iEXODDUS
THE FUTURE OF MOBILITY

A Comprehensive Dataset of construction sites, incidents and tunnel

Milestone Number	MS6
Dissemination Level	PU (Public)
Deliverable Responsible	THI
Planned Delivery Date	2025-2-28 (M15)
Actual Delivery Date	2025-3-09 (M15)
Document Version & Status	V1.1 | Final

Project Acronym	iEXODDUS
Project Title	EXtension of ODDs - applied in connected and aUtomated driving and Standardization procedures
Grant Agreement Number	101146091
Project Coordinator	Virtual Vehicle Research GmbH
Project Website	www.iexoddus-project.eu

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

iEXODDUS has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101146091. This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) for the Swiss partners.

Authors

Name	Organisation	Name	Organisation
Xie Xujun	THI	Romain Bellessort	CANON CRF
Alexander Wilczynski	ika, RWTH Aachen	Eric Nassor	CANON CRF
Silas Damaschke	ika, RWTH Aachen	Martin Kirchengast	ViF
Lukas Zanger	ika, RWTH Aachen	Jakob Reckenzaun	ViF
Tuğrul Can Erk	FORD OTOSAN	Leonardo Gonzalez	Tecnalia

Reviewers

Name	Organisation	Date

Change History

Version	Date	Name/Organisation	Description
V0.1	2025-10-01	Xujun - THI	Created the report
V1.0	2026-02-15	Xujun - THI	Finalized the report
V1.1	2026-03-09	Xujun - THI	Update report content based on internal review

Table of Contents

List of Figures

List of Tables

Abbreviations

Term	Definition
ADAS	Advanced Driver Assistance System
ALKS	Automated Lane Keeping System
AV	Automated Vehicle
CAM	Cooperative Awareness Message
CAV	Connected and Automated Vehicle
CCAM	Connected Cooperative and Automated Mobility
C-ITS	Cooperative Intelligent Transport Systems
CPM	Collective Perception Service Message
DEC	Websites, Patent Filings, Videos etc.
DEM	Demonstrator, Pilot, Prototype
DENM	Decentralized Environmental Notification Message
DT	Digital Twin
ETSI	European Telecommunications Standards Institute
GNSS	Global Navigation Satellite System
LiDAR	Light Detection and Ranging
I2V	Infrastructure to Vehicle
V2X	Vehicle to Everything
iEXODDUS	Extension of ODDs - applied in connected and automated driving and standardization procedures
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
IVIM	Infrastructure to Vehicle Information Message
LiDAR	Light Detection and Ranging
OBU	On-Board Unit
ODD	Operational Design Domain
PU	Public
R	Document, Report
RSU	Roadside Unit
SEN	Sensitive
UWB	Ultra-Wide Band
V2X	Vehicle to Everything
WP	Work Package
CODA dataset	Corner Case Dataset for Autonomous Driving dataset
DAD	Dashcam Accident Dataset
CCD	Car Crash Dataset
A3D framework	AnAn Accident Detection framework
DoTA	Detection of Traffic Anomalies
DADA	Driver Attention in Driving Accidents
TUMTraf-A	TUM Traffic Accident Dataset
TU-DAT	Temple University Dataset

CADP	CCTV Accident Dataset
NVIDIA STRIVE	NVIDIA Stress-Test Drive
DISC	Dataset for Analyzing Driving Styles in Simulated Crashes
ZOD	Zenseact Open Dataset
ONCE dataset	One Million Scenes dataset
PAVE	Production Autonomous Vehicle Evaluation
H-V2X	Highway V2X
VICAD	Vehicle-Infrastructure Collaborative Autonomous Driving

1. Executive Summary

The progression of Autonomous Vehicle (AV) technology has reached a pivotal inflection point. While early-stage development focused on nominal driving conditions—clear weather, standardized infrastructure, and compliant road users, the current frontier of research is defined by the "long tail" of operational scenarios. These scenarios, characterized by low probability but high consequences, represent the primary barrier to the widespread commercial deployment of Level 4 and Level 5 systems. Among the most challenging of this edge cases are Work Zones, Incidents/Anomalies, and Tunnel Environments.

This report provides an exhaustive analysis of the public dataset landscape relevant to these three critical domains. By synthesizing data from over 20 distinct real-world and synthetic. The analysis reveals a significant dichotomy in data availability. Real-world datasets, while rich in visual fidelity, suffer extreme class imbalance; safety-critical events such as accidents or complex construction zones often constitute less than 0.2% of labeled data. Conversely, synthetic environments generated by simulators like CARLA are emerging as essential tools for "densifying" these rare events, allowing for the procedural generation of collision scenarios that are ethically impossible to capture in the wild.

This document serves as a strategic guide for researchers and engineers, detailing the specifications, sensor modalities, and specific utility of these datasets in solving the perception challenges of the most hostile driving environments.

2. Introduction

The progression of Autonomous Vehicle technology has reached a pivotal inflection point. While early-stage development focused on nominal driving conditions—clear weather, standardized infrastructure, and compliant road users, the current frontier of research is defined by the "long tail" of operational scenarios. These scenarios, characterized by low probability but high consequences, represent the primary barrier to the widespread commercial deployment of Level 4 and Level 5 systems. Among the most challenging of this edge cases are Work Zones, Incidents/Anomalies, and Tunnel Environments.

This report provides an exhaustive analysis of the public dataset landscape relevant to these three critical domains. The analysis reveals a significant dichotomy in data availability. Real-world datasets, while rich in visual fidelity, suffer extreme class imbalance; safety-critical events such as accidents or complex construction zones often constitute less than 0.2% of labeled data [1]. Conversely, synthetic environments generated by simulators like CARLA are emerging as essential tools for "densifying" these rare events, allowing for the procedural generation of collision scenarios that are ethically impossible to capture in the wild.

This document serves as a strategic guide for researchers and engineers, detailing the specifications, sensor modalities, and specific utility of these datasets in solving the perception challenges of the most hostile driving environments.

2.1 The Statistical Challenge of Rare Events

The fundamental challenge in modern AV perception is not the detection of standard objects, but the interpretation of context in OOD scenarios. Standard supervised learning models require vast quantities of training data to generalize effectively. However, the most critical driving scenarios reside in the statistical "long tail."

Research analyzing the ROADWork dataset indicates that despite the ubiquity of road maintenance in the driver's experience, actual work zone frames appear with vanishing frequency in general-purpose datasets [1]. For instance, in the massive Mapillary Vistas and BDD100K collections, images containing active roadwork constitute only 2.3% and 0.4% of the total volume, respectively. This scarcity creates a "blind spot" for algorithms trained on these distributions, leading to systems that may perform exceptionally well on highways but fail catastrophically when confronted with the high-entropy environment of a construction site.

2.2 Defining the Critical Domains

To structure the analysis of the available data, this report categorizes the landscape into three distinct, high-friction ODDs.

- **Work Zones (Construction):** These are environments of "negotiated order." They violate the static map assumption, introducing temporary lanes, novel signal types (arrow boards, hand signals), and dynamic occlusions. The primary challenge is not just object detection, but semantic interpretation—understanding that a worker's gesture supersedes a traffic light.

- **Incidents and Anomalies:** This category encompasses the chaotic moments of traffic failure—collisions, near-misses, and the presence of foreign objects (debris, fallen cargo). Due to the ethical impossibility of intentionally creating accidents for data collection, this domain is the most data-starved, relying heavily on synthetic generation and opportunistic dashcam recording.

- **Tunnel Environments:** Tunnels represent a "denial of service" attack on AV sensors. The enclosed geometry blocks GNSS signals, forcing reliance on dead-reckoning or SLAM. Simultaneously, the lighting transitions at portals (the "black hole" effect) challenge camera dynamic range, while the repetitive wall structures can cause geometric degeneracy in LiDAR processing.

2.3 Methodology of Analysis

Data acquisition for accident analysis is categorized by the vantage point of the sensors and the nature of the reality captured. This report is structured around these three pillars:

- **Ego-Perspective (Onboard)**: Data captured from the sensors mounted on the AV itself (dashcams, LiDARs, IMUs). This represents the agent's localized view, crucial for developing self-reliant systems but limited by occlusion and field of view.
- **Infrastructure Perspective (Roadside/V2X):** Data captured from stationary sensors. This offers a top-down view eliminating blind spots and facilitating cooperative perception, but faces challenges in calibration, coverage, and privacy.
- **Simulation (Synthetic):** Data generated in virtual environments. This offers infinite scalability, safety, and controllability, allowing for the generation of adversarial scenarios that would be impossible to stage in the real world.

The following sections provide a deep-dive analysis of the datasets defining the state-of-the-art in each of these domains.

3. Work Zone and Construction Datasets: Navigating Entropy

Road work zones represent the concept of "entropy" in autonomous driving: they are environments where order dissolves, standard rules are suspended, and unpredictability is the only constant. Unlike the static structure of a highway or the rhythmic flow of a suburban intersection, construction zones are dynamic, featuring temporary signage, irregular lane shifts, and the presence of vulnerable human workers.

This chapter explores the landscape of Ego-Centric Datasets, which capture the sensory experience of a deployed AV using LiDAR, Radar, and Camera suites. We examine how these datasets train models to handle the geometric and semantic challenges of altered road topologies. We will analyze three critical benchmarks: ROADWork [1], which addresses the specific semantics of construction; CODA [2], which targets the "long tail" of corner cases and irregularities; and the Waymo Open Dataset [3–6], which offers high-fidelity geometric data in dense urban environments. Together, these resources form the foundation for training perception algorithms to bring order to the chaos of the construction site.

3.1 Ego-Centric Datasets: The Vehicle's Perspective

Ego-perspective datasets mimic the sensory inputs of a deployed autonomous vehicle. They are typically collected using a sensor suite comprising LiDAR, Radar, and Cameras mounted on a moving platform. The primary utility of these datasets in the context of work zones is to train models for object detection (cones, signs), semantic segmentation (drivable area), and path planning.

3.1.1 ROADWork: The Benchmark for Construction Semantics

The ROADWork dataset [1] addresses one of the most challenging scenarios in self-driving: navigating through roadwork zones. Despite significant advances in autonomous driving technology, work zones remain a major obstacle due to their rarity, high variability, and frequent deviation from standard traffic rules.

To bridge this critical gap, ROADWork provides a massive-scale, multi-modal collection of real-world scenarios sourced from 18 diverse U.S. cities. As illustrated in Figure 1, the dataset goes beyond simple object detection by providing:

- **Rich Semantic Annotations:** Dense segmentation masks for traffic drums, cones, and work vehicles.
- **Signage Intelligence:** OCR and classification of Temporary Traffic Control (TTC) signs, such as "Utility Work Ahead" or directional arrows.
- **Path Planning Support:** Delineated drivable paths that guide models through complex lane shifts and temporary bypasses.
- **Natural Language Descriptions:** Detailed scene narratives that provide context for the spatial relationships between construction elements and the roadway.

The data was collected through extensive field campaigns specifically designed to capture the diversity and complexity of real-world roadwork scenarios. This large-scale collection effort ensures that the dataset represents a wide geographical distribution and various types of construction activities encountered in urban and suburban environments.

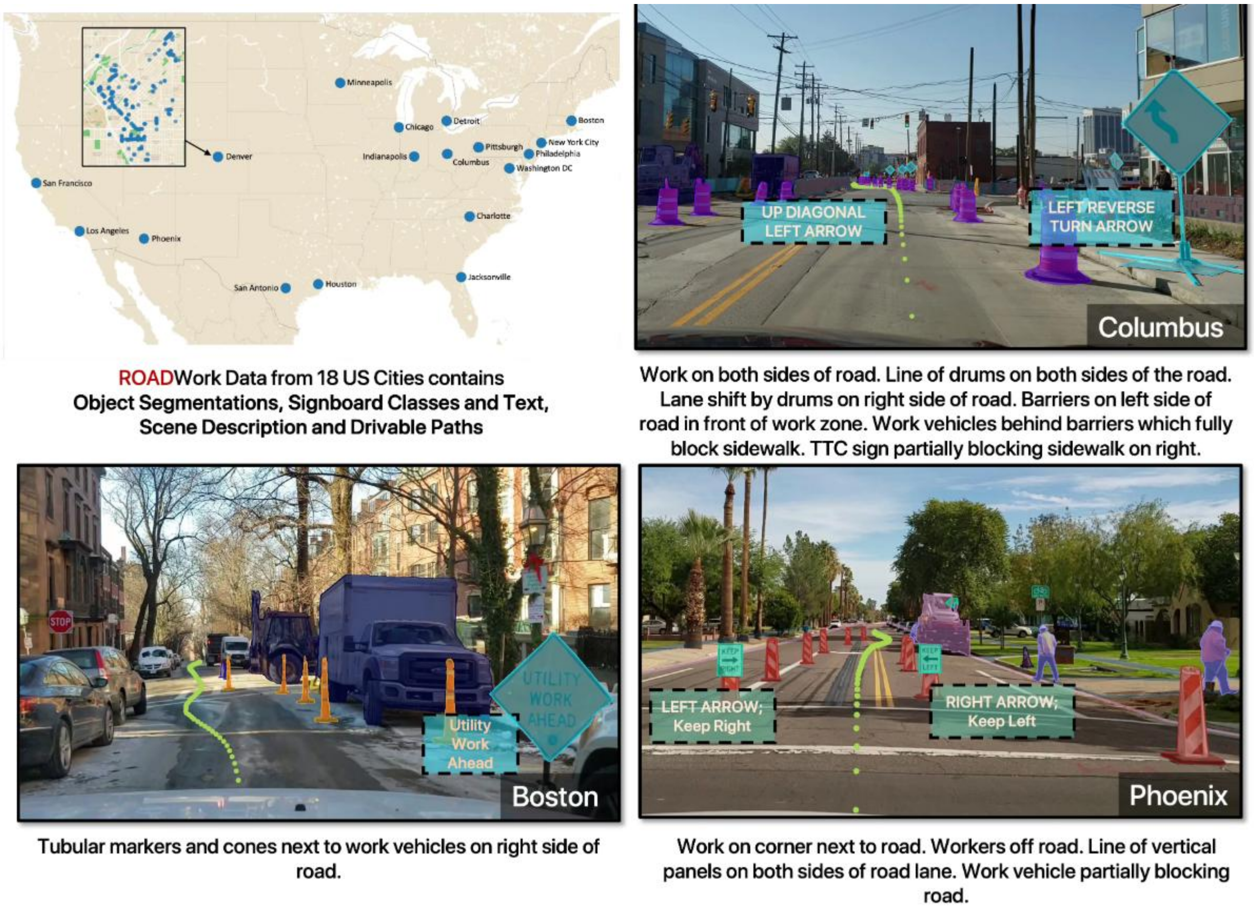

Figure 1 Overview of the ROADWork Dataset. A multi-city collection (18 US locations) featuring comprehensive annotations for autonomous navigation in work zones [1].

The dataset comprises:

- 4,375 thirty-second videos capturing dynamic work zone scenarios
- 9,650 richly annotated keyframes providing detailed static analysis
- 129,017 images with automatically computed pathways from more than 5,000 work zones
- Coverage of 15 types of objects and 360 types of TTC (Temporary Traffic Control) signs

3.1.1.1 Data Annotation Schema

The ROADWork dataset features a multi-level annotation schema that mirrors human cognition in understanding work zones. The annotations are organized into several comprehensive levels:

4.1 Level (a): Work Zone Recognition and Object Detection

- Bounding box annotations for constituent objects within work zones
- Segmentation labels providing pixel-level object delineation
- Detection of 15 different object types commonly found in roadwork zones

4.2 Level (b): Fine-Grained Object Annotations

- Detailed object-level attributes stored in annotation structures within JSON files

- Fine-grained classification of work zone elements including specific types of barriers, cones, and equipment
- Image-level attributes stored in image structures for contextual understanding

4.3 Level (c): TTC Sign Recognition

- 360 types of TTC signs annotated with their specific meanings and placements
- Sign text interpretation capabilities for understanding temporary traffic instructions
- Sign graphics analysis for recognizing visual patterns and symbols

4.4 Level (d): Pathway and Navigation Annotations

- Automatically computed pathways showing navigable routes through work zones
- Geo-localization data for precise positioning of work zone elements
- Navigation guidance annotations for understanding how to safely traverse work zones

3.1.2 CODA Dataset - A Corner Case Dataset with Roadwork Zone Relevance

The CODA (Corner Case Dataset for Autonomous Driving) [2] represents a significant step forward in autonomous driving research, specifically targeting the "long tail" of driving scenarios. Unlike standard datasets that focus on routine traffic, CODA is designed to expose the limitations of object detection algorithms by curating rare, high-risk, and irregular scenes.

While not exclusively dedicated to construction, CODA is a vital resource for roadwork analysis because construction zones are inherently corner cases. They break standard traffic rules and introduce unusual objects. As illustrated in the provided image, CODA captures these complexities with rich annotations:

- **Construction Machinery:** Heavy equipment like excavators and loaders (seen in the left-hand panels) are annotated as distinct vehicle classes.
- **Traffic Facilities & Obstructions:** The dataset distinguishes between static barriers (purple boxes) and temporary obstructions like traffic cones (orange/cyan boxes).
- **Vulnerable Road Users:** It includes critical miscellaneous classes, such as the wheelchair user shown in the bottom-right panel, ensuring models can detect unexpected pedestrians near work zones.

In Figure 2, four urban scenes illustrating the dataset's focus on detecting rare and irregular objects. The annotated bounding boxes highlight standard roadwork elements—such as **excavators and barriers** (top-left, bottom-left)—alongside critical "long-tail" anomalies like **wheelchairs** (bottom-right) and scattered **traffic cones** (top-right), demonstrating the dataset's relevance for complex construction zone navigation.

To ensure high-quality and diverse data, CODA was constructed by mining approximately 1,500 distinct corner cases from major benchmarks like KITTI, nuScenes, and ONCE. This rigorous curation process filters out the mundane, leaving only the most challenging scenarios to help researchers improve the safety and robustness of autonomous systems in unpredictable environments.

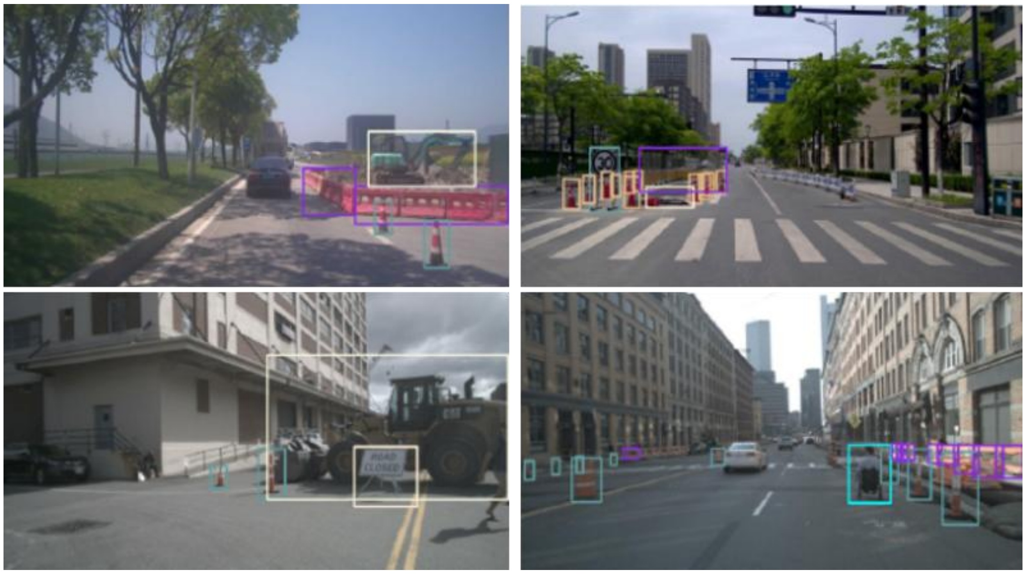

Figure 2 Representative Corner Cases in the CODA Dataset [2].

3.1.2.1 Detailed Dataset Statistics

CODA utilizes multimodal sensor data including:

- Camera: 2D RGB images for visual perception
- LiDAR: 3D point clouds for spatial understanding

Dataset Scale and Composition

- Total Scenes: 1,500 carefully selected real-world driving scenes
- Total Frames: Approximately 10,000 road driving scenes
- Annotated Objects: 80,180 annotated objects spanning 43 object categories
- Corner Cases: Nearly 6,000 high-quality annotated road corner cases
- Average Cases per Scene: 4 object-level corner cases per scene (on average)

CODA provided annotations are mainly **2D bounding boxes** with 43 object categories. Annotation can be grouped into 7 super-categories including *pedestrian*, *cyclist*, *vehicle*, *animal*, *traffic facility*, *obstruction* and *misc*, which can be further divided into fine-grained categories.

- **Pedestrian**: pedestrian
- **Cyclist**: Cyclist, Tricycle
- **Vehicle**: Car, Truck, Bus, Motorcycle, Construction vehicle, Recreational Vehicle, Moped, Trailer
- **Animal**: Dog, animal
- **Traffic Facility:** Traffic Light, Traffic Sign, Traffic Cone, Traffic Island, Traffic Box, Sentry Box, Warning Sign
- **Obstruction**: Barrier, Bollard, Concrete Block, Debris, Dustbin
- **Misc:** Basket, Chair Machinery, Misc, Phone Booth, Stroller, Suitcase, Wheelchair, Cart

3.1.3 Waymo Open Dataset: Geometric Fidelity in Altered Zones

The Waymo Open Dataset [3–6] serves as a premier benchmark for geometric perception and autonomous motion planning. Updated substantially through 2024 and 2025, the dataset leverages its massive collection volume in dense urban centers like San Francisco and Phoenix to capture critical long-tail events, including roadwork.

While not exclusively a construction dataset, Waymo's focus on complex urban driving naturally encompasses frequent roadwork scenarios. As illustrated in Figure 3, the dataset provides high-resolution sensor data for diverse construction topologies—ranging from active sites with workers and aggressive lane merges (top) to passive traffic control using static cones (bottom). These real-world examples are critical for training models to handle the "specialized scenarios" found in the End-to-End Driving component.

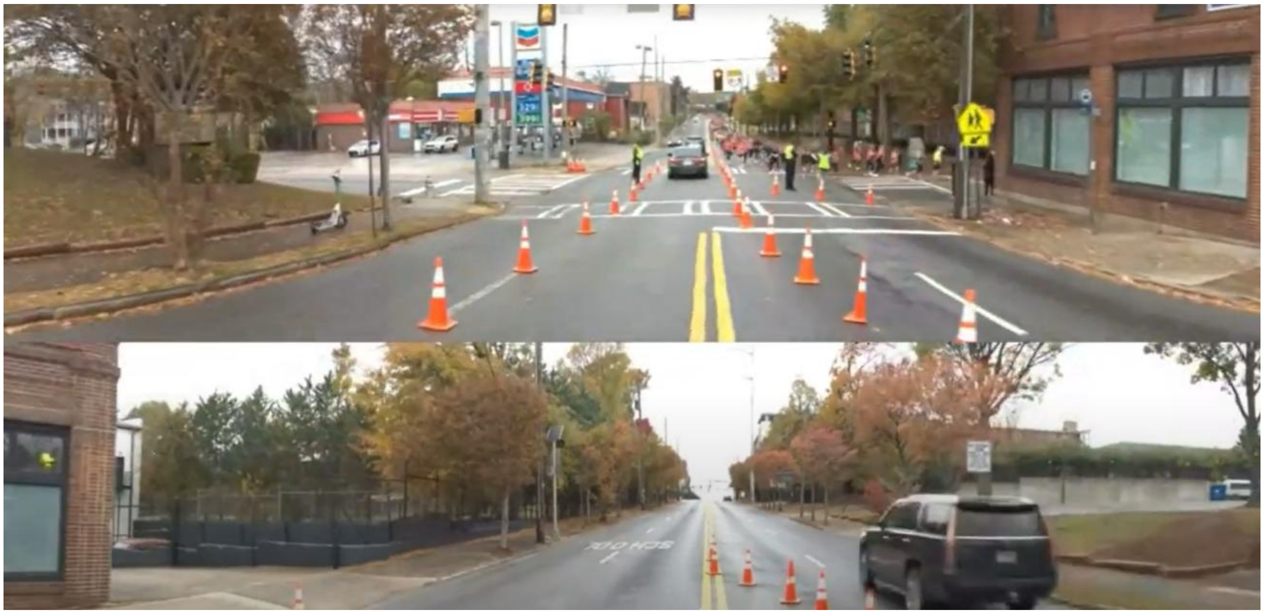

Figure 3 Construction Scenarios in the Waymo Open Dataset [5].

The Waymo Open Dataset comprises three major components:

- Perception Dataset: 2,030 driving segments with high-resolution sensor data and labels Motion [4, 6]
- Dataset: 103,354 segments containing object trajectories and 3D maps [3, 4]
- End-to-End Driving Dataset: Includes specialized scenarios like construction zones [5]

Key Metrics:

- 12.6 million 3D bounding box labels across 1,200 segments
- 172.6K object annotations with 2D key points for camera objects
- 10K object annotations with 3D key points for LiDAR objects
- 28 fine-grained object categories for detailed perception tasks

3.1.3.1 Annotation Schema and Sensor Modalities

The dataset provides labels for 4 main object classes: *Vehicles*, *Pedestrians*, *Cyclists*, and *Signs; and* 28 Fine-Grained Categories. Including: *Car, Bus, Truck, Other Large Vehicle, Trailer, Ego Vehicle, Motorcycle, Bicycle, Pedestrian, Cyclist*, plus additional specialized categories

Annotation Types

- **3D Bounding Boxes:** LiDAR-based annotations for spatial localization
- **2D Bounding Boxes:** Camera-based annotations for vehicles, pedestrians, and cyclists

- **Key Points:** 2D key points for camera objects, 3D key points for LiDAR objects
- **Tracking IDs:** Unique identifiers for object trajectories across frames

Sensor Modalities

The dataset features multi-modal sensor fusion with five synchronized sensor types:

- **LiDAR System:** 1 mid-range LiDAR (top-mounted), 4 short-range LiDARs (front, side left, side right, rear). Provides dense 3D point cloud data for geometric perception
- **Camera System:** 5 high-resolution RGB cameras covering 360° field of view, including front, side, and rear perspectives
- **RADAR:** For velocity and long-range detection
- **IMU & GPS:** For precise localization and motion tracking.

3.2 Summary and Future Outlook

This chapter has provided a detailed examination of the current state of ego-centric datasets used for work zone navigation. We explored how datasets like ROADWork provide the necessary benchmarks for identifying construction-specific semantics, while CODA plays a crucial role in exposing models to the "long tail" of rare, high-risk scenarios that define the irregularity of road work. Furthermore, we analyzed the Waymo Open Dataset, highlighting its utility in providing high-resolution geometric fidelity across complex, urban construction topologies in cities like San Francisco and Phoenix.

Collectively, these datasets enable significant strides in object detection, semantic segmentation, and path planning from the perspective of the vehicle itself. However, while the ego-centric view is well-documented, a holistic solution for work zone navigation remains incomplete. Our analysis reveals two significant voids in the current data landscape:

The Infrastructure and V2X Gap: Existing public datasets are almost exclusively built from the perspective of the ego-vehicle. There is a distinct lack of datasets capturing work zones from an infrastructure perspective (roadside units, surveillance cameras) or utilizing V2X (Vehicle-to-Everything) communication. To fully resolve occlusion issues and foresee hazards beyond the ego-vehicle's line of sight, the industry requires data that correlates on-board sensors with external infrastructure feeds.

The Simulation Deficit: We currently lack comprehensive simulation datasets specifically tailored for work zone environments. Because real-world construction zones are dangerous and ephemeral, collecting sufficient real-world failure data is risky. High-fidelity synthetic datasets—capable of procedurally generating endless variations of cone placements, worker movements, and weather conditions—are essential to bridge the gap between rare real-world events and robust model generalization.

Future research must pivot toward these multi-modal and synthetic data sources to move from simply detecting the work zone to mastering its entropy.

4. Public Accident Datasets for Autonomous Driving Safety Validation

The validation of AD systems has entered a critical phase where the primary challenge is no longer nominal operation, but the robust handling of safety-critical "corner cases." Among these, traffic accidents represent the most significant hurdle due to their statistical rarity, physical complexity, and severe consequences. This chapter presents an exhaustive analysis of public datasets containing car accidents, structured around three fundamental sensing domains: **Ego-Perspective** (onboard sensors), **Infrastructure Perspective** (roadside units/surveillance), and **Simulation Environments** (synthetic data).

The research identifies a paradigm shift in dataset composition between 2018 and 2025. Early benchmarks, such as the **Dashcam Accident Dataset (DAD) [7]** and **Car Crash Dataset (CCD) [8]**, relied primarily on monocular RGB video for temporal anomaly detection. In contrast, emerging datasets from 2024–2025, including **TUMTraf-A** and **MM-AU**, integrate multi-modal sensor fusion (LiDAR, Radar, Camera), rigorous standardization (OpenLABEL), and cognitive annotations (textual causal reasoning) to support higher-level safety tasks.

Furthermore, the analysis highlights the indispensable role of **Simulation Environments** like CARLA and NVIDIA STRIVE. Given the ethical and logistical impossibility of staging sufficient real-world crashes to achieve statistical significance, synthetic environments have evolved to provide "End-to-End" V2X simulation and adversarial scenario generation. This report evaluates over 15 major datasets, detailing their sensor modalities, annotation granularity, licensing, and utility for safety validation, ultimately arguing for a hybrid data strategy that leverages the strengths of all three domains to bridge the gap to Level 5 autonomy.

4.1 Ego-Centric Datasets: The Vehicle’s Perspective

Ego-perspective datasets are the most prevalent form of accident data, driven by the ubiquity of commercial dashcams and the relatively low cost of data collection. These datasets focus on equipping the ego-vehicle with the ability to anticipate accidents and understand their causes, forming the first line of defense in AD safety.

4.1.1 Dashcam Accident Dataset (DAD)

The Dashcam Accident Dataset (DAD)[7] stands as a seminal benchmark in the field of temporal anomaly detection. Sourced primarily from dashcam footage in Taiwan, it was one of the first large-scale attempts to curate diverse crash scenarios for computer vision research.

- Dataset Composition: DAD contains approximately 1,750 video clips, split between 620 accident sequences and 1,130 normal driving sequences. The videos are typically sampled into 100-frame sequences, standardizing the temporal input for machine learning models.

- Scenario Diversity: The dataset captures a wide array of crash modalities, including rear-end collisions, side impacts (T-bone), and vehicle-pedestrian accidents. The geographic specificity (Taiwan) introduces a unique traffic pattern characterized by high scooter density, offering a distinct challenge compared to Western-centric datasets.

- Annotation & Utility: DAD provides frame-level annotations that demarcate the onset of the accident. This granularity enables the training of models to predict "Time-to-Accident" (TTA), a critical metric for collision warning systems. More recent research has utilized DAD to train Vision-

Language Models (VLMs) like Video-LLaMA.8 By feeding the video sequence into a VLM, researchers aim to move beyond simple detection to a semantic understanding of the crash evolution (e.g., "The scooter swerved left into the path of the bus").

- Limitations: The primary limitation of DAD lies in its monocular nature. Without depth information (LiDAR) or precise calibration, estimating the 3D distance and velocity of objects is challenging, limiting its utility for Level 4/5 planning tasks.

Figure 4 shows geographic and scenario diversity in the Dashcam Accident Dataset (DAD). The visualization contrasts the structured, lane-compliant environments of the KITTI benchmark (left) with the chaotic, high-density traffic scenarios captured in DAD (right). The map highlights the dataset's collection across major Taiwanese cities (e.g., Taipei, Taichung, Kaohsiung), introducing distinct "scooter-dominant" traffic patterns. The callouts illustrate diverse collision modalities—such as "motorbike hits motorbike" and "car hits car" providing the unstructured accident data necessary for training anomaly detection models.

Figure 4 Geographic and Scenario Diversity in the Dashcam Accident Dataset (DAD) [7].

4.1.2 Car Crash Dataset (CCD)

Building upon the foundation of DAD, the Car Crash Dataset (CCD) [8] represents a significant leap in annotation richness and environmental context. While DAD focused on the presence of a crash, CCD introduces the context of the crash. CCD comprises 1,500 trimmed accident videos collected from YouTube and 3,000 normal driving videos sampled from the BDD100K dataset. Each video is standardized to 50 frames at 10 Hz.

CCD provides a comprehensive semantic breakdown of the event. Shown as Figure 5, the annotation schema (left and right panels) details critical context:

- Ego-Involvement: A critical binary label indicating whether the recording vehicle was involved in the collision or merely a witness. This distinction is vital for "ego-risk" assessment—an AV needs to react differently to a threat targeting itself versus a crash occurring in adjacent lanes.

- Environmental Attributes: Annotations include weather conditions (Snowy, Rainy, Clear) and lighting (Day, Night), allowing for the evaluation of perception of robustness under adverse conditions.
- Participant & Reason: The dataset includes descriptions of accident participants (e.g., SUV, Truck) and the reason for the accident, facilitating causal analysis.

Technical Implementation: To support efficient research, the CCD repository provides pre-extracted feature vectors using VGG-16 and Cascade R-CNN.12 This allows researchers to focus on the temporal modeling aspects (e.g., using LSTMs or Transformers) without the computational overhead of processing raw video frames.

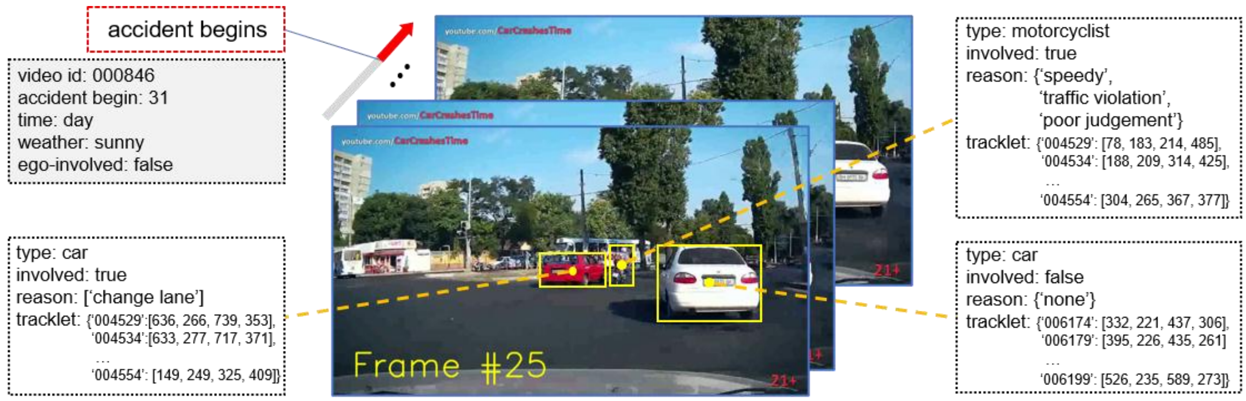

Figure 5 Rich Semantic Annotations in the Car Crash Dataset (CCD) [8].

4.1.3 AnAn Accident Detection (A3D)

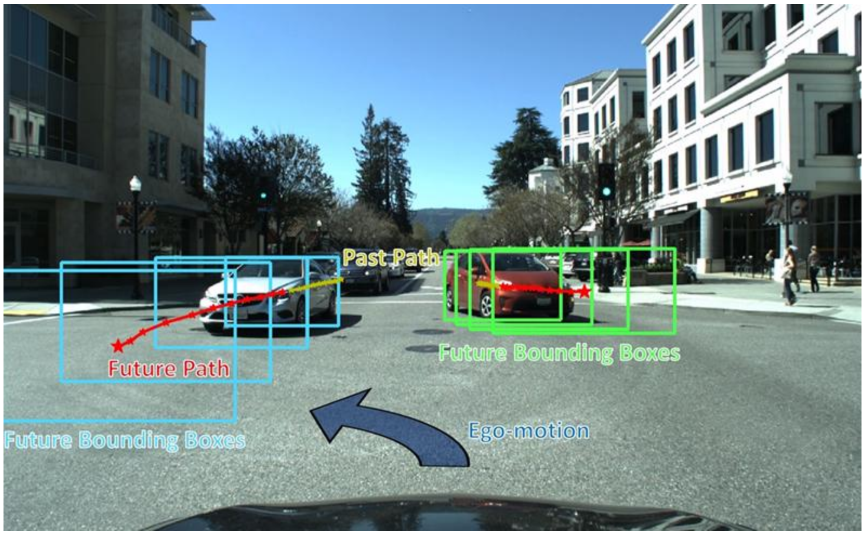

Figure 6 Trajectory Prediction as a Proxy for Safety in the AnAn Accident Detection (A3D) Framework [9].

While DAD and CCD rely on supervised learning—teaching the model to recognize labeled crash examples—the AnAn Accident Detection (A3D) dataset [9] introduces a novel methodological approach: Unsupervised Anomaly Detection. This shift is particularly relevant for work zones, which present an "open set" of potential hazards that are impossible to fully catalogue.

Figure 6 demonstrates the core mechanism of unsupervised anomaly detection. By analyzing the **Past Path** (yellow trace) and **Ego-motion**, the model projects **Future Bounding Boxes** (cyan and green wireframes) and a **Future Path** (red trace). In the A3D methodology, an accident is not detected by recognizing a specific "crash pattern," but is mathematically defined as a significant deviation from these predicted trajectories. This allows the system to flag sudden stops or erratic swerves without prior training on those specific accident types.

The "Deviation" Hypothesis: The core hypothesis of A3D research is that an accident can be mathematically defined as a deviation from a predicted trajectory. Instead of training a binary classifier on "crash" vs. "no-crash," A3D supports frameworks that learn the physics of normal traffic flow. The model predicts the future location of traffic participants based on past frames (as seen in Figure 6). When the actual future location deviates significantly from the prediction—for example, a car suddenly creating a collision angle or a worker stepping into a lane—an anomaly is flagged.

Dataset Content: The dataset provides approximately 1,500 clips of diverse traffic accidents captured from ego-view cameras, serving as a testbed for these prediction-based models.

This unsupervised approach is crucial for navigating the entropy of construction zones. Since it is impossible to train a model on every conceivable edge case (e.g., a falling sign, a rolling barrel, or a specific type of construction vehicle), models that learn "normality" (smooth flow) and detect deviations are theoretically more robust. They do not need to know what the object is to know that its behavior is dangerous.

4.1.4 Detection of Traffic Anomalies (DoTA)

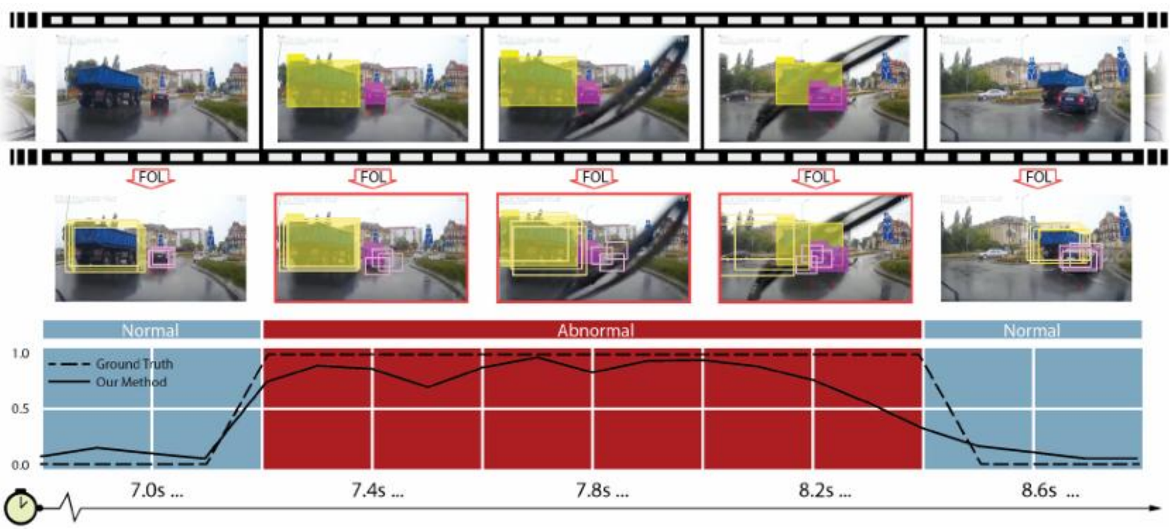

Figure 7 Overview of DoTA proposed method) using a sample video from DoTA dataset based on future object localization (FOL).

The DoTA [10] dataset 14 scales up the concept of anomaly detection, providing a broader and more diverse benchmark. Figure 7 demostrated an sample of annotated bounding boxes (filled) and predicted boxes are presented.

- Scale: DoTA contains 4,677 videos, significantly larger than DAD or A3D.
- Annotation Granularity: It provides temporal, spatial, and categorical annotations for anomalies. This allows models to answer "When" (temporal start/end), "Where" (spatial bounding box), and "What" (accident type) simultaneously.

- Evaluation Metric: DoTA introduced the Spatial-Temporal Area Under Curve (STAUC) metric, which evaluates detection performance across both space and time, penalizing models that correctly detect a crash time but fail to localize the crashing vehicle.

4.1.5 DADA-2000: Modeling Driver Attention

DADA-2000 (Driver Attention in Driving Accidents) [11] 16 addresses the human factor in accident avoidance. In many critical scenarios, the difference between a near-miss and a collision is where the driver was looking.

- Data Content: 2,000 video clips totaling over 658,000 frames.

- Unique Modality: The dataset includes Saliency Maps (eye-tracking data) collected from human observers viewing the accident clips. This data captures fixation points, saccade paths, and focusing time.

This allows for the training of "Attentive" neural networks. By mimicking human gaze patterns, AV perception systems can learn to prioritize processing resources on high-risk regions (e.g., an intersection corner) rather than irrelevant background details, potentially improving reaction times.

4.1.6 MM-AU: Multi-Modal Accident Understanding (2024)

Released in 2024, MM-AU [12] represents the state-of-the-art in "Abductive Accident Video Understanding." It bridges Computer Vision with Natural Language Processing.

- Scale: 11,727 in-the-wild ego-view videos, making it one of the largest datasets in this domain.

- Multi-Modal Annotations: Visual: Over 2.23 million annotated object bounding boxes.

- Textual: 58,650 pairs of Question-Answer covering accident reasons ("Why did the crash happen?"), prevention solutions ("How could it have been avoided?"), and categorization.

- Cognitive Tasks: MM-AU supports 8 distinct tasks, including Ego-view Accident Reason Answering and Causal Inference.

This dataset is pivotal for Explainable AI (XAI). An AV that makes an emergency stop must be able to justify its decision to passengers and regulators. MM-AU trains systems to output reasoning (e.g., "I stopped because the pedestrian on the right was obscured by a truck and entered the lane") rather than black-box probabilities.

4.1.7 Nexar Dashcam Collision Prediction Dataset

The Nexar dataset [13] introduces industrial-scale data to the academic research community. Sourced from a vast commercial network of connected dashcams, it captures rare, high-energy events. First column represents Before accident interval, columns 2-4 represent Alert interval, and fifth column represent After accident interval. Within the alert interval, column 2 represents the ”time-of-alert” which is the earliest moment that the driver could intervene to prevent the accident. column 4 represents the ”time-of-accident”, and column 3 represents an intermediate frame between “time-of-alert” and “time-of-accident”.

- Trigger Mechanism: Nexar's data is captured using IMU triggers. The recording is automatically saved when the device detects a g-force spike indicative of hard braking or impact. This ensures a high density of "true positive" crash events.
- Scale: 1,500 curated video clips (400 confirmed collisions, 350 near-collisions).
- Annotation Rigor: To establish ground truth for accident timing, the dataset employs a consensus mechanism where 10 human annotators label the "Time of Alert" and "Time of Impact." The final label is the median of these annotations, providing a robust baseline for human reaction time.

Figure 8 Nexar dataset samples from before accident interval, alert interval, and after accident interval.

4.1.8 Comparative Summary of Ego-Perspective Datasets

Table 1 Comparative Summary of Ego-Perspective Public Accident Datasets

Dataset	Year	Source	Modality	Scale	Primary Focus
DAD	2018	Dashcams (Taiwan)	Mono Video	1,750	Anomaly Detection, TTA
A3D	2019	Dashcams	Mono Video	1,500	Unsupervised Learning
DADA-2000	2019	Crowd-sourced	Video + Gaze	2,000	Driver Attention/Saliency
CCD	2020	YouTube/BDD	Mono Video	4,500	Contextual Attributes
DoTA	2022	YouTube	Mono Video	4,677	Spatio-Temporal Anomalies
MM-AU	2024	Web/Public	Video + Text	11,727	Causal Reasoning/VQA
Nexar	2025	Commercial Fleet	Video + IMU	1,500	Collision Prediction

4.2 Infrastructure Perspective Datasets: The "God's Eye" View

Infrastructure-based perception involves sensors mounted on roadside infrastructure (gantries, poles, traffic lights). This perspective addresses the fundamental physical limitation of ego-vehicles: occlusion. An infrastructure sensor can perceive a collision course between two vehicles (e.g., at a blind intersection) seconds before the vehicles can see each other. This domain is critical for Vehicle-to-Everything (V2X) safety systems.

4.2.1 TUM Traffic Accident Dataset (TUMTraf-A)

Figure 9 Visualization of the TUM Traffic Accident dataset with 3D box annotations, track IDs and trajectories.

The TUMTraf-A dataset [14] is widely regarded as the gold standard for infrastructure-based accident research. Collected at the A9 Digital Test Field in Munich, Germany, it provides high-fidelity, sensor-fused data of real-world highway accidents. In Figure 9, The left figure illustrates a vehicle is in the process of overturning following collision. The right figure shows a vehicle has pulled over after catching fire following a collision.

- **Sensor Suite:** The infrastructure setup is exceptionally comprehensive, featuring:
- **Cameras:** 4x Basler ace (1920x1200 resolution @ 50Hz).
- **LiDAR:** 1x Valeo SCALA B2 (16 layers, 200m range @ 25Hz).
- **Radars:** Integrated into the fusion pipeline for velocity estimation.
- **Scenario Content:** The dataset captures 10 sequences of severe real-world accidents, including high-speed crashes, vehicle rollovers, and vehicle fires. While the number of sequences is low, the data fidelity is unmatched.
- **Annotation Standards:**
- **3D Labels:** Unlike ego-datasets that rely on 2D boxes, TUMTraf-A provides 2.6 million labeled 3D bounding boxes and instance segmentation masks.
- **OpenLABEL:** The dataset utilizes the OpenLABEL standard, ensuring interoperability with modern MLOps pipelines and standardized safety evaluation tools.

TUMTraf-A enables the creation of "Digital Twins" for highway safety. By fusing camera and LiDAR data, researchers can reconstruct the precise 3D trajectory and energy of a crash, training RSU-based systems to detect accidents and alert incoming traffic via V2X messages.

4.2.2 TU-DAT (Temple University Dataset)

Released in mid-2025, TU-DAT [15] focuses on urban anomalies and aggressive driving behaviors using a hybrid collection methodology.

- **Data Sources:** It combines real-world footage from roadside CCTV cameras with high-fidelity synthetic data generated using the BeamNG.drive physics engine.

- **Behavioral Focus:** The dataset specifically targets aggressive behaviors that often precede accidents, such as tailgating, weaving, and speeding.
- **Cognitive Annotations:** TU-DAT includes spatiotemporal annotations designed to support logic-based reasoning frameworks. This aligns with the trend toward "neuro-symbolic" AI, where deep learning detects the object, and symbolic logic reasons about the rule violation (e.g., "Vehicle A is within stopping distance of Vehicle B at speed X").
- **Access:** The dataset is freely available for research purposes, addressing the scarcity of public CCTV data due to privacy regulations.

Figure 10 Some frames of TU-DAT dataset of accident scenarios.

4.2.3 CADP (CCTV Accident Dataset)

CADP [16] provides a dedicated benchmark for accident forecasting from surveillance cameras.

- **Scope:** 1,416 video segments collected from CCTV footage, with a focus on urban intersections and traffic monitoring.

- **Methodological Innovation:** The researchers introduced "Context Mining" techniques to improve the detection of small objects (e.g., pedestrians) that are often pixel-limited in wide-angle CCTV views.

- **Forecasting Metrics:** CADP establishes a benchmark for accident prediction horizons, achieving an average Time-to-Accident (TTA) prediction of ~1.68 seconds. This timeframe is critical for activating intersection safety systems (e.g., holding a red light).

4.2.4 Simulation and Synthetic Environments: Generating the Impossible

The statistical validation of AD safety faces a "rarity paradox": real-world accidents are too rare to capture in sufficient quantity for training, yet too dangerous to ignore. Simulation environments bridge this gap by allowing for the ethical, scalable, and controllable generation of crash scenarios.

4.2.5 DeepAccident: End-to-End V2X Simulation

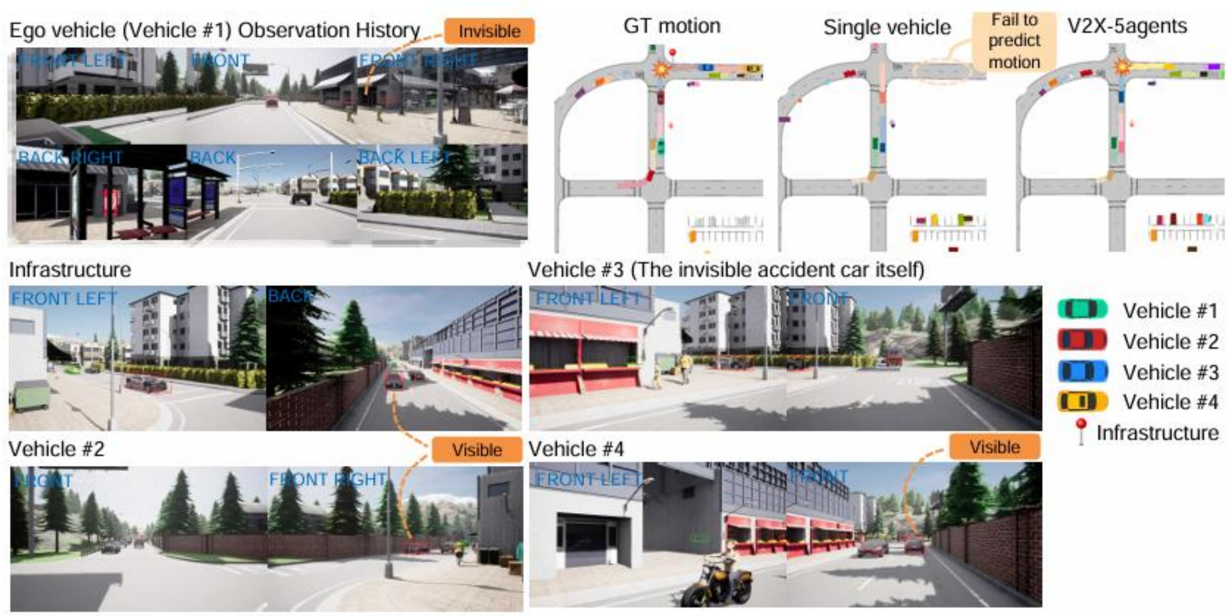

Figure 11 Illustration of the proposed end-to-end motion and accident prediction task.

DeepAccident [17] is a comprehensive synthetic benchmark generated using the **CARLA** simulator. It is explicitly designed to support V2X research by providing a holistic view of crash scenarios. In Figure 11, given the history camera observations, the single vehicle model (vehicle # 1) fails to predict any motion or accident on the forward right side due to occlusion from buildings. In contrast, the V2X model communicates with other vehicles and infrastructure, thereby successfully anticipating the upcoming accident. The red and green bounding boxes in the images, respectively, represent the colliding vehicles and the other V2X vehicles behind them.

- **Scenario Design:** For every simulated accident, DeepAccident records data from five distinct agents:

- **Vehicles 1 & 2:** The colliding vehicles.

- **Vehicles 3 & 4:** "Witness" vehicles following the crash.
- **Infrastructure:** A roadside RSU observing the scene.
- **Scale:** The dataset contains 57,000 annotated frames and 285,000 samples.
- **New Tasks:** It introduces the task of "End-to-End Motion and Accident Prediction," evaluating how well a model can predict the future trajectories of all agents and the probability of collision.
- **Key Insight:** By providing perfectly synchronized multi-view data, DeepAccident allows researchers to quantify the "V2X Gain"—the specific improvement in safety margins achieved by adding infrastructure sensors compared to single-vehicle perception.[31]

4.2.6 NVIDIA STRIVE: Adversarial Scenario Generation

STRIVE (Stress-Test Drive) [18] represents a shift from static datasets to dynamic **adversarial generation**.

- **Concept:** Instead of manually scripting a crash, STRIVE uses a generative model to perturb the trajectories of background traffic agents in a realistic manner. It essentially "attacks" the AV planner, searching for specific combinations of behaviors (e.g., a cut-in followed by hard braking) that cause the planner to fail.
- **Realism:** Unlike naive attacks that might spawn a brick wall in front of the car, STRIVE optimizes adversarial noise within the latent space of a traffic model, ensuring that the generated scenarios remain physically and behaviorally plausible.
- **Utility:** STRIVE functions as a continuous validation tool, actively discovering new failure modes in the planner logic.

4.2.7 DISC: Human Behavior in Mixed Autonomy

Figure 12 NVIDIA STRIVE: Scenarios from the driver's point of view.

DISC (Dataset for Analyzing Driving Styles in Simulated Crashes) [19] addresses the "Human Element" in simulation. Figure 12 shows the scenarios in the dataset. Specifically, the left image depicts the deer-crossing scenario, where the participant encounters a running deer while driving on a country road. The middle image shows the running red lights scenario, where a participant encounters a car at an intersection that crosses the road after ignoring the red traffic signal. The right image displays a jaywalking pedestrian from the driver's point of view.

- **Platform:** Collected using a VR-based driving simulator (TRAVERSE).
- **Focus:** It analyzes how different human driving styles (e.g., aggressive, timid, distracted) react to imminent crash scenarios.

- **Data Richness:** In addition to vehicle kinematics (steering, braking), DISC captures **eye-tracking** data.

- **Application:** This data is critical for "Mixed Autonomy" environments. AVs must be able to predict the irrational or panicked reactions of human drivers. DISC provides the ground truth for modeling these non-optimal human behaviors.

4.2.8 SynSHRP2

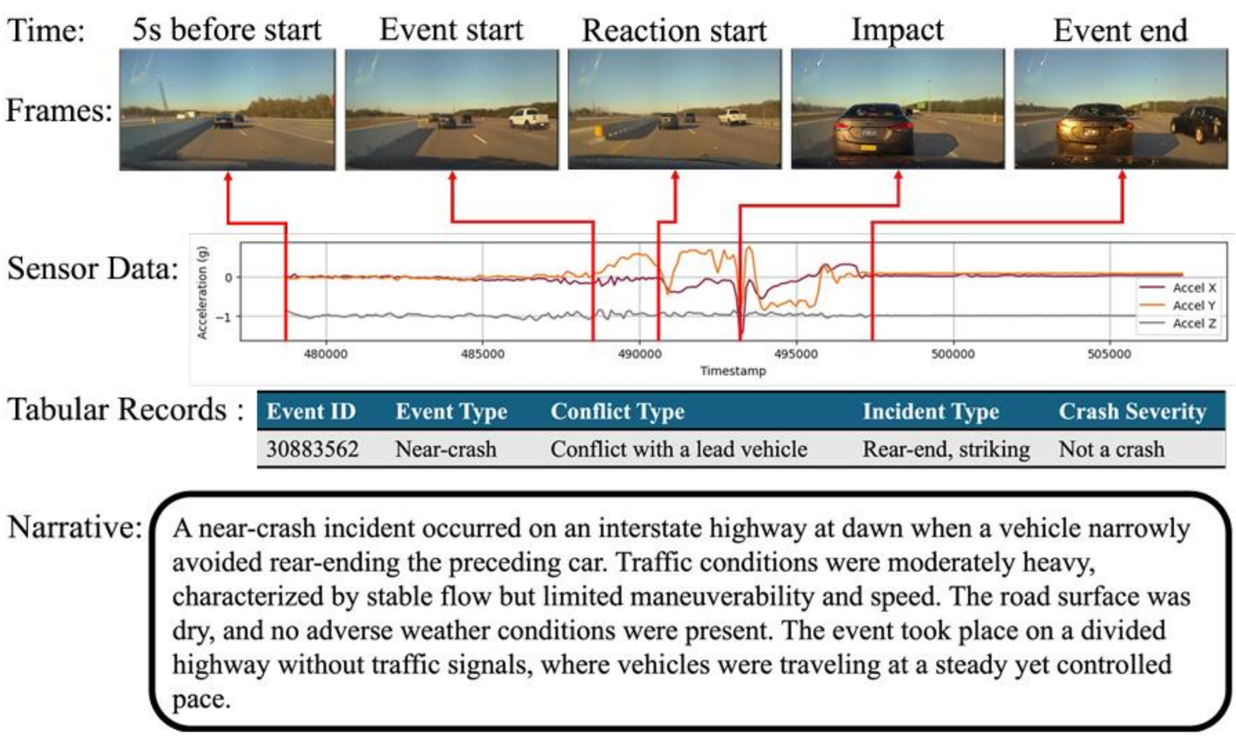

Figure 13 Example illustrating all data types in SynSHRP2.

SynSHRP2 [20] utilizes Generative AI to unlock valuable naturalistic data. The original SHRP2 study captured thousands of real crashes, but the data is restricted due to privacy concerns (faces, license plates). Figure 13 shows four components of the dataset: tabular records, sensor data, keyframe images and corresponding times tamps, and comprehensive narrative descriptions of events.

- **Methodology:** Researchers used diffusion models (Stable Diffusion, ControlNet) to "re-synthesize" the SHRP2 keyframes. This process generates new, synthetic images that preserve the semantic and kinematic truth of the accident (vehicle positions, damage, lighting) while completely hallucinating the specific identity details (faces, plates).

- **Outcome:** A fully anonymized, public dataset that retains the statistical properties of real-world crashes without legal encumbrance.

4.2.9 Simulation Platforms

The generation of these datasets relies on sophisticated simulation engines:

- **CARLA:** The open-source standard for academic research, supporting flexible sensor configurations and Python APIs.

- **BeamNG.drive:** A soft-body physics engine used (e.g., in TU-DAT) to simulate realistic vehicle damage and deformation, which rigid-body simulators often miss.
- **PreScan:** A physics-based simulation platform used heavily in industry for sensor validation (e.g., modeling radar reflections or camera lens flares).

4.3 Conclusion

The landscape of autonomous driving accident datasets has matured significantly, evolving from simple collections of dashcam videos to sophisticated, multi-modal benchmarks that span the ego, infrastructure, and simulation domains.

For Ego-Perception: MM-AU and Nexar represent the cutting edge, pushing the boundaries of causal understanding and industrial-scale validation.

For Infrastructure: TUMTraf-A provides the high-fidelity 3D ground truth necessary to validate V2X safety concepts, offering a blueprint for future smart highways.

For Simulation: DeepAccident and STRIVE demonstrate that synthetic data is not merely a substitute for real data, but a superior tool for generating the multi-view, adversarial scenarios required to rigorously stress-test autonomous planners.

Researchers and engineers aiming to build holistic safety systems must adopt a hybrid data strategy: utilizing TUMTraf-A for precise 3D perception training, MM-AU for developing high-level causal logic, and DeepAccident/STRIVE for validating the planner against millions of synthetic edge cases. Only through the convergence of these three perspectives can the industry bridge the gap between "demonstration-level" autonomy and "safety-critical" deployment.

Table 2 Feature Comparison of Key Datasets of Public Accident Datasets

Dataset	Year	Modality	With IMU	Primary Focus
DAD	Ego (Real)	Mono Video	No	Anomaly Detection, TTA
CCD	Ego (Real)	Mono Video	No	Ego-Risk, Environment
MM-AU	Ego (Real)	Video + Text	No	Causal Reasoning, VQA
Nexar	Ego (Real)	Video + IMU	No	Large-Scale Prediction
TUMTraf-A	Infra (Real)	Camera + LiDAR	**Yes**	3D Detection, Fusion
DeepAccident	Sim (V2X)	Cam + LiDAR	**Yes**	V2X Motion Prediction
DISC	Sim (VR)	Kinematics + Eye	No	Human Behavior

5. Public Datasets for Autonomous Navigation in Tunnel and Urban Canyon Environments

The robust operation of Autonomous Navigation Systems (ANS) is predicated on the availability of high-fidelity, diverse, and corner-case-rich datasets. While the last decade has seen a proliferation of datasets covering open highways and standard urban grids, the specific Operational Design Domains (ODD) of **tunnels** and **urban canyons** remain critically under-represented. These environments constitute "denied" zones—areas where the primary localization anchor, the Global Navigation Satellite System (GNSS), is either severely degraded by multipath effects or entirely absent. Furthermore, these environments present unique perception challenges, including extreme luminance dynamic ranges at tunnel portals, repetitive geometric structures that confuse Lidar-based odometry, and severe occlusion in narrow urban corridors.

An "urban canyon" is defined as a street flanked by tall buildings that significantly restrict the Sky View Factor (SVF). In cities like Hong Kong, New York, or Tokyo, this results in two primary failure modes for GNSS:

- **Non-Line-of-Sight (NLOS) Reception:** The satellite signal is blocked directly but reflected off a glass skyscraper. The receiver locks onto the reflection, which has traveled a longer path, resulting in pseudorange errors that can shift the calculated position by 50 to 100 meters.
- **Reduced Satellite Geometry:** The "dilution of precision" (DOP) increases as satellites are only visible in a narrow strip of sky directly overhead.

Generic datasets often treat tunnels as short, anomalous segments to be discarded or flagged. Specialized datasets, however, treat them as the primary domain. They employ specialized hardware—such as Ring Laser Gyroscopes (RLG) or Thermal Cameras—to establish ground truth and maintain perception where standard sensors fail. The following sections detail these datasets, categorized by their operational perspective.

5.1 Ego-Perspective Datasets: The Quest for Localization and Mapping

The ego-perspective—data collected from the autonomous vehicle itself—is the traditional bedrock of autonomous driving research. In tunnel and urban canyon scenarios, the primary value of a dataset lies in its ability to support **Long-Term Precision Localization** without GNSS. The "Gold Standard" for these datasets is defined by the rigor of their Ground Truth (GT) generation.

5.1.1 The Odyssey Dataset: Benchmarking Lidar-Inertial Odometry

Released in late 2025, the **Odyssey** dataset [21] represents a watershed moment for GNSS-denied navigation research. Figure 14 demostrate a subset of the variety of environments included in the Odyssey dataset: Standard urban, suburban and rural environments as well as long-term GNSS-denied situations such as indoor parking garages and tunnels.

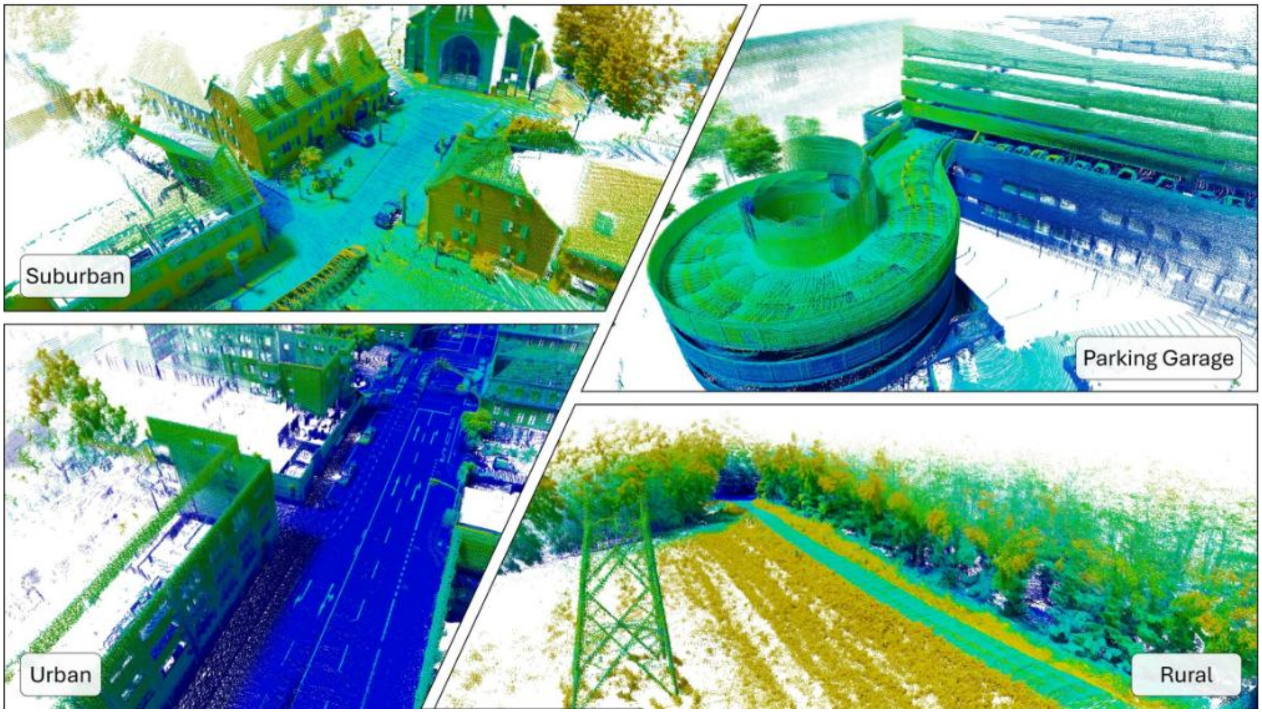

Figure 14 Example subset of the variety of environments included in the Odyssey dataset.

5.1.1.1 Motivation and Ground Truth Innovation

Most existing datasets (e.g., KITTI, newer College) rely on GNSS/INS systems to generate ground truth trajectories. In tunnels, these systems rely on MEMS-based IMUs to "coast" between GNSS fixes. However, MEMS IMUs suffer from bias instability (drift) over time. If a tunnel is 3 km long, a MEMS-based ground truth will drift significantly, making it impossible to accurately benchmark high-performance SLAM algorithms—effectively, the "ruler" becomes less accurate than the "measure."

Odyssey solves this by utilizing a navigation-grade **Inertial Navigation System (INS)** equipped with a **Ring Laser Gyroscope (RLG)**.

- **Mechanism:** RLGs utilize the Sagnac effect, where counter-propagating laser beams in a ring cavity allow for the measurement of rotation rates with extreme precision, independent of mechanical vibration or temperature fluctuations.
- **Impact:** This setup allows Odyssey to provide sub-meter accurate ground truth trajectories even after kilometers of driving in tunnels and underground parking garages without a single GPS satellite lock.

5.1.1.2 Scenario Composition

The dataset is explicitly designed to stress-test Lidar-Inertial Odometry (LIO) systems. It features:

- **Long-Duration GNSS-Denied Zones:** Tunnels and multi-story underground parking structures where the vehicle drives for extended periods.

- **Dynamic Characteristics:** Unlike slow-moving robot datasets, Odyssey includes automotive-grade speeds and dynamics, including "stop-and-go" traffic within tunnels—a notoriously difficult scenario for IMU bias estimation due to the lack of excitation.

- **Sensors:** It includes a high-density Ouster OS1-128 LiDAR, providing dense geometry capture essential for loop closure detection in feature-poor tunnel environments.

Table 3 Statistic of Odyssey Dataset

Feature	Odyssey Dataset Specification
Platform	Automotive (Passenger Vehicle)
Primary LiDAR	Ouster OS1-128 (128 beams)
Ground Truth	NovAtel INS with **Ring Laser Gyroscope** (RLG)
Key Scenarios	Tunnels, Underground Parking, Urban Canyons, Bumpy Roads
Data Volume	7,000+ frames (annotated subset), multiple long sequences

5.1.2 UrbanLoco: The Urban Canyon Benchmark

While Odyssey focuses on the total denial of GNSS (tunnels), **UrbanLoco [22]** focuses on the degradation of GNSS in **Urban Canyons**.

5.1.2.1 Geographic Uniqueness

Collected in the hyper-dense urban environments of **Hong Kong** and **San Francisco**, UrbanLoco captures the verticality of modern megacities. Hong Kong poses a unique challenge: narrow streets flanked by 50+ story buildings create a "deep urban canyon" effect where the sky view is a mere slit. The UrbanLoco dataset focuses on highly urbanized areas in San Francisco and Hong Kong with a full sensor suite: 360 degree camera (San Francisco), fish eye sky camera (Hong Kong), LIDAR, GNSS receivers and IMU. The dataset covers various road conditions including tunnels, urban canyons, construction sites, sharp maneuvers, hills, etc. (see Figure 15)

Figure 15 An Overview of the UrbanLoco Dataset.

5.1.2.2 The Tunnel-Canyon Transition

One of the most critical aspects of UrbanLoco is its capture of **transition zones**.

- **Scenario:** A vehicle exits a deep urban canyon (high multipath error) and immediately enters a cross-harbor tunnel (zero signal).
- **Research Value:** This forces localization algorithms to manage the "handover" between GNSS-aided fusion and pure dead-reckoning. Algorithms must detect the degradation of the GNSS signal quality *before* it disappears entirely to prevent injecting errors into the state estimator just before entering the tunnel.

5.1.2.3 Sensor Suite and Synchronization

The dataset utilizes a full sensor suite tailored for "tightly coupled" fusion:

- **LiDAR:** RoboSense RS-LiDAR-32 (providing 360-degree coverage).
- **Camera:** 6x cameras providing a panoramic view, essential for visual-inertial odometry (VIO) which can complement LiDAR in geometrically repetitive canyons.
- **Ground Truth:** Generated using a high-end NovAtel SPAN-CPT system, post-processed with local base station data to mitigate multipath effects as much as possible.

5.1.3 MSD-VMMS-HK: Mobile Mapping in Tunnels

The **MSD-VMMS-HK** dataset [23] approaches the problem from a **Mobile Mapping System (MMS)** perspective. While similar to autonomous driving datasets, its primary goal is the creation of static High-Definition (HD) Maps.

Tunnel Mapping Methodology: Constructing a map of a tunnel requires extreme precision. This dataset employs a **Factor Graph Optimization (FGO)** approach. The trajectory is "anchored" at the tunnel entrance and exit where GNSS is valid. Inside the tunnel, high-precision LiDAR odometry (LO) constraints are added to the graph. The final optimization distributes the drift error across the tunnel length, resulting in a centimeter-accurate map.

- **Utility:** This dataset is invaluable for "Map-Based Localization" research. Autonomous vehicles often navigate tunnels not by building a map on the fly, but by matching their live LiDAR scans to a pre-built map. MSD-VMMS-HK provides both the raw data to build the map and the resulting point cloud to test localization matching.

5.2 Simulation and Synthetic Environments: Generating the Impossible

The statistical validation of AD safety faces a "rarity paradox": real-world accidents are too rare to capture in sufficient quantity for training, yet too dangerous to ignore. Simulation environments bridge this gap by allowing for the ethical, scalable, and controllable generation of crash scenarios.

5.2.1 SHIFT: Continuous Domain Adaptation

The SHIFT Dataset [24] introduces a large-scale synthetic dataset designed to address a critical gap in autonomous driving research: the ability of perception models to adapt to continuously changing environmental conditions (domain shifts).

Continuous vs. Discrete: Most datasets have separate folders for "Day" and "Night." SHIFT simulates the process of driving into a tunnel. This continuous stream allows for the evaluation of Online Domain Adaptation algorithms, which must adjust their internal feature representations in real-time as the vehicle crosses the tunnel threshold.

Dataset Statistics

- The dataset is notable for its scale and the complexity of its sensor suite.
- **Total Size:** 4,800+ video sequences.
- **Environments:** 8 different locations/towns.
- **Capture Frequency:** 10 Hz.

Shift Types (see Figure 16)**:**

- **Discrete Shifts:** Fixed conditions per sequence (e.g., specific weather or time).
- **Continuous Shifts:** Dynamic changes within a sequence (e.g., a sunny day slowly turning into a heavy storm, or day transitioning to night).

Perception Tasks: Supports 13 distinct tasks, including:

- 2D & 3D Object Detection
- Semantic & Instance Segmentation
- Monocular & Stereo Depth Estimation
- Optical Flow
- Object Tracking (2D & 3D MOT)

Sensor Suite (11 Sensors total):

- **RGB Cameras:** 5 cameras (Front, Left, Right, etc.) at 1280 × 800 resolution (90° FOV).
- **Stereo Cameras:** 1 stereo pair.
- **LiDAR:** 128-channel sensor.
- **Other:** Depth camera, Optical flow sensor, GNSS, and IMU.

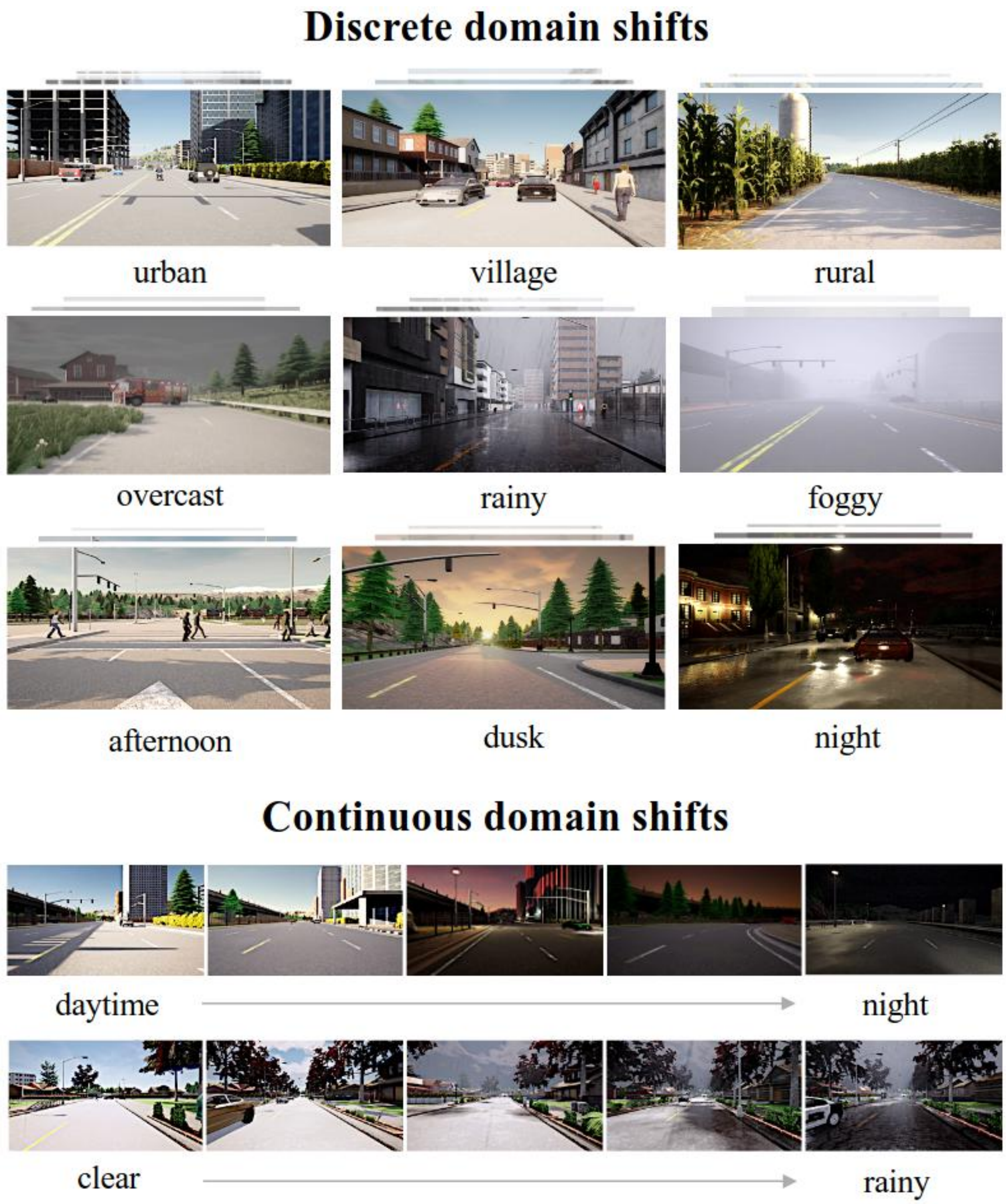

Figure 16 The Spectrum of Environmental Entropy in the SHIFT Dataset.

5.2.2 Truck V2X

TruckV2X [25] is the a large-scale, cooperative perception dataset specifically designed for autonomous heavy trucks. TruckV2X addresses the unique challenges of trucking, such as dynamic trailer movements, huge blind spots, and self-occlusion (where the trailer blocks the tractor's view during turns).

The dataset is generated using the CARLA simulator and Unreal Engine to create high-fidelity scenarios. It focuses on V2X (Vehicle-to-Everything) cooperation, enabling the truck to "see" through occlusions by utilizing data from nearby CAVs and RSUs. Shown in Figure 17, the top row shows RGB images from agent perspectives; the middle row displays BEV projections of LiDAR point clouds; the bottom row highlights occluded areas in close-up views, where dashed ellipses annotate critical blind zones for each agent (trailer obstructs tractor and truck obstructs others). This scenario presents the agent sequence from left to right astractor, trailer, CAV, and RSU.

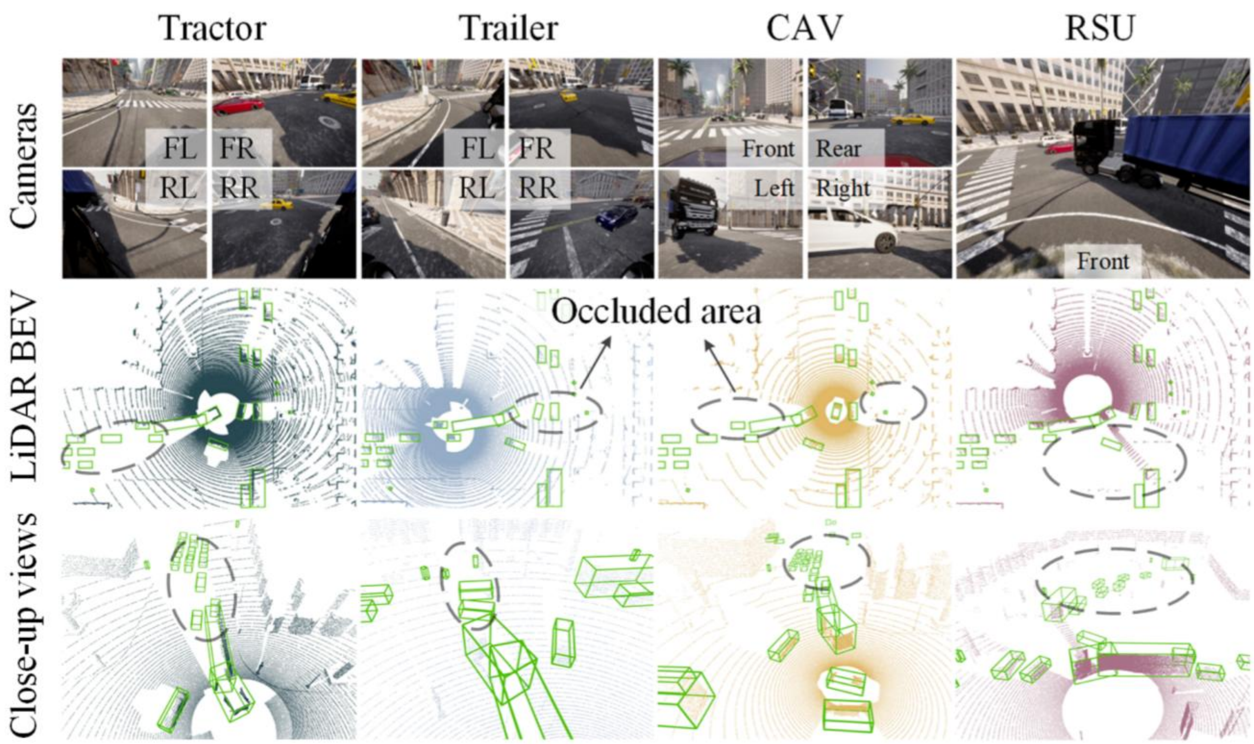

Figure 17 T-intersection scenarios in Truck V2X.

5.2.2.1 Dataset Statistics

The dataset is large-scale and multi-modal, covering diverse urban and highway scenarios.

Metric	Count / Description
Total Scenarios	**64** diverse scenarios
LiDAR Frames	**88,396** point clouds
Camera Images	**1,000,000+** (1 Million) images
3D Annotations	**1.18 Million** 3D bounding boxes
Agents per Scene	Tractor, Trailer, CAV, RSU
Sensor Modalities	LiDAR, RGB Camera, IMU
Data Splits	Train, Validation, Test

Sensor Configuration

- The sensor suite is extensive to cover the truck's massive blind spots:
- Tractor: 2x 64-channel LiDARs, 5x Surround-view Cameras.
- Trailer: 2x 64-channel LiDARs, 5x Surround-view Cameras (crucial for rear/side view during turns).
- CAV: 1x LiDAR, 4x Cameras.
- RSU: LiDAR and Camera coverage of the intersection/road.

Annotation Format

- 3D Bounding Boxes: Location (x, y, z), Dimensions (l, w, h), Rotation (yaw).
- Object ID: Unique tracking ID for traffic participants.

- Attributes: Speed, class label (e.g., vehicle, pedestrian).
- Metadata: Timestamp, ego-vehicle pose.

5.3 Synthesis and Comparative Analysis

The following table summarizes the key public datasets available for Tunnel and Urban Canyon autonomy, categorized by their primary utility.

Table 4 Summary the key public datasets available for Tunnel and Urban Canyon autonomy

Dataset	Perspective	Sensor Modality	Key Environment / Scenario	Primary Research Application
Odyssey	Ego (Vehicle)	**RLG INS**, LiDAR, Camera	**Tunnels**, Parking Garages, Urban Canyons	**LIO Benchmarking**, Drift Analysis in GNSS-Denied Zones
UrbanLoco	Ego (Vehicle)	LiDAR, Camera, GNSS/INS	**Urban Canyons** (HK/SF), Tunnel Transitions	Multipath Mitigation, Tightly-coupled Fusion
RCooper	Infrastructure	LiDAR, Camera	**Roadside Corridors**, Intersections	**Cooperative Perception**, Multi-Target Tracking
MSD-VMMS-HK	Ego (Mobile Mapping)	MLS (LiDAR+Img)	**Mountain/Harbor Tunnels** (HK)	**HD Map Construction**, Factor Graph Optimization
SHIFT	Ego (Vehicle)	LiDAR, Camera, GNSS/INS	**CARLA**	**Perception models** continuously changing environmental conditions
Truck V2X	V2X	LiDAR, Camera, GNSS/INS	**CARLA**	**cooperative perception** for autonomous heavy trucks

5.3.1 Gap Analysis and Future Directions

Despite the richness of recent releases, several gaps remain:

- **High-Speed Tunnel Dynamics:** Most tunnel datasets (Odyssey, UrbanLoco) are collected at moderate urban speeds. There is a lack of open data capturing high-speed (>100 km/h) dynamics in highway tunnels, where aerodynamic effects and vibration significantly impact sensor stability.

- **Acoustic Data:** Tunnels have unique acoustic signatures. "Audio for Autonomy" is an emerging field (detecting sirens or crashes by sound), yet no major tunnel dataset currently includes synchronized audio streams.

End-to-End (E2E) in Tunnels: E2E driving models require massive data. While nuScenes and Waymo support E2E, the specific "corner cases" of tunnel driving (e.g., changing lanes inside a tunnel based on variable speed limits) are not densely represented enough to train robust E2E policies without simulation augmentation.

6. Public Datasets for General Highway Navigation

In this chapter, an exhaustive analysis of publicly available autonomous driving datasets that feature highway scenarios is provided. The analysis synthesizes data from over 200 datasets, with a specific focus on recent releases from 2024–2026, including the Zenseact Open Dataset (ZOD), DAIR-V2X, and the newly released PAVE and AGC-Drive datasets. Key findings indicate a paradigmatic shift toward multi-modal sensor fusion (LiDAR, Camera, Radar, and UAVs) and the integration of "End-to-End" (E2E) learning frameworks that utilize raw sensor inputs for direct trajectory planning. Furthermore, the emergence of datasets specifically targeting adverse weather on highways and "black-box" autonomous mode evaluation signals a maturation of the field from simple object detection to complex behavioral safety assessment.

6.1 Ego-Centric Datasets: The Vehicle's Perspective

Ego-centric datasets form the backbone of autonomous driving research. They simulate the exact inputs an autonomous vehicle (AV) receives during operation. For highway scenarios, the value of an ego-centric dataset is determined by the diversity of the environments (weather, lighting), the range of the sensors, and the volume of high-speed interaction data (merges, cut-ins, overtakes).

6.1.1 Zenseact Open Dataset (ZOD)

The Zenseact Open Dataset (ZOD) [26] represents a significant leap forward in dataset fidelity, specifically targeting the "short-range bias" prevalent in earlier academic datasets like nuScenes or KITTI. Collected over a two-year period across European road networks, ZOD focuses extensively on highway segments in Sweden, covering a geographical area nine times larger than comparable datasets.

Sensor Fidelity and Range

ZOD is distinguished by its high-resolution sensor suite, designed specifically for long-range detection. The ego vehicle is equipped with an 8 Megapixel (MP) forward-facing camera, high-density LiDAR, and high-precision GNSS/IMU.

- **Long-Range Annotation:** The dataset provides annotations for objects up to 245 meters away. This is a critical differentiator for highway validation. In many other datasets, annotations stop at 100 meters, which is insufficient for decision-making at 120 km/h. At that speed, 100 meters is covered in less than 3 seconds. ZOD's extended range allows researchers to train detection models that can identify hazards with enough lead time for smooth braking maneuvers.
- **Sensor Modality:** The dataset includes data from cameras, LiDAR, and GNSS/IMU. Recent updates have also integrated radar data, which is crucial for determining the velocity of distant objects via the Doppler effect, a capability that camera-only systems struggle with.

Sensor Fidelity and Range

ZOD is structured to support multiple AD tasks simultaneously:

- **Frames:** 100,000 curated images accompanied by 2 seconds of supporting sensor data (LiDAR/GPS). These are selected to maximize diversity, covering distinct scenarios rather than repetitive frames from a single drive.
- **Sequences:** 1,473 sequences, each 20 seconds long. These capture dynamic maneuvers such as high-speed overtaking, lane changes, and highway merging.
- **Drives:** Full drives lasting several minutes. These are essential for continuous localization (SLAM) and mapping tasks, allowing the vehicle to build a consistent map of a long highway stretch.

Environmental Diversity

Unlike datasets collected in the sunny climates of California or Arizona (e.g., Waymo, Argoverse), ZOD captures the variability of European weather. It includes extensive data in snow, rain, and fog, as well as varying lighting conditions (night, twilight). This is particularly relevant for highway autonomy, where spray from other vehicles on wet roads can blind sensors—a phenomenon explicitly captured in ZOD.

6.1.2 ONCE (One Million Scenes) Dataset

The ONCE (One Million Scenes) dataset [27] addresses the "data hunger" of deep learning models. While supervised learning requires expensive human annotation, self-supervised and semi-supervised learning can leverage vast amounts of unlabeled data. ONCE provides 1 million LiDAR scenes and 7 million corresponding camera images, with a smaller subset (16k scenes) fully annotated.

- **Scale and Diversity** The dataset covers 144 driving hours across a range of environments including downtowns, suburbs, bridges, tunnels, and highways.
- **Highway Specifics**: The metadata (meta_info) associated with the dataset allows researchers to filter specifically for highway and bridge scenarios. This is useful for pre-training backbone networks on the visual and geometric statistics of highway environments (e.g., the visual rhythm of lane markings, the geometric structure of guardrails) before fine-tuning on annotated data.
- **Weather Conditions:** ONCE emphasizes diversity in weather (sunny, cloudy, rainy) and time of day (morning, noon, afternoon, night). The inclusion of night-time highway driving is valuable for testing the robustness of perception systems against headlight glare and low-contrast lane markings.

6.1.3 PAVE: Production Autonomous Vehicle Evaluation

The PAVE dataset [28] introduces a novel paradigm in dataset creation. PAVE consists entirely of data collected while the vehicle was in autonomous mode. PAVE contains over 100 hours of naturalistic data from production of autonomous vehicles. The data is segmented into key frames with 20 Hz trajectories spanning 6 seconds into the past and 5 seconds into the future.

- **End-to-End Evaluation:** The dataset is designed to evaluate the safety of the AV's decision-making. By analyzing the vehicle's trajectory in relation to the surrounding traffic, researchers can assess the quality of the "end-to-end" motion planning.
- **Scenario Attributes:** Key frames are richly annotated with scenario-level attributes, including area type (highways, urban roads, residential), driver intent, and traffic density. This allows for specific benchmarking of highway performance—for example, measuring the smoothness of the AV's trajectory during a high-speed lane change or its reaction time to a braking event ahead.

6.1.4 MAN TruckScenes

Highway automation is economically critical for the logistics industry. However, autonomous trucks (ATs) have different kinematic constraints than passenger cars. They require longer stopping distances and have different blind spots. The MAN TruckScenes dataset [29] addresses this by providing multi-modal data from the perspective of an autonomous truck.

- **Long-Range Perception:** The dataset utilizes 4D radar and cameras to achieve the necessary perception ranges. It provides data annotations extending beyond 230 meters, which is essential for heavy vehicles that cannot brake aggressively.

- **Truck-Specific Ontology:** The dataset includes 27 object categories specific to the challenges faced by trucks. This likely includes categories for diverse trailer types, other heavy machinery, and infrastructure elements that might pose clearance risks.

6.1.5 comma2k19

Unlike the research-grade datasets discussed above, comma2k19 [30] focuses on "democratizing" autonomous driving research using consumer-grade hardware (smartphones/EONs) mounted on production vehicles.

- **Dataset Composition:** The dataset consists of over 33 hours of commute driving on California's Highway 280 (San Jose to San Francisco). It is segmented into 2019 clips of 1 minute each.
- **End-to-End Lateral Control:** The dataset includes CAN bus data (steering angle, speed) and GNSS traces. This makes it ideal for training "behavioral cloning" models for lateral control (lane keeping) and longitudinal control (adaptive cruise control). It demonstrates that effective highway autonomy (Level 2+) can be achieved with relatively low-cost sensors, provided the data volume is sufficient.

6.2 Infrastructure-Centric Datasets: A Top-Down Perspective

Infrastructure-centric datasets decouple perception from the vehicle, placing sensors on static infrastructure such as gantries, bridges, or poles. This perspective provides a top-down view that eliminates the blind spots inherent to ego-centric systems. These datasets are foundational for developing "Smart Highway" concepts and C-ITS.

6.2.1 H-V2X (Highway V2X) Perspective: Infrastructure-Only

While many infrastructure datasets focus on intersections (which are geometrically complex but low speed), H-V2X [31] is the first large-scale dataset dedicated to highway roadside perception.

- **Scale and Scope:** The dataset covers over 100 kilometers of highway and includes 1.9 million annotated samples in Bird's-Eye-View (BEV) space. This scale is necessary to capture the variety of highway geometries, including curvature changes, on/off-ramps, and variable lane configurations.

- **Vector Maps:** A unique feature of H-V2X is the inclusion of vector maps. These maps provide the topological logic of the road (lane connectivity, speed limits, merge areas). Combining sensor data with vector maps allows for sophisticated "map-aware" perception tasks, such as predicting which exit a vehicle will take based on its lane position 500 meters prior.

The dataset benchmarks three specific tasks: BEV detection, BEV tracking, and trajectory prediction. These are the core components of a roadside perception system designed to feed data to connected vehicles.

6.2.2 HighD and NGSIM Perspective: Infrastructure-Only

These datasets differ from the above in that they do not primarily provide raw sensor data (like LiDAR point clouds) for perception training. Instead, they provide high-precision vehicle trajectories.

- **NGSIM (Next Generation Simulation):** Collected in the mid-2000s on US highways (US-101, I-80), this dataset utilized synchronized cameras to extract vehicle trajectories. Despite its age, it remains a benchmark for traffic flow theory and driver behavior modeling.

- **HighD [32]:** This dataset uses drone footage to extract vehicle trajectories on German highways. The use of drones eliminates the "observer effect" (where drivers change behavior because they see a test vehicle or roadside camera). HighD provides a naturalistic view of highway interactions, risk indices, and lane change dynamics. It is crucial for training the "prediction" and "planning" components of an AV stack, ensuring the AV behaves in a way that is predictable to human drivers.

6.3 Cooperative & Hybrid Datasets (V2X)

The frontier of autonomous driving lies in cooperative perception, where vehicles and infrastructure share raw or processed sensor data to create a unified world model. These datasets are the most complex to create, requiring precise temporal and spatial synchronization between distinct entities (vehicles and infrastructure). They address the physical limitations of single-agent perception.

6.3.1 DAIR-V2X

DAIR-V2X [33] is a pioneering dataset that systematically captures the same scene from both the vehicle and the infrastructure perspective. It is designed to facilitate Vehicle-Infrastructure Collaborative Autonomous Driving (VICAD).

- **Dataset Composition:** The dataset includes 10 km of highway driving and approximately 100 km of city roads. It is divided into three subsets: DAIR-V2X-C (Cooperative), DAIR-V2X-I (Infrastructure-only), and DAIR-V2X-V (Vehicle-only).
- **Temporal Alignment (V2X-Seq):** A major challenge in V2X is latency—the time it takes for data to travel from the infrastructure to the vehicle. The V2X-Seq update addresses this by providing sequential data that allows researchers to study tracking and forecasting over time. This helps in developing algorithms that can compensate for communication delays.
- **Re-Identification (DAIR-V2XReid):** Released in 2024, this sub-dataset focuses on the problem of Vehicle Re-Identification. It ensures that a vehicle seen by "Camera A" on the highway is correctly identified as the same vehicle by "Camera B" kilometers down the road, or by the vehicle's own sensors. This is essential for maintaining a persistent track of objects across a large-scale network.

6.4 Summary of Key Datasets

Table 5 Summary of Key Datasets for General Highway Navigation

Dataset Name	Perspective	Highway Focus	Key Sensors/Data	Notable Feature
Zenseact (ZOD)	Ego-Only	High (European)	8MP Cam, LiDAR	245m range, Winter weather
ONCE	Ego-Only	Medium (China)	LiDAR, Cam	1M scenes, Semi-supervised
PAVE	Ego-Only	High (Global)	Cam, GNSS	End-to-End, Autonomous Mode
MAN TruckScenes	Ego-Only	High (Trucks)	4D Radar, Cam	Truck-specific, long stopping
comma2k19	Ego-Only	High (US)	Cam, GNSS, CAN	Consumer hardware, 33hrs highway
H-V2X	Infra-Only	Very High	Cam, Radar	100km coverage, BEV focus
DAIR-V2X	Cooperative	Low/Med (10km)	Veh+Infra LiDAR	VIC synchronization, Sequential

7. Conclusion

7.1 Synthesis of Findings

The analysis of autonomous driving datasets across Work Zones (Chapter 3), Accident Scenarios (Chapter 4), Tunnels/Urban Canyons (Chapter 5), and Highway Autonomy (Chapter 6) reveals a critical inflection point in the industry. We have successfully moved past the era of simple data collection—characterized by disparate dashcam footage and basic object detection—into an era of structural validation.

While ego-centric perception has reached a high degree of maturity through datasets like Waymo Open and ZOD, our investigation identifies that single-viewpoint data is no longer sufficient to solve the remaining "long-tail" of safety-critical edge cases. Whether it is the irregularity of a construction zone, the GNSS-denied environment of a tunnel, or the adversarial causality of a collision, the current data landscape is defined not by what is present, but by what is missing: the Infrastructure Perspective and High-Fidelity Simulation.

7.2 The Three Pillars of Future Dataset Development

Based on the gaps identified across all domains (Infrastructure/V2X gaps in Ch. 3, Simulation deficits in Ch. 4, and High-Speed/Acoustic voids in Ch. 5), future research and dataset creation must pivot toward three integrated pillars:

Cooperative Perception Integration: To resolve occlusion in work zones and complex intersections, datasets must evolve from "ego-only" to "ego + infrastructure." The blueprint provided by TUMTraf-A and DAIR-V2X must be expanded to include diverse environments. Safety cannot be guaranteed solely by what the vehicle sees; it requires the "God's eye view" of the road network.

Procedural & Adversarial Simulation: Real-world data collection is too risky for accident scenarios and too ephemeral for construction zones. As demonstrated by DeepAccident and STRIVE, the industry must embrace synthetic data not merely as a stopgap, but as the primary engine for training planners against millions of variations of entropy (e.g., erratic worker movements, extreme weather, and rare collision trajectories).

Multi-Modal Sensory Expansion: Standard sensor suites (Camera/LiDAR) are insufficient for the unique physics of specific domains. To master high-speed tunnels and adverse highway weather, future datasets must integrate acoustic sensors (for siren/crash detection in tunnels) and capture high-speed vehicle dynamics (aerodynamics/vibration) that current urban-focused datasets miss.

7.3 Final Outlook

The path to robust, safety-critical autonomy does not lie in simply collecting more data, but in collecting better connected data.

The "Golden Era" of massive, uncurated raw logs is ending. We are entering the "Era of Hybrid Intelligence," where the strongest models will be trained on a convergence of real-world ego-data for texture, infrastructure data for occlusion reasoning, and synthetic data for behavioral robustness. Only by bridging these domains can we move from demonstrating autonomy in controlled environments to deploying it safely in the chaotic reality of the open road.

References

[1] A. Ghosh *et al.,* "ROADWork: A Dataset and Benchmark for Learning to Recognize, Observe, Analyze and Drive Through Work Zones," 2024. [Online]. Available: https://arxiv.org/pdf/2406.07661

[2] K. Li *et al.,* "CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving," 2022. [Online]. Available: https://arxiv.org/pdf/2203.07724

[3] J. Mei *et al.,* "Waymo Open Dataset: Panoramic Video Panoptic Segmentation," 2022. [Online]. Available: https://arxiv.org/pdf/2206.07704

[4] P. Sun *et al.,* "Scalability in Perception for Autonomous Driving: Waymo Open Dataset," 2019. [Online]. Available: https://arxiv.org/pdf/1912.04838

[5] R. Xu *et al.,* "WOD-E2E: Waymo Open Dataset for End-to-End Driving in Challenging Long-tail Scenarios," 2025. [Online]. Available: https://arxiv.org/pdf/2510.26125

[6] S. Ettinger *et al.,* "Large Scale Interactive Motion Forecasting for Autonomous Driving : The Waymo Open Motion Dataset," 2021. [Online]. Available: https://arxiv.org/pdf/2104.10133

[7] F.-H. Chan, Y.-T. Chen, Y. Xiang, and M. Sun, "Anticipating Accidents in Dashcam Videos," in *Lecture Notes in Computer Science*, vol. 10114, *Computer Vision – ACCV 2016*, S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds., Cham: Springer International Publishing, 2017, pp. 136–153.

[8] W. Bao, Q. Yu, and Y. Kong, "Uncertainty-based Traffic Accident Anticipation with Spatio-Temporal Relational Learning," 2020. [Online]. Available: https://arxiv.org/pdf/2008.00334

[9] H. Liao *et al.,* "Real-time Accident Anticipation for Autonomous Driving Through Monocular Depth-Enhanced 3D Modeling," 2024. [Online]. Available: https://arxiv.org/pdf/2409.01256

[10] Y. Yao *et al.,* "DoTA: Unsupervised Detection of Traffic Anomaly in Driving Videos," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 444–459, 2023, doi: 10.1109/TPAMI.2022.3150763.

[11] J. Fang, D. Yan, J. Qiao, J. Xue, H. Wang, and S. Li, "DADA-2000: Can Driving Accident be Predicted by Driver Attention? Analyzed by A Benchmark," 2019. [Online]. Available: https://arxiv.org/pdf/1904.12634

[12] J. Fang *et al.,* "Abductive Ego-View Accident Video Understanding for Safe Driving Perception," 2024. [Online]. Available: https://arxiv.org/pdf/2403.00436

[13] D. C. Moura, S. Zhu, and O. Zvitia, "Nexar Dashcam Collision Prediction Dataset and Challenge," 2025. [Online]. Available: https://arxiv.org/pdf/2503.03848

[14] W. Zimmer *et al.,* "Safety-Critical Learning for Long-Tail Events: The TUM Traffic Accident Dataset," 2025. [Online]. Available: https://arxiv.org/pdf/2508.14567

[15] P. Pradeep Kumar and K. Kant, "TU-DAT: A Computer Vision Dataset on Road Traffic Anomalies," *Sensors (Basel, Switzerland)*, vol. 25, no. 11, 2025, doi: 10.3390/s25113259.

[16] A. Shah, J. B. Lamare, T. N. Anh, and A. Hauptmann, "CADP: A Novel Dataset for CCTV Traffic Camera based Accident Analysis," 2018. [Online]. Available: https://arxiv.org/pdf/1809.05782

[17] T. Wang *et al.,* "DeepAccident: A Motion and Accident Prediction Benchmark for V2X Autonomous Driving," 2023. [Online]. Available: https://arxiv.org/pdf/2304.01168

[18] D. Rempe, J. Philion, L. J. Guibas, S. Fidler, and O. Litany, *Generating Useful Accident-Prone Driving Scenarios via a Learned Traffic Prior*, 2022.

[19] S. S. S. Kumar, S. Thalapanane, Peethambari, Guru Nandhan Appiya Dilipkumar, S. SriHari, L. Zheng, and M. C. Lin, *DISC: Dataset for Analyzing Driving Styles In Simulated Crashes for Mixed Autonomy*, 2025.

[20] *SynSHRP2: A Synthetic Multimodal Benchmark for Driving Safety-critical Events Derived from Real-world Driving Data*, 2025. [Online]. Available: https://openaccess.thecvf.com/content/iccv2025w/wdfm-ad/html/shi_synshrp2_a_synthetic_multimodal_benchmark_for_driving_safety-critical_events_derived_iccvw_2025_paper.html

[21] A. Kurda, S. Steuernagel, L. Jung, and M. Baum, *Odyssey: An Automotive Lidar-Inertial Odometry Dataset for GNSS-denied situations*, 2025.

[22] W. Wen *et al.,* "UrbanLoco: A Full Sensor Suite Dataset for Mapping and Localization in Urban Scenes," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, May. 2020 - Aug. 2020, pp. 2310–2316.

[23] S. Lu, S. Bao, W. Shi, Y. Wei, S. Zhang, and D. Yang, "Multimodal sensor dataset from vehicle-mounted mobile mapping system for comprehensive urban scenes," *Scientific data*, vol. 12, no. 1, p. 1411, 2025, doi: 10.1038/s41597-025-05471-1.

[24] T. Sun *et al.,* "SHIFT: A Synthetic Driving Dataset for Continuous Multi-Task Domain Adaptation," 2022. [Online]. Available: https://arxiv.org/pdf/2206.08367

[25] T. Xie, Z. Song, F. Wen, J. Li, G. Liu, and Z. Zhao, "TruckV2X: A Truck-Centered Perception Dataset," *IEEE Robot. Autom. Lett.*, vol. 10, no. 9, pp. 9312–9319, 2025, doi: 10.1109/lra.2025.3592884.

[26] M. Alibeigi *et al.,* "Zenseact Open Dataset: A large-scale and diverse multimodal dataset for autonomous driving," in 2023, pp. 20121–20131. Accessed: Nov. 14 2024. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1913276

[27] J. Mao *et al.,* "One Million Scenes for Autonomous Driving: ONCE Dataset," 2021. [Online]. Available: https://arxiv.org/pdf/2106.11037

[28] X. Li, C. Wang, Y. Liu, D. He, J. Zhang, and K. Ma, *PAVE: An End-to-End Dataset for Production Autonomous Vehicle Evaluation*, 2025.

[29] F. Fent *et al.,* "MAN TruckScenes: A multimodal dataset for autonomous trucking in diverse conditions," in *Advances in Neural Information Processing Systems 37*, San Diego, California, USA, 2024, pp. 62062–62082.

[30] H. Schafer, E. Santana, A. Haden, and R. Biasini, *A Commute in Data: The comma2k19 Dataset*, 2018.

[31] C. Liu, M. Zhu, and C. Ma, "H-V2X: A Large Scale Highway Dataset for BEV Perception," in 2025, pp. 139–157. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-73232-4_8

[32] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems," 2018. [Online]. Available: https://arxiv.org/pdf/1810.05642

[33] H. Yu *et al.,* "DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection," 2022. [Online]. Available: https://arxiv.org/pdf/2204.05575