HetroD: A High-Fidelity Drone Dataset and Benchmark for Heterogeneous Traffic in Autonomous Driving

Yu-Hsiang Chen[1,2] Wei-Jer Chang[2] Wei Zhan[2] Masayoshi Tomizuka[2] Yi-Ting Chen[1]

[1]National Yang Ming Chiao Tung University [2]UC Berkeley

Project Page: https://hetrod-data.github.io/HetroD/

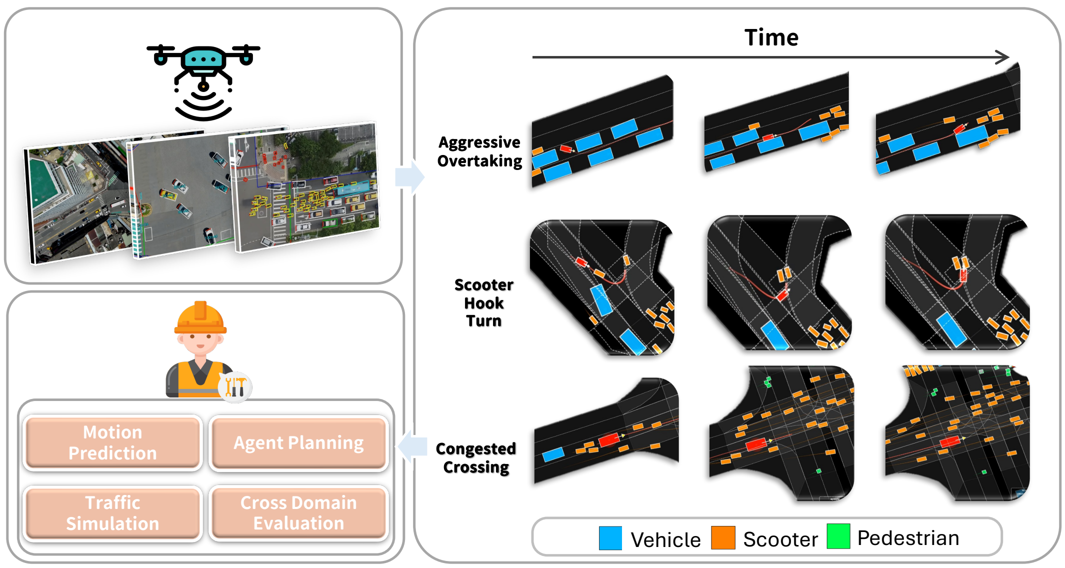

Figure 1. *HetroD* is a high-fidelity, drone-captured dataset that highlights culturally grounded maneuvers such as hook turns, aggressive overtakes, queue cutting, and congested crossings and illustrates dense, fine-grained interactions among cars, scooters, cyclists, and pedestrians. We establish a unified benchmark and systematically evaluate the performance of existing algorithms in agent planning, motion prediction, traffic simulation, and cross-domain evaluation in dense and heterogeneous traffic environments.

Abstract

We present HetroD, a dataset and benchmark for developing autonomous driving systems in dense, heterogeneous environments. Unlike many prior datasets focused on lane-disciplined traffic, HetroD captures culturally grounded behaviors such as hook turns, lane splitting, and informal right-of-way negotiation. It comprises over 65.4 k agent trajectories (cars, scooters, buses, cyclists, and pedestrians) with centimeter-accurate annotations, HD maps, traffic signal states, and a modular toolchain for extracting per-agent scenarios. This work enables modeling behaviors of vulnerable road users (VRUs) in heterogeneous traffic and constructs standardized benchmarks for forecasting, planning, simulation, and multi-agent behavior modeling. Preliminary results show that state-of-the-art models trained on existing datasets struggle to generalize, revealing key limitations in handling heterogeneous traffic.

1. Introduction

Navigating heterogeneous traffic remains one of the core challenges in the development of autonomous driving systems. In many dense urban centers worldwide, cars, scooters, bicycles, and pedestrians compete for limited road space and negotiate right-of-way through subtle and often culturally embedded cues [6]. Yet most publicly available datasets still primarily capture lane-disciplined traffic and model agents as independent [24], thereby underrepresenting culturally grounded, interaction-rich behaviors. As a result, downstream models and widely used simulators inherit these biases: they either hard-code simplified vulnerable

Table 1. **Comparison of Datasets on Interaction, Density & Diversity Metrics.** We report key statistics across vehicle and drone-view datasets. *Interaction Density*[1] measures the number of agent pairs within a scenario whose time-to-collision (TTC) [39] is below a threshold τ. *Interaction Scale*[2] is the total number of such interactions, computed collectively over all datasets. *Heterogeneous Interaction Scale*[3] counts interactions between agents of different types, computed collectively over all datasets. *Geographical Density*[4] represents the average number of agents per unit area A within an 8-second window. *Type Diversity*[5] captures the type-level diversity within a scene using the Gini–Simpson index. *All metric values are normalized to [0,1] across datasets for direct comparison.*

Dataset	**Platform**	**Tracks**	**Duration**	**Interaction Density**	**Interaction Scale**	**Heterogeneous Inter. Scale**	**Geo. Density**	**Type Diversity**
NuScenes [4]	On-board	$\sim$90k†	320h	—	0.642	0.286	—	—
Waymo [36]	On-board	7.6M	574h	—	**1.000**	0.423	—	—
Argoverse2 [41]	On-board	13.9M	763h	—	0.567	0.151	—	—
NuPlan [5]	On-board	$\sim$5M†	1282h	—	0.891	0.202	—	—
INTERACTION [50]	Drone	40k	16.5h	0.008	0.183	—	0.011	—
inD [2]	Drone	13.5k	10h	0.010	0.122	0.164	0.023	0.584
SinD [45]	Drone	13.2k	7.02h	0.012	0.175	0.344	0.016	**0.742**
HetroD	Drone	**65.4k**	**17.5h**	**0.029**	<u>0.718</u>	**1.000**	**0.026**	<u>0.642</u>

† Estimated values based on official statistics.
— Metric not available.
[1] $\mathcal{D}_{\text{inter}} = \sum_{i,j} \mathbb{1}_{\text{TTC}_{i,j} < 2\,\text{s}}$.
[2] $\mathcal{S}_{\text{inter}} = \sum_{\text{scenarios}} \mathcal{D}_{\text{inter}}$.
[3] $\mathcal{S}_{\text{het}} = \sum_{\text{scenarios}} \sum_{i,j} \mathbb{1}_{(\text{TTC}_{i,j} < 2\,\text{s} \wedge \text{type}_i \neq \text{type}_j)}$.
[4] $\mathcal{D}_{\text{geo}} = N/A$, where N is the number of agents within an 8s window and A is the corresponding area.
[5] $\mathcal{H}_{\text{type}} = 1 - \sum_c p_c^2$, where p_c is the proportion of agents of type c.

road user (VRU) templates [25] or merely replay recorded trajectories [5, 15], which limits their ability to capture reactive dynamics. These limitations are further analyzed in Section 2.

This gap between current datasets and real-world scenes calls for data that captures the informal, high-density interactions typical of mixed-agent traffic environments. To bridge this gap, we introduce *HetroD*, a drone-captured dataset collected across six topologically diverse, high-traffic urban locations in Taiwan. Compared to prior datasets, HetroD offers: **High interaction density**: records up to three-fold higher cross-agent interaction counts than any drone dataset; **Motion diversity**: culturally grounded behaviors such as hook turns, lane splitting, and aggressive overtakes; and **Topological breadth**: six intersection archetypes with centimeter-accurate HD maps, bounding boxes, and signal states. Together, these traits close the data gap and position HetroD as a realistic testbed for developing autonomous driving systems in dense, culturally heterogeneous traffic.

Our contributions are: (i) a 17.5-hour drone dataset featuring over 65.4 k agent tracks and centimeter-level annotations of heterogeneous traffic; (ii) a benchmark suite comprising heterogeneous scenarios, task baselines, and a plug-and-play conversion toolchain; and(iii) evidence that HetroD reveals common failure modes of current state-of-the-art methods in heterogeneous-traffic environments.

2. Related Work

Autonomous driving datasets vary by sensing modality and deployment context. We group related work into four categories: on-board, infrastructure-view, drone-view datasets, and unified learning frameworks built upon them.

On-Board Sensor Datasets. [4, 9, 13, 14, 16, 31, 36, 41, 43, 47, 51] offer rich multimodal data but suffer from occlusions and limited VRU coverage in dense traffic [4, 13, 36]. While METEOR [7] pioneered heterogeneous traffic capture, it lacks HD maps and comprehensive annotations, resulting in car-centric data that underrepresents reactive interactions.

Infrastructure-View Datasets. [21, 30, 42, 44, 46, 48, 49, 53, 54] use fixed cameras or V2X sensors to reduce occlusion, but often lack resolution [30, 49], calibration [42, 53], or class diversity [21, 48], limiting their utility for modeling cross-type agent behaviors.

Drone-View Datasets. [2, 3, 11, 19, 20, 26–28, 32, 35, 37, 45, 50, 52] provide occlusion-free, global views ideal for interaction analysis. However, many are collected in lane-disciplined settings [19, 35]; lack standardized scenario formats; and exhibit fragmented VRU tracks due to small-object tracking limitations [1, 11]. They also underrepresent fine-grained behaviors such as informal yielding, weaving, or reverse flows. To the best of our knowledge, no existing public drone dataset provides both per-agent, centimeter-accurate ground truth and wide-area coverage across diverse, heterogeneous urban environments, limiting their applicability to safety-critical or VRU-aware tasks.

Unified Learning Frameworks. [8, 12, 17, 18, 22,

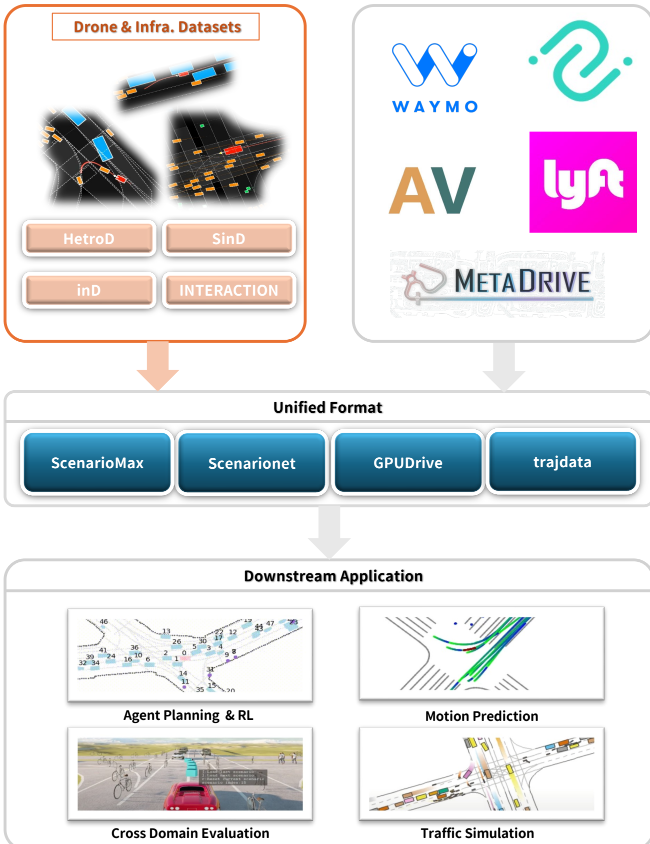

Figure 2. We develop a standardized toolkit that converts a wide range of traffic scene datasets into standardized, agent-centric data formats [8, 17, 18, 23], enabling seamless comparisons across datasets for forecasting, planning, and simulation.

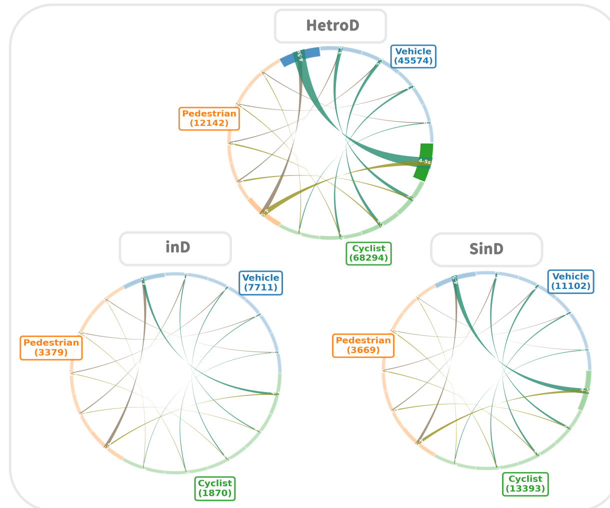

Figure 3. **Cross-Type Interaction Chord Diagrams.** With equal-length sampled scenarios, each chord shows the number of interactions between agent types, grouped by time-to-collision (TTC) [39] ranges (0–1, 1–2, 2–3, 3–4, 4–5 s). HetroD exhibits denser, riskier cross-type interactions, especially among vehicles, cyclists, and pedestrians.

23, 40] standardize interfaces for learning and simulation, rely on high-quality upstream data (especially fine-grained, interaction-aware trajectories and agent-centric scenarios). However, the toolchains required to produce such structured annotations remain underdeveloped in existing drone datasets, limiting their utility for downstream tasks.
In contrast, HetroD is designed to close these gaps by offering dense, heterogeneous scenarios with high-fidelity annotations, cultural motion diversity, and plug-and-play compatibility (see Figure 2).

3. The HetroD Dataset

HetroD is a large-scale drone-view dataset comprising 17.5 hours of ultra-high-resolution (5.4K) video, collected across six topologically and behaviorally distinct urban sites in Taiwan. It includes over 65.4 k unique trajectories spanning signalized intersections, unsignalized merging zones, and densely mixed corridors, traffic archetypes rarely represented together in existing datasets.

To meaningfully quantify traffic complexity in dense, heterogeneous settings, we introduce five normalized principled metrics that capture spatial, behavioral, and interaction-level diversity (Table 1). These metrics (ranging from *interaction density* to *type diversity*) are computed from 1000 uniformly sampled 8-second scenarios per dataset (per-scenario metrics), or over full dataset coverage (scale metrics). Together, they support robust comparison across platforms and traffic domains.

Key insights. *(1) Dataset scale and interaction complexity:* HetroD contains the largest number of unique agent tracks among existing drone-view datasets, and exhibits the highest *interaction density* and *heterogeneous interaction scale* (Table 1). *(2) Cultural and behavioral richness:* While SinD [45] offers a balanced distribution of agent types, HetroD presents a unique setting where *scooters are as prevalent as cars*, reflecting traffic patterns not captured in prior datasets. These agents demonstrate culturally grounded behaviors such as weaving, filtering, and informal negotiation, rarely modeled at scale. Risk indicators like TTC [39] and DRAC [33] reveal significantly higher latent conflict rates (see Figure 3). *(3) Modeling utility:* Motion prediction experiments show large distributional shifts triggered by rare maneuvers (e.g., informal U-turns, reverse flows), highlighting the need for datasets that capture such diversity.

Leveraging this diversity, HetroD fills a long-standing gap in heterogeneous-traffic modeling and unlocks two pivotal research axes: *(i) High-fidelity heterogeneous-traffic simulation* (from full-scene replay to reactive VRU modeling); *(ii) VRU motion prediction and cross-domain gener-*

alization, enabling out-of-distribution testing on rare, culturally grounded maneuvers.

4. Evaluation

We construct a set of challenging per-agent scenarios from HetroD. Specifically, we sample agents exhibiting non-trivial behavior such as long traversals, abrupt heading changes, and dense interactions within multi-agent contexts. These selected agents are used to instantiate per-agent scenarios for evaluation.

4.1. Motion Forecasting

We evaluate cross-dataset generalization of two state-of-the-art predictors (*MTR* [34] and *Wayformer* [29]) on HetroD. Following the UniTraj [12] protocol, models trained on Waymo, Argoverse 2, and NuScenes are directly evaluated on HetroD using the Brier FDE [41] metric. As shown in Table 2, both models exhibit significant performance drops; MTR performs better likely due to its anchor-based decoding, which offers greater robustness in uncertain, cluttered scenes. In contrast, Wayformer heavily over-predicts, particularly for agile agents such as cyclists, indicating transformer-based methods are sensitive to dense, visually cluttered scenes.

These results demonstrate a fundamental limitation of existing forecasting models when confronted with heterogeneous, culturally diverse traffic, emphasizing the need for methods explicitly designed to capture dense agent interactions.

Table 2. **Cross-dataset evaluation (Brier FDE ↓)** of **MTR** [34] and **Wayformer** [29], trained on large-scale datasets and evaluated on **HetroD**. Bold indicates worst-case generalization.

	MTR-NuScenes	**MTR-Waymo***	**MTR-AV2**
NuScenes	2.82	3.16	4.17
SinD	5.87	4.61	4.28
HetroD	**10.98**	**9.05**	**4.92**
	Wayformer-NuScenes	**Wayformer-Waymo***	**Wayformer-AV2**
Argoverse2	4.02	2.74	2.41
SinD	5.00	3.75	3.60
HetroD	**12.69**	**16.04**	**12.15**

Waymo* uses only 30% of original training data due to resource constraints.

4.2. Agent Planning

We evaluate planning performance on HetroD, comparing two rule-based planners: the standard Intelligent Driver Model (IDM) [38] and PDM [10], a top-performing planner from the NuPlan benchmark [5, 15], implemented using the V-Max [8] framework. To better reflect the challenges of VRU interactions, we extend the evaluation by incorporating a VRU-specific collision metric (tracking side and lateral collisions involving cyclists and pedestrians), crucial in scenes with overtakes and unstructured flows where forward-collision checks fall short.

As shown in Table 3, both rule-based planners exhibit performance drops on HetroD compared to NuPlan, including increased VRU collisions and reduced comfort scores, despite high centerline compliance and speed. As detailed in Table 4, collisions in HetroD frequently involve lateral interactions, which traditional planners (e.g., PDM, IDM), optimized for structured, car-centric settings are unable to anticipate. These findings highlight a critical gap: rule-based planners fail to account for lateral VRU interactions required for safe navigation in dense, heterogeneous traffic, emphasizing the need for interaction-aware planning methods. *Further experiments and detailed ablations will appear in the full version of this paper.*

Table 3. **Closed-Loop Non-Reactive Planning Evaluation in Heterogeneous Traffic** Our experiments show that rule-based planners (IDM and PDM) have difficulty handling lateral interactions and avoiding collisions with vulnerable road users in the **HetroD** scenario. To ensure a fair comparison, we disabled the off-road penalty in NuPlan's aggregate score, because in dense, high-flow situations these planners rigidly follow map centerlines without adaptive behavior, making off-road violations far more likely.

Dataset	**Planner**	**NuPlan Score ↑**	**TTC Within Bound ↑**	**Progress Ratio ↑**	**Multiple Lane Score ↑**	**Comfort ↑**	**At-Fault Collision ↓**
NuPlan	IDM	0.81	0.94	0.92	0.98	0.46	0.02
	PDM-Closed	0.82	0.97	0.91	0.99	0.29	0.008
HetroD	IDM	0.75	0.88	0.89	0.94	0.29	0.066
	PDM-Closed	0.71	0.93	0.85	0.95	0.03	0.051

Table 4. **VRU Collision Types in Planning Results.** HetroD exposes planners to a high rate of lateral VRU collisions, reflecting unstructured, high-density scenarios such as lane splitting and parallel overtaking. These cases reveal key blind spots of rule-based planning methods, demonstrating HetroD's value for detailed diagnostic analysis and the development of interaction-aware traffic policies.

Planner	**At-Fault Collisions**	**VRU Front Collisions**	**VRU Lateral Collisions**
IDM	0.066	0.008	0.036
PDM-Closed	0.051	0.005	0.036

5. Conclusions

HetroD enables structured benchmarking in heterogeneous traffic with vulnerable road users (VRUs), addressing key gaps in existing datasets. It pairs high-fidelity annotations with a modular toolchain for agent-centric scenario extraction. Results show that state-of-the-art models struggle to generalize in dense, mixed-agent settings, highlighting the need for interaction-aware learning and simulation in heterogeneous traffic.

Acknowledgements

We sincerely thank leveLXData by fka GmbH for their invaluable support in helping us generate high-quality data. Their expertise in mobility engineering and commitment to innovation were instrumental to the success of our data processing efforts.

References

[1] Joshua Andle, Nicholas Soucy, Simon Socolow, and Salimeh Yasaei Sekeh. The Stanford Drone Dataset is More Complex than We Think: An Analysis of Key Characteristics, 2022.

[2] Julian Bock, Robert Krajewski, Tobias Moers, Steffen Runde, Lennart Vater, and Lutz Eckstein. The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections. In *IV*, pages 1929–1934, 2020.

[3] Antonia Breuer, Jan-Aike Termöhlen, Silviu Homoceanu, and Tim Fingscheidt. openDD: A Large-Scale Roundabout Drone Dataset. In *ITSC*, pages 1–6, 2020.

[4] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A Multimodal Dataset for Autonomous Driving. In *CVPR*, pages 11618–11628, 2020.

[5] Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. NuPlan: A closed-loop ML-based planning benchmark for autonomous vehicles, 2022.

[6] Rohan Chandra, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. TraPHic: Trajectory Prediction in Dense and Heterogeneous Traffic Using Weighted Interactions. In *CVPR*, 2019.

[7] Rohan Chandra, Xijun Wang, Mridul Mahajan, Rahul Kala, Rishitha Palugulla, Chandrababu Naidu, Alok Jain, and Dinesh Manocha. METEOR:A Dense, Heterogeneous, and Unstructured Traffic Dataset With Rare Behaviors, 2022.

[8] Valentin Charraut, Thomas Tournaire, Waël Doulazmi, and Thibault Buhet. V-Max: A Reinforcement Learning Framework for Autonomous Driving, 2025.

[9] Kan Chen, Runzhou Ge, Hang Qiu, Rami Ai-Rfou, Charles R. Qi, Xuanyu Zhou, Zoey Yang, Scott Ettinger, Pei Sun, Zhaoqi Leng, Mustafa Baniodeh, Ivan Bogun, Weiyue Wang, Mingxing Tan, and Dragomir Anguelov. WOMD-LiDAR: Raw Sensor Dataset Benchmark for Motion Forecasting. In *ICRA*, pages 4766–4773, 2024.

[10] Daniel Dauner, Marcel Hallgarten, Andreas Geiger, and Kashyap Chitta. Parting with Misconceptions about Learning-based Vehicle Motion Planning, 2023.

[11] Oussema Dhaouadi, Johannes Meier, Luca Wahl, Jacques Kaiser, Luca Scalerandi, Nick Wandelburg, Zhuolun Zhou, Nijanthan Berinpanathan, Holger Banzhaf, and Daniel Cremers. Highly Accurate and Diverse Traffic Data: The DeepScenario Open 3D Dataset, 2025.

[12] Lan Feng, Mohammadhossein Bahari, Kaouther Messaoud Ben Amor, Éloi Zablocki, Matthieu Cord, and Alexandre Alahi. UniTraj: A Unified Framework for Scalable Vehicle Trajectory Prediction, 2024.

[13] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *CVPR*, pages 3354–3361, 2012.

[14] Luis Gressenbuch, Klemens Esterle, Tobias Kessler, and Matthias Althoff. MONA: The Munich Motion Dataset of Natural Driving. In *ITSC*, pages 2093–2100, 2022.

[15] Cole Gulino, Justin Fu, Wenjie Luo, George Tucker, Eli Bronstein, Yiren Lu, Jean Harb, Xinlei Pan, Yan Wang, Xiangyu Chen, John D. Co-Reyes, Rishabh Agarwal, Rebecca Roelofs, Yao Lu, Nico Montali, Paul Mougin, Zoey Yang, Brandyn White, Aleksandra Faust, Rowan McAllister, Dragomir Anguelov, and Benjamin Sapp. Waymax: An Accelerated, Data-Driven Simulator for Large-Scale Autonomous Driving Research, 2023.

[16] John Houston, Guido Zuidhof, Luca Bergamini, Yawei Ye, Long Chen, Ashesh Jain, Sammy Omari, Vladimir Iglovikov, and Peter Ondruska. One Thousand and One Hours: Self-driving Motion Prediction Dataset. In *CoRL*, pages 409–418, 2020.

[17] Boris Ivanovic, Guanyu Song, Igor Gilitschenski, and Marco Pavone. trajdata: A Unified Interface to Multiple Human Trajectory Datasets, 2023.

[18] Saman Kazemkhani, Aarav Pandya, Daphne Cornelisse, Brennan Shacklett, and Eugene Vinitsky. GPUDrive: Data-driven, multi-agent driving simulation at 1 million FPS, 2025.

[19] Robert Krajewski, Julian Bock, Laurent Kloeker, and Lutz Eckstein. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. In *ITSC*, pages 2118–2125, 2018.

[20] Robert Krajewski, Tobias Moers, Julian Bock, Lennart Vater, and Lutz Eckstein. The rounD Dataset: A Drone Dataset of Road User Trajectories at Roundabouts in Germany. In *ITSC*, pages 1–6, 2020.

[21] Guido Küppers, Jean-Pierre Busch, Lennart Reiher, and Lutz Eckstein. V2AIX: A Multi-Modal Real-World Dataset of ETSI ITS V2X Messages in Public Road Traffic. *CoRR*, 2024.

[22] Quanyi Li, Zhenghao Peng, Lan Feng, Qihang Zhang, Zhenghai Xue, and Bolei Zhou. MetaDrive: Composing Diverse Driving Scenarios for Generalizable Reinforcement Learning, 2022.

[23] Quanyi Li, Zhenghao Peng, Lan Feng, Zhizheng Liu, Chenda Duan, Wenjie Mo, and Bolei Zhou. ScenarioNet: Open-Source Platform for Large-Scale Traffic Scenario Simulation and Modeling, 2023.

[24] Mingyu Liu, Ekim Yurtsever, Jonathan Fossaert, Xingcheng Zhou, Walter Zimmer, Yuning Cui, Bare Luka Zagar, and Alois C. Knoll. A Survey on Autonomous Driving Datasets: Statistics, Annotation Quality, and a Future Outlook, 2024.

[25] Sumbal Malik, Manzoor Ahmed Khan, Aadam, Hesham El-Sayed, Farkhund Iqbal, Jalal Khan, and Obaid Ullah. CARLA+: An Evolution of the CARLA Simulator for Complex Environment Using a Probabilistic Graphical Model. *Drones*, 7(2), 2023.

[26] Johannes Meier, Luca Scalerandi, Oussema Dhaouadi, Jacques Kaiser, Araslanov Nikita, and Daniel Cremers.

CARLA Drone: Monocular 3D Object Detection from a Different Perspective. In *GCPR*, 2024.

[27] Tobias Moers, Lennart Vater, Robert Krajewski, Julian Bock, Adrian Zlocki, and Lutz Eckstein. The exiD Dataset: A Real-World Trajectory Dataset of Highly Interactive Highway Scenarios in Germany. In *IV*, pages 958–964, 2022.

[28] Awad Mukbil, Yasin Maan Yousif, Sakif Hossain, and Jörg P. Müller. CTV-Dataset: A Shared Space Drone Dataset for Cyclist-Road User Interaction Derived from Campus Experiments. In *ITSC*, pages 3186–3191, 2023.

[29] Nigamaa Nayakanti, Rami Al-Rfou, Aurick Zhou, Kratarth Goel, Khaled S. Refaat, and Benjamin Sapp. Wayformer: Motion Forecasting via Simple Efficient Attention Networks, 2022.

[30] U. S. Department of Transportation Federal Highway Administration. Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data. https://data.transportation.gov/d/8ect-6jqj, 2016. [Accessed 16-12-2024].

[31] Abhishek Patil, Srikanth Malla, Haiming Gang, and Yi-Ting Chen. The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes, 2019.

[32] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes. In *ECCV*, pages 549–565, 2016.

[33] Jiajun Shen and Guangchuan Yang. Crash Risk Assessment for Heterogeneity Traffic and Different Vehicle-Following Patterns Using Microscopic Traffic Flow Data. *Sustainability*, 12(23), 2020.

[34] Shaoshuai Shi, Li Jiang, Dengxin Dai, and Bernt Schiele. Motion Transformer with Global Intention Localization and Local Movement Refinement, 2023.

[35] Paul Spannaus, Peter Zechel, and Kilian Lenz. AUTOMATUM DATA: Drone-based highway dataset for the development and validation of automated driving software for research and commercial applications. In *IV*, pages 1372–1377, 2021.

[36] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *CVPR*, pages 2443–2451, 2020.

[37] Pavlo Tkachenko, Novel Certad, Gunda Singer, Cristina Olaverri-Monreal, and Luigi del Re. The JKU DORA Traffic Dataset. *IEEE Access*, 10:92673–92680, 2022.

[38] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, 62(2):1805–1824, 2000.

[39] James R. Ward, Gabriel Agamennoni, Stewart Worrall, Asher Bender, and Eduardo Nebot. Extending Time to Collision for probabilistic reasoning in general traffic scenarios. *Transportation Research Part C: Emerging Technologies*, 51:66–82, 2015. Cited by: 112.

[40] Theodor Westny, Björn Olofsson, and Erik Frisk. Toward Unified Practices in Trajectory Prediction Research on Bird's-Eye-View Datasets, 2025.

[41] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting. In *NeurIPS*, 2021.

[42] Hao Xiang, Zhaoliang Zheng, Xin Xia, Runsheng Xu, Letian Gao, Zewei Zhou, Xu Han, Xinkai Ji, Mingxi Li, Zonglin Meng, Li Jin, Mingyue Lei, Zhaoyang Ma, Zihang He, Haoxuan Ma, Yunshuang Yuan, Yingqian Zhao, and Jiaqi Ma. V2X-Real: a Large-Scale Dataset for Vehicle-to-Everything Cooperative Perception, 2024.

[43] Pengchuan Xiao, Zhenlei Shao, Steven Hao, Zishuo Zhang, Xiaolin Chai, Judy Jiao, Zesong Li, Jian Wu, Kai Sun, Kun Jiang, Yunlong Wang, and Diange Yang. PandaSet: Advanced Sensor Suite Dataset for Autonomous Driving, 2021.

[44] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. OPV2V: An Open Benchmark Dataset and Fusion Pipeline for Perception with Vehicle-to-Vehicle Communication, 2022.

[45] Yanchao Xu, Wenbo Shao, Jun Li, Kai Yang, Weida Wang, Hua Huang, Chen Lv, and Hong Wang. SIND: A Drone Dataset at Signalized Intersection in China. In *ITSC*, pages 2471–2478, 2022.

[46] Dongfang Yang, Linhui Li, Keith A. Redmill, and Ümit Özgüner. Top-view Trajectories: A Pedestrian Dataset of Vehicle-Crowd Interaction from Controlled Experiments and Crowded Campus. In *IV*, pages 899–904, 2019.

[47] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning, 2020.

[48] Haibao Yu, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, Hanyu Li, Xing Hu, Jirui Yuan, and Zaiqing Nie. DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection, 2022.

[49] Haibao Yu, Wenxian Yang, Hongzhi Ruan, Zhenwei Yang, Yingjuan Tang, Xu Gao, Xin Hao, Yifeng Shi, Yifeng Pan, Ning Sun, Juan Song, Jirui Yuan, Ping Luo, and Zaiqing Nie. V2X-Seq: A Large-Scale Sequential Dataset for Vehicle-Infrastructure Cooperative Perception and Forecasting. In *CVPR*, pages 5486–5495, 2023.

[50] Wei Zhan, Liting Sun, Di Wang, Haojie Shi, Aubrey Clausse, Maximilian Naumann, Julius Kümmerle, Hendrik Königshof, Christoph Stiller, Arnaud de La Fortelle, and Masayoshi Tomizuka. INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps. *CoRR*, 2019.

[51] Yuxin Zhang, Cheng Wang, Ruilin Yu, Luyao Wang, Wei Quan, Yang Gao, and Pengfei Li. The AD4CHE Dataset and Its Application in Typical Congestion Scenarios of Traffic Jam Pilot Systems. *IEEE Trans. Intell. Veh.*, 8(5):3312–3323, 2023.

[52] Ou Zheng, Mohamed Abdel-Aty, Lishengsa Yue, Amr Abdelraouf, Zijin Wang, and Nada Mahmoud. CitySim: A Drone-Based Vehicle Trajectory Dataset for Safety-Oriented Research and Digital Twins. *Transportation Research Record*, 2678(4):606–621, 2024.

[53] Walter Zimmer, Gerhard Arya Wardana, Suren Sritharan, Xingcheng Zhou, Rui Song, and Alois C. Knoll. TUMTraf V2X Cooperative Perception Dataset, 2024.

[54] Walter Zimmer, Ross Greer, Daniel Lehmberg, Marc Pavel, Holger Caesar, Xingcheng Zhou, Ahmed Ghita, Mohan Trivedi, Rui Song, Hu Cao, et al. Towards Vision Zero: The Accid3nD Dataset. *arXiv preprint arXiv:2503.12095*, 2025.