

Towards Vision Zero: The TUM Traffic Accid3nD Dataset

Walter Zimmer ¹✉ , Ross Greer ² , Xingcheng Zhou ¹ , Rui Song ^{1,4} , Hu Cao ¹ ,
 Daniel Lehmberg ¹ , Marc Pavel ¹ , Ahmed Ghita ³ , Akshay Gopalkrishnan ⁵ ,
 Holger Caesar ⁶ , Mohan Trivedi ⁵ , Alois C. Knoll ¹ 

¹ Technical University of Munich, ² University of California Merced, ³ SETLabs Research GmbH,
⁴ Fraunhofer IVI, ⁵ University of California San Diego, ⁶ Delft University of Technology

accident-dataset.github.io



Figure 1. **Raw TUMTraf-Accid3nD dataset.** Accidents are recorded from roadside cameras on the A9 Test Field for Autonomous Driving in Munich, Germany. The dataset includes scenes with high-speed traffic, collisions, and overturning vehicles. Some vehicles are catching fire.

Abstract

Even though a significant amount of work has been done to increase the safety of transportation networks, accidents still occur regularly. They must be understood as unavoidable and sporadic outcomes of traffic networks. No public dataset contains 3D annotations of real-world accidents recorded from roadside camera and LiDAR sensors. We present the TUM Traffic Accid3nD (TUMTraf-Accid3nD) dataset, a collection of real-world highway accidents in different weather and lighting conditions. It contains vehicle crashes at high-speed driving with 2,634,233 labeled 2D bounding boxes, instance masks, and 3D bounding boxes with track IDs. In total, the dataset contains 111,945 labeled image and point cloud frames recorded from four roadside cameras and LiDARs at 25 Hz. The dataset contains six object classes and is provided in the OpenLABEL format. We propose an accident detection model that combines a rule-based approach with a learning-based one. Experiments and ablation studies on our dataset show the robustness of our proposed method. The dataset, model, and code are available on our website.

1. Introduction

Vision Zero is a worldwide initiative to reduce road deaths and serious injuries through innovative, data-driven interventions. Autonomous driving (AD) and intelligent infrastructure play a significant role in making roads safer by preventing accidents before they happen. Collecting data on events that occur most rarely (*long-tail events*) is important for robust machine learning of data-driven perception, planning, and control models in robotic systems [1]. In the case of AD, however, these events are especially costly to collect [2, 3, 4]. Long-tail events such as accidents and near-misses [5] come at great risk to human life and are otherwise difficult to stage and capture in the natural driving environment [6, 7]. The time between an accident and the arrival of medical assistance significantly impacts whether the passengers of a vehicle survive an accident. Automatic accident detection reduces this time and has the potential to save lives.

Toward the goal of Vision Zero, we propose *TUMTraf-Accid3nD*, a 3D perception dataset specifically curated for accident scenarios. We present novel labeling methods alongside newly introduced tasks, including 3D object

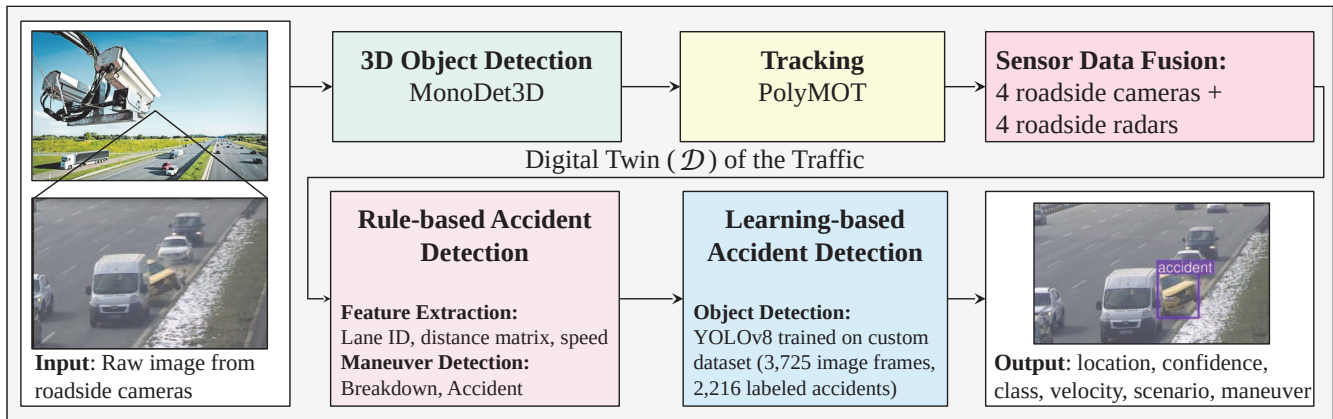


Figure 2. **Accident detection pipeline.** We use advanced 3D perception techniques and multi-sensor data fusion to create a real-time digital twin of the traffic. Starting with raw camera images, the framework first performs 3D object detection using MonoDet3D to identify and localize vehicles in three dimensions. Following detection, Poly-MOT tracking is applied to maintain continuity across frames, while sensor data fusion combines inputs from four roadside cameras and four radars. The digital twin is then used in two accident detection modules: 1) The **Rule-based Accident Detection** module extracts features such as lane IDs, distance matrices, and velocities, identifying potential accidents through predefined maneuver detection rules. 2) The **Learning-based Accident Detection** module employs a YOLOv8 object detector, trained on a custom dataset, to detect accident events. The final output includes the object’s location, confidence score, class, velocity, and detected scenario or maneuver.

detection, tracking, and accident detection. Furthermore, we demonstrate how vision-language models can enhance safety analysis to improve overall situational awareness. Our dataset offers 3D accident labels with tracking information. It supports the development of safer autonomous systems through multiple directions: (1) the data supports improved learning on a variety of perception-related tasks, such as detection [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19], tracking, and segmentation [20]. (2) The nature of the dataset allows for the study of methods of cooperative perception between roadside sensors observing the same scene from different view angles. Cooperative perception allows for a reduction of occlusion through shared information. (3) Roadside sensors allow the creation of digital twins of the traffic scene [21, 22, 23, 24]. These digital twins expand the visibility range beyond the egocentric view, which may be used to provide adequate warning lead time for safe approaches and control transitions for risky traffic events [25].

In addition to standard computer vision tasks—such as 2D object detection, instance segmentation, 3D object detection, sensor data fusion, tracking, trajectory prediction, and accident classification, our dataset also supports accident analysis. This involves reconstructing the sequence of events to understand how the accident occurred and what contributing factors may have played a role. For instance, was the vehicle speeding? Was there a traffic jam a short distance ahead? Did the driver appear inattentive or react too late?

Furthermore, this dataset can be utilized by Vision-Language Models (VLMs) [26] to interpret complex scenes. Given an image as input, VLMs can provide a textual description of the scene, identifying if an accident has occurred, is actively

unfolding, or if there are signs of an imminent collision. This capability enhances the dataset’s potential for developing sophisticated tools for accident detection and scene analysis in real-world traffic environments.

Contributions. To summarize, our main contributions are:

- We present the TUMTraf-Accid3nD dataset, a dataset curated specifically for rare and hazardous traffic incidents, providing a groundbreaking resource for accident-specific 3D perception research.
- TUMTraf-Accid3nD features 2,634,233 labeled 3D boxes, instance masks, and 2D box annotations with track IDs of real accidents across various lighting and weather conditions. This makes it the largest dataset focused on real-world accident scenarios captured in 3D.
- We introduce an annotation method that enables highly accurate and detailed labeling of accidents and near-miss events. The annotation process includes precise 2D and 3D bounding boxes, instance segmentation, and tracking of all traffic participants in the scene.
- We propose a framework to detect accidents in real-time and in different weather and lighting conditions.
- We support eight tasks with our dataset: object recognition, 2D object detection, accident classification, instance segmentation, 3D object detection, tracking, accident anticipation, and trajectory prediction.
- We provide baseline results for the first six tasks and open source our dataset, detection framework, and dev kit.

2. Problem Statement

High-quality datasets focused on accident scenarios are essential for advancing road safety. However, existing datasets

Table 1. **Overview of publicly available accident datasets.** We compare the TUMTraf-Accid3nD dataset with other available synthetic and real accident datasets based on the following criteria: year, type (synthetic or real), perspective, number of image frames (#Img), available point clouds (PCs), number of 2D bounding boxes (#2D BB), number of 2D instance masks (#Masks), number of 3D bounding boxes (#3D BB), and track IDs (T). Entries with * indicate approximation based on average video frame counts reported in respective papers.

Dataset	Year	Type	Perspect.	#Img	PCs	#2D BB	#Masks	#3D BB	T
◦ VIENA ² [28]	2019	synth	vehicle	2,250,000	×	×	×	×	×
◦ GTACrash [29]	2019	synth	vehicle	751,610	×	×	×	×	×
◦ MP-RAD [30]	2023	synth	roadside	366,600	×	×	×	×	×
◦ RiskBench [31]	2024	synth	vehicle	N/A	×	×	×	×	✓
◦ DeepAccident (DA) Dataset [32]	2024	synth	V2V & V2I	57,000	✓	285,000	×	285,000	✓
◦ Accident Image Analysis (AIAD) [33]	2018	real	variable	10,480	×	×	×	×	×
◦ Car Accident Det. & Pred. (CADP) [34]	2018	real	roadside	75,030*	×	×	×	×	×
◦ Causality in Traffic Accident (CTA) [35]	2020	real	vehicle	342,495*	×	×	×	×	×
◦ Car Crash Dataset (CCD) [36]	2020	real	vehicle	75,000	×	×	×	×	×
◦ Acc. Det. CCTV Footage (CCTVF) [37]	2020	real	roadside	990	×	×	×	×	×
◦ Argus Dataset [38]	2021	real	roadside	120,000	×	×	×	×	×
◦ YoutubeCrash [39]	2021	real	vehicle	7,720	×	×	×	×	×
◦ IITH Road Accident Dataset [40]	2022	real	roadside	127,138	×	×	×	×	×
◦ TAD [41]	2022	real	roadside	24,810	×	×	×	×	×
◦ Accident Detection Model (ADM) [42]	2023	real	roadside	3,250	×	×	×	×	×
◦ MM-AU Dataset [43]	2024	real	vehicle	2,190,000	×	2,233,683	2,233,683	×	×
◦ TUMTraf-Accid3nD Dataset (ours)	2025	real	roadside	111,656	✓	2,634,233	2,634,233	2,634,233	✓

rarely cover real accident cases in sufficient detail, making this an underrepresented but critical area of research. Most autonomous driving datasets focus on normal driving conditions to avoid the complexities of actual accident scenarios. This limits the potential for models to effectively learn and predict high-risk events. To improve safety in autonomous systems, there is a pressing need for a specialized dataset that captures the complexity of accident scenarios in diverse conditions. Such a dataset would provide a robust foundation for developing algorithms capable of understanding, detecting, and ultimately helping to prevent accidents.

3. Related Work

Existing accident detection methods have never been tested on real roadside traffic data of an ITS test stretch. Real accident datasets are rare and do not contain enough data to train deep learning models [27], as shown in Table 1.

3.1. Accident Datasets

Several accident datasets [28, 29, 32, 33, 34, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45] have recently been released. However, most are limited to 2D annotations, lacking realistic 3D labeling needed for comprehensive accident analysis. Although some 3D datasets exist, they are synthetic and not representative of real-world conditions.

The DeepAccident [32] dataset contains 691 synthetic accident scenarios in the CARLA simulator. These accidents were generated based on crash reports published by the National Highway Traffic Safety Administration (NHTSA).

The dataset contains labeled data from four vehicles and one roadside infrastructure camera. One limitation of this work is that all accidents are generated in a simulation environment and do not represent realistic crash scenes. Hence, the sim-to-real gap must be addressed to improve the generalization capabilities of the perception models.

MM-AU [43] is a dataset for multi-modal accident understanding in videos. It contains 11,727 ego-view accident videos and 2.23 million 2D object boxes but lacks 3D box annotations, instance masks, and track IDs. Our dataset addresses this by providing high-quality 3D labeled real-world accident data in high-speed highway scenarios.

3.2. Accident Detection

Accident detection aims to identify the time and location of accidents within video frames, which is challenging due to rapid object motions, visual occlusions, and viewpoint changes caused by camera movements during crashes [46]. Early detection methods [47] primarily rely on frame-level appearance changes or simple motion cues to identify accidents, but these approaches struggled in complex environments. More recent methods emphasize spatiotemporal modeling, capturing motion consistency and scene evolution across frame sequences to improve robustness [43, 48]. Supervised methods train deep networks to classify accident vs. non-accident frames, while unsupervised techniques, such as DoTA [49], detect abnormal motion patterns by predicting future trajectories and flagging deviations between predicted and actual motions. Despite these advancements, most existing approaches rely on monocular 2D video data [50], losing



Figure 3. **Visualization of the labeled TUMTraF-Accid3nD dataset** with 3D box annotations, instance masks, track IDs, and trajectories. Accidents are recorded from roadside cameras on a test bed for autonomous driving. The dataset includes scenes with collisions and overturning vehicles, with some vehicles even catching fire.

important depth and spatial cues. This limits their ability to estimate object distances and collision risk accurately, reinforcing the need for real-world 3D accident datasets to enhance detection performance and generalizability.

3.3. Accident Anticipation

Accident anticipation aims to predict accidents before they occur, providing early warnings based on evolving scene dynamics. These methods typically process sequential video frames and estimate future accident likelihood using recurrent or temporal convolutional networks. To improve interpretability, models like DSA-RNN [51] introduce attention mechanisms that dynamically focus on critical objects and regions contributing to accident risk. Later works further enhance these attention mechanisms by incorporating driver behavior, such as gaze direction or steering patterns, to capture human reactions to emerging hazards. Methods like DRIVE [52] explicitly model interactions between traffic participants using spatiotemporal graphs, improving anticipation through relational reasoning among vehicles, pedestrians, and infrastructure. However, existing approaches rely heavily on annotated accident time windows, which are expensive to obtain and prone to ambiguity. Moreover, the absence of large-scale, realistic 3D accident datasets hinders generalization across diverse driving scenarios, emphasizing the urgent need for diverse 3D benchmarks to enable robust and transferable anticipation systems.

4. The TUM Traffic Accid3nD Dataset

This section presents the TUM Traffic Accid3nD (TUMTraF-Accid3nD) dataset, a real-world dataset of rare, hazardous traffic incidents for accident-specific 3D perception tasks.

4.1. Sensor Suite

The roadside infrastructure sensor setup is designed to continuously monitor traffic flow and detect accident scenarios using a diverse suite of sensors. Positioned at strategic locations, the sensor suite used for the dataset includes nine sensors: four high-definition cameras, four radars, and one LiDAR. These sensors are mounted on two sensor stations, such as the one shown in Fig. 4, to capture real-time traffic data from multiple perspectives. Each sensor type contributes uniquely to a multi-modal data fusion framework. The cameras provide high-resolution visual information, enabling object detection, tracking, and scene interpretation. Radars add velocity and range information, capturing the motion dynamics of vehicles, even in challenging weather conditions like fog or heavy rain. The LiDAR offers precise 3D spatial data, creating a detailed map of object positions and surroundings. Together, these sensors deliver a rich, fused dataset that enhances the accuracy of accident detection and analysis by combining visual, motion, and depth information. Through careful calibration, the data from each sensor is aligned in a common coordinate system to ensure a reliable multi-modal fusion. This setup enables scene understanding in various lighting and weather conditions and supports advanced applications like trajectory prediction, cooperative perception, and accident anticipation.

4.2. Data Collection Process

The data collection process includes several key stages: data recording, extraction, and anonymization. We capture high-quality data and ensure privacy and usability.

Data Recording. Data is continuously captured from roadside cameras, LiDAR, and radar sensors to monitor traffic flow and detect accident events during the day and night. Each sensor records data at 25 FPS to provide details for high-speed accident analysis. The recorded data is com-

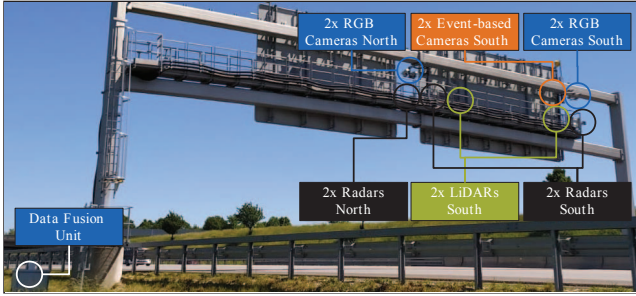


Figure 4. **Visualization of the sensor setup.** We show one of the two gantry bridges used to record the highway data with accidents.

pressed and saved in *rosbag* files on secure servers in real-time. Each sensor stream is time-synchronized, ensuring that data across cameras, LiDAR, and radar align accurately.

Data Selection. Our data selection system uses a rule-based accident detection framework to prioritize capturing accident and near-miss events. By monitoring indicators like lane deviations, sudden speed changes, and unusual vehicle interactions, the system flags high-priority events and focuses on relevant accident scenarios. The selected data is then extracted into individual frames or point cloud scans.

Data Anonymization. To protect privacy, all data undergoes an anonymization process. License plates and personally identifiable information like faces are blurred. A YOLOv8 [53] network was trained on local license plates to detect them in real-time.

The dataset development kit, which includes pipelines for processing, loading, and managing data, enables researchers to efficiently access, preprocess, and work with the dataset.

4.3. Labeling Process

We use 3D BAT [54], an automatic 3D bounding box annotation toolbox, to annotate our dataset. It includes a detection and tracking step and a manual quality check to enhance accuracy. This tool automates the labeling process by first applying a custom 2D object detector [14] (based on YOLOv7 [55]) and a 3D object detector (MonoDet3D [14]). The objects are tracked with the PolyMOT [56] tracker and fused using a late fusion approach [14]. Each labeled traffic participant contains a 2D box, a corresponding instance mask, and 3D box information (position, dimensions, and rotations) with a track ID and speed value. To ensure high labeling quality, we manually inspect the generated annotations in the labeling tool, adjust them accordingly, and finally export them in the OpenLABEL [57] standard.

4.4. Coverage and Scenario Diversity

Our dataset features 111,945 labeled camera and LiDAR frames with over 2.6 million 2D and 3D box annotations, each accompanied by track IDs, trajectory data, and classification across six different instance types: cars, trucks, buses,

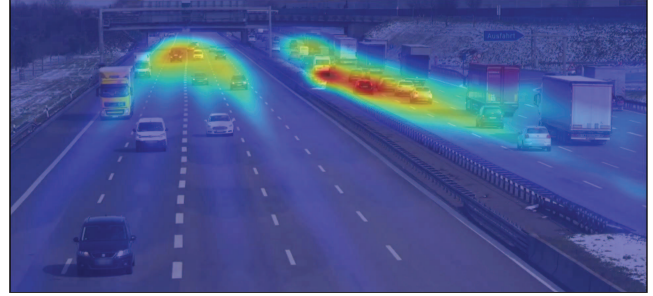


Figure 5. **Heatmap visualization of traffic participant locations.** The left lane on the highway towards the north direction indicates a high traffic volume.

pedestrians, motorcycles, and bicycles. It captures a diverse range of accident scenarios, including high-speed lane changes leading to rear-end collisions, vehicle overturning upon impact, and vehicles catching fire. Additionally, it includes instances involving emergency response vehicles and near-miss events. Example cases are illustrated in Figure 1. The dataset serves as an essential resource for developing and validating AI-based detection, tracking, data fusion, and trajectory prediction algorithms, as well as understanding the occurrence and after-effects of naturally occurring high-speed crash incidents and other accidents on highways.

4.5. Annotation Schema

Our annotation schema is based on the OpenLABEL [57] standard, structured to store 3D labels with detailed tracking information, event data, and attributes that document accident scenarios. Each 3D label includes tracking IDs to uniquely identify objects across frames and to capture the trajectories of road users over time. Each annotation holds attributes that mark the sensor ID, track history, speed, and number of 3D points inside the 3D box. This schema ensures an accurate, multi-dimensional representation of accidents and enables detailed analysis of incident sequences and individual participant behavior throughout each event.

4.6. Dataset Statistics

Our dataset provides a large collection of annotated accident events and diverse object types across 111,945 camera and LiDAR frames, with over 2.6 million 3D box annotations with track IDs. This large-scale dataset includes six object classes: cars, trucks, buses, motorcycles, bicycles, and pedestrians. The occurrence of labeled accident events allows for the evaluation of detection models under realistic, high-speed highway scenarios.

Figs. 5, 6, 7, 8, 9, 10, and 11 provide a detailed statistical overview: Fig. 5 illustrates a heatmap of traffic participants to highlight busy lanes with a high traffic density. The left lane contains a high traffic volume because of a van that has a breakdown on that lane. Figure 6 presents the distribution of object classes. Cars and trucks are the most commonly

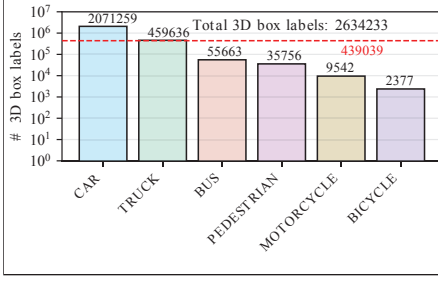


Figure 6. Distribution of object classes.

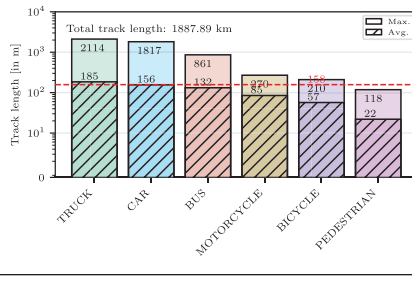


Figure 7. Average and max. track lengths of all labeled object classes.

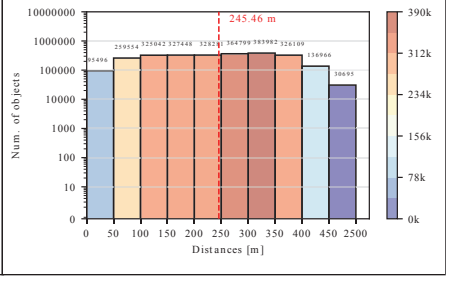


Figure 8. Histogram of labeling distances.

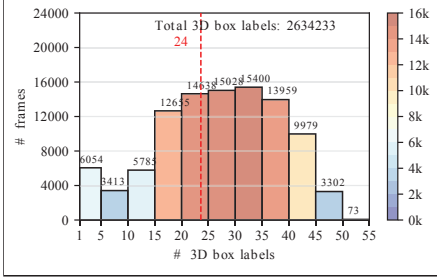


Figure 9. Histogram of the number of 3D box labels.

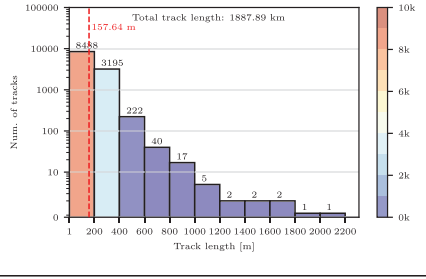


Figure 10. Histogram of the track lengths.

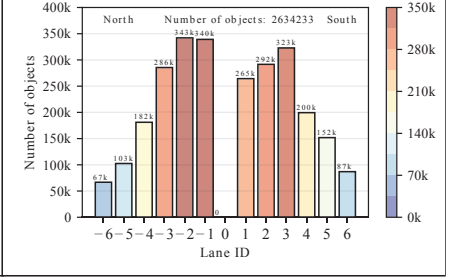


Figure 11. Lane distribution of all labeled objects on the highway.

labeled objects in the dataset. Fig. 7 shows average and maximum track lengths. Track lengths vary, with most up to 200 meters (see Fig. 10) and one truck reaching a 2,114-meter-long track. The total accumulated track length of all labeled objects is 1887.89 km on the highway. Fig. 8 shows the distribution of the labeling distances. On average, objects are 245 meters away from the sensor. Fig. 9 illustrated the histogram of the number of 3D box labels. Most frames contain 15–45 labeled objects, with up to 52 objects in a single frame and an average of 24 objects per image frame. The lane distribution analysis in Fig. 11 shows that the majority of labeled objects are in the southbound lanes, particularly in lane -2. The highway has 12 lanes in total, six lanes heading north and six lanes heading south.

4.7. Comparative Analysis

Unlike most 3D perception datasets, which typically focus on general traffic and non-accident scenarios, our dataset is dedicated to real-world accidents and near-miss events. This emphasis on critical situations allows for detailed modeling and analysis of incidents, setting it apart from popular 3D perception datasets like KITTI [58, 59], nuScenes [60], and Waymo Open Dataset [61]. While these datasets capture diverse traffic objects and scenes, they include no labeled accident events, limiting their use in the development of safety-focused AVs. Our dataset’s unique strengths include its rich 3D annotations of accidents, detailed tracking information, and a variety of accident types such as high-speed collisions, multi-vehicle pile-ups, and emergency responses. The inclu-

sion of labeled multi-sensor data enables robust multi-modal data fusion for scene understanding. We provide long tracks, capturing trajectories across multiple frames and supporting applications in trajectory prediction and risk assessment, which are essential for accident anticipation. This dataset offers a great resource for advancing research in safety-critical situations, filling a gap in the field with data curated specifically for accident detection and prevention.

5. Methodology for Accident Detection

Our framework consists of two modules, the detection and tracking module and the accident detection module. Fig. 2 visualizes the process from detecting objects in raw images to finally detecting accident events.

5.1. 3D Object Detection and Tracking

Our framework uses 3D perception techniques and multi-sensor data fusion to create a real-time digital twin of the traffic. Starting with raw camera images, the framework first performs 3D object detection using MonoDet3D [14] to identify and localize vehicles in three dimensions. Following detection, PolyMOT tracking [56] is applied to maintain continuity across frames, while sensor data fusion combines inputs from four roadside cameras and four radars, leading to generated trajectories and track histories.

Table 2. Evaluation of two object detection methods on our test set.

Model	mAP@[.5]	mAP@[.5:.95]	mIoU	FPS
◦ YOLOv7x-seg	78.50	53.40	85.72	22
◦ YOLOv8x-seg	80.10 +1.60	58.50 +5.10	88.50 +2.78	62

Table 3. Maneuver detection rules for the rule-based approach.

$v_i \geq \frac{15 \text{ km/h}}{3.6}$	Vel. of vehicle (i)
$v_i > v_{\text{lead},i}$	$v_{\text{lead},i}$: Vel. of lead vehicle
$v_i \geq v_j \quad \forall i < j \leq N$	$d_{\text{lead},i}$: Distance to lead vehicle
$d_{\text{lead},i} \geq d_{\text{thresh}}$	d_{thresh} : Predef. distance threshold
$d_{\text{lead},i} < \left(\frac{v_i - v_{\text{lead},i}}{30}\right)^2$	$\text{TTC}_{\text{lead},i}$: Time-to-Coll. lead veh.
$\text{TTC}_{\text{lead},i} \leq \text{TTC}_{\text{thresh}}$	$\text{TTC}_{\text{thresh}}$: Time-to-Coll. threshold

5.2. Accident Detection Pipeline

Our accident detection pipeline combines rule-based and learning-based methods to detect and classify accidents in real-time using roadside infrastructure data. This dual-module approach uses raw images and the trajectories of the digital twins that were produced with 3D perception, tracking, and fusion modules to find accident events.

Rule-Based Accident Detection. This module analyzes tracked object trajectories, which are generated by the MonoDet3D and PolyMOT frameworks, as outlined in the previous section. It extracts features like lane IDs, distance matrices, and object velocities, identifying potential accidents through predefined maneuver detection rules (see Table 3). If all six rules apply simultaneously, the corresponding traffic participant is classified as an accident event. For example, sudden changes in speed or trajectory angle, unsafe lane changes, and close proximity situations are flagged as potential incidents. These rule-based detections provide immediate alerts and enable rapid incident response.

Learning-Based Accident Detection. To enhance accuracy, the pipeline also includes a learning-based YOLOv8 object detector trained on our custom accident dataset. When the rule-based module signals a potential accident, the learning-based detector validates the event by analyzing the camera feed. YOLOv8 is fine-tuned to detect accident-specific cues such as collisions, overturned vehicles, and vehicle fires, producing classifications with associated confidence scores, object locations, and velocities. Detected incidents must appear in at least three consecutive frames with a confidence score above 0.8 to minimize false positives. Additionally, results from all available cameras in a scene are fused to enhance robustness. The output of this integrated accident detection pipeline includes a real-time accident classification for each detected vehicle, along with its location and other critical metadata. This pipeline, illustrated in Fig. 2, supports both training data preparation and live accident monitoring on highways.

Table 4. Evaluation of the PolyMOT 3D tracker on our test set.

Model	FN	FP	MT	PT	ML	IDS	FRAG	HOTA	MOTA	MOTP
PolyMOT	1313	657	20	58	50	19	50	0.45	0.18	1.63

Table 5. 3D detection results of MonoDet3D on our test set.

Model	3D mAP@[.1]
◦ MonoDet3D + YOLOv7 (baseline)	15.20
◦ MonoDet3D + YOLOv7 + PolyMOT	16.23 +1.03
◦ MonoDet3D + YOLOv8	17.77 +2.57
◦ MonoDet3D + YOLOv8 + PolyMOT	18.24 +3.04

6. Experimental Results

First, we outline the benchmark tasks and establish baseline performance metrics for 3D object detection, 3D tracking, and 3D accident detection on our dataset. Baseline performance metrics include 3D mean average precision (mAP_{3D}) for 3D detection, Multi-Object Tracking Accuracy (MOTA) for tracking, and F1 score values for accident classification.

6.1. Object Detection and Tracking Performance

Moreover, we evaluate two 2D detection and segmentation models on our test set using an input size of 1280^2 px and TensorRT acceleration on an NVIDIA RTX 3090 GPU (see Table 2). The YOLOv8x-seg performs best in all metrics. We further evaluate the PolyMOT [56] 3D tracker on our test set and report the results in Table 4.

Ablation Study. We provide an ablation study on our test set for multiple 3D object detection baselines of MonoDet3D [14] in combination with different YOLO models and trackers. In Table 5 we can see that MonoDet3D performs best when employing the YOLOv8 model and PolyMOT tracker.

6.2. Accident Classification Evaluation

We recorded camera images and the fused perception results for 128 days, stored them in *rosbag* files, and processed these recordings. The automatic accident analysis was executed on 12,290 15-minute videos. Figure 12 shows the quantitative evaluation results. In total, 831,969 unique vehicles were identified (5,547 per day). 26.08% of vehicles were driving faster than the allowed speed limit of 120 km/h in the south direction. We found that outbound (north) traffic is often driving faster than inbound (south). The maximum detected speed was 264 km/h in the north direction of the highway where no speed limit is set. We detected 3,748 (0.45%) standing vehicles in a driving lane, 138 standing vehicles in a shoulder lane, and 120 breakdown events. Qualitative results of the rule-based accident detection are shown in Figure 13.

Classification evaluation. We first evaluate the accuracy and runtime performance of both accident classification models on our TUMTraf-Accid3nD dataset (see Table 6). The rule-based approach was able to detect 120 breakdown

Table 6. **Accident classification** results and runtime evaluation of the RBA and LBA approach.

Approach	Accuracy $\uparrow$	Runtime $\downarrow$ [s]	
	F1-Score $\uparrow$	2 cam.	4 cam.
Rule-based Approach (RBA)	0.667	0.086	0.127
Learning-based Approach (LBA)	0.889	4.072	7.995

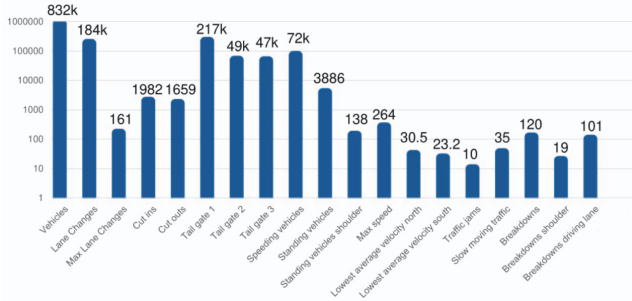


Figure 12. Quantitative results of the RBA accident detection module over a 128-day monitoring period. The statistics include the detected vehicles, maneuvers, scenarios, and accident events.

events. Four were false positives due to inaccurate object detections. This leads to a precision rate of 96.67%. On the other hand, the learning-based accident detection approach achieves a precision rate of 75.00%, which is limited by the relatively small training dataset, which consists of 3,725 images and 2,216 labeled accident events.

Runtime evaluation. The rule-based approach (RBA) runs at 95.05 FPS (10.41 ms per frame) on an NVIDIA RTX 3090 GPU. The total runtime for a 15-minute *rosbag* file with 22,500 ROS messages recorded at 25 FPS is 234.25 seconds. This includes the extraction of the lane ID, the distance calculation between all vehicles, and the scenario classification. In Table 7 we compare our two approaches based on the runtime. On average the RBA approach is about 57 times faster than our learning-based approach (LBA).

7. Conclusion and Future Work

Traffic accidents remain a leading cause of death worldwide, and the ability to rapidly detect accidents via roadside infrastructure sensors holds the potential to accelerate emergency response times and save lives. In this work, we introduced the TUMTraf-Accid3nD dataset, a resource focused on accident detection, and evaluated two detection methods on real-world data. Our dataset, detection framework, and dev kit are available as open-source tools on our project website to encourage widespread research and collaboration across academia and industry. By establishing TUMTraf-Accid3nD as a benchmark for accident detection, we aim to foster advancements in accident anticipation and detection, ultimately supporting the *Vision Zero* goal of eliminating traffic fatalities by 2050.

Table 7. **Runtime comparison.** We compare our rule-based and learning-based accident detection based on the runtime.

Sequence	Runtime $\downarrow$ [s]	
	Rule-based	Learning-based
Sequence S01, part I	8.63	484.69
Sequence S01, part II	6.60	474.69
Sequence S13, part I	5.02	240.43
Sequence S13, part II	5.33	248.16
Avg. (with 2 cameras)	5.17	244.30
Avg. (with 4 cameras)	7.61	479.69
Average (overall)	6.39	361.99



Figure 13. **Qualitative results of our accident detection framework on the TUMTraf-Accid3nD test set.** Left: The rule-based approach detected a rear-end collision. Right: The learning-based approach detected a car crash with a confidence score of 0.8.

Future Directions. We will expand our dataset to include more accident events under varied lighting and weather conditions, enhancing its applicability to complex real-world scenarios. We also plan to incorporate additional types of accidents in diverse environments, particularly urban areas where high-risk intersections often involve vulnerable road users (VRUs). By capturing events where VRUs are at risk due to behaviors like crossing against traffic signals or smartphone distraction, we aim to support additional safety applications. To increase the dataset’s variety, we intend to expand the number of accidents recorded by five sensors to include data from 12 cameras on highways and another 12 in urban settings. Our design allows for integration with VLMs (see supplementary) to enable the interpretation and analysis of critical events, thereby supporting advancements in AV safety and intelligent traffic management. Finally, we will explore accident video diffusion methods [62] for generating realistic accident scenarios [63].

Impact on AD Safety. The TUMTraf-Accid3nD dataset and our detection framework will contribute to autonomous driving safety by enabling more accurate detection and prediction of accident scenarios. The availability of this dataset supports the development of robust algorithms aimed at reducing accident rates, improving system responses, and ultimately enhancing the safety of AVs on the road.

Limitations. Our RBA module currently focuses on rear-end collisions. We plan to address that by expanding the detection framework to capture additional accident types. Future improvements will enhance the adaptability to detect a broader range of accidents, further strengthening its value as a resource for autonomous safety research.

8. Acknowledgment

This research was supported by the Federal Ministry of Education and Research in Germany within the AUTOTech.agil project, Grant Number: 01IS22088U.

References

- [1] Filippou Christianos, Peter Karkus, Boris Ivanovic, Stefano V. Albrecht, and Marco Pavone. Planning with Occluded Traffic Agents using Bi-Level Variational Occlusion Models, in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pages 5558–5565. doi: [10.1109/ICRA48891.2023.10160604](https://doi.org/10.1109/ICRA48891.2023.10160604).
- [2] Ahmed Ghita, Bjørk Antoniussen, Walter Zimmer, Ross Greer, Christian Creß, Andreas Møgelmoose, Mohan M. Trivedi, and Alois C. Knoll. ActiveAnno3D - An Active Learning Framework for Multi-Modal 3D Object Detection, in *2024 IEEE Intelligent Vehicles Symposium (IV)*, 2024, pages 1699–1706. doi: [10.1109/IV55156.2024.10588452](https://doi.org/10.1109/IV55156.2024.10588452).
- [3] N Kulkarni, A Rangesh, J Buck, J Feltracco, M Trivedi, N Deo, R Greer, S Sarraf, and S Sathyanarayana. Create a large-scale video driving dataset with detailed attributes using Amazon SageMaker Ground Truth, 2021.
- [4] Tim Fingscheidt, Hanno Gottschalk, and Sebastian Houben. *Deep neural networks and data for automated driving: Robustness, uncertainty quantification, and insights towards safety*. Springer Nature, 2022. doi: [10.1007/978-3-031-01233-4](https://doi.org/10.1007/978-3-031-01233-4).
- [5] Hirokatsu Kataoka, Teppei Suzuki, Shoko Oikawa, Yasuhiro Matsui, and Yutaka Satoh. Drive Video Analysis for the Detection of Traffic Near-miss Incidents, in *IEEE Int. Conf. on robotics and automation (ICRA)*, 2018, pages 3421–3428.
- [6] Huanan Wang, Xinyu Zhang, Zhiwei Li, Jun Li, Kun Wang, Zhu Lei, and Ren Haibing. IPS300+: a Challenging multi-modal data sets for Intersection Perception System, in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pages 2539–2545. doi: [10.1109/ICRA46639.2022.9811699](https://doi.org/10.1109/ICRA46639.2022.9811699).
- [7] Ross Greer, Bjørk Antoniussen, Mathias V Andersen, Andreas Møgelmoose, and Mohan M Trivedi. The Why, When, and How to Use Active Learning in Large-Data-Driven 3D Object Detection for Safe Autonomous Driving: An Empirical Exploration, *preprint arXiv:2401.16634*, 2024.
- [8] 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](https://doi.org/10.1109/CVPR52733.2024.02139).
- [9] Xingcheng Zhou, Deyu Fu, Walter Zimmer, Mingyu Liu, Venkatnarayanan Lakshminarasimhan, Leah Strand, and Alois C. Knoll. WARM-3D: A Weakly-Supervised Sim2Real Domain Adaptation Framework for Roadside Monocular 3D Object Detection, in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, 2024, pages 3489–3496. doi: [10.1109/ITSC58415.2024.10919929](https://doi.org/10.1109/ITSC58415.2024.10919929).
- [10] Mingyu Liu, Ekim Yurtsever, Marc Brede, Jun Meng, Walter Zimmer, Xingcheng Zhou, Bare Luka Zagar, Yuning Cui, and Alois C. Knoll. GraphRelate3D: Context-Dependent 3D Object Detection with Inter-Object Relationship Graphs, in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, 2024, pages 3481–3488. doi: [10.1109/ITSC58415.2024.10919972](https://doi.org/10.1109/ITSC58415.2024.10919972).
- [11] Sondas Mohamed, Walter Zimmer, Ross Greer, Ahmed Alaaeldin Ghita, Modesto Castrillón-Santana, Mohan Trivedi, Alois Knoll, Salvatore Mario Carta, and Mirko Marras. Transfer Learning from Simulated to Real Scenes for Monocular 3D Object Detection, in *Computer Vision – ECCV 2024 Workshops*, A. Del Bue, C. Canton, J. Pont-Tuset, and T. Tommasi, Eds., Cham: Springer Nature Switzerland, 2025, pages 309–325. doi: [10.1007/978-3-031-91813-1_20](https://doi.org/10.1007/978-3-031-91813-1_20).
- [12] Salvatore Carta, Modesto Castrillón-Santana, Mirko Marras, Sondas Mohamed, Alessandro Sebastian Podda, Roberto Saia, Marco Sau, and Walter Zimmer. RoadSense3D: A Framework for Roadside Monocular 3D Object Detection, in *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, in UMAP Adjunct '24. Cagliari, Italy: Association for Computing Machinery, 2024, pages 452–459. doi: [10.1145/3631700.3665236](https://doi.org/10.1145/3631700.3665236).
- [13] Walter Zimmer, Christian Creß, Huu Tung Nguyen, and Alois C. Knoll. TUMTraf Intersection Dataset: All You Need for Urban 3D Camera-LiDAR Roadside Perception, in *2023 IEEE 26th Int. Conf. on Intelligent Transportation Systems (ITSC)*, Sep. 2023, pages 1030–1037. doi: [10.1109/ITSC57777.2023.10422289](https://doi.org/10.1109/ITSC57777.2023.10422289).
- [14] Walter Zimmer, Joseph Birkner, Marcel Brucker, Huu Tung Nguyen, Stefan Petrovski, Bohan Wang, and Alois C. Knoll. InfraDet3D: Multi-Modal 3D Object Detection based on Roadside Infrastructure Camera and LiDAR Sensors, in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pages 1–8. doi: [10.1109/IV55152.2023.10186723](https://doi.org/10.1109/IV55152.2023.10186723).
- [15] Walter Zimmer, Jialong Wu, Xingcheng Zhou, and Alois C. Knoll. Real-Time And Robust 3D Object Detection with Roadside LiDARs, in *Proceedings of the 12th International Scientific Conference on Mobility and Transport*, C. Antoniou, F. Busch, A. Rau, and M. Hariharan, Eds., Singapore: Springer Nature Singapore, 2023, pages 199–219. doi: [10.1007/978-981-19-8361-0_13](https://doi.org/10.1007/978-981-19-8361-0_13).
- [16] Walter Zimmer, Emec Ercelik, Xingcheng Zhou, Xavier Jair Diaz Ortiz, and Alois Knoll. A Survey of Robust 3D Object Detection Methods in Point Clouds, *preprint arXiv:2204.00106*, 2022.
- [17] Walter Zimmer, Marcus Grabler, and Alois Knoll. Real-time and robust 3D object detection within road-side LiDARs using domain adaptation, *preprint arXiv:2204.00132*, 2022.
- [18] Mark Philip Philipsen, Morten Bornø Jensen, Andreas Møgelmoose, Thomas B. Moeslund, and Mohan M. Trivedi. Traffic Light Detection: A Learning Algorithm and Evaluations on Challenging Dataset, in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pages 2341–2345. doi: [10.1109/ITSC.2015.378](https://doi.org/10.1109/ITSC.2015.378).
- [19] Hala Abualsaud, Sean Liu, David B. Lu, Kenny Situ, Akshay Rangesh, and Mohan M. Trivedi. LaneAF: Robust Multi-Lane Detection With Affinity Fields, *IEEE Robotics and*

- Automation Letters*, vol. 6, no. 4, pages 7477–7484, 2021, doi: [10.1109/LRA.2021.3098066](https://doi.org/10.1109/LRA.2021.3098066).
- [20] Ross Greer, Akshay Gopalkrishnan, Maitrayee Keskar, and Mohan Trivedi. Patterns of vehicle lights: Addressing complexities of camera-based vehicle light datasets and metrics, *Pattern Recognition Letters*, vol. 178, pages 209–215, 2024.
 - [21] Philipp Kremer, Navid Nourani-Vatani, and Sangyoung Park. A digital twin for teleoperation of vehicles in urban environments, in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2023, pages 12521–12527.
 - [22] Annkathrin Krämer, Christoph Schöller, Dhiraj Gulati, Venkatnarayanan Lakshminarasimhan, Franz Kurz, Dominik Rosenbaum, Claus Lenz, and Alois Knoll. Providentia - A Large-Scale Sensor System for the Assistance of Autonomous Vehicles and Its Evaluation, *Journal of Field Robotics*, pages 1156–1176, Feb. 2022, Accessed: Nov. 15, 2024. [Online]. Available: <https://elib.dlr.de/135631/>
 - [23] Christian Creß, Zhenshan Bing, and Alois C. Knoll. Intelligent Transportation Systems Using Roadside Infrastructure: A Literature Survey, *IEEE Transactions on Intelligent Transportation Systems*, pages 1–0, 2023, doi: [10.1109/TITS.2023.3343434](https://doi.org/10.1109/TITS.2023.3343434).
 - [24] Raphael van Kempen *et al.* AUTotech.agil: Architecture and Technologies for Orchestrating Automotive Agility, *32nd Aachen Colloquium Sustainable Mobility*. RWTH Aachen University, page 49, Oct. 2023. doi: [10.18154/RWTH-2023-09783](https://doi.org/10.18154/RWTH-2023-09783).
 - [25] Ross Greer, Nachiket Deo, Akshay Rangesh, Mohan Trivedi, and Pujitha Gunaratne. Safe Control Transitions: Machine Vision Based Observable Readiness Index and Data-Driven Takeover Time Prediction, in *27th Int. Technical Conf. on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration*, 2023.
 - [26] 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](https://doi.org/10.1109/TIV.2024.3402136).
 - [27] Mingyu Liu, Ekim Yurtsever, Jonathan Fossaert, Xingcheng Zhou, Walter Zimmer, Yuning Cui, Bare Luka Zagar, and Alois C Knoll. A Survey on Autonomous Driving Datasets: Statistics, Annotation Quality, and a Future Outlook, *IEEE Trans. on Intelligent Vehicles*, 2024.
 - [28] Mohammad Sadegh Aliakbarian, Fatemeh Sadat Saleh, Mathieu Salzmann, Basura Fernando, Lars Petersson, and Lars Andersson. VIENA²: A Driving Anticipation Dataset, in *Computer Vision – ACCV 2018*, C. V. Jawahar, H. Li, G. Mori, and K. Schindler, Eds., Cham: Springer Int. Publishing, 2019, pages 449–466. doi: [10.1007/978-3-030-20887-5_28](https://doi.org/10.1007/978-3-030-20887-5_28).
 - [29] Hoon Kim, Kangwook Lee, Gyeongjo Hwang, and Changho Suh. Crash to Not Crash: Learn to Identify Dangerous Vehicles Using a Simulator, *Proceedings of the AAAI Conf. on Artificial Intelligence*, vol. 33, no. 1, pages 978–985, Jul. 2019, doi: [10.1609/aaai.v33i01.3301978](https://doi.org/10.1609/aaai.v33i01.3301978).
 - [30] Thakare Kamalakar Vijay *et al.* Detection of Road Accidents Using Synthetically Generated Multi-Perspective Accident Videos, *IEEE Trans. on Intelligent Transport. Systems*, 2023.
 - [31] Chi-Hsi Kung, Chieh-Chi Yang, Pang-Yuan Pao, Shu-Wei Lu, Pin-Lun Chen, Hsin-Cheng Lu, and Yi-Ting Chen. RiskBench: A Scenario-based Benchmark for Risk Identification, in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pages 14800–14807. doi: [10.1109/ICRA57147.2024.10610270](https://doi.org/10.1109/ICRA57147.2024.10610270).
 - [32] Tianqi Wang, Sukmin Kim, Ji Wenxuan, Enze Xie, Chongjian Ge, Junsong Chen, Zhenguo Li, and Ping Luo. DeepAccident: A Motion and Accident Prediction Benchmark for V2X Autonomous Driving, *Proceedings of the AAAI Conf. on Artificial Intelligence*, vol. 38, no. 6, pages 5599–5606, Mar. 2024, doi: [10.1609/aaai.v38i6.28370](https://doi.org/10.1609/aaai.v38i6.28370).
 - [33] Mehdi Ghatee. Accident Images Analysis Dataset (AIAD). Accessed: Oct. 13, 2024. [Online]. Available: <https://github.com/mghatee/Accident-Images-Analysis-Dataset>
 - [34] Ankit Parag Shah, Jean-Bapstite Lamare, Tuan Nguyen-Anh, and Alexander Hauptmann. CADP: A Novel Dataset for CCTV Traffic Camera based Accident Analysis, in *2018 15th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Nov. 2018, pages 1–9. doi: [10.1109/AVSS.2018.8639160](https://doi.org/10.1109/AVSS.2018.8639160).
 - [35] Tackgeun You and Bohyung Han. Traffic Accident Benchmark for Causality Recognition, in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., Cham: Springer International Publishing, 2020, pages 540–556. doi: [10.1007/978-3-030-58571-6_32](https://doi.org/10.1007/978-3-030-58571-6_32).
 - [36] Wentao Bao, Qi Yu, and Yu Kong. Uncertainty-based Traffic Accident Anticipation with Spatio-Temporal Relational Learning, in *Proceedings of the 28th ACM Int. Conf. on Multimedia*, in MM '20. Seattle, WA, USA: Association for Computing Machinery, 2020, pages 2682–2690. doi: [10.1145/3394171.3413827](https://doi.org/10.1145/3394171.3413827).
 - [37] Charan Kumar. Accident Detection From CCTV Footage. [Online]. Available: <https://www.kaggle.com/dsv/1379553>
 - [38] Khaled Sabry and Mohamed Emad. Road Traffic Accidents Detection Based On Crash Estimation, in *2021 17th Int. Computer Engineering Conf. (ICENCO)*, Dec. 2021, pages 63–68. doi: [10.1109/ICENCO49852.2021.9698968](https://doi.org/10.1109/ICENCO49852.2021.9698968).
 - [39] Hoon Kim, Kangwook Lee, Gyeongjo Hwang, and Changho Suh. Predicting Vehicle Collisions Using Data Collected From Video Games, *Machine Vision and Applications*, vol. 32, no. 4, page 93, Jun. 2021, doi: [10.1007/s00138-021-01217-2](https://doi.org/10.1007/s00138-021-01217-2).
 - [40] Karishma Pawar and Vahida Attar. Deep Learning based Detection and Localization of Road Accidents from Traffic Surveillance Videos, *ICT Express*, vol. 8, no. 3, pages 379–387, Sep. 2022, doi: [10.1016/j.icte.2021.11.004](https://doi.org/10.1016/j.icte.2021.11.004).
 - [41] Yajun Xu, Huan Hu, Chuwen Huang, Yibing Nan, Yuyao Liu, Kai Wang, Zhaoxiang Liu, and Shiguo Lian. TAD: A Large-Scale Benchmark for Traffic Accidents Detection From Video Surveillance, *IEEE Access*, vol. 13, no. , pages 2018–2033, 2025, doi: [10.1109/ACCESS.2024.3522384](https://doi.org/10.1109/ACCESS.2024.3522384).
 - [42] Shubhankar Shandilya. Accident Detection Model (ADM) Dataset. [Online]. Available: <https://universe.roboflow.com/accident-detection-model/accident-detection-model>

- [43] Jianwu Fang, Lei-lei Li, Junfei Zhou, Junbin Xiao, Hongkai Yu, Chen Lv, Jianru Xue, and Tat-Seng Chua. Abductive Ego-View Accident Video Understanding for Safe Driving Perception, in *2024 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2024, pages 22030–22040. doi: [10.1109/CVPR52733.2024.02080](https://doi.org/10.1109/CVPR52733.2024.02080).
- [44] Sachin Khandelwal. YOLO Accident. Accessed: Nov. 13, 2024. [Online]. Available: https://drive.google.com/drive/u/0/folders/1_4p1mW0BzDX0ZoHdaMIU5c7Y0ni0Kz02?direction=a
- [45] Xingyuan Chen, Huahu Xu, Mingyang Ruan, Minjie Bian, Qishen Chen, and Yuzhe Huang. SO-TAD: A Surveillance-oriented Benchmark for Traffic Accident Detection, *Neuro-computing*, vol. 618, page 129061, 2025, doi: <https://doi.org/10.1016/j.neucom.2024.129061>.
- [46] Jianwu Fang, Jiahuan Qiao, Jianru Xue, and Zhengguo Li. Vision-Based Traffic Accident Detection and Anticipation: A Survey, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pages 1983–1999, 2024, doi: [10.1109/TCSVT.2023.3307655](https://doi.org/10.1109/TCSVT.2023.3307655).
- [47] Yu Yao, Mingze Xu, Yuchen Wang, David J. Crandall, and Ella M. Atkins. Unsupervised Traffic Accident Detection in First-Person Videos, in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pages 273–280. doi: [10.1109/IROS40897.2019.8967556](https://doi.org/10.1109/IROS40897.2019.8967556).
- [48] P. Mehrannia, S. S. G. Bagi, B. Moshiri, and O. Basir. Deep representation of imbalanced spatio-temporal traffic flow data for traffic accident detection, *IET Intelligent Transport Systems*, vol. 17, no. 3, pages 606–619, 2022, doi: [10.1049/itr2.12287](https://doi.org/10.1049/itr2.12287).
- [49] Yu Yao, Xizi Wang, Mingze Xu, Zelin Pu, Yuchen Wang, Ella Atkins, and David J. Crandall. DoTA: Unsupervised Detection of Traffic Anomaly in Driving Videos, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pages 444–459, Jan. 2023, doi: [10.1109/TPAMI.2022.3150763](https://doi.org/10.1109/TPAMI.2022.3150763).
- [50] Daniel Bogdoll, Jan Imhof, Tim Joseph, Svetlana Pavlitska, and J Marius Zöllner. Hybrid Video Anomaly Detection for Autonomous Driving, *RROW Workshop at the British Machine Vision Conference (BMVC) 2024*, 2024.
- [51] Fu-Hsiang Chan, Yu-Ting Chen, Yu Xiang, and Min Sun. Anticipating Accidents in Dashcam Videos, in *Computer Vision – ACCV 2016*, S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds., Cham: Springer Int. Publishing, 2017, pages 136–153. doi: [10.1007/978-3-319-54190-7_9](https://doi.org/10.1007/978-3-319-54190-7_9).
- [52] Wentao Bao, Qi Yu, and Yu Kong. DRIVE: Deep Reinforced Accident Anticipation with Visual Explanation, in *2021 IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, Oct. 2021, pages 7599–7608. doi: [10.1109/ICCV48922.2021.00752](https://doi.org/10.1109/ICCV48922.2021.00752).
- [53] Glenn Jocher, Jing Qiu, and Ayush Chaurasia. Ultralytics YOLO. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [54] Walter Zimmer, Akshay Rangesh, and Mohan Trivedi. 3D BAT: A Semi-Automatic, Web-based 3D Annotation Toolbox for Full-Surround, Multi-Modal Data Streams, in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pages 1816–1821. doi: [10.1109/IVS.2019.8814071](https://doi.org/10.1109/IVS.2019.8814071).
- [55] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors, in *2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pages 7464–7475. doi: [10.1109/CVPR52729.2023.00721](https://doi.org/10.1109/CVPR52729.2023.00721).
- [56] Xiaoyu Li, Tao Xie, Dedong Liu, Jinghan Gao, Kun Dai, Zhiqiang Jiang, Lijun Zhao, and Ke Wang. Poly-MOT: A Polyhedral Framework For 3D Multi-Object Tracking, in *2023 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2023, pages 9391–9398. doi: [10.1109/IROS55552.2023.10341778](https://doi.org/10.1109/IROS55552.2023.10341778).
- [57] Nicola Croce. OpenLABEL Concept Paper. Accessed: May 14, 2024. [Online]. Available: <https://www.asam.net/index.php?eID=dumpFile&t=f&f=3876&token=413e8c85031ae64cc35cf42d0768627514868b2f>
- [58] 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](https://doi.org/10.1109/CVPR.2012.6248074).
- [59] 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](https://doi.org/10.1177/0278364913491297).
- [60] 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 Multi-modal 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](https://doi.org/10.1109/CVPR42600.2020.01164).
- [61] 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](https://doi.org/10.1109/CVPR42600.2020.00252).
- [62] Cheng Li, Keyuan Zhou, Tong Liu, Yu Wang, Mingqiao Zhuang, Huan-ang Gao, Bu Jin, and Hao Zhao. AVD2: Accident Video Diffusion for Accident Video Description, *Accepted for Int. Conf. on Robotics and Automation*, 2025.
- [63] Yuping Wang, Shuo Xing, Cui Can, Renjie Li, Hongyuan Hua, Kexin Tian, Zhaobin Mo, Xiangbo Gao, Keshu Wu, Su-long Zhou, Hengxu You, Juntong Peng, Junge Zhang, Zehao Wang, Rui Song, Mingxuan Yan, Walter Zimmer, Xingcheng Zhou, Peiran Li, Zhaohan Lu, Chia-Ju Chen, Yue Huang, Ryan A. Rossi, Lichao Sun, Hongkai Yu, Zhiwen Fan, Frank Hao Yang, Yuhao Kang, Ross Greer, Chenxi Liu, Eun Hak Lee, Xuan Di, Xinyue Ye, Liu Ren, Alois Knoll, Xiaopeng Li, Shuiwang Ji, Masayoshi Tomizuka, Marco Pavone, Tianbao Yang, Jing Du, Ming-Hsuan Yang, Hua Wei, Ziran Wang, Yang Zhou, Jiachen Li, and Zhengzhong Tu. Generative AI for Autonomous Driving: Frontiers and Opportunities. Accessed: Jun. 30, 2025. [Online]. Available: <http://arxiv.org/abs/2505.08854>