Anomaly Detection in Autonomous Driving: A Survey

Daniel Bogdoll[1,2*], Maximilian Nitsche[1,2*], J. Marius Zöllner[1,2]
[1]FZI Research Center for Information Technology, Karlsruhe, Germany
[2]KIT Karlsruhe Institute of Technology, Karlsruhe, Germany
{bogdoll, nitsche, zoellner}@fzi.de

Abstract

Nowadays, there are outstanding strides towards a future with autonomous vehicles on our roads. While the perception of autonomous vehicles performs well under closed-set conditions, they still struggle to handle the unexpected. This survey provides an extensive overview of anomaly detection techniques based on camera, lidar, radar, multimodal and abstract object level data. We provide a systematization including detection approach, corner case level, ability for an online application, and further attributes. We outline the state-of-the-art and point out current research gaps.

1. Introduction

Anomalies, also called corner cases, occur everyday on the street, which is why autonomous vehicles need to cope with them. This "long tail of rare events" [45] is seen by many as the core obstacle towards large scale deployments of autonomous vehicles [1,50,81]. While there are exciting advances in handling the rare and unknown [47,91,92], it remains crucial to detect anomalies, which is still challenging [55]. In autonomous driving (AD), there are many levels of corner cases and multiple sensor modalities, including camera, lidar, and radar. While an extensive survey regarding camera-based approaches [14] exists, there is little to no research regarding other sensors or corner cases on higher levels of abstraction, including surveys. Here, we provide an overview of anomaly detection methods in the domain of AD for different sensor modalities, including methods not explicitly developed for AD, but which we deem applicable.

We characterize the anomaly detection techniques in Tables 1-5 across the modalities camera, lidar, radar, multimodal, and abstract object level. They are further characterized by their general detection approach, type of corner case, evaluation dataset or simulation, as well as regarding their possible online application. We classify the detection approaches following Breitenstein et al. in five concepts: "reconstruction, prediction, generative, confidence scores, and feature extraction" [14]. Confidence score techniques are often derived by post-processing without interfering with the training of a neural network and subdivided into Bayesian approaches, learned scores, and scores obtained by post-processing. Reconstructive approaches try to reconstruct normality and consider any kind of deviation from it as anomalous. Generative approaches are closely related to the former reconstructive approaches, but also take into account the discriminator's decision or the distance to the training data. Feature extraction can be based on hand-crafted or learned features to determine a class label or compare modalities on various feature levels. Prediction based techniques predict the next frame(s) expected under normality. An overview can be found in Figure 1.

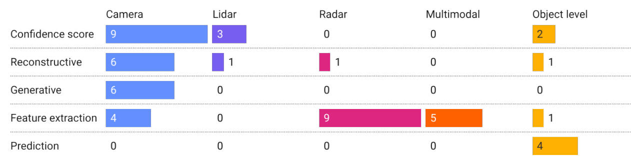

Figure 1. Overview of anomaly detection approaches based on camera, lidar, radar, multimodal, and abstract object level data.

We follow Breitenstein et al. [13] for the systematization of corner cases with the levels pixel, domain, object, scene and scenario, each being harder to detect. Heidecker et al. [40] extended these camera-based levels to incorporate lidar and radar sensors. Similar to their work, we use the terms "anomaly" and "corner case" interchangeably. In this survey, we focus on natural, external corner cases. Thus, we exclude anomalies on the sensor layer [40]; anomalies on the pixel level; and anomalies due to adversarial attacks.

We list all datasets or simulation environments used and label techniques as online capable if they or similar approaches of equal computational complexity (or higher, denoted by **) are reported as such or name a frame rate above 10 FPS. Methods marked with * are not providing inference performance measurements and are thus labeled as offline.

*These authors contributed equally

2. Anomaly Detection on Camera Data

Autonomous vehicles are often equipped with different camera systems, like stereo, mono, and fisheye cameras, to ensure a rich perception of the environment. Thus, anomaly detection in camera data holds great potential for more robust visual perception. For this section, we introduce two more criteria following the Fishyscapes (FS) benchmark [31]: auxiliary data and retraining. The former indicates whether an approach requires anomalous data during training. Retraining, however, specifies whether methods cannot use pretrained models, but require a special loss or the retraining, which might decrease the performance [31]. All camera-based methods can be found in Table 1.

Confidence score. Approaches on the basis of confidence scores constitute a baseline for the detection of anomalies based on the estimation of uncertainty in neural networks. As one of the earlier works, Kendall et al.'s *Bayesian SegNet* [51] derives the uncertainty of the *semantic segmentation* (SemSeg) network *SegNet* by Monte Carlo dropout sampling, where higher variance of the classes indicates higher uncertainty. The uncertainty can be interpreted as a pixel-wise anomaly score to detect obstacles on roads [69,89]. A similar approach to detect unknown obstacles on the road is proposed by Jung et al. [48]. They obtain class-conditioned standardized max logits of a segmentation network. This procedure is motivated by the finding that max logits have their own ranges for different predicted classes. The mean and standard deviations are thereby determined from the training samples. Thus, the standardization can be categorized as a *learned confidence score* approach. In addition to the standardization, they suppress class boundaries and apply a dilated smoothing to consider local semantics in broad receptive fields. Heidecker et al. [41] model the epistemic uncertainty of *Mask R-CNN* [39] and quantify the class and positional uncertainty of instances. They outline a criterion to detect anomalies based on the position and class uncertainty. Anomalies due to positional uncertainty are defined by the standard deviation of scaled bounding boxes exceeding a predefined threshold. In addition, instances are considered anomalous due to class uncertainty whenever the standard deviation of any class is above the predefined threshold. But Bayesian segmentation networks are slow in inference due to their multiple forward passes through the network with Monte Carlo dropout for each frame. Therefore, Huang et al. [44] simulate the sampling procedure via *region-based temporal aggregation* in frame sequences and retain the network's online capability. To ensure the correct uncertainty estimation of moving objects, the previous segmentation is warped via optical flow. Bevandić et al. [6] present a multi-task network to simultaneously segment the input frame into semantics as well as output an anomaly probability map. The latter overrides the SemSeg whenever a probability exceeds a threshold to calibrate the confidence score when the model faces outliers. Most recently, Du et al. [28] presented the general learning framework *Virtual Outlier Synthesis* (VOS), which contrastively shapes the decision boundary of neural networks by synthesizing virtual outliers. At first, they estimate a class-conditioned multivariate Gaussian distribution in the penultimate latent space. Afterwards, outliers are sampled from a sufficiently small ϵ-likelihood region of this learned distribution. These virtual outliers near the class-boundary encourage the model to form a compact decision boundary between in-distribution (ID) and out-of-distribution (OOD) data. Furthermore, they propose a novel training objective with free energy as an uncertainty measurement, where ID data has negative and the virtual outliers positive energy. During inference, OOD objects are detected with a *logistic Regressor* based on the uncertainty score.

While the former approaches concentrate on anomalies on the object level, Breitenstein et al. [12] are the first to detect collective anomalies. They learn the normal quantity of class-instances based on a reference dataset. The class-instances themselves are predicted via a *Mask R-CNN* [39] to end up with a discrete distribution of classes. Furthermore, they introduce a variation of the earth-mover's distance (EMD) for inference, namely the earth-mover's deviation (EMDEV). Besides the comparison of distributions, the EMDEV is a signed value which indicates whether the scene contains more or less instances of a class than usual.

Reconstructive. Reconstructive and generative approaches are predominantly used for anomaly detection on the object level, since the models learn to reproduce the normality of the training data without any auxiliary data of anomalous objects. For instance, a recent work by Vojir et al. [89] proposes the reconstruction module *JSR-Net* to detect road anomalies based on a pixel-wise score. They enhance trained SemSeg networks by incorporating their information from known classes into the anomaly score. The network architecture consists of a reconstruction and a semantic coupling module. The former is connected to the backbone of the SemSeg network and reconstructs the road in a discriminative way, meaning it reduces the reconstruction loss of the road while increasing the loss for the remaining environment. In the subsequent module, the resulting pixel-error map is coupled with the output logits of the SemSeg to end up with a pixel-wise anomaly score. The extension module is trained on augmented road images, where patches of noise or a part of the input image are randomly positioned on the road and labeled as anomalous. The evaluation on various datasets shows the superiority of *JSR-Net* in comparison to others [5, 23, 56, 57] while preserving the closed-set segmentation performance.

A similar approach is evaluated by Ohgushi et al. [69] against the LaF benchmark on a highway dataset with real and synthetic road obstacles. In contrast to Vojir et al., they

Table 1. Overview of anomaly detection techniques on camera data.

Author(s)	Year	Ref	Technique	Approach	Aux Data	Retraining	Corner Case Level	Dataset / Simulation	Online
Du et al.	2022	[28]	VOS	Confidence — Learned	✗	✓	Object — Single-Point	PASCAL-VOC [30], BDD100K [99]	✗*
Jung et al.	2021	[48]	Standardized Max Logits	Confidence — Learned	✗	✗	Scene — Contextual	FS Lost-and-Found (LaF) [7], RA [57]	✓(13.3 FPS)
Heidecker et al.	2021	[41]	MC Dropout	Confidence — Bayesian	✗	✗	Object — Single Point	A2D2 [34]	✗
Chan et al.	2021	[20]	Entropy Maximization	Confidence — Learned	✓	✓	Object — Single-Point	LaF [70], Cityscapes (CS) [22], FS [7]	✓
Breitenstein et al.	2021	[12]	EMDEV	Confidence — Post-processed	✗	✓	Scene — Collective	CS [22] as D_{in}, ECP [11] & A2D2 [34] as D_{out}	✓
Bevandić et al.	2021	[6]	LDN-BIN	Confidence — Learned	✓	✓	Object — Single-Point	Vistas [66] & CS [22] as D_{in}, ImageNet [24] & WD [4] as D_{out}	✗*
Malinin and Gales	2018	[64]	Dirichlet Prior Networks	Confidence — Bayesian	✓	✓	Object — Contextual	FS LaF [7]	✗*
Huang et al.	2018	[44]	RTA	Confidence — Bayesian	✗	✗	Object — Single-Point	CamVid [15]	✓
Kendall et al.	2016	[51]	Bayesian SegNet	Confidence — Bayesian	✗	✗	Object — Single-Point	CamVid [15]	✓(11.11 FPS)
Vojir et al.	2021	[89]	JSR-Net	Reconstruction	✗	✗	Scene — Contextual	LaF [70], RA [57] & RO [56], FS LaF [7]	✗*
Ohgushi et al.	2021	[69]	Autoencoder + SemSeg	Reconstruction	✗	✗	Scene — Contextual	LaF [70], Highway dataset	✗
Lis et al.	2021	[56]	Erasing	Reconstruction	✗	✓	Scene — Contextual	FS LaF [7], RO (daylight) [56]	✗*
Di Biase et al.	2021	[25]	SynBoost	Reconstruction	✓	✗	Object — Single-Point	FS LaF, Static, Web (Oct. 2020) [7]	✗
Blum et al.	2021	[8]	NF + Logistic regression	Reconstruction	✓	✗	Object — Single-Point	FS LaF [7], FS Web & Static [7]	✗
Creusot and Munawar	2015	[23]	Compressive RBM	Reconstruction	✗	✓	Scene — Contextual	Recordings & YT Japanese highways	✓(10 FPS)
Nitsch et al.	2021	[68]	GAN + Post hoc statistics	Generative	✗	✓	Object — Single-Point	KITTI [33] & nuScenes [16] as D_{in}, ImageNet [24] as D_{out}	✓
Grcić et al.	2021	[35]	NFlowJS	Generative	✗	✓	Object — Single-Point	WD-Pascal [4], LaF [70], SMIYC [19], StreetHazards [42]	✓(18.4 FPS)
Xia et al.	2020	[95]	SynthCP	Generative	✗	✗	Object — Single-Point	CS [22], StreetHazards [42]	✗*
Löhdefink et al.	2020	[60]	Autoencoder DM	Generative	✗	✓	Domain — Domain Shift	CS [22], BDD100K [99], KITTI [33]	✓
Lis et al.	2019	[57]	Resynthesis	Generative	✗	✗	Object — Single-Point	LaF [70], RA [57]	✗*
Haldimann et al.	2019	[38]	Semantic cGAN	Generative	✗	✗	Scene — Single-Point	CS [22] as D_{in}, Vistas [66] as D_{out} [67]	✗*
Xue et al.	2019	[96]	Multi-layer Occlusion	Feature Extraction	✓	✓	Scene — Contextual	LaF [70]	✓
Bolte et al.	2019	[10]	Feature MSE	Feature Extraction	✗	✓	Domain — Domain Shift	KITTI [33], CS [22], BDD100K [99]	✓
Zhang et al.	2018	[101]	DeepRoad	Feature Extraction	✗	✗	Domain — Domain Shift	Udacity [87]	✓
Bai et al.	2018	[2]	SVM	Feature Extraction	✓	✗	Scene — Contextual & Collective	Urban roads	✓

combine the entropy loss of the SemSeg with the perceptual loss between the real and reconstructed image to form an anomaly map. They outline a set of post-processing steps where the final obstacle score map depends on the semantic information, the aforementioned anomaly map, and a superpixel division to refine local regions.

Di Biase et al. [25] leverage image re-synthesis [57] by combining the reconstruction error with two uncertainty maps of the segmentation network. The network outputs the softmax entropy and distance additionally to the segmentation output. Similar to [69], the perceptual difference is used as the reconstruction loss between the input and synthesized image. All predicted maps and the input image are fused in a spatial-aware dissimilarity module with three parts: encoder, fusion module, and decoder. In the fusion module, the encoded and re-synthesized inputs and the semantic image are concatenated and fused with a 1x1 convolution. The resulting feature map is evaluated against the jointly encoded uncertainty and perceptual difference via point-wise correlation. The final pixel-wise anomaly segmentation is provided by decoding the fused features and spatial-aware normalization with the semantic information.

Generative. According to the FS, LaF, and Segment Me If You Can (SMIYC) obstacle track benchmarks, the dense anomaly detection with *NFlowJS* of Grcić et al. [35] outperforms all contemporary techniques and represents the current state-of-the-art of camera-based anomaly detection. *NFlowJS* is jointly trained to generate synthetic negative patches with *normalizing flows* (NF) atop regular images alongside training the dense prediction network based on these created mixed-content images. The generated negative patches are thereby defined as the anomaly mask. During training, the discriminative model is encouraged to yield a uniform predictive distribution for the generated patch. This induces the generative distribution of the NF to move away from the inliers. At the same time, it is trained to maximize the likelihood of inliers. These opposing objectives support the generation of images at the boundary of the training data while sensitizing the discriminative model for anomalies. Especially, due to the former facet, the synthesized anomaly patches are likely to contain parts similar to inliers where the model predicts with high confidence. A strong penalizing of this behavior demolishes the model's confidence on actual inlier pixels. Therefore, they find the Jensen-Shannon (JS) divergence as a mildly penalizing loss of high confidence predictions. During inference, the closed-set segmentation is masked by the anomaly map generated by a threshold exceeding temperature scaled softmax and the JS divergence between output probability and uniform distribution. In contrast to former generative models, the *NFlowJS* does only rely on the anomaly synthesis during training, resulting in a real-time inference speed. Blum et al. [8] also evaluate an NF based approach with logistic regression on their FS benchmark. However, the results are incomparable with *NFlowJS*.

Nitsch et al. [68] adopt and enhance a generative approach of Lee et al. [54] for the detection of object anomalies. Lee et al. propose an auxiliary *generative adversarial network* (GAN) which encourages an object classifier to provide low confidence for samples outside the training distribution. Nitsch et al. extend the approach by a post hoc network statistic, which estimates a class-conditioned Gaussian distribution over the network's weights of the bottleneck layer. A cosine similarity metric determines the distribution distance and classifies a given sample based on an empirical threshold. Since they only perform classification, the localization of objects has to be done in advance.

Similarly, Lis et al. [57] adopt GANs to re-synthesize the input image and detect anomalies on the object level by the difference in appearance. However, the image generation is based on the final SemSeg map in contrast to [69, 89], where a decoder reconstructs the image based on the intermediate feature space of the SemSeg. As the SemSeg preserves the scene layout but loses the precise scene's appearance, regular reconstruction errors, like the perceptual loss, would output a high overall difference without informative results. Thus, they propose a discrepancy network which encodes the input and the re-synthesized image via multiple *VGG16* [83] networks with shared weights. The features are collectively concatenated with the convolutional encoded semantic map and correlated on all extraction levels and fed into the final decoding CNN on the respective feature level. The semantic-to-image synthesis is also adopted and evaluated by [38, 95] in form of a *conditional GAN* (cGAN) with a subsequent dissimilarity scoring.

Löhdefink et al. [60] present an approach for the detection of domain shifts. An autoencoder learns the domain of a given dataset in a self-supervised manner. The approach characterizes the training data domain via the distribution of the autoencoder's peak signal-to-noise ratio (PSNR). During inference, the domain mismatch (DM) is estimated by comparing the learned and incoming PSNR distribution of the data via the EMD. The evaluation shows a strong rank order correlation between the autoencoder's DM metric and the decrease of SemSeg performance when faced with target domains different than the source domain. While the inference is real-time capable, the approach has to accumulate a certain number of images, as it uses batches as input.

Feature Extraction. Another domain shift detection is proposed by Bolte et al. [10], where the mean-squared error (MSE) of feature maps is compared. The MSE is evaluated over entire datasets or batches. Similarly, Zhang et al. [101] propose the *DeepRoad* framework to validate single input images based on the distance to the training embedding of *VGGNet* features [83]. Bai et al. [2] detect anomalies in urban road scenes and classify entire input scenes as anomalous. They identify a set of representatives for normal urban scenes via the *k-means clustering* of *scale-invariant feature transform* (SIFT) features. Finally, images are classified by a *one-class support vector machine* (one-class SVM).

Overall, many of the previously outlined techniques work without external data, but require a retraining of the proposed extension module or entire detection architecture.

3. Anomaly Detection on Lidar Data

Most often, autonomous vehicles do not solely rely on camera data. Although, camera data has the highest resolution of the three sensor modalities, it lacks an accurate measurement of depth. Therefore, *light detection and ranging* (lidar) sensors, which provide a three-dimensional depth map of the environment, are often found in sensor setups. While there is much research about local denoising of lidar point clouds on the pixel level [3, 74], we are interested in anomalies on object and domain level, where an entire cluster of points or a large and constant shift in appearance is considered as anomalous. Especially weather conditions like rain, snow, and fog heavily influence the data. All lidar-based methods can be found in Table 2.

Confidence score. Recent research by Zhang et al. [100] shows that rain affects the lidar measurement quality, as resulting point clouds are sparser, noisier, and the average intensity is lower. Therefore, they aim to quantify the lidar degradation with the *Deep Semi-supervise Anomaly Detection* (DeepSAD) approach [77]. They first project 3D lidar data into a 2D intensity image. DeepSAD then transforms the images into a latent space, where all normal images, i.e., the scans without rain, fall into a hypersphere and all abnormal, i.e., rain affected, images are mapped away from the hypersphere's center. Finally, the distance of a transformed test image to the learned center of the hypersphere is interpreted as the anomaly score. As the model architecture defines anomalies as those who fall out of the hypersphere, we list the proposed methodology as a *learned confidence* detection approach in Table 2. The trained DeepSAD reaches a Spearman's correlation of up to 0.82 between the rainfall intensity and degradation score on dynamic, simulated test data. This indicates a considerably accurate quantification of anomaly detection due to weather conditions. Although the approach is developed for rainy and normal weather conditions, we suspect that the proposed method is transferable to other weather conditions, such as snow and fog.

In the past, several architectures have been proposed to detect objects in point clouds, like *VoxelNet* [102], *PointR-CNN* [82], and *PointNet++* [73]. However, these are based on a closed-set setting, thus being only capable of detecting classes that were included in the training set. In contrast, open-set detection methods are able to explicitly classify objects outside the closed-set as unknown upon the regular detection of the predefined classes. The open-set setting therefore loosens the constraint to classify all detections as one of the predefined classes. Consequently, one expects the false positive rate to improve and the model to acknowledge the novelty of objects upon never seen instances.

The idea of an open-set detector for 3D point clouds was first implemented by Wong et al. [94]. They propose an *Open-Set Instance Segmentation* (OSIS) network, which learns a category-agnostic embedding to cluster points into instances regardless of their semantics. The inference is based on a bird's eye view (BEV) lidar frame and consists of two stages: the closed-set and open-set perception. In the first stage, a backbone of 2D convolutions extracts multi-scale features, which are then fed into a detection and an embedding head. The latter is the core of *OSIS* and learns the category-agnostic embedding space. Moreover, the embedding head yields the prototypes of possible closed-set classes. Points are then associated to prototypes of known categories by the learned embedding space. In the second stage, the remaining unassociated points are

Table 2. Overview of anomaly detection techniques on automotive lidar data

Author(s)	Year	Ref	Technique	Approach	Corner Case Level	Dataset / Simulation	Online
Zhang et al.	2021	[100]	DeepSAD [77]	Confidence — Learned	Domain — Domain Shift	Simulation & static / dynamic real envs.	✗*
Cen et al.	2021	[17]	MLUC	Confidence — Learned	Object — Single-Point	UDI [17] & KITTI [33]	✗*
Wong et al.	2019	[94]	OSIS	Confidence — Learned	Object — Single-Point	TOR4D [94] & Rare4D	✗*
Masuda et al.	2021	[65]	VAE (FoldingNet)	Reconstruction	Object — Single-Point	ShapeNet [21]	✗*

considered as unknown. Those are clustered into instances of unknown objects via *density-based spatial clustering of applications with noise* (DBSCAN) [29]. The outlined approach falls into the category of *learned confidence* scores, as the prototypes are learned during training and unknown objects are identified by their uncertainty of class association. *OSIS* is evaluated on two large-scale, non-public datasets. Here, the technique outperforms other adapted deep learning based instance segmentation algorithms for the detection of single-point anomalies on the object level.

The *OSIS* network is later used as a baseline for comparison of the *Metric learning with Unsupervised Clustering* (MLUC) network developed by Cen et al. [17]. They focus on two primary challenges: identifying regions of unknown objects with high probability and enclosing these regions' points with proper bounding boxes. In context of the first problem, the paper shows that the euclidean distance sum (EDS), based on metric learning, is more suitable than a naive softmax probability metric to differentiate between regions of known and unknown objects. They replace the classifier of closed-set detections with the euclidean distance representation to all prototypes of the embedding space. The euclidean distance-based probability is incorporated into the loss function, such that the embedding vector of known classes is close to the corresponding prototypes of the respective class. However, unknown objects are mapped close to the center of the embedding, having a smaller EDS. The EDS measures the uncertainty of closed-set detections. Therefore, boxes with an EDS lower than a threshold λ_{EDS} are considered as regions of unknown objects. Similarly to *OSIS*, these bounding boxes of low confidence are then refined by unsupervised depth clustering. The *MLUC* considerably outperforms *OSIS*.

Reconstructive. Masuda et al. [65] show an approach to detect whether an object point cloud is anomalous or not. In contrast to the preceding methods, this technique is based on point clouds of single encapsulated objects. Since automotive lidars provide full environment scans, single objects or regions of interest would need to be extracted by detection or clustering approaches first. The proposed VAE is based on the *FoldingNet* decoder [98] and learns to reconstruct the set of known objects which are considered as normal. The point cloud is then classified as anomalous based on the reconstruction and the Chamfer distance as an anomaly score. The approach is evaluated on the ShapeNet [21] dataset, which also includes a variety of objects outside the AD domain. The results are promising, as the model achieves an average AUC of 76.3%, where known classes were defined as anomalies.

Overall, anomaly detection on the object level in lidar data is just gaining momentum, after research has already led to various closed-set detection architectures.

4. Anomaly Detection on Radar Data

Radar is the third sensor modality often used in AD. It has a higher range at the cost of a lower resolution and less detailed spatial information than lidar sensors. In comparison to both previous modalities, radar is more robust to changing weather and daytime conditions [90]. In the following, we concentrate on anomaly detection techniques designed for radar systems installed in the automotive industry, like surround, long, and short range radars and exclude techniques based on ultra-wideband and through-the-wall radars. We additionally characterize approaches by the method set (mathematical, feature engineering, Machine Learning (ML) or Deep Learning (DL)) used for detection. All radar-based methods can be found in Table 3.

Radar estimates an objects' position by measuring the time of flight of electromagnetic multipath waves and their reflections. Due to the multipath propagation, radar can detect even occluded objects [86]. However, this advantage is mitigated by the fact that this also causes noise, reflections and artifacts. Especially reflective surfaces, like guardrails on highways or smooth walls, produce non-existing artifacts, often refereed to as "ghost targets" [18, 79, 90]. These are a long-standing challenge affecting automotive radars [58]. For this reason, and as this survey focuses on anomalies above the pixel level, we specifically concentrate on methods to detect ghost targets and alike.

Feature Extraction. Most recent work by Liu et al. [58] proposes a model of multipath propagation to identify and remove ghosts based on the targets' range difference, based on reflections from a guardrail. The established model and numerical results show that the range difference between each real vehicle and its corresponding ghost target only differs slightly. In contrast, the range differences between two, even closely located, real targets are usually far greater. The proposed ghost removal algorithm leverages this finding as it distinguishes between real and ghost targets based on a maximum range difference threshold Δr, which is nu-

Table 3. Overview of anomaly detection techniques on automotive radar data

Author(s)	Year	Ref	Technique	Approach	Method set	Corner Case Level	Dataset / Simulation	Online
Liu et al.	2021	[58]	Range difference	Feature Extraction	Mathematical	Scene — Contextual	Numerical simulation ($\leq 50m$)	✓
Griebel et al.	2021	[36]	MFG PointNet++ [73]	Feature Extraction	DL	Scene — Contextual (Pixel — Local Outlier)	Hand-labeled 2D data ($\leq 70m$)	✓(42.7 FPS)
Chamseddine et al.	2021	[18]	PointNet [73]	Feature Extraction	DL	Scene — Contextual	Lidar-labeled 3D data	✓**
Kraus et al.	2020	[52]	PointNet++ [73]	Feature extraction	DL	Scene — Contextual	NLOS [80]	✓**
Prophet et al.	2019	[72]	Features + RF	Feature Extraction	Feat. Eng. & ML	Scene — Contextual	Hand-labeled 2D data	✓
Ryu et al.	2018	[79]	MLP	Feature Extraction	ML	Scene — Contextual	City center intersection	✓
Kamann et al.	2018	[49]	Geometric propagation	Feature Extraction	Mathematical	Scene — Contextual	Experimental setup	✓
Visentin et al.	2017	[88]	Pauli decomposition	Feature Extraction	Mathematical	Scene — Contextual	Experimental setup	✓
Roos et al.	2017	[76]	Orientation & motion	Feature Extraction	Mathematical	Scene — Contextual	Simulation	✗
Garcia et al.	2019	[32]	OGM + CNN	Reconstructive	DL	Scene — Contextual	Hand-labeled 2D data	✗

merically determined in advance. While this mathematical approach is simple and effective in simulation, one has to consider its constraints, as it is limited to a highway-like driving scene with three lanes of fixed size. Moreover, the distance between the target and the reflective guardrail takes only three values and does not simulate lane changes of real vehicles. Similar work was done by Holder et al. [43], Kamann et al. [49], Visentin et al. [88], and Roos et al. [76].

The latest ML algorithms are utilized to detect radar anomalies in a greater variety of driving scenes without the aforementioned constraints of a mathematical model to work. In this context, ghost targets are often defined as a separate class. For instance, Griebel et al. [36] implement a DL method utilizing the *PointNet++* architecture. The original architecture uses multi-scale grouping (MSG) layers to extract features on different scales in a point cloud. The MSG module uses a circular form to query a point's neighboring information. They introduce an extension of the original grouping module, hypothesising that anomalous radar targets occur in a ring-shaped region around the radar sensor origin within the same range as car targets. The so-called multi-form grouping (MFG) module is a combination of the original circular as well as the new ring querying form. Hence, the module incorporates the neighborhood information of both forms at multiple scales. Moreover, they do not solely focus on the detection of multi-path anomalies, like ghost targets, but also on other single target anomalies caused by the Doppler velocity ambiguities or errors in the direction of arrival estimation. The latter are local outliers and fall into the pixel level.

Kraus et al. [52] utilize *PointNet++* to not only differentiate between real and ghost objects, but also classify them as (ghost) pedestrians or (ghost) cyclists. Therefore, the evaluation is limited to the Non-Line-of-Sight (NLOS) dataset [80], including only vulnerable road users. They tackle the challenge of sparse radar data by accumulating measurements over a period of 200ms.

While the former approaches detect anomalies in single-shot 2D radar data, Chamseddine et al. [18] evaluate the *PointNet++* architecture to detect ghost targets in dense 3D radar data. The *PointNet++* architecture is, in contrast to other common 3D detection networks [53, 102], able to learn individual point features and therefore well suited to classify single radar points into real or ghost targets. Ablation studies show that the form of representation of spatial information matter as the additional encoding of points in spherical coordinates boosts the network's performance.

Another noteworthy approach to detect ghost anomalies regardless of their causes is the procedure of Prophet et al. [72], where initially moving targets are identified by the scanned radial velocity and a threshold value to improve scene understanding. Afterwards, a set of handcrafted features is defined for each detection under test (DUT). These features comprehend the DUT parameters, the vehicles motion state, the error value calculated in the first step, the number of static and moving neighbors, as well as the calculation of an occupancy grid map (OGM) around the DUT. Moreover, they include a Boolean feature indicating the presence of a moving neighbor detection around the DUT in the previous frame. Consequently, this technique is the first to incorporate temporal data to improve detection. Finally, these features are fed into ML algorithms like a SVM, *k nearest neighbor classifier* (KNN), or RF. According to the subsequent evaluation on a data set of 36,916 detections, the RF outperforms all other algorithms with a success rate of 91.2%. Similarly, Ryu et al. [79] train a *multilayer perceptron* (MLP) on a set of six features to remove ghost targets from a tracking algorithm.

Reconstructive. Garcia et al. [32] use a two-channel image consisting of the aforementioned occupancy grid and moving detections map as an input of a *fully convolutional network* (FCN). The proposed architecture is segmented in an encoder and a decoder part. While the former extracts the semantic information into a lower resolution representation, the latter reconstructs the spatial information and maps the extracted representation back to the original image size. In the resulting map of probabilities, a moving target is considered a ghost detection. The technique achieves a binary classification accuracy of 92% on a test set of 50 images.

Overall, many approaches assume that ghost and real targets can be differentiated by their feature set in contrast to conventional, i.e., reconstructive or confidence based tech-

Table 4. Overview of anomaly detection on multimodal sensor data

Author(s)	Year	Ref	Technique	Approach	Corner Case Level	Dataset / Simulation	Online
Wang et al.	2021	[90]	Multimodal transformers	Feature Extraction	Scene — Contextual	Auto-labeled nuScenes [16]	✗*
Sun et al.	2020	[85]	RGB-D network	Feature Extraction	Scene — Contextual	CS [22]	✓(22 FPS)
Ji et al.	2020	[46]	SVAE	Feature Extraction	Scene — Contextual	TerraSentia	✓
Gupta et al.	2018	[37]	MergeNet	Feature Extraction	Scene — Contextual	LaF [70]	✗(5 FPS)
Pinggera et al.	2016	[71]	FPHT	Feature Extraction	Scene — Contextual	LaF [70]	✓(20 FPS)

niques as shown in Table 3. Despite that, we expect future work to further improve by taking into account temporal information, as indicated in [72].

5. Anomaly Detection on Multimodal Data

Autonomous vehicles are typically equipped with multiple modalities. In the following, we provide an overview of techniques which identify anomalies based on irregularities between the individual sensors or by fusing information. All multimodal methods can be found in Table 4.

Feature Extraction. Following on from the previous detection of ghost targets in radar data, Wang et al. [90] propose a multimodal technique. Transformers are well suited for 3D point clouds as their attention mechanism is permutation invariant, which is hard for conventional neural networks. Moreover, transformers explicitly model a point's interactions, in contrast to the aforementioned architectures like *PointNet++*. The authors adopt a multimodal transformer network to detect radar ghost targets by referencing lidar points. Radar point clouds are way sparser than lidar point clouds, which hinders the data matching. Therefore, individual radar points query for surrounding lidar points by KNN and provide local feature information, like a "magnifying lens". They apply self-attention for the unstructured radar data itself to identify ghost targets, as these show high affinity to the corresponding real targets. The attention modules are stacked to a network. Lastly, the fully connected segmentation head of *PointNet++* is utilized to classify individual radar points as possible ghost targets. The proposed method is evaluated on the nuScenes dataset [16]. Worth mentioning, the ground truth of ghost targets was generated by comparing radar and lidar data.

Sun et al. [85] present a real-time fusion network for SemSeg based on RGB-D data. The primary goal of the multimodal architecture is to improve image segmentation by incorporating depth information. Furthermore, they argue that the multi-source segmentation framework is also capable to detect unexpected road obstacles, providing a unified pixel-wise scene understanding. However, the evaluation on the CS dataset [22] does not provide detection performance measures for the unexpected obstacles, as the approach concentrates on the SemSeg of closed-set classes. Another RGB-D based detection of road obstacles is implemented by Gupta et al. [37] in form of *MergeNet*. As the architecture's name suggests, the model merges two networks, the *Stripe-net* and *Context-net*, via a third meta *Refiner-net*. The *Stripe-net* extracts low-level features of the RGB and depth data in parallel, based on images split in stripes. This forces the network to learn discriminative features within narrow bands of information and a small subset of parameters. Moreover, this allows for a more reliable detection of small road obstacles. In contrast, the *Context-net* is trained on the entire RGB image and is determined to learn high-level features. The *Refiner-net* acts as a meta network to combine the complementary features and end up with a form of curriculum learning. As a result, *MergeNet* is trained to discriminate between road, off-road, and small obstacle, where we consider the latter as abnormal.

Ji et al. [46] propose a *supervised VAE* (SVAE) to merge multiple sensor modalities of different dimensionality. This is especially useful for the fusion of dense lidar data and radar data of lower resolution. They abandon the decoder after training and use the learned encoder as a feature extractor. The modalities' latent representation is then – along with other encoded modalities – fed into a fully connected layer to identify an anomalous operation mode of the vehicle. Even though the method was designed for field robots, we expect it to be transferable to other driving scenes.

In summary, one can see in Figure 1, that all of the multimodal anomaly detection techniques are based on the comparison of the individual modalities' *extracted features*. We argue, that multimodal detection could become much more relevant, as it broadens the search space for potential anomalies, while reducing the risk of false positives.

6. Anomaly Detection on Abstract Object Data

The previous sections gave an overview of anomaly detection techniques suitable for specific sensor modalities. The following approaches are focusing on a more abstract level of pattern analysis, i.e., the detection of anomalous behavior in scenarios, which are not necessarily bound to a sensor modality. Thus, the approaches are designed to detect anomalies on the scenario level [13] and deal with risky and abnormal driving behavior of non-ego vehicles. All abstract object-level based methods can be found in Table 5.

Prediction. Yang et al. [97] assess the behavior of driving vehicles based on *Hidden Markov Models* (HMM) to detect anomalous scenarios. The observation states

Table 5. Overview of anomaly detection techniques on abstract object level data

Author(s)	Year	Ref	Technique	Approach	Corner Case Level	Dataset / Simulation	Online
Yang et al.	2019	[97]	HMM	Prediction	Scenario — Risky	CARLA [27]	✓
Bolte et al.	2019	[9]	Adversarial AE	Prediction	Scenario — Anomalous, Novel, Risky	CS [22]	✗*
Liu et al.	2018	[59]	U-Net [75] + Flownet [26]	Prediction	Scenario — Anomalous, Novel	CUHK [61], UCSD [63], ST [62]	✓(25 FPS)
Yuan et al.	2018	[101]	Bayes Model	Prediction Confidence — Bayesian	Scenario — Anomalous	Driving videos	✗
Zhang et al.	2018	[101]	DeepRoad	Feature Extraction	Scenario — Anomalous	Udacity [87]	✓
Stocco et al.	2020	[84]	SelfOracle	Reconstructive Confidence — Learned	Scenario — Risky	Udacity [87]	✓

of the Markov model are provided by the *Conditional Monte Carlo Dense Occupancy Tracker* (CMCDOT) framework [78] and comprise real-time velocity as well as vehicle position through probabilistic occupancy grids. The framework derives these observations based on point cloud and odometry data. As a result, the pipeline can reliably infer risky and abnormal driving behaviors in simulated multi-lane highway scenarios with two non-ego vehicles.

Bolte et al. [9] propose an anomaly detection on the scenario level, where patterns are observed over a sequence of sensor data, i.e., camera images. They consider all subtypes: anomalous, novel, and risky scenarios [13]. They quantify the anomalous behavior for moving objects, such as pedestrians or cars, due to the nature of scenario anomalies. The error between the real and a predicted frame is considered as the anomaly score. The predicted frame is generated by an adversarial autoencoder and based on the past sequence of input frames. Hence, the anomaly score can also be interpreted as the non-predictability of the model. The model is evaluated with MSE, PSNR, and structural similarity index measure (SSIM) [93] metrics, and anomalous scenarios are determined by a threshold. They localize anomalous behaving objects by dividing the input image into grid cells of user-specific size and weight close objects higher, as those pose a higher risk of collision.

A similar, but more comprehensive, approach is outlined in the paper of Liu et al. [59]. They adopt *U-Net* [75] as an image-to-image translation model to predict the next frame based on the past sequence of frames. In contrast to the former approach [9], their framework considers also temporal information of scenarios. They extend their objective function by an optical flow constraint to retain the motion information of moving objects. The optical flow is calculated via *Flownet* [26]. They leverage adversarial training to discriminate between real and fake images to further boost the performance of the future frame prediction. Anomalous scenarios are again identified by the PSNR of the real and predicted frame exceeding a predefined threshold.

Reconstructive. Stocco et al. propose *SelfOracle* [84] for the detection of safety-critical misbehavior, like collisions and out-of-bound episodes. The architecture uses a VAE to reconstruct a set of preceding input images of a current scene and calculates the corresponding reconstruction errors. During the training on normal data, the model is fitting a probability distribution to the observed reconstruction errors via maximum likelihood estimation. The estimated distribution can then be used to determine a threshold value θ to distinguish between anomalous and normal behavior. The parameter ϵ corresponds to the probability of the tail and thus θ controls for the false positive rate of the detection. In addition, *SelfOracle* implements a time-aware anomaly scoring by applying a simple autoregressive filter on the sequence of reconstruction errors, as the current error might be susceptible to single-frame outliers. While they evaluate *SelfOracle* only in a simulation environment, the approach seems promising and even outperforms the author's implementation of the *DeepRoad* framework.

Finally, anomaly detection on the object level heavily depends on human driving behavior. Therefore, with the rise of autonomous vehicles on the road, AD will experience a large concept drift in behavior prediction.

7. Conclusion

In this paper, we provide an extensive survey of techniques for the detection of anomalies in the field of autonomous driving. While the survey by Breitenstein et al. that we build upon [14] is limited to camera data, we characterize techniques across different sensor modalities. Most of the recent advancements are concerned with image-based anomaly detection, while lidar- and radar-based approaches are still struggling to gain momentum. One reason for this is the absence of benchmarks, which so far only exist in the camera sector. The community misses common datasets of labeled anomalies, which leaves the unified comparison of detection techniques difficult. Tables 1-5 show that each modality might be more suitable for the detection of one or only few types of corner cases, as e.g., lidar-based techniques focus strongly on single-point anomalies. Overall, the state-of-the-art especially detects contextual anomalies on the scene level, while collective anomalies lack behind.

Acknowledgment

This work results from the project KI Data Tooling (19A20001J), funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK).

References

[1] Drago Anguelov. Taming the Long Tail of Autonomous Driving Challenges. https://www.youtube.com/watch?v=Q0nGo2-y0xY, 2019. 1

[2] Shuang Bai, Chao Han, and Shan An. Recognizing Anomalies in Urban Road Scenes Through Analysing Single Images Captured by Cameras on Vehicles. *Sensing and Imaging*, 19, 2018. 3, 4

[3] Haris Balta, Jasmin Velagic, Walter Bosschaerts, Geert De Cubber, and Bruno Siciliano. Fast Statistical Outlier Removal Based Method for Large 3D Point Clouds of Outdoor Environments. *IFAC-PapersOnLine*, 51, 2018. 4

[4] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Simultaneous Semantic Segmentation and Outlier Detection in Presence of Domain Shift. *Pattern Recognition*, 11824, 2019. 3

[5] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Simultaneous Semantic Segmentation and Outlier Detection in Presence of Domain Shift. *Pattern Recognition*, 11824, 2019. 2

[6] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Dense outlier detection and open-set recognition based on training with noisy negative images. *arXiv:2101.09193*, 2021. 2, 3

[7] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. Fishyscapes: A Benchmark for Safe Semantic Segmentation in Autonomous Driving. In *IEEE/CVF International Conference on Computer Vision Workshop*, 2019. 3

[8] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. The Fishyscapes Benchmark: Measuring Blind Spots in Semantic Segmentation. *International Journal of Computer Vision*, 129, 2021. 3

[9] Jan-Aike Bolte, Andreas Bar, Daniel Lipinski, and Tim Fingscheidt. Towards Corner Case Detection for Autonomous Driving. In *IEEE Intelligent Vehicles Symposium*, 2019. 8

[10] Jan-Aike Bolte, Markus Kamp, Antonia Breuer, Silviu Homoceanu, Peter Schlicht, Fabian Huger, Daniel Lipinski, and Tim Fingscheidt. Unsupervised Domain Adaptation to Improve Image Segmentation Quality Both in the Source and Target Domain. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 3, 4

[11] Markus Braun, Sebastian Krebs, Fabian Flohr, and Dariu M. Gavrila. EuroCity Persons: A Novel Benchmark for Person Detection in Traffic Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41, 2019. 3

[12] Jasmin Breitenstein, Andreas Bär, Daniel Lipinski, and Tim Fingscheidt. Detection of collective anomalies in images for automated driving using an earth mover's deviation (emdev) measure. In *IEEE Intelligent Vehicles Symposium Workshops*, 2021. 2, 3

[13] Jasmin Breitenstein, Jan-Aike Termohlen, Daniel Lipinski, and Tim Fingscheidt. Systematization of Corner Cases for Visual Perception in Automated Driving. In *IEEE Intelligent Vehicles Symposium*, 2020. 1, 7, 8

[14] Jasmin Breitenstein, Jan-Aike Termöhlen, Daniel Lipinski, and Tim Fingscheidt. Corner Cases for Visual Perception in Automated Driving: Some Guidance on Detection Approaches. *arXiv:2102.05897*, 2021. 1, 8

[15] Gabriel J. Brostow, Julien Fauqueur, and Roberto Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30, 2009. 3

[16] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A Multimodal Dataset for Autonomous Driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 3, 7

[17] Jun Cen, Peng Yun, Junhao Cai, Michael Yu Wang, and Ming Liu. Open-set 3D Object Detection. *arXiv:2112.01135*, 2021. 5

[18] Mahdi Chamseddine, Jason Rambach, Didier Stricker, and Oliver Wasenmuller. Ghost Target Detection in 3D Radar Data using Point Cloud based Deep Neural Network. In *25th International Conference on Pattern Recognition*. IEEE, 2021. 5, 6

[19] Robin Chan, Krzysztof Lis, Svenja Uhlemeyer, Hermann Blum, Sina Honari, Roland Y. Siegwart, Mathieu Salzmann, P. Fua, and Matthias Rottmann. SegmentMeIfYouCan: A benchmark for anomaly segmentation. *arXiv:2104.14812*, 2021. 3

[20] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *IEEE/CVF International Conference on Computer Vision*, 2021. 3

[21] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository. *arXiv:1512.03012*, 2015. 5

[22] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 3, 7, 8

[23] Clement Creusot and Asim Munawar. Real-time small obstacle detection on highways using compressive RBM road reconstruction. In *IEEE Intelligent Vehicles Symposium*, 2015. 2, 3

[24] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 3

[25] Giancarlo Di Biase, Hermann Blum, Roland Siegwart, and Cesar Cadena. Pixel-wise Anomaly Detection in Complex Driving Scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 3

[26] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning Optical Flow with Convolutional Networks. In *IEEE International Conference on Computer Vision*, 2015. 8

[27] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, 2017. 8

[28] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. VOS: Learning What You Don't Know by Virtual Outlier Synthesis. *arXiv:2202.01197*, 2022. 2, 3

[29] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1996. 5

[30] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88, 2010. 3

[31] Fishyscapes. Results - The Fishyscapes Benchmark. https://fishyscapes.com/results, 2022. 2

[32] Javier Martinez Garcia, Robert Prophet, Juan Carlos Fuentes Michel, Randolf Ebelt, Martin Vossiek, and Ingo Weber. Identification of Ghost Moving Detections in Automotive Scenarios with Deep Learning. In *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility*, 2019. 6

[33] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32, 2013. 3, 5

[34] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S. Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, Tiffany Fernandez, Martin Jänicke, Sudesh Mirashi, Chiragkumar Savani, Martin Sturm, Oleksandr Vorobiov, Martin Oelker, Sebastian Garreis, and Peter Schuberth. A2D2: Audi Autonomous Driving Dataset. *arXiv:2004.06320*, 2020. 3

[35] Matej Grcić, Petra Bevandić, Zoran Kalafatić, and Siniša Šegvić. Dense anomaly detection by robust learning on synthetic negative data. *arXiv:2112.12833*, 2021. 3

[36] Thomas Griebel, Dominik Authaler, Markus Horn, Matti Henning, Michael Buchholz, and Klaus Dietmayer. Anomaly Detection in Radar Data Using PointNets. *arXiv:2109.09401*, 2021. 6

[37] Krishnam Gupta, Syed Ashar Javed, Vineet Gandhi, and K. Madhava Krishna. MergeNet: A Deep Net Architecture for Small Obstacle Discovery. In *IEEE International Conference on Robotics and Automation*, 2018. 7

[38] David Haldimann, Hermann Blum, Roland Siegwart, and Cesar Cadena. This is not what I imagined: Error Detection for Semantic Segmentation through Visual Dissimilarity. *arXiv:1909.00676*, 2019. 3, 4

[39] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *IEEE International Conference on Computer Vision*, 2017. 2

[40] Florian Heidecker, Jasmin Breitenstein, Kevin Rösch, Jonas Löhdefink, Maarten Bieshaar, Christoph Stiller, Tim Fingscheidt, and Bernhard Sick. An Application-Driven Conceptualization of Corner Cases for Perception in Highly Automated Driving. *arXiv:2103.03678*, 2021. 1

[41] Florian Heidecker, Abdul Hannan, Maarten Bieshaar, and Bernhard Sick. Towards Corner Case Detection by Modeling the Uncertainty of Instance Segmentation Networks. *Pattern Recognition. ICPR International Workshops and Challenges*, 12664, 2021. 2, 3

[42] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joe Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling Out-of-Distribution Detection for Real-World Settings. *arXiv:1911.11132*, 2022. 3

[43] Martin Holder, Clemens Linnhoff, Philipp Rosenberger, Christoph Popp, and Hermann Winner. Modeling and simulation of radar sensor artifacts for virtual testing of autonomous driving. In *9. Tagung Automatisiertes Fahren*, 2019. 6

[44] Po-Yu Huang, Wan-Ting Hsu, Chun-Yueh Chiu, Ting-Fan Wu, and Min Sun. Efficient Uncertainty Estimation for Semantic Segmentation in Videos. *Computer Vision – ECCV*, 11205, 2018. 2, 3

[45] Ashesh Jain, Luca Del Pero, Hugo Grimmett, and Peter Ondruska. Autonomy 2.0: Why is self-driving always 5 years away? *arXiv:2107.08142*, 2021. 1

[46] Tianchen Ji, Sri Theja Vuppala, Girish Chowdhary, and Katherine Driggs-Campbell. Multi-Modal Anomaly Detection for Unstructured and Uncertain Environments. *arXiv:2012.08637*, 2020. 7

[47] K J Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards Open World Object Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 1

[48] Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, and Jaegul Choo. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *IEEE/CVF International Conference on Computer Vision*, 2021. 2, 3

[49] Alexander Kamann, Patrick Held, Florian Perras, Patrick Zaumseil, Thomas Brandmeier, and Ulrich T. Schwarz. Automotive Radar Multipath Propagation in Uncertain Environments. In *21st International Conference on Intelligent Transportation Systems*. IEEE, 2018. 6

[50] Andrej Karpathy. Tesla Autonomy Day. https://www.youtube.com/watch?v=Ucp0TTmvqOE, 2019. 1

[51] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. *arXiv:1511.02680*, 2016. 2, 3

[52] Florian Kraus, Nicolas Scheiner, Werner Ritter, and Klaus Dietmayer. Using Machine Learning to Detect Ghost Images in Automotive Radar. In *IEEE 23rd International Conference on Intelligent Transportation Systems*, 2020. 6

[53] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. PointPillars: Fast Encoders for Object Detection From Point Clouds. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 6

[54] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution

samples and adversarial attacks. *Advances in Neural Information Processing Systems*, 31, 2018. 3

[55] Kaican Li, Kai Chen, Haoyu Wang, Lanqing Hong, Chao Ye, Jianhua Han, Yukuai Chen, Wei Zhang, Chunjing Xu, Dit-Yan Yeung, Xiaodan Liang, Zhenguo Li, and Hang Xu. CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving. *arXiv:2203.07724*, 2022. 1

[56] Krzysztof Lis, Sina Honari, Pascal Fua, and Mathieu Salzmann. Detecting Road Obstacles by Erasing Them. *arXiv:2012.13633*, 2021. 2, 3

[57] Krzysztof Lis, Krishna Kanth Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the Unexpected via Image Resynthesis. In *IEEE/CVF International Conference on Computer Vision*, 2019. 2, 3

[58] C. Liu, S. Liu, C. Zhang, Y. Huang, and H. Wang. Multipath propagation analysis and ghost target removal for FMCW automotive radars. In *IET International Radar Conference*, 2021. 5, 6

[59] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future Frame Prediction for Anomaly Detection – A New Baseline. *arXiv:1712.09867*, 2018. 8

[60] Jonas Lohdefink, Justin Fehrling, Marvin Klingner, Fabian Huger, Peter Schlicht, Nico M. Schmidt, and Tim Fingscheidt. Self-Supervised Domain Mismatch Estimation for Autonomous Perception. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020. 3, 4

[61] Cewu Lu, Jianping Shi, and Jiaya Jia. Abnormal event detection at 150 FPS in MATLAB. In *IEEE International Conference on Computer Vision*, 2013. 8

[62] Weixin Luo, Wen Liu, and Shenghua Gao. A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework. In *IEEE International Conference on Computer Vision*, 2017. 8

[63] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. Anomaly detection in crowded scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010. 8

[64] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in Neural Information Processing Systems*, 31, 2018. 3

[65] Mana Masuda, Ryo Hachiuma, Ryo Fujii, Hideo Saito, and Yusuke Sekikawa. Toward Unsupervised 3d Point Cloud Anomaly Detection Using Variational Autoencoder. In *IEEE International Conference on Image Processing*, 2021. 5

[66] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In *IEEE International Conference on Computer Vision*, 2017. 3

[67] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In *IEEE International Conference on Computer Vision*, 2017. 3

[68] Julia Nitsch, Masha Itkina, Ransalu Senanayake, Juan Nieto, Max Schmidt, Roland Siegwart, Mykel J. Kochenderfer, and Cesar Cadena. Out-of-Distribution Detection for Automotive Perception. In *IEEE International Intelligent Transportation Systems Conference*, 2021. 3

[69] Toshiaki Ohgushi, Kenji Horiguchi, and Masao Yamanaka. Road Obstacle Detection Method Based on an Autoencoder with Semantic Segmentation. *Computer Vision –*, 12627, 2021. 2, 3

[70] Peter Pinggera, Sebastian Ramos, Stefan Gehrig, Uwe Franke, Carsten Rother, and Rudolf Mester. Lost and Found: Detecting small road hazards for self-driving vehicles. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016. 3, 7

[71] Peter Pinggera, Sebastian Ramos, Stefan Gehrig, Uwe Franke, Carsten Rother, and Rudolf Mester. Lost and Found: Detecting Small Road Hazards for Self-Driving Vehicles. *arXiv:1609.04653*, 2016. 7

[72] Robert Prophet, Javier Martinez, Juan-Carlos Fuentes Michel, Randolf Ebelt, Ingo Weber, and Martin Vossiek. Instantaneous Ghost Detection Identification in Automotive Scenarios. In *IEEE Radar Conference*, 2019. 6, 7

[73] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017. 4, 6

[74] Yousra Regaya, Fodil Fadli, and Abbes Amira. Point-Denoise: Unsupervised outlier detection for 3D point clouds enhancement. *Multimedia Tools and Applications*, 80, 2021. 4

[75] O. Ronneberger, T. Brox, and P. Fischer. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, 9351, 2015. 8

[76] Fabian Roos, Mohammadreza Sadeghi, Jonathan Bechter, Nils Appenrodt, Jurgen Dickmann, and Christian Waldschmidt. Ghost target identification by analysis of the Doppler distribution in automotive scenarios. In *18th International Radar Symposium*. IEEE, 2017. 6

[77] Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection. In *International Conference on Learning Representations*, 2020. 4, 5

[78] Lukas Rummelhard, Amaury Negre, and Christian Laugier. Conditional Monte Carlo Dense Occupancy Tracker. In *IEEE 18th International Conference on Intelligent Transportation Systems*, 2015. 8

[79] In-hwan Ryu, Insu Won, and Jangwoo Kwon. Detecting Ghost Targets Using Multilayer Perceptron in Multiple-Target Tracking. *Symmetry*, 10, 2018. 5, 6

[80] Nicolas Scheiner, Florian Kraus, Fangyin Wei, Buu Phan, Fahim Mannan, Nils Appenrodt, Werner Ritter, Jurgen Dickmann, Klaus Dietmayer, Bernhard Sick, and Felix Heide. Seeing Around Street Corners: Non-Line-of-Sight Detection and Tracking In-the-Wild Using Doppler Radar. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 6

[81] Jeff Schneider. Self Driving Cars & AI: Transforming our Cities and our Lives. https://www.youtube.com/watch?v=jTio_MPQRYc, 2019. 1

[82] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. PointRCNN: 3D Object Proposal Generation and Detection From Point Cloud. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 4

[83] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2015. 4

[84] Andrea Stocco, Michael Weiss, Marco Calzana, and Paolo Tonella. Misbehaviour prediction for autonomous driving systems. In *Proceedings of the ACM/IEEE International Conference on Software Engineering*, 2020. 8

[85] Lei Sun, Kailun Yang, Xinxin Hu, Weijian Hu, and Kaiwei Wang. Real-time Fusion Network for RGB-D Semantic Segmentation Incorporating Unexpected Obstacle Detection for Road-driving Images. *arXiv:2002.10570*, 2020. 7

[86] Khac-Phuc-Hung Thai, Olivier Rabaste, Jonathan Bosse, Dominique Poullin, Israel Hinostroza, Thierry Letertre, and Thierry Chonavel. Around-the-corner radar: Detection and localization of a target in non-line of sight. In *IEEE Radar Conference*, 2017. 5

[87] Udacity. Final leaderboard of Udacity Challenge 2. `https://github.com/udacity/self-driving-car`, 2016. 3, 8

[88] Tristan Visentin, Jurgen Hasch, and Thomas Zwick. Analysis of multipath and DOA detection using a fully polarimetric automotive radar. In *European Radar Conference*. IEEE, 2017. 6

[89] Tomas Vojir, Tomáš Šipka, Rahaf Aljundi, Nikolay Chumerin, Daniel Olmeda Reino, and Jiri Matas. Road anomaly detection by partial image reconstruction with segmentation coupling. In *IEEE/CVF International Conference on Computer Vision*, 2021. 2, 3

[90] LeiChen Wang, Simon Giebenhain, Carsten Anklam, and Bastian Goldluecke. Radar Ghost Target Detection via Multimodal Transformers. *IEEE Robotics and Automation Letters*, 6, 2021. 5, 7

[91] Xin Wang, Thomas E Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. Frustratingly simple few-shot object detection. *arXiv:2003.06957*, 2020. 1

[92] Xinlong Wang, Zhiding Yu, Shalini De Mello, Jan Kautz, Anima Anandkumar, Chunhua Shen, and Jose M Alvarez. FreeSOLO: Learning to Segment Objects without Annotations. *arXiv:2202.12181*, 2022. 1

[93] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13, 2004. 8

[94] Kelvin Wong, Shenlong Wang, Mengye Ren, Ming Liang, and Raquel Urtasun. Identifying Unknown Instances for Autonomous Driving. *arXiv:1910.11296*, 2019. 4, 5

[95] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, and Alan L. Yuille. Synthesize Then Compare: Detecting Failures and Anomalies for Semantic Segmentation. *Computer Vision – ECCV 2020*, 12346, 2020. 3, 4

[96] Feng Xue, Anlong Ming, Menghan Zhou, and Yu Zhou. A Novel Multi-layer Framework for Tiny Obstacle Discovery. *arXiv:1904.10161*, 2019. 3

[97] Chule Yang, Alessandro Renzaglia, Anshul Paigwar, Christian Laugier, and Danwei Wang. Driving Behavior Assessment and Anomaly Detection for Intelligent Vehicles. In *IEEE International Conference on CIS and IEEE Conference on RAM*, 2019. 7, 8

[98] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. FoldingNet: Point Cloud Auto-Encoder via Deep Grid Deformation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 5

[99] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. *arXiv:1805.04687*, 2020. 3

[100] Chen Zhang, Zefan Huang, Marcelo H. Ang, and Daniela Rus. LiDAR Degradation Quantification for Autonomous Driving in Rain. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021. 4, 5

[101] Mengshi Zhang, Yuqun Zhang, Lingming Zhang, Cong Liu, and Sarfraz Khurshid. DeepRoad: GAN-based metamorphic testing and input validation framework for autonomous driving systems. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, 2018. 3, 4, 8

[102] Yin Zhou and Oncel Tuzel. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 4, 6