Supplementary Material:
CARLANE: A Lane Detection Benchmark for Unsupervised Domain Adaptation from Simulation to multiple Real-World Domains

Julian Gebele[*,1] **Bonifaz Stuhr**[*,1,2] **Johann Haselberger**[*,1,3]
[1]University of Applied Science Kempten
[2]Autonomous University of Barcelona
[3]Technische Universität Berlin
`carlane.benchmark@gmail.com`

A Appendix

A.1 Example Usage of the CARLANE Benchmark

A Jupyter Notebook with a tutorial to read the datasets for usage in PyTorch can be found at https://carlanebenchmark.github.io.

A.2 Model Vehicle Description

In Figure 1, the self-built 1/8th model vehicle is shown, which we used to gather the images for the 1/8th scaled target domain. A NVIDIA Jetson AGX is the central computation unit powered by a separate Litionite Tanker Mini 25000mAh battery. For image collection, we utilize the software framework ROS Melodic and a Stereolabs ZEDM stereo camera with an integrated IMU. The camera is directly connected to the AGX and captures images with a resolution of 1280×720 pixels and a rate of 30 FPS.
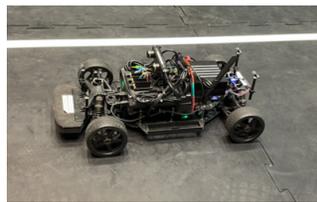

Figure 1: Picture of the 1/8th model vehicle we built to capture images in our 1/8th target domain.

A.3 Reproducibility of the Baselines

To ensure reproducibility, we strictly follow UFLD [1] and the corresponding UDA method for model architecture and hyperparameters. Thereby, we utilize UFLD as an encoder for the UDA method. We provide a detailed table of the tuned hyperparameters, architecture changes, and objectives in the main text. In addition, the trained weights of our baselines, their entire implementation, and the configuration files of our baselines are made publicly available at https://carlanebenchmark.github.io.
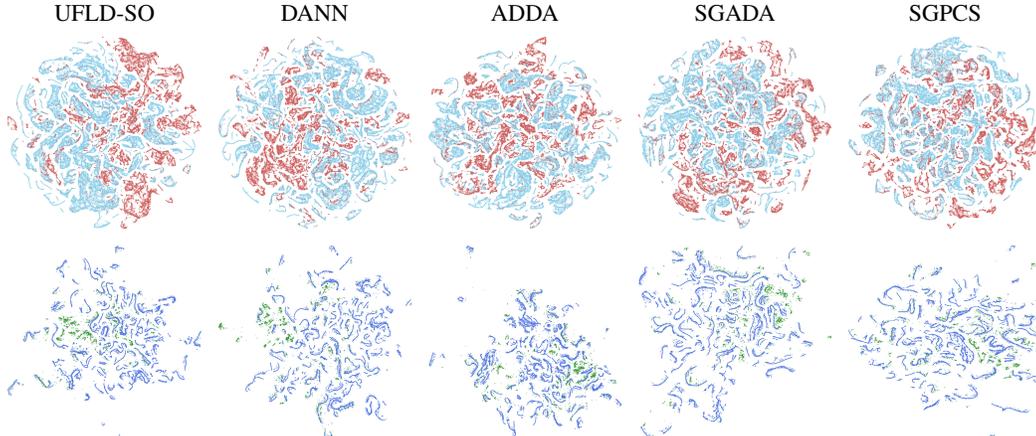
---

[*]Equal contribution

Figure 2: t-SNE visualizations of the MoLane dataset (top) and the TuLane dataset (bottom). The source domain is marked in blue, the real-world model vehicle target domain in red, and TuLane's target domain in green.

Initialization. We initialize convolutional layer weights with kaiming normal and their biases with 0.0. Linear layer weights are initialized with normal (mean = 0.0, std = 0.01), batch normalization weights and biases are initialized with 1.0.

A.4 Additional Results

t-SNE feature clustering. Figure 2 shows the t-SNE feature clustering of the trained baselines for the MoLane and TuLane dataset, respectively. We observe that few features of both domains spread over the entire plot for higher-performing UDA methods. However, there are still large clusters of features from one domain, indicating that the domain adaptation only occurred slightly.

Qualitative results. We randomly sample results from our baselines and show them in Figures 4, 5, and 6. Compared to UFLD-SO, the UDA baselines ADDA, SGADA, and SGPCS increase performance consistently. UFLD-TO samples show the best results on the target domain.

A.5 Comparison to Related Work

In Table 1, we compare CARLANE with the datasets created by related work. The main differentiators are that our dataset contains three distinct domains, including a scaled model vehicle, and is publicly available. To further compare our synthetic datasets with related work, the applied variations during the data collection process are summarized in Table 2. Additionally, we highlight noticeable differences in the visual quality of the simulation engines in Figure 3. Scenes captured in Carla are more realistic and detailed.

Table 1: Comparison of CARLANE (ours) with datasets created by related work.

Dataset	Year	Publicly Available	Domains	Simulation	Resolution	Total Images	Annotations
[2]	2019	✗	sim, real	blender	480×360	391K	3D
[3]	2020	✗	sim, real	blender	480×360	586K	3D
[4]	2022	✗	sim, real	Carla	1280×720	23K	2D
ours	2022	✓	sim, real, scaled	Carla	1280×720	163K	2D

A.6 Author Statement

In accordance with the Apache License Version 2.0, January 2004, the authors bear all responsibility in case of violation of rights. The descriptions made in the paper and its supplementary material are accurate and agreed upon by all authors.

Table 2: Comparison of applied variations for the collection of the synthetic datasets.

Dataset	Ego Vehicle	Camera Position	Lane Deviation	Traffic	Pedestrians	World Objects	Daytime	Weather	City	Rural	Highway	Terrain	Lane Topology	Road Appearance
[2]	✗	✓	✓	✓	✗	✓	✓	✗	✗	✓	✗	✓	✓	✓
[3]	✗	✓	✓	✓	✗	✓	✓	✗	✗	✓	✗	✓	✓	✓
[4]	✗	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
ours	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 3: Visual comparison of simulation images from the custom blender simulation used in [2, 3] and the Carla simulation used by [4] and our work. We observe that scenes captured in Carla are more detailed and realistic.

References

[1] Z. Qin, H. Wang, and X. Li, "Ultra Fast Structure-aware Deep Lane Detection," in *The European Conference on Computer Vision (ECCV)*, 2020.

[2] N. Garnett, R. Cohen, T. Pe'er, R. Lahav, and D. Levi, "3D-LaneNet: End-to-End 3D Multiple Lane Detection," in *ICCV*, pp. 1013 – 1021, 2019.

[3] N. Garnett, R. Uziel, N. Efrat, and D. Levi, "Synthetic-to-Real Domain Adaptation for Lane Detection," in *ACCV*, 2020.

[4] C. Hu, S. Hudson, M. Ethier, M. Al-Sharman, D. Rayside, and W. Melek, "Sim-to-Real Domain Adaptation for Lane Detection and Classification in Autonomous Driving," 2022.
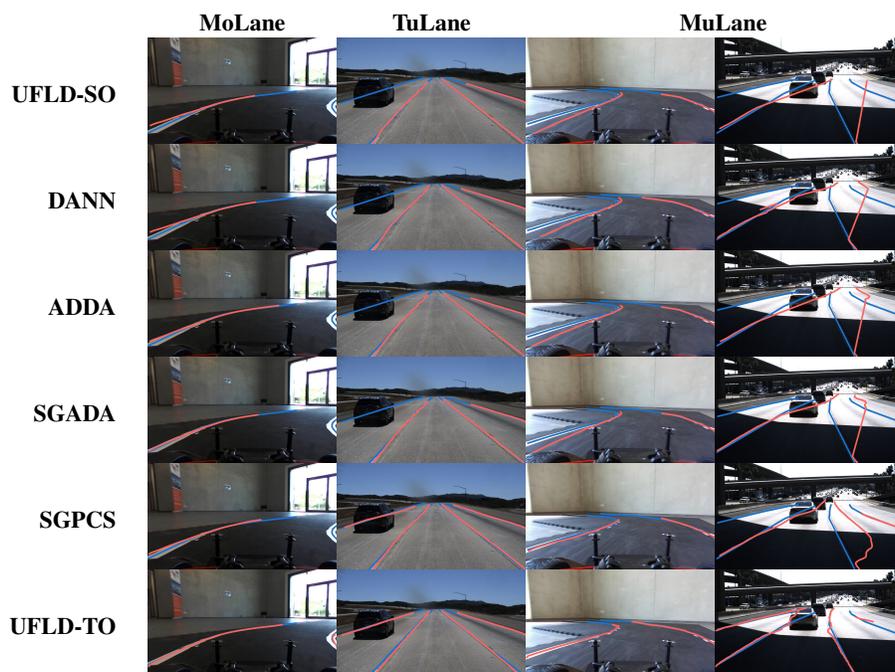
Figure 4: Qualitative results of target domain predictions. Images are randomly sampled. Ground truth lane annotations are marked in blue, predictions in red.
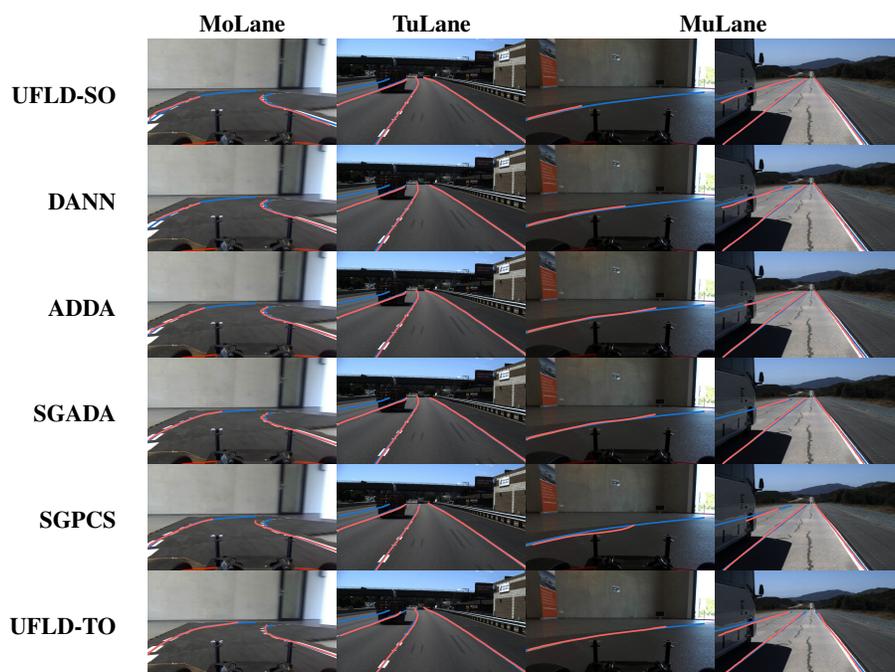

Figure 5: Qualitative results of target domain predictions. Images are randomly sampled. Ground truth lane annotations are marked in blue, predictions in red.
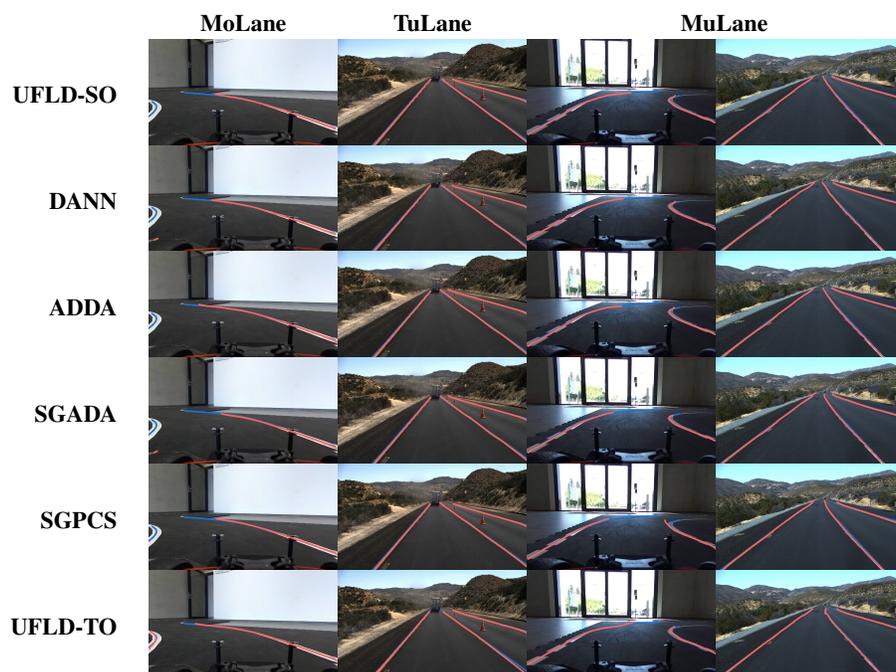
Figure 6: Qualitative results of target domain predictions. Images are randomly sampled. Ground truth lane annotations are marked in blue, predictions in red.

B Datasheet for the CARLANE Benchmark

Motivation

For what purpose was the dataset created? *Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

CARLANE was created to be the first publicly available single- and multi-target Unsupervised Domain Adaptation (UDA) benchmark for lane detection to facilitate future research in these directions. However, in a broader sense, the datasets of CARLANE were also created for unsupervised and semi-supervised learning and partially for supervised learning. Furthermore, a real-to-real transfer can be performed between the target domains of our datasets.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

As released on June 17, 2022, the initial version of CARLANE was created by Julian Gebele, Bonifaz Stuhr, and Johann Haselberger from the Institute for Driver Assistance Systems and Connected Mobility (IFM). The IFM is a part of the University of Applied Sciences Kempten. Furthermore, CARLANE was created by Bonifaz Stuhr as part of his Ph.D. at the Autonomous University of Barcelona (UAB) and by Johann Haselberger as part of his Ph.D. at the Technische Universität Berlin (TU Berlin).

Who funded the creation of the dataset? *If there is an associated grant, please provide the name of the grantor and the grant name and number.*

There is no specific grant for the creation of the CARLANE Benchmark. The datasets were created as part of the work at the IFM and the University of Applied Sciences Kempten.

Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? *Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

The instances are drives on diverse roads in simulation, in an abstract 1/8th real world, and in full-scale real-world scenarios, along with lane annotations of the up to four nearest lanes to the vehicle.

How many instances are there in total (of each type, if appropriate)?

Table 3 shows the per-domain and per-subset breakdown of CARLANE instances. TuSimple is available at `https://github.com/TuSimple/tusimple-benchmark` under the Apache License Version 2.0, January 2004.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? *If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).*

The datasets of CARLANE contain samples of driving scenarios and lane annotations encountered in simulation and the real world. The datasets are not representative of all these driving scenarios, as the distribution of the latter is highly dynamic and diverse. Instead, the motivation was to resemble the variety and shifts of different domains in which such scenarios occur to strengthen the systematic study of UDA methods for lane detection. Therefore, CARLANE should be considered as an UDA benchmark with step-by-step testing possibility across three domains, which makes it possible to include an additional safety mechanism for real-world testing.

What data does each instance consist of? *"Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.*

Each labeled instance consists of the following components:

(1) A single 1280×720 image from a driving scenario.

(2) A .json file entry for the corresponding subset containing lane annotations following TuSimple. The lanes' y-values discretized by 56 raw anchors, the lanes' x-values to 101 gridding cells, with the last gridding cell representing the absence of a lane. The file path to the corresponding image is also stored in the .json file.

(3) A .png file containing lane segmentations following UFLD (`https://github.com/cfzd/Ultra-Fast-Lane-Detection`), where each of the four lanes has a different label.

(4) A .txt file entry for the corresponding subset containing the linkage between the raw image and its segmentation as well as the presence and absence of a lane.

Table 3: Dataset overview. Unlabeled images denoted by *, partially labeled images denoted by **.

Dataset	domain	total images	train	validation	test	lanes
MoLane	CARLA simulation	84,000	80,000	4,000	-	≤ 2
	model vehicle	46,843	43,843*	2,000	1,000	≤ 2
TuLane	CARLA simulation	26,400	24,000	2,400	-	≤ 4
	TuSimple	6,408	3,268	358	2,782	≤ 4
MuLane	CARLA simulation	52,800	48,000	4,800	-	≤ 4
	model vehicle + TuSimple	12,536	6,536**	4,000	2,000	≤ 4

Each unlabeled instance consists of an 1280×720 image from a driving scenario and a .txt file entry for the corresponding subset.

Is there a label or target associated with each instance? If so, please provide a description.

As described above, the labels per instance are discretized lane annotations and lane segmentations.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Everything is included. No data is missing.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

There are no relationships made explicit between instances. However, some instances are part of the same drive and therefore have an implicit relationship.

Are there recommended data splits(e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

Each domain is split into training and validation subsets. Details are shown in Table 3. The target domains for UDA additionally include test sets, which were recorded from separate tracks (model vehicle) or driving scenarios (TuSimple). Since UDA aims to adapt models to target domains, only the target domains include a test set.

Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

CARLANE was recorded from different drives through simulation and real-world domains. Therefore there are images captured from the same drive, which result in similar scenarios for consecutive images. Target domain samples were annotated by hand and may include human labeling errors. However, we double-checked labels and cleaned TuSimple's test set with our labeling tool.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

CARLANE is entirely self-contained.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.

The full-scale real-world target domain contains open-source images with unblurred license plates and people from the TuSimple dataset. This data should be treated with respect and in accordance with privacy policies. The other domains do not contain data that might be considered confidential since there where recorded in simulations or a controlled 1/8th real-world environment.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

CARLANE includes driving scenarios; therefore, its datasets could cause anxiety in people with driving anxiety.

Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how

these subpopulations are identified and provide a description of their respective distributions within the dataset.

No.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.

Yes, individuals could be identified in the full-scale real-world target domain from TuSimple, since it contains unblurred license plates and people. However, the remaining domains do not contain identifiable individuals.

Does the dataset contain data that might be considered sensitive in anyway(e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.

The full-scale real-world target domain from TuSimple could implicitly reveal sensitive information printed or put on the vehicles or people's wearings.

Collection Process

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

The source domain images of driving scenarios and the corresponding lane annotations were directly recorded from the simulation. Lanes were manually labeled for the directly recorded real-world images. For the images collected from the model vehicle, the authors annotated the data with a labeling tool created for this task. The labeling tool is publicly available at `https://carlanebenchmark.github.io`. The labeling tool is utilized to clean up the annotations of the test set in the real-world domain. The authors do not have information about the labeling process of the full-scale target domain since its data is derived from the TuSimple dataset.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?

The source domain data was collected using the CARLA simulator and its APIs with a resolution of 1280×720 pixels. The real-world 1/8th target domain was collected with a Stereolabs ZEDM camera with 30 FPS and a resolution of 1280×720 pixels. The lane distributions were additionally balanced with a bagging approach, and lanes were annotated with a labeling tool. More information can be found in the corresponding paper and the implementation. The implementation and all used tools are publicly available at `https://carlanebenchmark.github.io`.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

Source domain dataset entries are sampled based on the relative angle β of the agent to the center lane. For MoLane, five lane classes are defined for the bagging approach: strong left curve ($\beta \leq -45°$), soft left curve ($-45° < \beta \leq -15°$), straight ($-15° < \beta < 15°$), soft right curve ($15° \leq \beta < 45°$) and strong right curve ($45° \leq \beta$).

For TuLane, three lane classes are defined for the bagging approach: left curve ($-12° < \beta \leq 5°$), straight ($-5° < \beta < 5°$) and right curve ($5° \leq \beta < 12°$).

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

Only the authors were involved in the collection process. The authors do not have information about the people involved in collecting the TuSimple dataset.

Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

MoLane's data was collected and annotated from June 2021 to August 2021. Data for TuLane's source domain was collected in February 2022.

Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

No ethical reviews have been conducted to date. However, an ethical review may be conducted as part of the paper review process.

Preprocessing/cleaning/labeling

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.

As described above, lane annotations were labeled or cleaned using a labeling tool and sampled based on the relative angle β of the agent to the center lane.

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.

No.

Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.

Yes, the software is available at `https://carlanebenchmark.github.io`.

Uses

Has the dataset been used for any tasks already? If so, please provide a description.

The datasets were used to create UDA baselines for the corresponding paper presenting the CARLANE Benchmark.

Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

Yes, the baselines presented in the corresponding paper are available at `https://carlanebenchmark.github.io`.

What(other) tasks could the dataset be used for?

In a broader sense, the datasets of CARLANE can also be used for unsupervised and semi-supervised learning and partially for supervised learning.

Is there anything about the composition of the dataset or the way it was collected and pre-processed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?

Yes, TuLane and MuLane contain open-source images with unblurred license plates and people. This data should be treated with respect and in accordance with privacy policies. In general, CARLANE contributes to the research in the field of autonomous driving, in which many unresolved ethical and legal questions are still being discussed. The step-by-step testing possibility across three domains makes it possible for our benchmark to include an additional safety mechanism for real-world testing. This can help the consumer to mitigate the risks and harms to some extent.

Are there tasks for which the dataset should not be used? If so, please provide a description.

Since CARLANE focuses on UDA for lane detection and spans a limited number of driving scenarios, consumers should not solely really on this dataset to train models for fully autonomous driving.

Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

Yes, CARLANE is publicly available on the internet for anyone interested in using it.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?

CARLANE is distributed through kaggle at `https://www.kaggle.com/datasets/carlanebenchmark/carlane-benchmark`

DOI: 10.34740/kaggle/dsv/3798459

When will the dataset be distributed?

The datasets have been available on kaggle since June 17, 2022.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms o fuse (ToU)? If

so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

CARLANE is licensed under the Apache License Version 2.0, January 2004.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

TuSimple, which is used for TuLanes and Mu-Lanes target domains, is licensed under the Apache License Version 2.0, January 2004.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.

Unknown to authors of the datasheet.

<div align="center">

Maintenance

</div>

Who will be supporting/hosting/maintaining the dataset?

CARLANE is hosted on kaggle and supported and maintained by the authors.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

The curators of the datasets can be contacted under carlane.benchmark@gmail.com.

Is there an erratum? If so, please provide a link or other access point.

No.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?

New versions of CARLANE's datasets will be shared and announced on our homepage (`https://carlanebenchmark.github.io`) and at kaggle if corrections are necessary.

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.

Yes, we plan to support versioning of the datasets so that all the versions are available to potential users. We maintain the history of versions via our homepage (`https://carlanebenchmark.github.io`) and at kaggle. Each version will have a unique DOI assigned.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.

Others can extend/augment/build on CARLANE with the support of the open-source tools provided on our homepage. Besides these tools, there will be no mechanism to validate or verify the extended datasets. However, others are free to release their extension of the CARLANE Benchmark or its datasets under the Apache License Version 2.0.