daytime scenes) as in [33]. The second (CoBEVT-ConVeX) is obtained by fine-tuning the first model on the ConVeX training set for 30 additional epochs with learning rate of 0.00001. All results are reported in terms of AP@50, i.e., Average Precision computed at an Intersection-over-Union (IoU) threshold of 0.5. Table 3 presents the results, broken down by operating conditions. Consistently with expectations, results for both models demonstrate the negative impact of adverse operating conditions on the perception task, with clear weather and daytime scenarios achieving the highest scores, while snowy nighttime conditions emerge as the most challenging. Comparing the two models, CoBEVT-V2X-Sim achieves AP@50 scores ranging from 28 to 44, with a maximum gap of 15.8. Fine-tuning with ConVeX leads to substantial performance improvements across all scenarios, ranging from +13.77% to +32.3%, with a maximum gap of 12.7. These results indicate that fine-tuning on the ConVeX dataset effectively enhances the robustness of CoBEVT under adverse conditions. Notably, the largest gains are observed in rainy and snowy scenarios, which aligns with the fact that the V2X-Sim dataset consists predominantly of clear weather and daytime scenes.

RECORDS AND STORAGE

The collected data is stored in a modified version of the *nuScenes* [34] storage format, adapted to support multiple agents and varying operating conditions per frame. As a result, the dataset can be parsed using the *nuScenes devkit* [35]. Our storage format differs from the original nuScenes format in the following:

•We do not store map data (e.g., images, semantic maps);
•We do not record the visibility level of instances (e.g., the fraction of a detection target appearing in an image);
•We do record the base scene and the operating conditions associated with each scene, by encoding both pieces of information in the `description` field of the scene, using the following syntax: `BASE_SCENE`; `WEATHER`; `TIME_OF_DAY`; `DESCR`, where `BASE_SCENE` is the base scene identifier as reported in Table 2, `WEATHER` and `TIME_OF_DAY` are labels describing the scene's operating conditions, as reported in Fig. 2, and `DESCR` contains any additional description of the scene.

All annotations and metadata – including sensor parameters, maps, ego poses, and operating conditions – are stored in JSON files and organized within a relational database, as detailed in the *nuScenes* documentation [35]. Sensor data, annotations, and metadata are structured in files and folders as outlined in Table 4.

INSIGHTS AND NOTES

We remark that, by providing both the raw data and the source code, we enable other researchers to expand the ConVeX dataset with additional annotations, facilitating the support for further downstream tasks such as semantic segmentation.

SOURCE CODE AND SCRIPTS

The source code used to generate the dataset includes a collection of Python scripts for the sampling, processing, and annotation of raw data and for the generation of the final samples in the modified nuScene format described in Section . Scripts take as input the raw data we generated for each scene. Such data is organized in a folder hierarchy, with folders labelled by the base scene identifier at the top level (level 0), the weather conditions at level 1, and the time of day at level 2. Each scene in level-2 folder includes two subfolders: `data` and `videos`. The former contains GNSS data (`gpsData.csv`), LiDAR data (`lidarData.csv`), and ego pose data (`vehData.csv`) for every agent in the scene, all in CSV format. The latter contains the video feeds (in AVI format) for each agent.

Both the scripts and the raw data used to generate the ConVeX dataset are publicly available at [36], which also

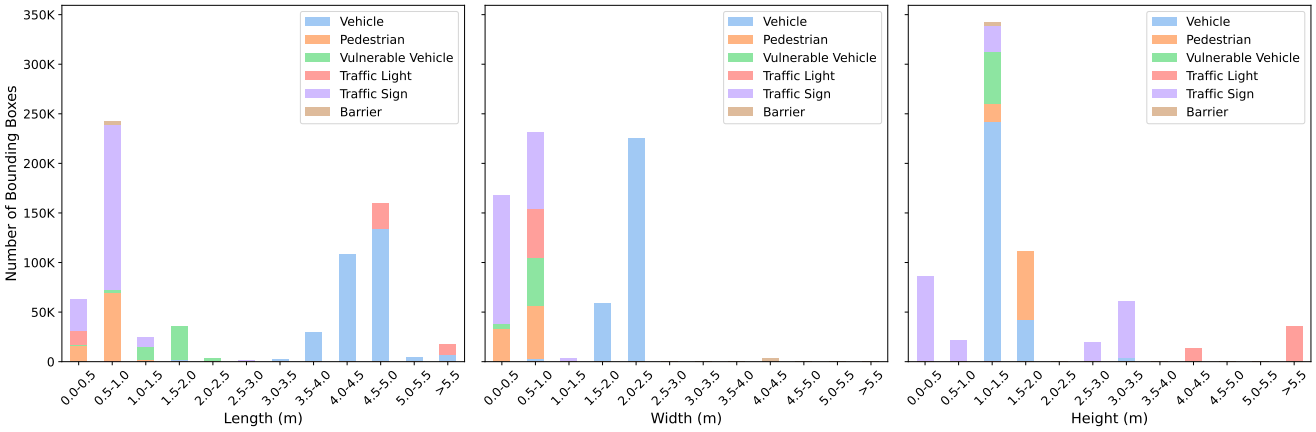

FIGURE 6. **Distribution of the length, width, and height of bounding box annotations across different classes of dynamic and static objects.**

TABLE 3. **Robustness assessment of the CoBEVT CP algorithm under adverse conditions, using the ConVeX dataset. CoBEVT-V2X-Sim: CoBEVT model trained on the V2X-Sim dataset. CoBEVT-ConVeX: obtained by fine-tuning CoBEVT-V2X-Sim on ConVeX. All results are reported in terms of AP@50, with percentage increase relative to CoBEVT-V2X-Sim between parentheses.**

	CoBEVT-V2X-Sim				CoBEVT-ConVeX			
Time	**Clear**	**Cloudy**	**Rainy**	**Snowy**	**Clear**	**Cloudy**	**Rainy**	**Snowy**
Twilight	40.32	39.61	30.08	32.24	46.38 (+15.0%)	46.55 (+17.5%)	37.85 (+25.8%)	39.27 (+21.8%)
Daytime	44.00	41.08	34.18	34.46	50.03 (+13.7%)	47.36 (+15.3%)	41.72 (+22.1%)	42.58 (+23.6%)
Nighttime	38.15	35.37	31.02	28.21	44.83 (+17.5%)	42.38 (+19.8%)	40.06 (+29.2%)	37.32 (+32.3%)

TABLE 4. **Folder structure of the ConVeX dataset.**

File or folder	**Description**
`sampled/`	Folder for sensor data at key-frames.
`sampled/frames`	Folder for sampled video frames (PNG files).
`sampled/gnss`	Folder for sampled GNSS data (NPY files).
`sampled/points`	Folder for sampled point clouds data (PCD files).
`v1.0-ConVeX/`	Folder for dataset annotations and metadata (JSON files).
`v1.0-ConVeX/scene.json`	Sequences of consecutive frames, from X to Y seconds long, extracted from a log.
`v1.0-ConVeX/sample.json`	Annotated snapshots of a scene at a particular timestamp (frames).
`v1.0-ConVeX/sample_data.json`	Data collected from a particular sensor at a given frame.
`v1.0-ConVeX/sample_annotation.json`	Bounding boxes defining the position of an object seen in a sample.
`v1.0-ConVeX/instance.json`	Instances of detection target classes observed in samples.
`v1.0-ConVeX/category.json`	Taxonomy of detection target classes.
`v1.0-ConVeX/attribute.json`	Properties of an instance that can change while the category remains the same.
`v1.0-ConVeX/sensor.json`	Sensor types.
`v1.0-ConVeX/calibrated_sensor.json`	Definition of particular sensors (camera/LiDAR) as calibrated on a particular vehicle.
`v1.0-ConVeX/ego_pose.json`	Ego vehicle poses at a particular timestamp.
`v1.0-ConVeX/log.json`	Logs/recordings from which the data was extracted.

provides additional documentation on how to navigate the data and run the scripts.

ACKNOWLEDGEMENTS AND INTERESTS

This work was supported by CARS@Polito, Torino, Italy, the Evolution of Electric Vehicles with V2I, Energy Management, and Smart Mobility (EVOLVE) project (Bando SWIch - Programma Regionale Piemonte F.E.S.R. 2021/2027), and PNRR PNC -A.1-N-1 – Progetto Living Lab To Move Project – MAAS4ITALY (Grant No. C15C22007220001). This work was also supported by "WEBFARE" (Nr. FISA2022-00908), funded by the Ministero dell'Universitá e della Ricerca - with the FISA 2022 (D.D. n. 1405 DEL 13/9/2022 Fondo Italiano per le Scienze Applicate (FISA)) program. This manuscript reflects only the authors' views and opinions, and the Ministry cannot be considered responsible for them. M. Palena contributed to the scene and dataset design, data annotation, and paper writing. D. C. Selvaraj designed the experiments to obtain the dataset. C. F. Chiasserini and T. Cerquitelli supervised the work and contributed to the paper writing.

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 146–177, Feb. 2023.

[2] J. Wang, Z. Wu, Y. Liang, J. Tang, and H. Chen, "Perception Methods for Adverse Weather Based on Vehicle Infrastructure Cooperation System: A Review," *Sensors*, vol. 24, no. 2, p. 374, Jan. 2024.

[3] A. Carballo, J. Lambert, A. Monrroy, D. Wong, P. Narksri, Y. Kitsukawa, E. Takeuchi, S. Kato, and K. Takeda, "LIBRE: The Multiple 3D LiDAR Dataset," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 1094–1101.

[4] M. J. Mirza, C. Buerkle, J. Jarquin, M. Opitz, F. Oboril, K.-U. Scholl, and H. Bischof, "Robustness of object detectors in degrading weather conditions," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 2719–2724.

[5] C. Sakaridis, D. Dai, and L. Van Gool, "ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021.

[6] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017. [Online]. Available: http://dx.doi.org/10.1177/0278364916679498

[7] D. Liu, Y. Cui, Z. Cao, and Y. Chen, "A Large-scale Simulation Dataset: Boost the Detection Accuracy for Special Weather Conditions," in *2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–8.

[8] L. H. Pham, D. N.-N. Tran, and J. W. Jeon, "Low-Light Image Enhancement for Autonomous Driving Systems using DriveRetinex-Net," in *2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*. Seoul, Korea (South): IEEE, Nov. 2020, pp. 1–5.

[9] S. Sural, N. Sahu, and R. R. Rajkumar, "ContextualFusion: Context-Based Multi-Sensor Fusion for 3D Object Detection in Adverse Operating Conditions," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. Jeju Island, Korea, Republic of: IEEE, Jun. 2024, pp. 1534–1541.

[10] 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*, 2017, pp. 1–16.

[11] M. Karvat and S. Givigi, "Adver-City: Open-Source Multi-Modal Dataset for Collaborative Perception Under Adverse Weather Conditions," Mar. 2025.

[12] A. Simulations, "Scaner studio," 2019.

[13] B. Yang, W. Luo, and R. Urtasun, "PIXOR: Real-time 3D Object Detection from Point Clouds," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 7652–7660.

[14] S. Manivasagam, S. Wang, K. Wong, W. Zeng, M. Sazanovich, S. Tan, B. Yang, W. Ma, and R. Urtasun, "Lidarsim: Realistic lidar simulation by leveraging the real world," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 2020, pp. 11 164–11 173.

[15] 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," Jul. 2022.

[16] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018. [Online]. Available: https://elib.dlr.de/124092/

[17] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, "OPV2V: An Open Benchmark Dataset and Fusion Pipeline for Perception with Vehicle-to-Vehicle Communication," in *2022 International Conference on Robotics and Automation (ICRA)*, May 2022, pp. 2583–2589.

[18] R. Xu, Y. Guo, X. Han, X. Xia, H. Xiang, and J. Ma, "Opencda: an open cooperative driving automation framework integrated with co-simulation," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 1155–1162.

[19] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan, and Z. Nie, "DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 21 329–21 338.

[20] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2X-ViT: Vehicle-to-Everything Cooperative Perception with Vision Transformer," in *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*. Berlin, Heidelberg: Springer-Verlag, Oct. 2022, pp. 107–124.

[21] R. Mao, J. Guo, Y. Jia, Y. Sun, S. Zhou, and Z. Niu, "DOLPHINS: Dataset for Collaborative Perception Enabled Harmonious and Interconnected Self-driving," in *Computer Vision – ACCV 2022*, L. Wang, J. Gall, T.-J. Chin, I. Sato, and R. Chellappa, Eds. Cham: Springer Nature Switzerland, 2023, pp. 495–511.

[22] H. Yu, W. Yang, H. Ruan, Z. Yang, Y. Tang, X. Gao, X. Hao, Y. Shi, Y. Pan, N. Sun, J. Song, J. Yuan, P. Luo, and Z. Nie, "V2X-Seq: A Large-Scale Sequential Dataset for Vehicle-Infrastructure Cooperative Perception and Forecasting," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 5486–5495.

[23] R. Xu, X. Xia, J. Li, H. Li, S. Zhang, Z. Tu, Z. Meng, H. Xiang, X. Dong, R. Song, H. Yu, B. Zhou, and J. Ma, "V2V4Real: A Real-World Large-Scale Dataset for Vehicle-to-Vehicle Cooperative Perception," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 13 712–13 722.

[24] A. Kulkarni, J. Chrosniak, E. Ducote, F. Sauerbeck, A. Saba, U. Chirimar, J. Link, M. Behl, and M. Cellina, "RACECAR - The Dataset for High-Speed Autonomous Racing," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2023, pp. 11 458–11 463.

[25] Y. Wang, S. Chen, Z. Song, and S. Zhou, "WHALES: A Multi-Agent Scheduling Dataset for Enhanced Cooperation in Autonomous Driving," Aug. 2025.

[26] W. Zimmer, G. A. Wardana, S. Sritharan, X. Zhou, R. Song, and A. C. Knoll, "TUMTraf V2X Cooperative Perception Dataset," Mar. 2024.

[27] R. Hao, S. Fan, Y. Dai, Z. Zhang, C. Li, Y. Wang, H. Yu, W. Yang, J. Yuan, and Z. Nie, "RCooper: A Real-world Large-scale Dataset for Roadside Cooperative Perception," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, Jun. 2024, pp. 22 347–22 357.

[28] H. Xiang, Z. Zheng, X. Xia, R. Xu, L. Gao, Z. Zhou, X. Han, X. Ji, M. Li, Z. Meng, L. Jin, M. Lei, Z. Ma, Z. He, H. Ma, Y. Yuan, Y. Zhao, and J. Ma, "V2X-Real: A Largs-Scale Dataset for Vehicle-to-Everything Cooperative Perception," in *Computer Vision – ECCV 2024*, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, Eds. Cham: Springer Nature Switzerland, 2025, pp. 455–470.

[29] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *European Conference on Computer Vision (ECCV)*, ser. LNCS, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., vol. 9906. Springer International Publishing, 2016, pp. 102–118.

[30] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta *et al.*, "Lgsvl simulator: A high fidelity simulator for autonomous driving," *arXiv preprint arXiv:2005.03778*, 2020.

[31] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "Airsim: High-fidelity visual and physical simulation for autonomous vehicles," in *Field and Service Robotics*, M. Hutter and R. Siegwart, Eds. Cham: Springer International Publishing, 2018, pp. 621–635.

[32] H. X. W. S. B. Z. J. M. Runsheng Xu, Zhengzhong Tu, "Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers," in *Conference on Robot Learning (CoRL)*, 2022.

[33] Y. Lu, Y. Hu, Y. Zhong, D. Wang, S. Chen, and Y. Wang, "An extensible framework for open heterogeneous collaborative perception," in *The Twelfth International Conference on Learning Representations*, 2024.

[34] 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," *arXiv preprint arXiv:1903.11027*, 2019.

[35] Motional Inc. nuscenes™ devkit. [Online]. Available: https://github.com/nutonomy/nuscenes-devkit/

[36] M. Palena, D. C. Selvaraj, C. F. Chiasserini, and T. Cerquitelli. Convex github repository. [Online]. Available: https://github.com/marcopalena/convex/