Towards Vision Zero: The TUM Traffic Accid3nD Dataset
Supplementary Material

accident-dataset.github.io

Figure 1.**Visualization labeled sequences within the TUMTraf-Accid3nD dataset.** It contains 12 sequences each recorded from four roadside cameras on a highway test bed. The dataset includes accident scenes during different lighting and weather conditions.

Contents

A. Additional Related Work

A.1. VLMs for Accident Detection

Vision-language models (VLMs) [1, 2] are increasingly being explored for accident detection, as they enable a deeper understanding of complex scenes by linking visual data with language-based descriptions. They use natural language to describe objects, actions, and interactions in visual scenes. VLMs like AccidentGPT [3] can interpret and contextualize

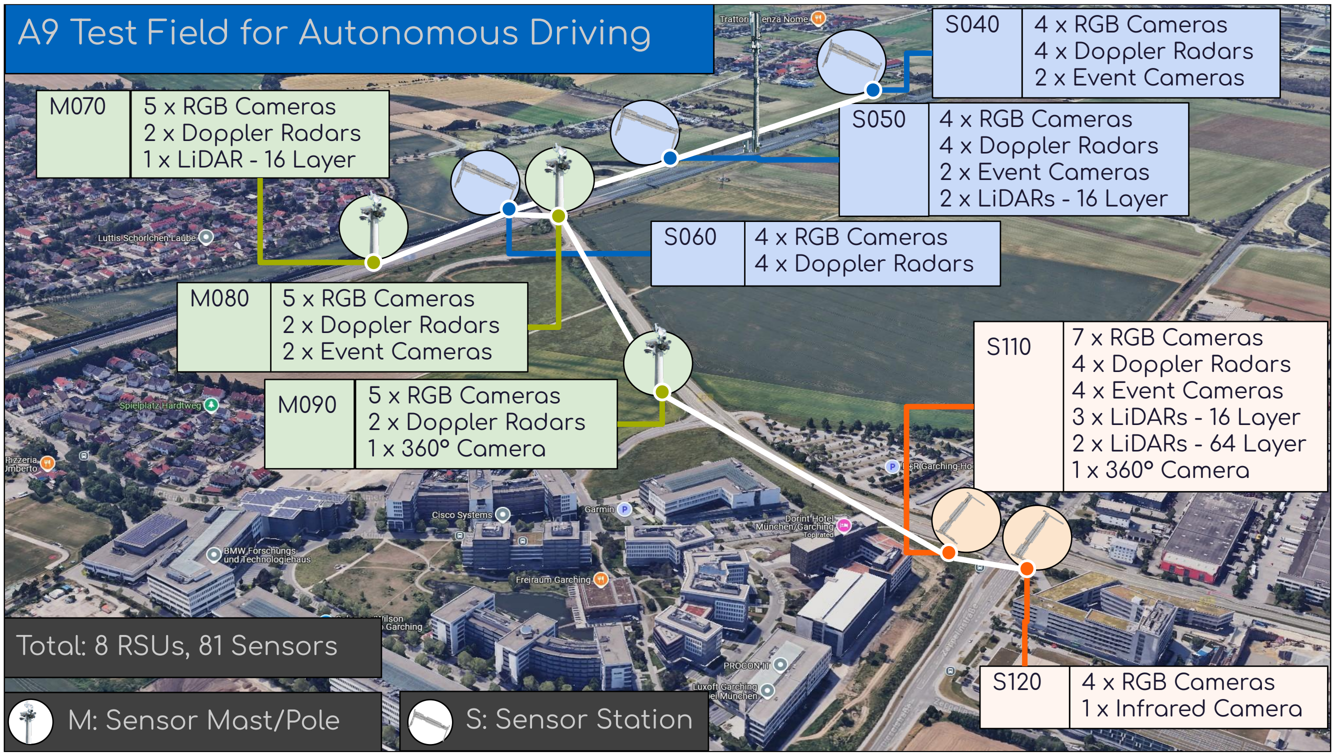

Figure 2. **Overview of the A9 Digital Test Field for Autonomous Driving.** The blue sensor stations (gantry bridges S040 and S050) on the highway were used to record the data with four roadside cameras, four radars, and one LiDAR.

accident scenarios with enhanced detail. The integration of VLMs aids in recognizing accident-related events, such as sudden braking, near-misses, or unusual traffic patterns. With vision-language models, systems can also provide interpretable explanations for detected incidents, offering insights that support safer decision-making in autonomous and intelligent transportation systems. Related approaches were developed in [4, 5] to detect both accidents and general traffic anomalies.

B. Extended Dataset Description

B.1. Data Collection Setup

The A9 Digital Test Field for Autonomous Driving is located on the A9 highway in Munich, Germany. Figure 2 shows the 3 km long test field from a bird's-eye view. It contains eight measurement stations in total: three on the highway (marked in blue), three in a rural area (marked in green), and another two in an urban area (marked in orange). Each station includes a robust sensor suite with roadside cameras, radars, and LiDAR units. The infrastructure supports seamless tracking of objects across multiple sensor stations using sensor data fusion and tracking. All traffic participants are continuously monitored, and the anonymized recordings are stored on secure servers. Sensors were positioned to optimize coverage, enabling detailed observation of accident-prone areas such as merging lanes and intersections. The dataset was recorded at a busy section of the highway, containing 12 lanes including two exit lanes. On this highway, various scenarios can occur in diverse weather (sunny, cloudy, foggy, rainy) and lighting conditions (day, dusk, dawn, and night time). The specific accident types are described in Section B.3.

B.2. Annotation Details

The dynamic highway environment posed significant challenges, particularly in ensuring accurate sensor calibration over a long time. Changes in harsh weather conditions, such as heavy rain, hail, or snow, have often required recalibrating the sensors to maintain an accurate data capture. Occlusion was another critical issue, as large trucks and dense traffic frequently obstructed smaller vehicles and pedestrians, complicating the annotation process. We apply novel tracking methods to track occluded objects. Additionally, accident scenarios often involved rolled-over vehicles, which made it difficult to annotate. We did not decide to annotate the roll and pitch angle of traffic participants to simplify the labeling guidelines. To ensure a high labeling quality, we adopted an iterative annotation process that incorporated feedback from domain experts.

Figure 4. **Visualization of labeled sequence S01.** A vehicle is speeding at 180 km/h and crashes into a van that has a breakdown on the left lane.

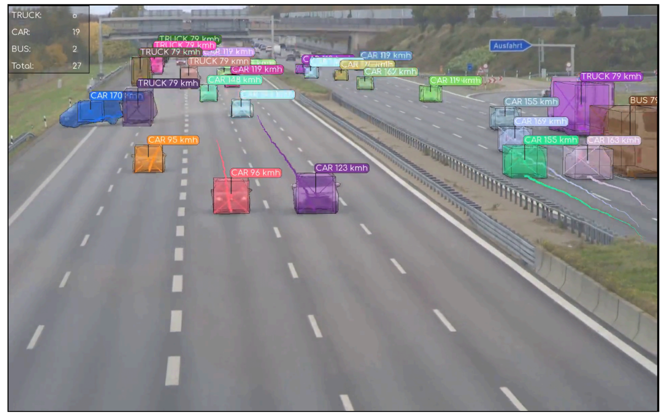

Figure 5. **Visualization of labeled sequence S02.** A blue van with a trailer is tipping over after a strong wind gust is hitting it.

Figure 6. **Visualization of labeled sequence S03.** A vehicle is not maintaining safety distance to the leading vehicle. It changes the lane to the right and hits another two vehicles. One of them is full 360-degree spin.

Figure 7. **Visualization of labeled sequence S04.** A post-accident scenario was labeled with arriving police, ambulance, and fire trucks. Some people get out of their cars to secure the site of an accident.

Figure 8. **Visualization of labeled sequence S05.** A van stops on the shoulder lane and starts to burn. The passengers are getting out of the van and secure their belongings. A man tries to extinguish the fire without any success.

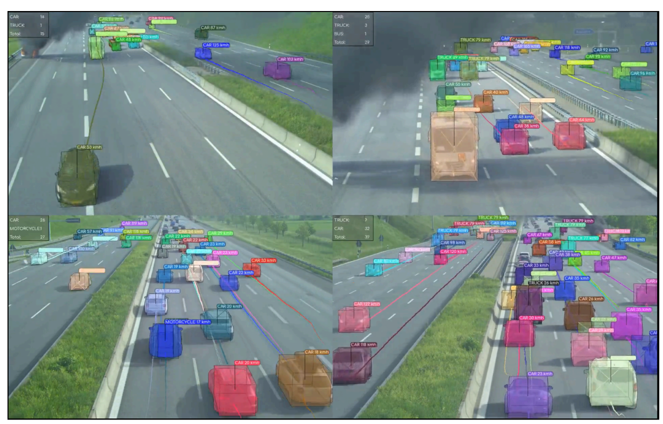

Figure 9. **Visualization of labeled sequence S06.** The van starts to burn heavily and produces a large smoke cloud that occludes the traffic participants in the roadside camera.

Figure 11. **Visualization of labeled sequence S07.** Emergency vehicles (police, ambulance, and fire trucks) arrive a van is burning for 15 minutes.

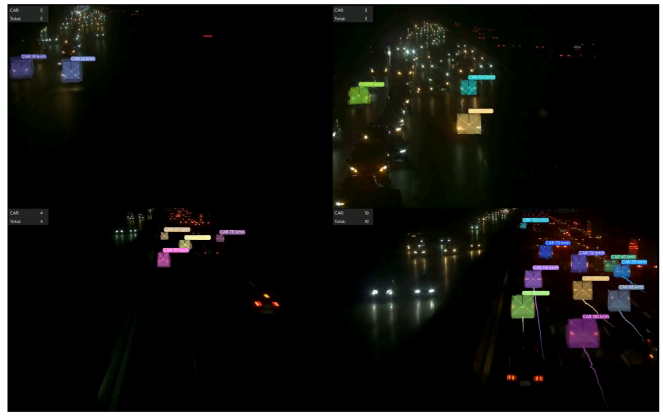

Figure 12. **Visualization of labeled sequence S08.** A traffic jam is forming on the highway. A vehicle changes lanes at night without turning on the indicator lights. Another vehicle is tail-gaiting that vehicle from the back.

Figure 13. **Visualization of labeled sequence S09.** A truck is changing lanes and tailgates a car from the back. Both vehicles stop on the shoulder lane to inspect the collision.

Figure 14. **Visualization of labeled sequence S10.** A truck is changing lanes and rams a passenger vehicle from the side. It is spinning 180-degree in one direction and then in the other.

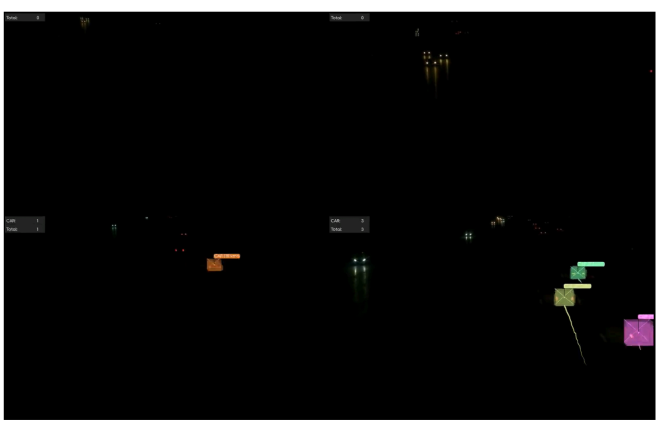

Figure 15. **Visualization of labeled sequence S11.** The driver of a pick-up truck falls asleep momentarily at night, changes three lanes, and flies over the guardrail. After landing behind the guardrail, it is rolling over three times.

Figure 16. **Visualization of labeled sequence S12.** A vehicle is speeding and crashes into a van that has a breakdown on the left lane.

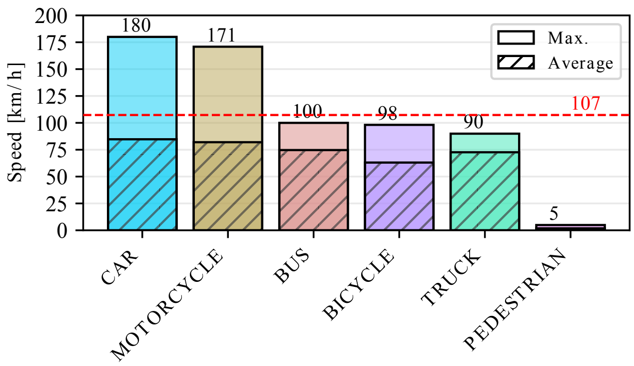

Figure 17. **Visualization of speed values for each labeled category.** We show the average and maximum speed values for all categories. The average speed is 107 km/h in the dataset.

B.3. Quality Assurance

A multi-level quality assurance process was implemented to ensure annotation precision and consistency. Expert reviewers conducted manual inspections of labeled frames to identify misalignments or inaccuracies in categories, instance masks, and 3D bounding boxes. Automated checks were employed to validate trajectory continuity and to ensure logical consistency in object movements. For example, any abrupt jumps in tracked objects' paths were flagged for review. This hybrid approach combined the advantages of human expertise and automated validation, significantly improving the reliability of the dataset while reducing annotation errors.

B.4. Detailed Dataset Visualization

Das dataset is visualized in Figs. 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, and 16. All 12 sequences recorded from four roadside cameras are displayed. The annotated frames in the dataset include the object category, speed, color-coded 2D instance masks and 3D bounding boxes to denote the object track, trajectory lines to indicate movement paths, and the total number of traffic participants in the current scene. These visualizations provide a view of interactions within a scene, such as the buildup to an accident or evasive maneuvers by other vehicles. These visualizations of traffic participants, coupled with speed values, make it easier to analyze and interpret complex traffic scenarios. Moreover, the speed and velocity of all traffic participants were calculated based on the 3D location and the time difference between the frames. Figure 17 visualized the average and max. speed values for each labeled category. The max. speed in the labeled dataset is 180 km/h (sequence S01). The overall maximum speed that was found in the recordings is 263 km/h in the north direction of the highway, where no speed limit is set.

C. Comparative Dataset Analysis

C.1. Quantitative Comparison

Our dataset is tailored for accident-centric research, addressing gaps left by popular datasets like KITTI [6, 7], nuScenes [8], Waymo Open [9], DAIR-V2X [10], and TUMTraf-V2X [11]. With over 111,000 labeled frames, it offers a large-scale dataset with a high density of safety-critical scenarios. Trajectories are notably longer with a maximum of 2,114 meters, allowing for the study of pre-accident behaviors such as abrupt lane changes or sudden decelerations. The dataset's multi-modal approach, combining data from four roadside cameras and one LiDAR, enhances its suitability for cooperative perception research. A comparative table (Table 1) illustrates these metrics, and emphasizes our dataset's strengths in annotation density, accident-specific focus, and environmental diversity.

C.2. Unique Features

Our dataset stands out with its accident-focused labeling and unique features tailored for safety-critical analysis. It includes high-speed accident scenarios (up to 180 km/h) and extensive trajectory lengths, with a cumulative track length of 2,250 km and individual tracks exceeding 2 km. Captured on a busy 12-lane highway, frames contain up to 55 labeled objects that are labeled at 25 Hz with synchronized camera and LiDAR data. With over 2.6 million 3D boxes, the dataset precisely records 3D vehicle locations and enables realistic accident reconstructions in simulations. It can also be used to train models for detecting small and distant objects (200–400 m away) or to better detect objects that are occluded by large vehicles like trucks. Finally, our dataset includes diverse scenarios, such as night-time accidents, which allows robust model training.

C.3. Prevalence of Accident Types

Accident scenarios are categorized into key types to support diverse research objectives. Rear-end and side collisions comprise 25% of accidents, often captured in stop-and-go traffic conditions. Breakdown events account for another 25%. They occur frequently on the highway, and vehicles stop on the shoulder lane to inspect the failure. Multi-vehicle pileups, where more than two vehicles were involved, also contained rear-end and side collisions. They occurred in 16.67% and are extensively annotated to capture their complex dynamics. In another two sequences (16.67%), the driver fell asleep. One sequence (8.3%) contains a scenario in which a heavy wind burst caused an accident. A trailer that was towed by a van tipped over, letting the vehicle tip over. The pie chart in Figure 18 illustrates the distribution of these accident types, offering a clear visualization of the dataset's focus areas and the breadth of scenarios included for safety analysis.

Figure 1. **VLM analysis results.** TUMTraf-Accid3nD scene images flagged as novel by the method of [4]. We note that the vehicle accident was successfully identified as novel by multiple cameras.

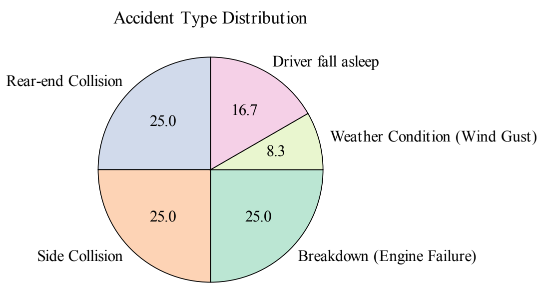

Figure 18. **Distribution of accident types.**

D. Benchmarking Experiments

D.1. Experimental Setup

Benchmarking experiments utilized high-performance NVIDIA GPUs (RTX 3090) and frameworks such as PyTorch for object detection and tracking. Key evaluation metrics include the mean Average Precision (mAP) for detection, F1-scores for classification tasks, and Multi-Object Tracking Accuracy (MOTA) for tracking performance. All benchmarks followed a standardized protocol to ensure reproducibility, including fixed dataset splits and uniform preprocessing pipelines. The experimental setup was validated to provide reliable baselines for future research.

D.2. Supported Tasks

The dataset supports a range of tasks critical to autonomous driving, including 2D object detection, 2D instance segmentation, 3D object detection, object tracking, sensor data fusion, trajectory prediction, and accident detection. Additionally, its multi-modal nature enables advanced use cases such as risk assessment, scene reconstruction, and cooperative perception. This allows researchers to address fundamental challenges in safety-critical scenarios.

D.3. Experimental Analysis

Performance varied significantly across object classes and accident scenarios. Cars, the most frequent object type, achieved the highest detection precision for accident detection, benefiting from abundant training samples. Trucks and motorcycles, while less frequent, showed lower precision due to challenges like size variability and partial occlusions. Scenarios involving sudden stops or lane changes exhibited higher accuracy compared to subtle incidents like slow drifts. This per-class breakdown informs researchers about class-specific challenges, guiding improvements in model architectures and dataset-balancing strategies.

D.4. Vision-Language Model Efficacy

As a demonstration of VLM analysis on the TUMTraf-Accid3nD benchmark, we apply the novelty detection method of [4] to the dataset. Of the 11,924 images in the test split, 12 are flagged as most novel, with 6 of these images depicting accidents and others depicting novel events such as stopped traffic or unusually patterned vans. This method prioritizes finding events unlike others in the dataset, so the relatively high frame rate challenges the method. This can be overcome by simple downsampling of the dataset. Interestingly, the method is able to pick up on novel accident scenes from each camera view. Example figures from this novel set are displayed in Fig. 1. We note that for future research, binary annotation of accident vs. non-accident scenes can allow for training and quantitative evaluation of this novelty detection method over the dataset.

E. Accident Detection Methodology

E.1. Rule-Based Detection

Our rule-based detection module uses a set of pre-defined thresholds to identify accidents. For example, a vehicle decelerating at a rate exceeding 5 m/s^2 or experiencing a sudden trajectory deviation of over 30° triggers an accident flag. These rules are grounded in real-world traffic data and validated through extensive simulations. Pseudocode and mathematical formulas outline the detection logic, providing transparency and replicability for researchers.

E.2. Learning-Based Detection

The YOLOv8 model [12], fine-tuned on our custom dataset, achieved a precision of 80% in detecting accident events. We used a balanced dataset split to train the model to mitigate the effects of class imbalance and to improve the detection sensitivity for less frequent accident types like side collisions. Hyperparameter tuning, such as adjusting learning rates and batch sizes, further optimized model performance. A detailed analysis of the model's sensitivity to dataset size and class distribution reveals its robustness in real-world scenarios, offering a reliable benchmark for future development.

F. Applications and Broader Implications

The dataset supports autonomous driving applications, particularly in safety-critical scenarios.

F.1. Autonomous Driving Use Cases

The dataset empowers autonomous driving systems by enhancing their ability to handle safety-critical situations. It supports tasks like real-time accident detection and trajectory prediction, enabling proactive measures like emergency braking or lane adjustments. For instance, recognizing abrupt braking patterns allows autonomous systems to react swiftly, minimizing collision risks.

F.2. Infrastructure-Based Applications

Beyond vehicle systems, the dataset has significant implications for Intelligent Transportation Systems (ITS). V2X communication applications use this data to send timely alerts to nearby vehicles and emergency services, reducing response times during accidents. These interventions use sensor data fusion to ensure precise and actionable insights for infrastructure-based safety measures.

F.3. Research Opportunities

The dataset unlocks avenues for advancing multi-modal learning, cooperative perception, and predictive modeling. Research can explore scene graph generation to map intricate relationships between objects or delve into accident anticipation models for preemptive interventions. These areas represent cutting-edge challenges that can reshape safety mechanisms in autonomous systems and transportation.

G. Limitations and Future Directions

G.1. Dataset Limitations

The dataset primarily focuses on highways, with limited representation of urban settings and adverse weather conditions like snow. Additionally, the scarcity of nighttime accident samples introduces biases that may impact model generalizability to diverse environments.

G.2. Future Expansion

Plans for expansion include capturing data in urban environments, incorporating additional sensor modalities such as 4D radar and thermal imaging for low-light scenarios, and increasing the diversity of accident types. These improvements aim to bridge current gaps, making the dataset more versatile for varied applications and conditions. Our goal is also to deploy the accident detection pipeline in real-world settings to notify traffic participants in real-time. In future work, we will also explore accident video diffusion approaches for generating realistic accident scenarios, enabling better training and evaluation of autonomous driving models in rare and safety-critical situations.

H. Development Kit Overview

The dataset is licensed under the Creative Commons License (CC BY-NC-SA 4.0) and is accompanied by a development kit comprising tools for annotation correction, model evaluation, and visualization. Pre-trained models and modular scripts are provided to simplify integration, enabling researchers to focus on experimentation and innovation.

References

[1] Xingcheng Zhou, Mingyu Liu, Ekim Yurtsever, Bare Luka Zagar, Walter Zimmer, Hu Cao, and Alois C. Knoll. Vision Language Models in Autonomous Driving: A Survey and Outlook, *IEEE Transactions on Intelligent Vehicles*, pages 1–20, 2024, doi: 10.1109/TIV.2024.3402136.

[2] Xingcheng Zhou and Alois C Knoll. GPT-4V as traffic assistant: an in-depth look at vision language model on complex traffic events, *preprint arXiv:2402.02205*, 2024.

[3] Lening Wang, Yilong Ren, Han Jiang, Pinlong Cai, Daocheng Fu, Tianqi Wang, Zhiyong Cui, Haiyang Yu, Xuesong Wang, Hanchu Zhou, Helai Huang, and Yinhai Wang. AccidentGPT: A V2X Environmental Perception Multi-modal Large Model for Accident Analysis and Prevention, in *2024 IEEE Intelligent Vehicles Symposium (IV)*, 2024, pages 472–477. doi: 10.1109/IV55156.2024.10588374.

[4] Ross Greer and Mohan Trivedi. Towards explainable, safe autonomous driving with language embeddings for novelty identification and active learning: Framework and experimental analysis with real-world data sets, *preprint arXiv:2402.07320*, 2024.

[5] Ross Greer, Samveed Desai, Lulua Rakla, Akshay Gopalkrishnan, Afnan Alofi, and Mohan Trivedi. Pedestrian Behavior Maps for Safety Advisories: CHAMP Framework and Real-World Data Analysis, in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pages 1–8. doi: 10.1109/IV55152.2023.10186648.

[6] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite, in *2012 IEEE Conf. on Computer Vision and Pattern Recognition*, Jun. 2012, pages 3354–3361. doi: 10.1109/CVPR.2012.6248074.

[7] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision Meets Robotics: The KITTI Dataset, *The Int. Journal of Robotics Research*, vol. 32, no. 11, pages 1231–1237, Sep. 2013, doi: 10.1177/0278364913491297.

[8] 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 *2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*

(CVPR), Jun. 2020, pages 11618–11628. doi: 10.1109/CVPR42600.2020.01164.

[9] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurélien 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 *2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pages 2443–2451. doi: 10.1109/CVPR42600.2020.00252.

[10] 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, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pages 21329–21338. doi: 10.1109/CVPR52688.2022.02067.

[11] Walter Zimmer, Gerhard Arya Wardana, Suren Sritharan, Xingcheng Zhou, Rui Song, and Alois C. Knoll. TUM-Traf V2X Cooperative Perception Dataset, in *2024 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2024, pages 22668–22677. doi: 10.1109/CVPR52733.2024.02139.

[12] Glenn Jocher, Jing Qiu, and Ayush Chaurasia. Ultralytics YOLO. [Online]. Available: https://github.com/ultralytics/ultralytics