SDAC: A Multimodal Synthetic Dataset for Anomaly and Corner Case Detection in Autonomous Driving

Lei Gong[1], **Yu Zhang**[1, 2*], **Yingqing Xia**[1], **Yanyong Zhang**[1,2], **Jianmin Ji**[1, 2]

[1]School of Computer Science and Technology, University of Science and Technology of China, Hefei, China
[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China
gleisa19@mail.ustc.edu.cn, yuzhang@ustc.edu.cn, xiayingqing@mail.ustc.edu.cn, {yanyongz, jianmin}@ustc.edu.cn

Abstract

Nowadays, closed-set perception methods for autonomous driving perform well on datasets containing normal scenes. However, they still struggle to handle anomalies in the real world, such as unknown objects that have never been seen while training. The lack of public datasets to evaluate the model performance on anomaly and corner cases has hindered the development of reliable autonomous driving systems. Therefore, we propose a multimodal **S**ynthetic **D**ataset for **A**nomaly and **C**orner case detection, called SDAC, which encompasses anomalies captured from multi-view cameras and the LiDAR sensor, providing a rich set of annotations for multiple mainstream perception tasks. SDAC is the first public dataset for autonomous driving that categorizes anomalies into object, scene, and scenario levels, allowing the evaluation under different anomalous conditions. Experiments show that closed-set models suffer significant performance drops on anomaly subsets in SDAC. Existing anomaly detection methods fail to achieve satisfactory performance, suggesting that anomaly detection remains a challenging problem. We anticipate that our SDAC dataset could foster the development of safe and reliable systems for autonomous driving.

Introduction

In recent years, Deep Neural Networks (DNNs) have achieved remarkable performance in various perception tasks on public autonomous driving datasets such as KITTI (Geiger, Lenz, and Urtasun 2012), nuScenes (Caesar et al. 2020) and Cityscapes (Cordts et al. 2016). Closed-set detection models are designed based on the assumption that all classes and situations in testing would be available during training. However, in open-world road scenes, autonomous vehicles may encounter anomalies and corner cases such as unknown objects or severe weather conditions that have never been seen while training. Closed-set methods still struggle to handle unexpected situations in real-world environments, which may lead to devastating consequences in some situations. For instance, Figure 1 depicts an anomaly example where an overturned white bus is located straight ahead of the road. A DeepLabv3+ (Chen et al. 2018) model trained on normal scenes without bus class labels incorrectly identifies the overturned white bus as background buildings, posing potential risks to autonomous vehicles in real-world scenarios. Therefore, to build a safe and reliable self-driving system, the perception modules should be capable of detecting anomalies and corner cases. Research in this domain is commonly named anomaly detection (Bogdoll, Nitsche, and Zöllner 2022), or open-set detection (Scheirer et al. 2012).

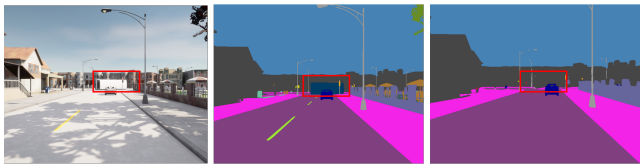


Figure 1: A DeepLabv3+ model trained without the bus class label incorrectly labels the overturned white bus as background buildings. From left to right are the RGB image, ground truth, and detection results.

Anomaly detection in autonomous driving is still in its early stages. One main reason that impedes the development is the lack of public datasets for performance evaluation. Table 1 lists publicly available datasets for anomaly detection in autonomous driving. They are primarily created for vision-based systems and provide either semantic masks or bounding boxes from images, which restricts the advancement of 3D anomaly detection techniques. Regarding the core concept, the terms "anomaly" and "corner case" are often used interchangeably, and there is no consistent definition across different datasets. A systematization of corner cases (Breitenstein et al. 2020) for visual perception in autonomous driving divides anomalies into different levels based on the theoretical complexity of their detection. Inspired by it, we propose to categorize anomalies into object, scene, and scenario levels: **object-level** anomalies represent unknown objects, **scene-level** anomalies contain unexpected patterns of objects and domain shifts (Ben-David et al. 2010), **scenario-level** corner cases denote patterns with temporal context that contain the potential for collision. According to the definition, datasets listed in Table 1 mainly focus on the object and scene level anomalies, and anomalies in these datasets are conflated due to the absence of systematization of corner cases. Furthermore, they do not provide scenario-level corner cases.

*Corresponding author

Dataset	Subsets	Sensors	Labels	Tasks	Size
Fishyscapes (Blum et al. 2021)	Lost & Found Static	Camera	Semantic Mask	Segmentation	375 1030
CAOS (Hendrycks et al. 2022)	StreetHazards BDD-Anomaly	Camera	Semantic Mask	Segmentation	1500 810
SegmentMeIfYouCan (Chan et al. 2021)	RoadAnomaly21 RoadObstacle21	Camera	Semantic Mask	Segmentation	100 327
CODA (Li et al. 2022)	CODA-KITTI CODA-nuScenes CODA-ONCE	Camera, LiDAR	Bounding Boxes	Object Detection	309 134 1057

Table 1: A number of public and available datasets for anomaly detection in autonomous driving.

To mitigate the above setbacks, we introduce SDAC, a multimodal synthetic dataset for anomaly and corner case detection in autonomous driving. Collected in the CARLA simulation environment (Dosovitskiy et al. 2017), SDAC provides data captured from multi-view RGB cameras and a 64-channel LiDAR sensor. Annotations in SDAC include 2D/3D bounding boxes as well as semantic/instance segmentation masks, enabling SDAC to support a variety of mainstream perception tasks for self-driving systems. SDAC is designed to be a comprehensive dataset that covers object, scene, and scenario-level corner cases. It is composed of two parts: SDAC-SNAP and SDAC-SCENARIO. SDAC-SNAP provides object and scene-level anomalies, which contains 5 different subsets: **Normal, TruckBus, Props, RainFog,** and **Accident**. Different anomaly subsets focus on different specific anomalous conditions. SDAC-SCENARIO offers the temporal sequences of pre-crash and crash scenarios, aiming to establish a common research foundation for scenario understanding and crash avoidance.

To the best of our knowledge, SDAC is the first public dataset that categorizes anomalies systematically into different levels for anomaly detection in autonomous driving, allowing the comprehensive evaluations of both closed-set and anomaly detection techniques across a diverse range of anomalous conditions. We evaluated both closed-set detection methods based on different modalities as well as anomaly detection methods on SDAC. Experimental results show that closed-set detection methods suffer significant performance degradation. Anomaly detection methods also fail to achieve satisfactory results in unknown identification, indicating that it still remains a challenging problem. We hope SDAC can help to facilitate the development of robust and reliable autonomous driving systems.

Related Work

Anomaly Datasets in Autonomous Driving

A number of public and available datasets for anomaly detection are listed in Table 1. The Fishyscapes benchmark (2021) is proposed for anomaly segmentation, which evaluates pixel-wise uncertainty estimates toward the detection of anomalous objects in front of the vehicle. Synthetic images in Fishyscapes are generated by overlaying images from Cityscapes with objects from Pascal VOC (Everingham et al. 2010) or crawled from the internet. The CAOS (2022) benchmark includes two datasets named StreetHazards and BDD-Anomaly. StreetHazards is a synthetic dataset based on the CARLA simulation environment. The anomalies are taken from the external model library and then inserted into the driving scenes. BDD-Anomaly considers infrequent object classes (*motorcycle, train, bicycle*) in the BDD100K (Yu et al. 2020) dataset as anomalies. Unlike the above two benchmarks that rely on synthetic images, the SegmentMeIfYouCan benchmark (2021) introduces two tracks entirely collected in the real world. The anomaly track RoadAnomaly21 is collected from web resources, in which anomalies can appear anywhere in the image. The obstacle track RoadObstacle21 focuses on the task of obstacle segmentation, whose goal is to identify all objects on the road. Different from the datasets mentioned above for anomaly segmentation and providing only images and semantic masks, CODA (2022) is a novel dataset of object-level corner cases designed for object detection. It employed a two-stage selection process to choose corner cases from KITTI, nuScenes, and ONCE (Mao et al. 2021) benchmarks for autonomous driving. While CODA is collected from datasets that include point cloud data, it only provides 2D anomaly bounding boxes in image space.

Anomaly Detection Methods

Anomaly detection approaches can be classified into different concepts, such as: "reconstruction, prediction, generative, confidence scores, and feature extraction." (Bogdoll, Nitsche, and Zöllner 2022). On the object level, two main perception tasks are object detection and semantic segmentation. Open-set or open-world object detection methods (Joseph et al. 2021; Gupta et al. 2022; Han et al. 2022; Zohar, Wang, and Yeung 2023) detect unknown objects without explicit supervision. They usually integrate their frameworks into traditional models, e.g. Faster R-CNN (Ren et al. 2015), Deformable DETR (Zhu et al. 2020), adapting them for the open-world setting. In the segmentation task, pixels belonging to unknown objects are associated with unknown class labels or high scores that indicate uncertainty. Uncertainty estimation methods could exploit the maximum softmax probability (Hendrycks and Gimpel 2016), max logits (Hendrycks et al. 2019), or standardized max logits (Jung et al. 2021) as anomaly score. Reconstructive approaches (Lis et al. 2019; Di Biase et al. 2021) leverage image re-synthesis and detect anomalies by measuring the dis-

similarity between the input and synthesized image. Generative approaches such as NFlowJS (Grcić et al. 2021) generate synthetic negative patches atop regular images as the anomaly mask.

On the scene level, collecting data from all possible domains is impractical in the open world. Therefore, measuring domain shifts during inference is critical to autonomous driving safety. Various domain mismatch metrics (Lohdefink et al. 2020; Zhang et al. 2018; Breitenstein et al. 2021) have been proposed to identify these scene-level anomalies.

For scenario-level corner case detection, the prediction-based approaches (Liu et al. 2018; Bolte et al. 2019) predict a future frame and compare it with the true frame to detect anomalies. Reconstructive method, for instance, SelfOracle (Stocco et al. 2020) uses autoencoder and time-series-based anomaly detection to reconstruct the driving scenarios seen by the car, predicting potential safety-critical misbehaviors.

SDAC Dataset

Properties of SDAC

SDAC is a multimodal dataset. Most recent advances are concerned with image-based anomaly detection, while 3D approaches are still struggling to gain momentum. One reason for this is the absence of datasets, which so far only exist in the camera sector. Therefore, SDAC provides not only images but also point clouds and 3D labels to facilitate the development of 3D anomaly detection techniques.

SDAC is a synthetic dataset generated in a CARLA simulation environment. Acquiring corner case scenarios in the real world is problematic as these cases are rare and also costly to annotate. Furthermore, due to the high risk involved, it is not feasible to record data in pre-crash or crash scenarios. However, in simulation environments, we can construct these crash scenarios in a safe manner and collect data from simulated sensors.

SDAC categorizes anomalies into object, scene, and scenario levels. The **object-level** anomalies represent unknown objects that have never been seen in training. Autonomous driving systems operate in the open world that contains a potentially infinite set of unknown classes. Miss-detection of anomalous obstacles, as shown in Figure 1, could result in serious consequences. **Scene-level** anomalies are defined as domain shifts in the environment and unexpected patterns of objects. Domain shift refers to the situation where the data distribution of training and testing domains differ significantly. Examples of domain shifts include changes in weather and lighting conditions, such as rain or fog. Unexpected patterns of objects refer to situations where objects appear in unusual locations or positions. For example, Figure 2d depicts an accident scene where the vehicle is overturned and a person is lying on the road. **Scenario-level** corner case denotes the pattern with temporal context. These patterns represent specific types of risky or anomalous situations that can occur during driving. Figure 2e illustrates an example with the pattern "something suddenly runs into the road". Those corner cases capture sequences of events that unfold over time, allowing researchers to gain insights into the potential risks and develop strategies to enhance crash avoidance.

Subset	Weather	Classes	Size
Normal	clear	normal*	2657‡
RainFog	rain,fog	normal	396
TruckBus	clear	truck,bus	377
Props	clear	unknown	375
Accident	clear,rain,fog	car,person,truck,bus	297

Table 2: Subsets in SDAC-SNAP. The Normal subset contains training and validation splits, while the others are for testing. *Normal classes include *car, person, motorcycle, bicycle*. ‡the Size of the Normal subset includes 1861 for training and 796 for testing.

Overview of SDAC

The SDAC sensor suite includes six cameras and a 64-channel LiDAR sensor. All cameras have a field of view (FoV) of 110° and a resolution of 2048×1024 pixel. The LiDAR sensor has a vertical Fov of range $[-24.8^\circ, 4^\circ]$, 80m maximum distance, and a scan rate of about 1.3M points per second. More details on sensor configuration and generation of SDAC are available on our website[1].

SDAC is composed of two parts. SDAC-SNAP provides RGB images from the front camera and point clouds, with annotations including 2D/3D bounding boxes as well as semantic/instance segmentation masks. There are 7 different categories in object detection tasks: *car, person, motorcycle, bicycle, truck, bus, unknown*. As for semantic segmentation, there are a total of 21 different classes in SDAC, with the first 19 categories being the same as those in Cityscapes. The last two categories, "Static" and "Dynamic", represent the obstacles in the Props subset.

In SDAC-SCENARIO, records of corner cases are grouped by patterns. Each record is about 10 seconds long, and all sensors are synchronized and captured in 10HZ. Objects hold the same id throughout the sequence. In most records, we provide the front camera images and point clouds as well as associated annotations. For scenarios such as a vehicle is going to merge from the entrance as shown in Figure 3, we provide images from 6 cameras, providing a 360° Fov and the box labels in each view.

SDAC-SNAP

Table 2 provides an overview of subsets in SDAC-SNAP. we introduce each of them in detail.

Normal is captured in the clear sunny environment with instances of normal classes. SDAC treats *car, motorcycle, bicycle* and *person* as normal classes. Existing anomaly datasets do not contain this subset since they consider datasets such as Cityscapes to be normal. In the open-world environment, it is unrealistic to assume what kind of anomaly the system might encounter. Therefore, anomaly detection techniques are generally required to be developed

[1] https://sdac-dataset.github.io/

(a) TruckBus (Object Level)

(b) Props (Object Level)

(c) RainFog (Scene Level)

(d) Accident (Object, Scene Level)

(e) A Scenario with the pattern "something suddenly run into the road" in SDAC-Scenario (Scenario Level)

Figure 2: Examples from different anomaly subsets in SDAC dataset.

Figure 3: A scenario where a vehicle is going to merge from the entrance ramp. From top left to right bottom are images captured from FrontLeft, Front, FrontRight, BackLeft, Back, and BackRight cameras. The vehicle is invisible in the Front view but can be seen in the FrontRight view.

based on the known "normal conditions" to gain generalizability to different corner cases. Thus, the Normal subset is indispensable and crucial for the SDAC dataset.

RainFog contains scene-level anomalies focusing on domain shifts in weather conditions. Autonomous driving systems are often designed on a specific Operational Design Domain (ODD) (ISO 2022), which refers to the specific set of conditions and operational boundaries in which an autonomous vehicle is intended to operate safely. For most self-driving systems, the weather condition in ODD is a clear sunny day. However, in the open world, weather conditions may change rapidly and turn rainy. Therefore, to investigate the impact of weather changes on DNNs performance, this subset collects driving scenes in rainy and foggy weather, holding the same class configuration as the Normal subset.

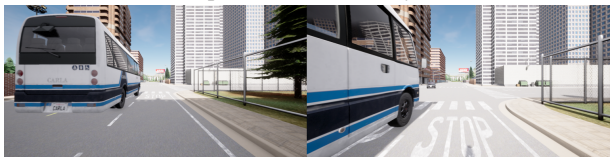
TruckBus and **Props** focus on object-level anomalies. They only contain objects of novel classes that have not been seen in Normal. TruckBus includes two novel classes: *truck* and *bus*. This subset is designed to investigate the model performance when the boundary of known and unknown classes is ambiguous, e.g. cars and trucks. We generate one single object within the drivable area in each frame. The Props subset contains various types of obstacles (e.g., box, traffic cone) that are labeled as *unknown*. Each frame contains one to three obstacles randomly placed on the road ahead of the ego vehicle. In contrast to TruckBus, obstacles in Props are characterized by their small size and limited number of point clouds, which significantly increases the detection complexity for anomaly detection algorithms.

The **Accident** is designed to simulate accident scenes in which objects appear in unusual locations or positions under different weather conditions. As shown in Figure 2d, scenes in this subset consist of an overturned vehicle and a person lying on the road positioned directly in front of the path of the ego vehicle. By mimicking real-world accident scenes, this subset introduces a combination of object and scene-level anomalies, which bring great challenges to the development of robust anomaly detection methods.

SDAC-SCENARIO

SDAC-SCENARIO focuses on pre-crash and crash scenarios. Critical scenario identification techniques (Zhang et al. 2022) have seen rapid development in support of safety research on autonomous driving systems. Safety-critical driving scenario typologies (Najm et al. 2007) provide valuable insights into the factors contributing to crashes, aiding in understanding crash patterns and developing safety measures. Recently, scenario-based test methods and standards (e.g., the under-development ISO 3450 series (ISO 2022) standards) have increasingly played a crucial role in evaluating the performance of autonomous driving systems. Motivated by this, SDAC-SCENARIO focuses on providing the data

(a) a person is obscured by bus



(b) a car in the junction area is obscured by bus

Figure 4: Different specific scenarios under the "visual obstruction" pattern where an obscured object suddenly appears in the view of ego vehicle.

captured in the pre-crash or crash scenarios to establish a common research foundation for scenario-level anomaly detection.

In (Menzel, Bagschik, and Maurer 2018), scenario representations are categorized into three levels of abstraction: functional scenario, logical scenario, and concrete scenario, ranging from semantic-level descriptions to concrete parameterized representations. We employ a similar approach that uses the *pattern*, *specific scenario*, and *concrete scenario* to describe corner cases. In SDAC-SCENARIO, corner cases are classified into distinct patterns, each named to directly reflect the causes, triggering conditions, contributing factors, or safety-critical operations associated with a crash. Specific scenarios within each pattern vary depending on the locations and activities of the ego vehicle and other objects prior to the critical event. Location refers to the agent's position on the road network, while activity represents the behavior, state, or action prior to the occurrence of a crash. To generate concrete instances of specific scenarios, we parameterize the scenarios and explore the parameter space to identify possible settings that can lead to dangerous situations. We use Scenic (Fremont et al. 2022), a domain-specific language for describing distributions over scenes and the behaviors of their agents over time, to generate these concrete scenarios in the CARLA simulation environment.

Taking the *pattern* "visual obstruction" as an example, it represents scenarios where the visibility of the ego vehicle is compromised due to the presence of occlusions. There are two different *specific scenario* examples as depicted in Figure 4. The specific scenario in Figure 4a is described as follows: ego vehicle is following the inner lane of a double-lane road, and a bus is parked ahead in the adjacent lane. A person obstructed by the bus is going to walk across the road. Core parameters in the parameterized scenario include the distance from the ego to the bus along the lane, the ego vehicle's speed, the person's speed, and the distance between the person and the ego vehicle when the person begins to walk.

The SDAC-SCENARIO subset currently comprises seven distinct patterns, each of which we will introduce briefly.

Something suddenly runs into the road. Unexpected object or entity, such as children, suddenly enters the road.

Accident ahead. A visible accident or collision ahead of the ego vehicle's path.

Visual obstruction. An obscured object suddenly appears in the view of the ego vehicle, leaving insufficient time for the ego vehicle to react with necessary safety measures.

Run the red light. A vehicle intentionally or unintentionally disregards a red traffic signal and proceeds through the intersection without stopping.

Lane changes suddenly. A vehicle suddenly changes the lane and drives into the current path of the ego vehicle.

Lead car brakes/slows down suddenly. Vehicle ahead of the ego vehicle suddenly applies the brakes or significantly reduces its speed.

Fails to Yield at Non-Signalized/Signalized Junctions. Ego fails to yield the right of way to another vehicle or pedestrian at intersections, crosswalks, or entrance/exit ramps.

Evaluation

We evaluated both closed-set detection methods and state-of-the-art open-set object detectors and anomaly segmentation methods on different anomaly subsets in SDAC. More details about the experiments can be found in the supplementary material.

For Closed-set detection task, we evaluated image-based, Lidar-based, and multimodal fusion methods to demonstrate the performance degradation of closed-set detection methods based on different modalities when they are exposed to the open-world environment. For image-based approaches, we selected anchor-based two-stage (Faster R-CNN), anchor-free one-stage (CenterNet (Zhou, Wang, and Krähenbühl 2019)), and Transformer-based (Deformable DETR (Zhu et al. 2020)) models. We trained these models on the Normal training set and evaluated their performance on the Normal validation set, RainFog, and Accident subset, employing COCO (Lin et al. 2014) average precision (AP) as the evaluation metric. AP is calculated by computing precision and recall values at different IoU thresholds (e.g., 0.5, 0.75, 0.95) and then averaging them. For Lidar-based 3D object detectors, we evaluated point-based (PointRCNN (Shi, Wang, and Li 2019)) and point-voxel integrated networks (PV-RCNN (Shi et al. 2020)) using metrics from the KITTI benchmark. Furthermore, we evaluated fusion approaches such as MVX-Net (Sindagi, Zhou, and Tuzel 2019) which combine the RGB images and point clouds. For semantic segmentation, we test PSPNet (Zhao et al. 2017), DeepLabV3+ (2018) and report the Cityscapes intersection over union (IoU) metrics for different classes.

For Open-set object detection, we evaluated ORE (Joseph et al. 2021) and OpenDet (Han et al. 2022) on SDAC. ORE is an open-world object detector using contrastive clustering and energy-based unknown identification. The energy-based identifier requires known and extra unknown data to fit Weibull distributions on the energy values. These learned distributions are then applied during inference to identify unknown samples. We trained ORE on the Normal training set and then combined the Normal validation set with data in TruckBus and Props, respectively, to model known and

Method	Normal	RainFog	Accident	
	mAP	mAP	AP_c	AP_p
Faster R-CNN	**59.5**	35.6	32.4	0.2
CenterNet	**47.0**	29.3	29.0	0.2
D-DETR	**52.6**	33.0	34.6	0.3

(a) 2D object detection results on SDAC. mAP(↑) is the COCO mean average precision. AP_c(↑) and AP_p(↑) represent the average precision of *car* and *person* classes.

Method	Modality	Normal	RainFog
		AP_M	AP_M
PointRCNN	LiDAR	**81.8**	81.6
PV-RCNN	LiDAR	**92.1**	92.0
MVX-Net	Camera, LiDAR	**90.1**	90.0

(b) 3D object detection results on SDAC. AP_M(↑) denotes the mean average precision of 3D box detection with 40 recall positions (AP_{40}) on the moderate level of KITTI-sytle validation set.

Classes	PSPNet			DeepLabV3+		
	IoU_N	IoU_R	IoU_A	IoU_N	IoU_R	IoU_A
car	**93.7**	29.2	37.6	**94.3**	15.4	28.2
person	**82.2**	67.5	3.4	**83.7**	71.9	2.1

(c) Semantic segmentation results on SDAC. IoU_N(↑), IoU_R(↑), IoU_A(↑) denote the Cityscapes intersection over union of corresponding class on Normal, RainFog, and Accident.

Table 3: Experimental results of closed-set detection methods. Models are all trained on the Normal training set.

Model	Extra Data*	Normal	TruckBus		Props	
		AP_k	AP_u	A-§	AP_u	A-
ORE	T	80.7	0.04	284	0.02	79
	P	80.7	0.04	284	0.0	77
OpenDet	-	81.1	**55.5**	**252**	**5.3**	**51**

Table 4: Performance of open-set object detectors on SDAC. * denotes extra data used in training, **T** is TruckBus, **P** is Props. AP_k(↑) is the Pascal VOC AP_{50} of known classes. AP_u(↑) denotes the AP_{50} of unknown class. §(↓) is a short representation of Absolute Open-Set Error(A-OSE).

unknown energy distributions. Then we evaluated the model with the utilization of the energy-based identifier on both TruckBus and Props subsets. Different from ORE, OpenDet is trained with only close-set data without complex post-processing. They propose Contrastive Feature Learner (CFL) and Unknown Probability Learner (UPL) to identify unknown objects by separating high/low-density regions in the latent space. The metrics AP_u and Absolute Open-Set Error (A-OSE) are used to quantify how effective the model is in handling unknown objects, where A-OSE denotes the number count of unknown objects that get wrongly classified as any of the known classes.

For anomaly segmentation task, we evaluated SML (Jung et al. 2021) and PEBAL (Tian et al. 2022). SML is one of the confidence score-based methods that do not rely on external data, re-training, or extra networks. It standardizes the max logits in a class-wise manner that reduces the false positives

Model	Normal		TruckBus		Props	
	$mIoU_1$	$mIoU_2$	AUC	FPR_{95}	AUC	FPR_{95}
SML	**85.2**	-	86.9	76.3	89.0	73.4
PEBAL	81.6	80.5	**96.2**	**18.4**	**95.4**	**23.1**

Table 5: Anomaly segmentation performance on different subsets. $mIoU_1$(↑) is the Cityscapes mIoU on known classes of pre-trained segmentation models. $mIoU_2$(↑) is the mIoU on known classes after retraining or fine-tuning. AUC(↑) is the area under receiver operating characteristics (AUROC). FPR_{95}(↓) denotes the false positive rate (FPR) at 95% true positive rate (TPR).

Model	Accident-Car		Accident-Person	
	AP_1	AP_2	AP_1	AP_2
Faster R-CNN	32.4	**45.5**	0.2	**21.7**
CenterNet	29.0	**51.2**	0.2	**7.2**
D-DETR	34.6	**36.2**	0.3	**6.1**

Table 6: Comparison of AP between original training configurations (AP_1) and with additional data augmentations (AP_2). Additional data augmentation, including vertical and diagonal flips and rotation.

when using a single threshold. PEBAL does not explore uncertainty measures and instead explicitly learns a new pixel-wise anomaly class. They fine-tuned the model using both inlier and outlier datasets $\mathcal{D}^{in}$ and $\mathcal{D}^{out}$. Following their configurations, we take the Normal training set as $\mathcal{D}^{in}$ and COCO as $\mathcal{D}^{out}$. The performance of anomalous pixel identification is quantified by the area under the receiver operating characteristics (AUROC) and the false positive rate at 95% true positive rate (FPR_{95}).

Discussion of Results

Experimental results of closed-set detection methods in Table 3 reveal that closed-set detection models are inadequate to meet safety requirements in the open world.

As shown in Table 3a and Table 3c, weather changes bring significant performance decline to the detection models on image space. While employing domain adaptation techniques may be an intuitive idea to address domain shift issues, these techniques are typically designed to adapt models to a specific target domain, which restricts their ability to generalize to diverse unseen conditions in the open world.

Closed-set detection models are highly sensitive to postural variation. Evaluation results on the Accident subset in Table 3a and Table 3c indicate that these models face significant challenges in accurately identifying unexpected patterns, such as people lying on the road. This difficulty arises because the training set only includes examples of persons in a standing posture. Some available remedies, such as data augmentation techniques as shown in Table 6, only have limited effects on certain models (e.g., Faster R-CNN). Besides, we cannot expect what kind of patterns can emerge in the open world.

As shown in Table 3b, 3D detection models exhibit better robustness and are less affected by weather changes. We

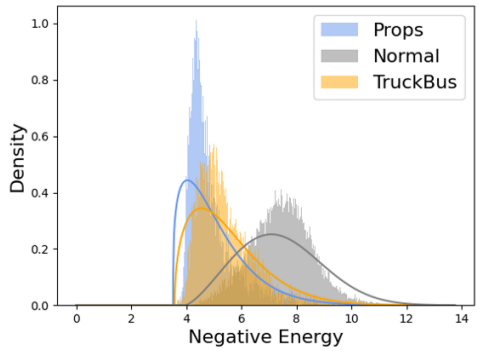


Figure 5: The energy distributions of known and unknown objects in SDAC dataset computed by ORE.

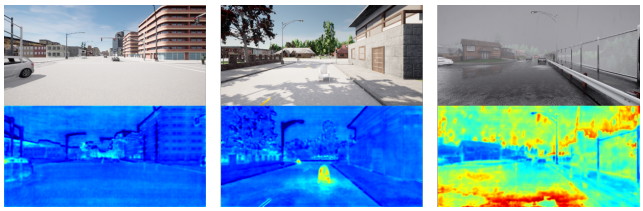


Figure 6: From left to right are anomaly segmentation results of SML on Normal, Props, and RainFog. The anomaly map has high scores (in yellow and red) for anomalous pixels.

did not evaluate 3D detection methods on Accident because these models are commonly based on the flat-world assumption, which means that they only consider encoding heading angles and eliminate the roll and pitch angles. Nevertheless, this assumption does not always hold in the real world, such as when driving through slopes or encountering accident scenes like those in the Accident subset. Therefore, there is a great need for full-degree-of-freedom object detectors and corresponding evaluation metrics.

Experimental results in Table 4 and Table 5 highlight the complexity and value of SDAC in enabling the comprehensive evaluations and analyses of models.

As shown in Table 4, ORE failed to recognize unknown objects in both TruckBus and Props subset, despite the clear distribution difference of energy between known and unknown objects as depicted in Figure 5. The primary reason could be that the proposals used to compute energy for unknown objects during training are selected based on IoU with unknown ground truth. However, most of these proposals will be discarded in the testing phase due to their low confidence scores because they are only used to calculate energy and not to update the parameters of the identifier during training. For unknown objects that have an ambiguous boundary of known classes such as trucks, the energy-based identifier fails to distinguish them from cars. In contrast, OpenDet can clearly distinguish between trucks and cars. This could be due to the CFL they proposed, which can compact features of known classes, leaving more low-density regions for unknown classes. However, it is also a challenge for OpenDet to detect obstacles in the Props subset.

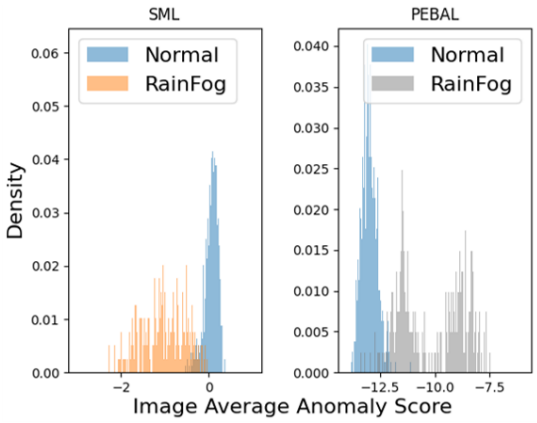


Figure 7: The distribution of average anomaly score of images in the Normal and RainFog subset.

Table 5 reports the results of anomaly segmentation models. SML and PEBAL are based on different methodologies, each having its strengths and weaknesses. SML does not use any external data, re-training, or extra networks, making them widely applicable to existing pre-trained segmentation models. As a result, its performance could be greatly influenced by the pre-trained model. Although PEBAL used the extra dataset (COCO) that has a significant domain gap between SDAC to train the model, it still achieved better performance. However, its performance also depends on the external data used in training. The retraining phase in OpenDet could somehow compromise the accuracy of known objects. We performed additional experiments that calculated the average anomaly scores of images in the RainFog subset and compared the distribution with that in Normal. As shown in Figure 6 and Figure 7, distributions in the RainFog and Normal subsets exhibit significant divergence. This indirectly indicates that semantic segmentation models are more susceptible to weather changes but also underscores the potential of anomaly segmentation methods in detecting different levels of anomaly in the open world.

Conclusion

In this work, we present SDAC, a multimodal synthetic dataset for anomaly and corner case detection in autonomous driving, which is composed of object-level, scene-level, and scenario-level anomalies. Experiments on closed-set detection methods reveal that they are unable to achieve reliable results in the open world, posing safety issues for autonomous vehicles. We further provide a comparison of different anomaly detection methods, demonstrating how the SDAC dataset facilitates a detailed analysis of these techniques. So far, the Lidar-based and scenario-level anomaly detection approaches are still struggling to gain momentum because of the absence of datasets. We anticipate that our SDAC dataset could foster the development of safe and reliable perception methods for autonomous driving.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.62272434) and the Anhui Provincial Key R&D Program Standardization Project (No. 202104h04020039).

References

Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Vaughan, J. W. 2010. A theory of learning from different domains. *Machine learning*, 79: 151–175.

Blum, H.; Sarlin, P.-E.; Nieto, J.; Siegwart, R.; and Cadena, C. 2021. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *Int J Comput Vis*, 129: 3119–3135.

Bogdoll, D.; Nitsche, M.; and Zöllner, J. M. 2022. Anomaly detection in autonomous driving: A survey. In *Proc. of CVPR*, 4488–4499.

Bolte, J.-A.; Bar, A.; Lipinski, D.; and Fingscheidt, T. 2019. Towards corner case detection for autonomous driving. In *Proc. of IV*, 438–445. IEEE.

Breitenstein, J.; Bär, A.; Lipinski, D.; and Fingscheidt, T. 2021. Detection of collective anomalies in images for automated driving using an earth mover's deviation (emdev) measure. In *Proc. of IV Workshops*, 90–97. IEEE.

Breitenstein, J.; Termöhlen, J.-A.; Lipinski, D.; and Fingscheidt, T. 2020. Systematization of corner cases for visual perception in automated driving. In *Proc. of IV*, 1257–1264. IEEE.

Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuScenes: A multimodal dataset for autonomous driving. In *Proc. of CVPR*, 11621–11631.

Chan, R.; Lis, K.; Uhlemeyer, S.; Blum, H.; Honari, S.; Siegwart, R.; Fua, P.; Salzmann, M.; and Rottmann, M. 2021. SegmentMeIfYouCan: A Benchmark for Anomaly Segmentation. In Vanschoren, J.; and Yeung, S., eds., *Proc. of NeurIPS Track on Datasets and Benchmarks*, volume 1. Curran.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proc. of ECCV*.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proc. of CVPR*, 3213–3223.

Di Biase, G.; Blum, H.; Siegwart, R.; and Cadena, C. 2021. Pixel-wise anomaly detection in complex driving scenes. In *Proc. of CVPR*, 16918–16927.

Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; and Koltun, V. 2017. CARLA: An open urban driving simulator. In *Proc. of CoRL*, 1–16. PMLR.

Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. *Int J Comput Vis*, 88: 303–338.

Fremont, D. J.; Kim, E.; Dreossi, T.; Ghosh, S.; Yue, X.; Sangiovanni-Vincentelli, A. L.; and Seshia, S. A. 2022. Scenic: A language for scenario specification and data generation. *Machine Learning*, 1–45.

Geiger, A.; Lenz, P.; and Urtasun, R. 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proc. of CVPR*, 3354–3361. IEEE.

Grcić, M.; Bevandić, P.; Kalafatić, Z.; and Šegvić, S. 2021. Dense anomaly detection by robust learning on synthetic negative data. *arXiv preprint arXiv:2112.12833*.

Gupta, A.; Narayan, S.; Joseph, K.; Khan, S.; Khan, F. S.; and Shah, M. 2022. Ow-detr: Open-world detection transformer. In *Proc. of CVPR*, 9235–9244.

Han, J.; Ren, Y.; Ding, J.; Pan, X.; Yan, K.; and Xia, G.-S. 2022. Expanding low-density latent regions for open-set object detection. In *Proc. of CVPR*, 9591–9600.

Hendrycks, D.; Basart, S.; Mazeika, M.; Zou, A.; Kwon, J.; Mostajabi, M.; Steinhardt, J.; and Song, D. 2019. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*.

Hendrycks, D.; Basart, S.; Mazeika, M.; Zou, A.; Kwon, J.; Mostajabi, M.; Steinhardt, J.; and Song, D. 2022. Scaling Out-of-Distribution Detection for Real-World Settings. In *Proc. of ICML*, volume 162, 8759–8773. PMLR.

Hendrycks, D.; and Gimpel, K. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*.

ISO. 2022. Road vehicles — Test scenarios for automated driving systems — Vocabulary. Technical report, ISO.

Joseph, K.; Khan, S.; Khan, F. S.; and Balasubramanian, V. N. 2021. Towards open world object detection. In *Proc. of CVPR*, 5830–5840.

Jung, S.; Lee, J.; Gwak, D.; Choi, S.; and Choo, J. 2021. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *Proc. of ICCV*, 15425–15434.

Li, K.; Chen, K.; Wang, H.; Hong, L.; Ye, C.; Han, J.; Chen, Y.; Zhang, W.; Xu, C.; Yeung, D.-Y.; et al. 2022. CODA: A real-world road corner case dataset for object detection in autonomous driving. In *Proc. of ECCV*.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Proc. of ECCV*, 740–755. Springer.

Lis, K.; Nakka, K.; Fua, P.; and Salzmann, M. 2019. Detecting the unexpected via image resynthesis. In *Proc. of ICCV*, 2152–2161.

Liu, W.; Luo, W.; Lian, D.; and Gao, S. 2018. Future frame prediction for anomaly detection–a new baseline. In *Proc. of CVPR*, 6536–6545.

Lohdefink, J.; Fehrling, J.; Klingner, M.; Huger, F.; Schlicht, P.; Schmidt, N. M.; and Fingscheidt, T. 2020. Self-supervised domain mismatch estimation for autonomous perception. In *Proc. of CVPR Workshops*, 334–335.

Mao, J.; Niu, M.; Jiang, C.; Liang, H.; Chen, J.; Liang, X.; Li, Y.; Ye, C.; Zhang, W.; Li, Z.; et al. 2021. One million scenes for autonomous driving: Once dataset. *arXiv preprint arXiv:2106.11037*.

Menzel, T.; Bagschik, G.; and Maurer, M. 2018. Scenarios for development, test and validation of automated vehicles. In *Proc. of IV*, 1821–1827. IEEE.

Najm, W. G.; Smith, J. D.; Yanagisawa, M.; et al. 2007. Pre-crash scenario typology for crash avoidance research. Technical report, United States. National Highway Traffic Safety Administration (NHTSA).

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in NeurIPS*, 28.

Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; and Boult, T. E. 2012. Toward open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(7): 1757–1772.

Shi, S.; Guo, C.; Jiang, L.; Wang, Z.; Shi, J.; Wang, X.; and Li, H. 2020. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proc. of CVPR*, 10529–10538.

Shi, S.; Wang, X.; and Li, H. 2019. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud. In *Proc. of CVPR*.

Sindagi, V. A.; Zhou, Y.; and Tuzel, O. 2019. Mvx-net: Multimodal voxelnet for 3d object detection. In *Proc. of ICRA*, 7276–7282. IEEE.

Stocco, A.; Weiss, M.; Calzana, M.; and Tonella, P. 2020. Misbehaviour prediction for autonomous driving systems. In *Proc. of ICSE*, 359–371.

Tian, Y.; Liu, Y.; Pang, G.; Liu, F.; Chen, Y.; and Carneiro, G. 2022. Pixel-wise energy-biased abstention learning for anomaly segmentation on complex urban driving scenes. In *Proc. of ECCV*, 246–263. Springer.

Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proc. of CVPR*, 2636–2645.

Zhang, M.; Zhang, Y.; Zhang, L.; Liu, C.; and Khurshid, S. 2018. DeepRoad: GAN-based metamorphic testing and input validation framework for autonomous driving systems. In *Proc. of ASE*, 132–142.

Zhang, X.; Tao, J.; Tan, K.; Törngren, M.; Sanchez, J. M. G.; Ramli, M. R.; Tao, X.; Gyllenhammar, M.; Wotawa, F.; Mohan, N.; et al. 2022. Finding critical scenarios for automated driving systems: A systematic mapping study. *IEEE Trans. Softw. Eng.*, 49(3): 991–1026.

Zhao, H.; Shi, J.; Qi, X.; Wang, X.; and Jia, J. 2017. Pyramid scene parsing network. In *Proc. of CVPR*, 2881–2890.

Zhou, X.; Wang, D.; and Krähenbühl, P. 2019. Objects as points. *arXiv preprint arXiv:1904.07850*.

Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; and Dai, J. 2020. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*.

Zohar, O.; Wang, K.-C.; and Yeung, S. 2023. Prob: Probabilistic objectness for open world object detection. In *Proc. of CVPR*, 11444–11453.