This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Learning to Identify Out-of-Distribution Objects for 3D LiDAR Anomaly Segmentation

Simone Mosco Daniel Fusaro Alberto Pretto
University of Padova, Italy
{moscosimon, fusarodani, alberto.pretto}@dei.unipd.it

Abstract

Understanding the surrounding environment is fundamental in autonomous driving and robotic perception. Distinguishing between known classes and previously unseen objects is crucial in real-world environments, as done in Anomaly Segmentation. However, research in the 3D field remains limited, with most existing approaches applying post-processing techniques from 2D vision. To cover this lack, we propose a new efficient approach that directly operates in the feature space, modeling the feature distribution of inlier classes to constrain anomalous samples. Moreover, the only publicly available 3D LiDAR anomaly segmentation dataset contains simple scenarios, with few anomaly instances, and exhibits a severe domain gap due to its sensor resolution. To bridge this gap, we introduce a set of mixed real-synthetic datasets for 3D LiDAR anomaly segmentation, built upon established semantic segmentation benchmarks, with multiple out-of-distribution objects and diverse, complex environments. Extensive experiments demonstrate that our approach achieves state-of-the-art and competitive results on the existing real-world dataset and the newly introduced mixed datasets, respectively, validating the effectiveness of our method and the utility of the proposed datasets. Code and datasets are available at https://simom0.github.io/lido-page/.

1. Introduction

Autonomous vehicles and robotic systems need to perceive the surrounding environment to operate effectively. LiDAR semantic segmentation [23] is a crucial task, which enables understanding of 3D scenes through point-wise classification. Several approaches [20, 27, 66] address this task by exploiting different LiDAR data representations. However, a crucial challenge in real-world scenarios is the ability to identify and segment unknown objects, not previously seen in training (*i.e.*, anomalies). Most methods are trained on standard closed-set assumptions with a fixed set of classes and struggle to identify unknown objects.

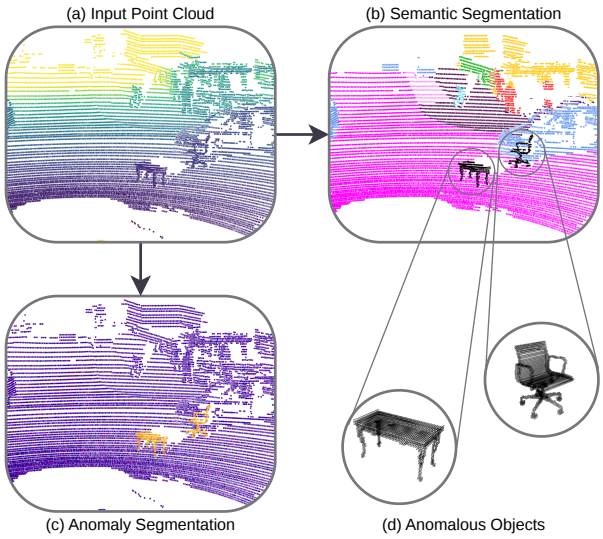


Figure 1. Overview of the LiDAR anomaly segmentation task. Given an input point cloud containing a previously unseen object (a), the purpose is to simultaneously perform semantic segmentation on known classes (b) and assign each point a score (c) indicating its probability of belonging to an anomalous object (d).

Recent approaches for anomaly detection and anomaly segmentation [16, 41, 42, 55] achieve promising progress in these tasks, supported by the availability of several datasets [6, 9, 31]. However, they are mainly related to the image domain, while most autonomous driving datasets [4, 7, 45] rely on LiDAR sensors for robust 3D perception. Some works [8, 61] address the open-set problem in 3D but struggle to segment anomalous objects or depend on unlabeled regions during training, considering them as anomalies. More recent baselines [43] adapt simple post-processing techniques or use model ensemble [30], which is computationally expensive and slow. In contrast, our method efficiently operates in the feature space, learning the distribution of inlier classes without relying on unlabeled regions or anomaly objects during training.

Furthermore, only a few LiDAR datasets for anomaly segmentation are available. The STU dataset [43] contains high-resolution point clouds, introducing a large domain gap with standard training data, and includes only binary anomaly masks for evaluation. Some datasets [61] are proprietary and thus not publicly available, while others [31] select unlabeled objects as anomalies, meaning that some still appear in the training data. To address these limitations, we construct a set of mixed real-synthetic datasets based on popular LiDAR segmentation benchmarks. We insert synthetic anomaly objects from ModelNet [67], ensuring a domain distribution distinct from the training classes. The resulting datasets include multiple LiDAR resolutions and frequent anomaly occurrences for diverse and challenging evaluation scenarios, with semantic labels.

The main contribution of this paper is a novel approach for LiDAR anomaly segmentation that operates directly in the feature space. Our method jointly performs semantic and anomaly segmentation by learning a distribution of inlier class prototypes, enabling the detection of unseen objects without requiring anomaly samples during training. Additionally, we introduce a new set of mixed real-synthetic datasets for LiDAR anomaly segmentation, constructed by combining real LiDAR scans and synthetic out-of-distribution objects. We design a strategy to maintain realism by aligning the geometric properties and point distribution of the inserted objects with the LiDAR beams measurement pattern and assigning consistent remission values based on reflection models. The key contributions of this work can be summarized as:

- We propose a novel deep learning approach for efficient 3D LiDAR anomaly segmentation that directly operates on the feature space.
- We introduce a new set of mixed real-synthetic LiDAR datasets with out-of-distribution objects, based on publicly available benchmarks.
- We demonstrate that our approach achieves state-of-the-art and competitive performance on the real and mixed datasets, respectively.

2. Related Work

2.1. Point Cloud Semantic Segmentation

Point cloud semantic segmentation aims to assign a class label to each 3D point. There exist four main categories to group these approaches, namely point-based, projection-based, voxel-based, and hybrid methods. Point-based methods directly process the raw 3D points. Pioneering works [47, 48] use MLPs and symmetric pooling functions to learn per-point features. Several methods [58, 60, 64] introduce point convolution, while others rely on the attention mechanism [65, 66, 75]. Recent works [21, 39, 46] combine both 3D and 2D operations with plane projections. Projection-based approaches project the point cloud onto a 2D surface. Several approaches [12, 15, 37, 62, 63, 69] operate on the range image representation, leveraging traditional 2D backbones, others [1, 74] exploit the bird-eye-view representation. Recent works [2, 27, 40] adapt attention-based backbones from the 2D field, improving results. Voxel-based methods represent the point cloud as a grid of 3D voxels and apply 3D convolutional operations. Initial approaches focus on computational efficiency, such as designing sparse 3D convolutions [14] or LiDAR-suitable 3D representations [78]. Some approaches [13, 29] integrate attention, while recent work [20] redesigns 3D sparse kernels for reduced complexity. Hybrid methods combine different point cloud representations or integrate data from other sensors. Several approaches [25, 57, 70] exploit different representations, while others [35, 71] leverage also RGB images. Recent trends focus on robust embedding learning [32] and on using 2D vision foundation models [73].

2.2. Anomaly Segmentation

Anomaly segmentation extends the task of anomaly detection by predicting if each individual pixel in an image or point in a point cloud belongs to an anomaly. Early approaches for 2D images [10, 24] develop post-processing techniques and directly work on softmax activations, some use an additional class in training for anomalies [6, 38], and others apply model ensembles [30] to measure disagreement. A line of work [18, 68, 76] employs generative models to resynthesize the input image and look at dissimilar areas for anomalies. Many unsupervised methods use synthetic anomaly data and train an anomaly detector [34, 51, 59]. Other approaches focus on confidence and uncertainty-based predictions [16, 22], entropy computation [42], or modeling the feature space to discriminate anomaly features from the inlier classes [55]. Recently, vision-language models based on CLIP [50, 77] have been applied in anomaly segmentation. Different methods [41, 49] explore the mask prediction mechanism to directly identify anomalies. Alongside, a large amount of data [3, 6, 9] is available when working with 2D images.

LiDAR anomaly segmentation is still not well explored. Several approaches are based on 2D post-processing techniques or model ensembles [43], while others focus on the open-set instance segmentation task [8, 61], performing poorly on anomaly objects. Moreover, few LiDAR datasets are available for anomaly segmentation. SOD [54] has a few low-resolution 16-beam LiDAR data, introducing a large domain gap. TOR4D and Rare4D [61] contain real-world LiDAR data but are proprietary and not publicly accessible. CODA [31] is derived from existing LiDAR datasets but faces the limitation that some anomaly objects may appear in training. Recent work STU dataset [43] provides

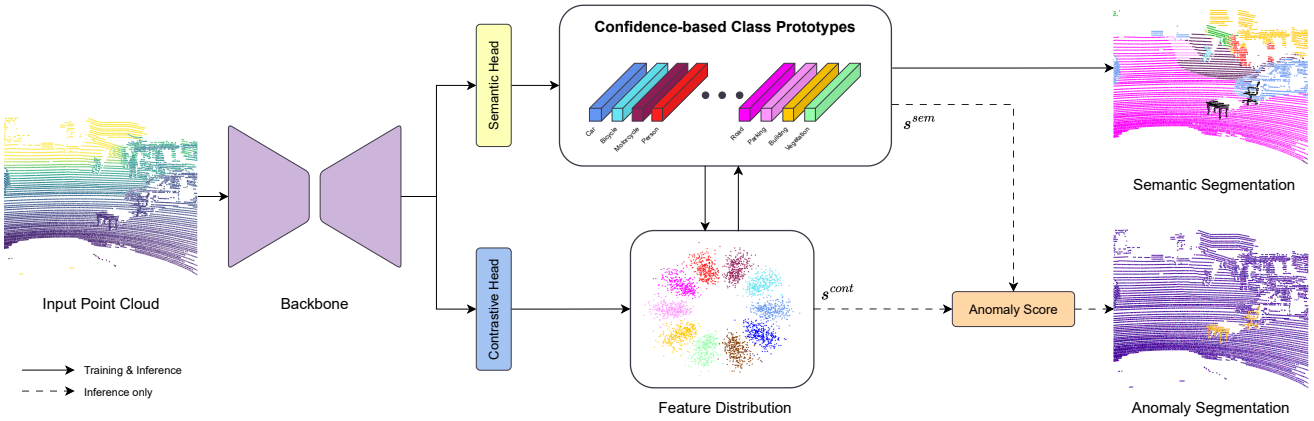


Figure 2. Overview of the proposed LIDO approach. A backbone extracts per-point features, which are then processed by two different heads. The segmentation head predicts point-wise class labels while constructing class prototypes in the feature space. The contrastive head models the distribution of inlier class features to better identify anomalies. During inference, the outputs of both heads are combined to produce the final anomaly segmentation scores.

large real-world data with anomaly annotation with high-resolution 128-beam LiDAR data. In this work, we propose a new approach for LiDAR anomaly segmentation, inspired by [55], along with three new datasets at different resolutions to address domain gaps in 3D anomaly segmentation.

3. Methodology

In this section, we present the architecture of our proposed approach, **LIDO** (**L**earning to **I**dentify Out-of-**D**istribution **O**bjects), illustrated in Fig. 2. In our setting, an *anomaly* denotes any point whose geometric characteristics do not correspond to any semantic class seen during training, i.e., an out-of-distribution point. LIDO is composed of a backbone to extract per-point features and two branches: the first produces semantic segmentation predictions and builds per-class prototypes, while the second directly models feature distribution to identify anomaly instances. We adopt MinkowskiNet [14] as a backbone for feature extraction, two linear layers as heads to produce final features for each branch, and a scoring mechanism to compute anomaly predictions in inference.

3.1. Problem Formulation

We represent a LiDAR scan $\mathbf{X} \in \mathbb{R}^{N\times 4}$ as a set of N points $\mathbf{p}_n = (x_n, y_n, z_n, i_n)$, where each point consists of 3D coordinates and intensity value, respectively. For each scan, we have a corresponding set of labels $\mathbf{Y} = \{y_n\}_{n=1}^{N}$, where each $y_n = \{1, \dots, C\}$, encodes a class c and C denotes the number of inlier classes that the network sees during training. The goal of LiDAR anomaly segmentation is to produce semantic predictions $\hat{\mathbf{Y}} = \{\hat{y}_n\}_{n=1}^{N}$ for inlier classes, $\hat{y}_n = \{1, \dots, C\}$, and jointly assign a score $s_n \in [0, 1]$ to each point $\mathbf{p}_n$, to predict the probability of a point being an anomaly or not. We denote with $\mathbf{X}_c = \{\mathbf{p}_n \in \mathbf{X} | y_n = c\}$ the set of points whose ground truth class is c, and with $\hat{\mathbf{X}}_c = \{\mathbf{p}_n \in \mathbf{X}_c | \hat{y}_n = y_n\}$ the set of true positive points for class c, where the ground truth and predicted class match.

3.2. Semantic Head

The semantic head is responsible for generating the semantic predictions and, at the same time, constructing a robust prototype for each inlier class. We use standard practice and optimize it with a weighted cross-entropy loss:

$$\mathcal{L}_{\text{ce}} = -\frac{1}{N}\sum_{n=1}^{N} w_c y_n \log(\sigma(f_n)), \tag{1}$$

where w_c corresponds to the class weight for class c, y_n is the ground truth class for point $\mathbf{p}_n$, σ denotes the softmax operation and f_n represents the pre-softmax feature vector predicted for point $\mathbf{p}_n$.

During training, we are also interested in the construction of class prototypes to gather all points belonging to a certain class closer in the feature space. To obtain this, we consider pre-softmax features of all true positives for each class and accumulate them. We extract a confidence value $\kappa_{\mathbf{p}}$ for each point by taking the maximum component of its pre-softmax feature vector $\kappa_{\mathbf{p}} = \max(f_{\mathbf{p}})$ and compute a confidence-based prototype (CP) for each class c, $\text{CP} = \{\text{CP}_1, \dots, \text{CP}_c\}$:

$$\text{CP}_c = \frac{\sum_{\mathbf{p}\in\hat{\mathbf{X}}_c} \kappa_{\mathbf{p}} f_{\mathbf{p}}}{\sum_{\mathbf{p}\in\hat{\mathbf{X}}_c} \kappa_{\mathbf{p}}}, \tag{2}$$

where $f_{\mathbf{p}}$ are the pre-softmax features of point $\mathbf{p}$.

At the beginning of each epoch e, we use the confidence-based prototypes computed in the previous epoch CP_c^{e-1} to

guide the network into producing feature vectors for points with ground truth class c, close to CP_c^{e-1}. We achieve this by introducing a prototype-based cosine embedding loss that enforces proximity between each feature and its corresponding prototype:

$$\mathcal{L}_{\text{prot}} = \frac{1}{N} \sum_{c \in C} \sum_{\mathbf{p} \in \mathbf{X}_c} \left(1 - \left\langle \text{CP}_c^{e-1}, f_{\mathbf{p}} \right\rangle\right). \tag{3}$$

In the first epoch, this loss is not active as there are no accumulated prototypes yet. Overall, the semantic head is optimized with a weighted sum of the above-mentioned losses, combined with the Lovasz Loss [5] $\mathcal{L}_{\text{lovasz}}$:

$$\mathcal{L}_{\text{shead}} = \lambda_1 \mathcal{L}_{\text{ce}} + \lambda_2 \mathcal{L}_{\text{lovasz}} + \lambda_3 \mathcal{L}_{\text{prot}}. \tag{4}$$

3.3. Contrastive Head

The aim of the contrastive head is to directly identify points belonging to anomalies by learning discriminative and distinguished per-class prototypes, and modeling their distribution in the feature space, similar to [55]. To achieve this, we adopt both the contrastive loss [11] and the objectosphere loss [17].

We denote with $f'_{\mathbf{p}}$ the pre-softmax features produced by the contrastive head for point $\mathbf{p}_n$ and compute the mean feature vector $\bar{f}_c$ for each class c as:

$$\bar{f}_c = \frac{1}{|\mathbf{X}_c|} \sum_{\mathbf{p} \in \mathbf{X}_c} f'_{\mathbf{p}}, \tag{5}$$

considering all the points whose ground truth label is c, with $|\mathbf{X}_c|$ the cardinality of $\mathbf{X}_c$. Then, we employ the contrastive loss $\mathcal{L}_{\text{cont}}$ to align the mean feature vector $\bar{f}_c$ to the normalized confidence-based prototype CP_c^{e-1} for the corresponding class c, at the previous epoch, and at the same time push them away from the feature vectors of the other classes:

$$\mathcal{L}_{\text{cont}} = -\sum_{c \in C} \log \frac{\exp(\langle \bar{f}_c, \text{CP}_c^{e-1} \rangle / \tau)}{\sum_{i=1}^{C} \exp(\langle \bar{f}_c, \text{CP}_c^{e-1} \rangle / \tau)}, \tag{6}$$

where τ is a temperature parameter. In addition, we use the objectosphere loss $\mathcal{L}_{obj}$ over each point $\mathbf{p} \in \mathbf{X}$, denoted as:

$$\mathcal{L}_{\text{obj}} = \begin{cases} \max(r - \|f'_{\mathbf{p}}\|^2, 0) & \text{if } \mathbf{p} \in \mathcal{D}_{in} \\ \|f'_{\mathbf{p}}\|^2 & \text{otherwise} \end{cases}, \tag{7}$$

where $f'_{\mathbf{p}}$ is the feature vector of a point $\mathbf{p}$ belonging to the set $\mathcal{D}_{in}$ of inlier classes, and r is a fixed threshold. Different from [55], we do not use unlabeled or void regions in training to learn anomaly features; thus, the goal of this loss is to push feature vectors of inlier classes far from the center of the C-dimensional hypersphere of radius r.

The contrastive head is then optimized with a weighted sum of these two losses as follows:

$$\mathcal{L}_{\text{chead}} = \lambda_4 \mathcal{L}_{\text{cont}} + \lambda_5 \mathcal{L}_{\text{obj}}. \tag{8}$$

3.4. Inference

In order to obtain both semantic segmentation and anomaly predictions, we combine the outputs of the two heads. The semantic head provides standard semantic segmentation predictions through cosine similarity among confidence-based prototypes CP_c for class c, accumulated during training, and per-point features f_n, for point $\mathbf{p}_n$:

$$\text{sim}_{n,c} = \langle f_n, \text{CP}_c \rangle. \tag{9}$$

The predicted inlier semantic class is obtained considering the maximum similarity values among all classes:

$$\hat{y}_n = \text{argmax}_c(\text{sim}_{n,c}). \tag{10}$$

The semantic head also produces a score that estimates if a point $\mathbf{p}_n$ belongs to an anomaly object, combining both the cosine distance and the entropy of the softmax values. First, we denote as s_n^{cos} the score corresponding to the maximum cosine distance:

$$s_n^{cos} = 1 - \max_c(\text{sim}_{n,c}), \tag{11}$$

where an high score considers the point $\mathbf{p}_n$ as anomaly. Second, we compute an entropy-based score s_n^{ent} using the normalized Shannon entropy [53] as:

$$s_n^{ent} = -\frac{1}{\log C} \sum_{c \in C} p_{n,c} \log(p_{n,c}), \tag{12}$$

where $p_{n,c}$ is the softmax probability of point $\mathbf{p}_n$ belonging to class c, obtained from $p_n = \sigma(f_n)$, with σ that denotes the softmax operation. Higher entropy values indicates the network uncertainty on assigning a point to a specific class, meaning that it may belong to an anomaly.

Then, we combine these two scores and obtain a point-wise score $s_n^{sem} = s_n^{cos} \cdot s_n^{ent}$ and normalize it with respect to the maximum value, to ensure values in range $[0, 1]$.

The contrastive head produces another anomaly prediction scores, based on the principle of the objectosphere loss [17]. We consider anomaly all the points whose feature vectors have norm below a certain threshold r, as specified in Eq. (7). Thus, we define a point-wise score s_n^{cont} for point $\mathbf{p}_n$ as:

$$s_n^{cont} = \max\left(0, \left(1 - \frac{\|f'_n\|^2}{r}\right)\right). \tag{13}$$

This score is 1 when the feature vector has norm equals to 0, while 0 when the norm is greater that the threshold r.

Finally, we obtain a per-point score for belonging to an anomaly, fusing the two scores from the heads as:

$$s_n = \frac{1}{2}\left(s_n^{sem} + s_n^{cont}\right). \tag{14}$$

Table 1. Comparison of publicly available LiDAR datasets for Anomaly Segmentation. BB indicates bounding box, SM semantic masks, and BM binary masks. Size refers to the number of scans in the dataset, #OoD Instances denotes the total number of anomaly objects.

Dataset	#Beams	Size	Labels	#OoD Instances
CODA-KITTI [31]	64	309	BB	399
CODA-nuScenes [31]	32	134	BB	1125
CODA-ONCE [31]	40	1057	BB	4413
SOD [54]	16	460/530	SM	-
STU [43]	128	8022/1960	BM	1965/-
nuScenes-OoD (Ours)	32	6019	SM	2398/7268
SemanticPOSS-OoD (Ours)	40	500	SM	196/586
SemanticKITTI-OoD (Ours)	64	4071	SM	1634/4894

4. Out-of-Distribution Datasets

In this section, we introduce our mixed real-synthetic Out-of-Distribution (OoD) datasets for 3D LiDAR anomaly segmentation. The datasets (see Tab. 1) are constructed from three autonomous driving benchmarks with different LiDAR sensor resolutions, complementary to the only available real-world dataset, STU [43]. We use ModelNet [67] as a source for synthetic anomaly objects, filtering its models to avoid overlap with categories and objects present in real-world LiDAR datasets (see Supplementary Material). To ensure realism, we also introduce a protocol for inserting synthetic objects into real LiDAR scans, manipulating point distributions, intensity values, and aligning them to the LiDAR sensor geometry in a beam-like format. The proposed OoD datasets are as follows:

nuScenes-OoD is constructed from the official validation set of nuScenes [7] dataset, comprising 6019 scans collected with a 32-beam LiDAR sensor.

SemanticPOSS-OoD is derived from the 500 samples in the validation sequence of SemanticPOSS [45], acquired with a 40-beam LiDAR sensor.

SemanticKITTI-OoD is based on the 4071 scans in the validation set of SemanticKITTI [4], generated with a 64-beam LiDAR sensor.

Each dataset contains two different versions, a *single* and a *multi* split, comprising respectively a single anomaly object or multiple ones in each single scan. To better resemble the real-world environment, where anomalies are not so frequent, but at the same time provide a strong evaluation setup, about 40% of the scans in the *single* split contain anomalies, while 60% for the *multi* split. Moreover, to set a different level of difficulty, anomaly objects in the *single* split are inserted only on the road, while in the *multi* split, also on other surfaces, as parking areas or sidewalks.

4.1. Protocol

Consider a real LiDAR point cloud $\mathbf{P} \in \mathbb{R}^{N\times 4}$ with N points, we randomly sample a model, from the filtered ModelNet dataset, $\mathbf{O} \in \mathbb{R}^{M\times 4}$, with M points and inject it into the LiDAR scan, resulting in a combined point cloud $\mathbf{S} = [\mathbf{P}, \mathbf{O}] \in \mathbb{R}^{(N+M)\times 4}$. For the insertion, we consider only planar surfaces: solely the road for the *single* split, and other available planar surfaces (e.g., road, sidewalks, parking areas, etc.) for the *multi* split. Standard augmentation techniques as rotation and scaling, are applied to the object points, ensuring a coherent size ratio with respect to the real-world scan. Points in $\mathbf{S}$ are then projected into a range image representation [37], and re-projected in the 3D space, resulting in a point cloud $\mathbf{S}' = [\mathbf{P}', \mathbf{O}']$, with $\mathbf{P}'$ and $\mathbf{O}'$ the filtered scan and object point clouds, respectively. The re-projection operation allows us to recover information from original points potentially occluded during object insertion and, at the same time, geometrically align the object points to the LiDAR spatial property, sampling points from the inserted object in the LiDAR beams format. Since ModelNet objects are not provided with intensity values, and earlier steps used temporary intensity values, we compute them using the Lambertian reflectance model [44]. Different from simulators as CARLA [19], which compute intensity in a simplified way, based solely on the distance from the sensor, our approach relies on physical model properties to better match the real-world behavior. The intensity i is computed as:

$$i = \frac{\rho \cdot \max(0, -\langle \mathbf{n}, \mathbf{r} \rangle)}{d^2}, \tag{15}$$

where ρ is the reflectivity value of the object, $\mathbf{n}$ is the surface normal at the point (estimated from its neighbors), $\mathbf{r}$ is the LiDAR beam direction towards the point, and d is the distance from the sensor. When the normal faces towards the sensor, the intensity value is higher, while it falls to 0 if the normal faces sideways or away. The reflectivity value ρ is an inherent property of the model's material, which measures how much light a surface reflects. We assign these values based on real object measurements (see Supplementary Material). To increase consistency with the target scan, we normalize the intensity values of the object with respect to the average intensity of the scan in which it is inserted and add a small per-point Gaussian perturbation as noise to better match the real-world properties. Finally, we obtain a mixed real-synthetic LiDAR scan with one or more injected ModelNet models representing anomalies, by just fusing points from the two refined point clouds and updating labels accordingly.

5. Experiments

The experimental section is designed to support our claims that: (i) our approach effectively works on the feature space to address the 3D LiDAR anomaly segmentation task; (ii) the proposed mixed real-synthetic datasets provide a valuable and challenging benchmark for this problem; (iii) the

Table 2. Anomaly Segmentation performance on STU dataset.

Method	Validation Set AUROC [%] ↑	Validation Set FPR@95 [%] ↓	Validation Set AP [%] ↑	Test Set AUROC [%] ↑	Test Set FPR@95 [%] ↓	Test Set AP [%] ↑
Mask4Former3D + MC Dropout [56]	65.76	79.82	0.17	61.51	82.37	0.11
Mask4Former3D + RbA [41]	73.00	100.0	1.64	66.38	100.0	0.81
Mask4Former3D + Max Logit [24]	87.27	68.76	2.02	84.53	81.49	0.95
Mask4Former3D + Void Classifier [6]	89.77	79.50	2.62	85.99	78.60	3.92
Mask4Former3D + Deep Ensemble [30]	90.93	37.34	6.94	86.74	58.05	5.17
LIDO (ours)	**95.05**	**34.86**	**27.53**	**93.67**	**34.29**	**14.99**

Table 3. Anomaly Segmentation performance on SemanticPOSS-OoD dataset.

Method	Single AUROC [%] ↑	Single FPR@95 [%] ↓	Single AP [%] ↑	Multi AUROC [%] ↑	Multi FPR@95 [%] ↓	Multi AP [%] ↑
Mask4Former3D + RbA [41]	49.09	100.0	0.13	50.42	100.0	0.26
Mask4Former3D + Max Logit [24]	61.05	91.85	0.19	63.77	88.74	0.40
Mask4Former3D + Deep Ensemble [30]	85.86	58.12	0.82	85.83	54.20	1.21
LIDO (ours)	**91.51**	**45.10**	**3.97**	**90.84**	**44.74**	**5.92**

proposed approach achieves competitive performance on both real and mixed datasets.

5.1. Implementation Details

Evaluation Setup. We evaluate our approach on four datasets: the STU [43] dataset and the three introduced mixed real-synthetic OoD datasets. For STU, we train our model solely on the SemanticKITTI [4] training split. We use the official STU split, with 19 sequences for validation and 51 for test. For our mixed real-synthetic OoD datasets, we follow standard training procedures used in LiDAR semantic segmentation. Models are trained on the corresponding base dataset, SemanticKITTI [4], SemanticPOSS [45], nuScenes [7], and evaluated on our modified validation split, which includes out-of-distribution objects (Sec. 4).

Training Setup. We train our model from scratch on a single NVIDIA A40 GPU for 64 epochs, with a batch size of 4 for all datasets. We use an SGD optimizer and a cosine annealing scheduler with linear warm-up. The learning rate increases to 2.4×10^{-1} over the first 5 epochs and then decreases to 1×10^{-2}, with weight decay 1×10^{-4}. We set $r = 5.0$, $\tau = 0.1$ and loss weights $\lambda_1 = 1.0$, $\lambda_2 = 1.5$, $\lambda_3 = 0.1$, $\lambda_4 = 0.5$, and $\lambda_5 = 0.5$. During training, we use standard augmentation techniques as rotation, flip, and scale. We do not use model ensembles nor test-time augmentations.

Metrics. Following [43], we adopt common metrics [6] for point-level evaluation on anomaly segmentation as Average Precision (AP), False Positive Rate at 95% True-Positive Rate (FPR@95), and Area Under the Receiver Operating characteristic Curve (AUROC). For semantic segmentation, we use the standard mean Intersection over Union (mIoU).

5.2. 3D LiDAR Anomaly Segmentation

The first experiment shows that our approach achieves state-of-the-art results on STU validation and test sets, as reported in Tab. 2. Our method surpasses all other approaches across all metrics, including ensemble-based [30] approaches (**+9.82%** AP). Despite the significant domain gap, the proposed method effectively models the feature distribution of inlier classes, improving the segmentation of anomaly objects, showing high AP and robust FPR metrics. We attribute this to the discriminative class-wise features learned during training, which mitigate the domain gap.

The second set of experiments highlights the value of the proposed mixed real-synthetic OoD datasets and shows that our method achieves competitive results on these benchmarks as well. Given the recent popularity of 3D LiDAR anomaly segmentation task, and the limited number of available approaches, we retrained several baselines from [43] built upon [72], whose implementations are publicly available, for a fair comparison. Extended experiments and further baseline comparisons are provided in the Supplementary Material.

Our approach significantly surpasses all other methods on both SemanticPOSS-OoD splits (Tab. 3), further demonstrating the effectiveness of modeling the feature space to distinguish between inlier and out-of-distribution classes. This benchmark, however, proves to be particularly challenging and the performance remains bounded, likely due to the large number of sparse instances in SemanticPOSS, which can be misinterpreted as anomalies during inference. The lower scan resolution also limits the number of features available for training, reducing overall performance.

Table 4 reports the results on SemanticKITTI-OoD. Our method achieves state-of-the-art performance on the *simple*

Table 4. Anomaly Segmentation performance on SemanticKITTI-OoD dataset.

Method	Single			Multi		
	AUROC [%] ↑	FPR@95 [%] ↓	AP [%] ↑	AUROC [%] ↑	FPR@95 [%] ↓	AP [%] ↑
Mask4Former3D + RbA [41]	59.68	100.0	4.56	71.39	100.0	13.63
Mask4Former3D + Max Logit [24]	76.49	99.63	4.84	84.05	99.32	**13.84**
Mask4Former3D + Deep Ensemble [30]	92.87	**27.69**	6.20	**92.19**	**28.64**	12.04
LIDO (ours)	**93.36**	31.19	**10.60**	89.89	39.04	9.42

Table 5. Anomaly Segmentation performance on nuScenes-OoD dataset

Method	Single			Multi		
	AUROC [%] ↑	FPR@95 [%] ↓	AP [%] ↑	AUROC [%] ↑	FPR@95 [%] ↓	AP [%] ↑
Mask4Former3D + RbA [41]	32.65	100.0	3.56	71.60	98.68	7.95
Mask4Former3D + Max Logit [24]	66.17	98.60	3.54	43.54	100.0	8.98
Mask4Former3D + Deep Ensemble [30]	**91.79**	**34.59**	**18.34**	**88.62**	**43.63**	**23.87**
LIDO (ours)	89.33	39.70	6.79	87.25	44.51	10.53

split and competitive results on the *multi* split, slightly behind the baselines from [43]. However, these approaches exhibit a high false positive rate (FPR), see Fig. 3, meaning that they tend to predict most points as anomalies. Moreover, the deep ensemble method, while effective, is computationally expensive and resource demanding compared to our method (see Tab. 6).

We report anomaly segmentation results on nuScenes-OoD in Tab. 5, where our approach achieves competitive performance despite the lower resolution of the LiDAR sensor. While it does not surpass the ensemble-based approach, it is important to note that these methods require significantly higher memory and computational resources (Tab. 6), leading to slower inference, whereas our approach remains lightweight and efficient. The high FPR values indicate a general model uncertainty on this challenging benchmark, partially explaining the advantage of model ensembles. These findings mark the effectiveness and practicality of the proposed feature-based approach while also suggesting room for improvement.

Overall, our approach improves upon previous methods, although metrics remain bounded, in particular, the average precision (AP). This limitation is mainly caused by the significant class imbalance inherent in LiDAR datasets [70] and the uncertainty issues observed in popular LiDAR semantic segmentation models [28]. Most models often misclassify similar surfaces, classify road objects as inlier classes [26], or struggle with sparse distant points [32], leading to false positives. Interestingly, the higher AP values on the *multi* splits may be explained by the greater number of anomaly points, which, combined with model uncertainty, tends to increase precision estimates.

We report standard semantic segmentation results in Tab. 7. The proposed approach, with the additional losses for anomaly segmentation, maintains competitive semantic segmentation performance, with slightly lower results compared to a baseline trained in the standard setting. The only exception is nuScenes-OoD, where the lower LiDAR resolution and fewer points per scan reduce effectiveness and class prototype building, reflecting the performance on the anomaly segmentation task. See Supplementary Material for detailed class-wise segmentation metrics.

Table 6. Model complexity and runtime comparison on nuScenes-OoD dataset. Tested on a NVIDIA A40 GPU.

Method	Params (M)	Runtime (ms)	Memory (GB)
Mask4Former3D [72]	39.6	168	1.8
Deep Ensemble (Sequential) [30]	118.8	861	1.9
Deep Ensemble (Parallel) [30]	118.8	287	5.7
LIDO (ours)	**21.7**	**38**	**0.6**

Table 7. Semantic segmentation results (mIoU, %) of standard baseline and our proposed approach. S and M denote the single and multi splits of our OoD datasets.

	KITTI-OoD		POSS-OoD		nuScenes-OoD		STU
	S	M	S	M	S	M	
Standard	**64.99**	**64.84**	**57.07**	**57.05**	**72.75**	**72.62**	**36.75**
LIDO (ours)	61.34	61.19	55.63	55.58	60.61	60.44	35.14

Table 6 reports computational complexity and runtime of different models. The results support our claim that the proposed approach is significantly more efficient and memory-friendly than ensemble models, achieving real-time performance (< 100 ms) and a general balanced trade-off between accuracy and computational cost.

5.3. Ablation Study

Finally, we conduct an ablation study to assess the contribution of each component in the proposed LiDAR anomaly segmentation pipeline. Starting from a baseline trained for

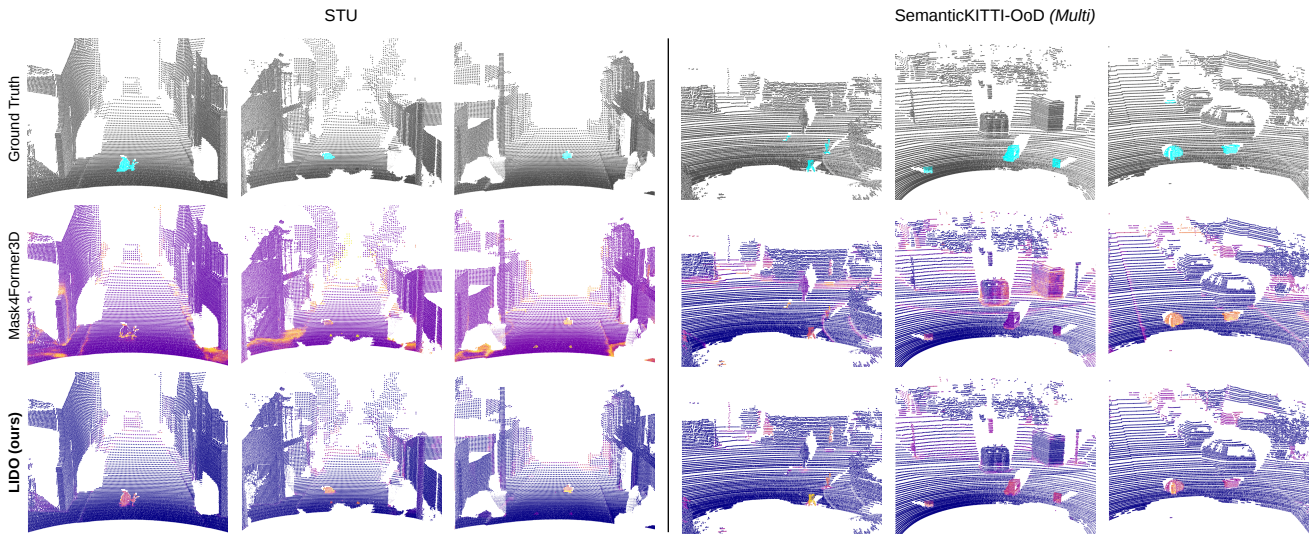


Figure 3. Anomaly Segmentation results on STU and our proposed SemanticKITTI-OoD dataset (Multi split). Ground truth anomaly objects are shown in cyan. Predicted anomaly scores range from blue to orange, indicating increasing probability of a point being anomalous.

standard LiDAR semantic segmentation, we analyze the effect of the additional loss functions introduced in the semantic and contrastive heads, together with the impact of various post-processing techniques. In Table 8 we see that simple thresholding (A) leads to poor results, while adding the prototype loss slightly improves the results, both with thresholding (B) and cosine distance (C), though the low AP and FPR values suggest that most scores are close to zero (*i.e.*, few anomalies detected). A significant improvement comes when combining prototype loss and contrastive loss, particularly when using the semantic head score (F), which benefits from incorporating entropy of Eq. (12) in the score computation. Even thresholding (D) and cosine distance (E) confirm the benefit of per-class contrastive learning, pushing features apart in embedding space. The addition of the objectosphere loss yields further improvements when using the corresponding scores (G and H), although the latter shows a higher false positive rate, indicating that relying only on the features norm is suboptimal. The combined score (I) achieves the best overall results, demonstrating the complementarity of the proposed losses.

Table 8. Ablation study of the LiDAR anomaly segmentation pipeline on STU validation set. $\mathcal{L}_{\text{prot}}$ refers to the prototype loss in Eq. (3), $\mathcal{L}_{\text{cont}}$ to the contrastive loss of Eq. (6), and $\mathcal{L}_{\text{obj}}$ to the objectosphere loss in Eq. (7). "INF" indicates the inference technique: "ML" for max logit, "C" for cosine distance, "SH" for the semantic head score, "CH" for the contrastive head score, and s_n for the proposed score described in Sec. 3.4.

	$\mathcal{L}_{\text{prot}}$	$\mathcal{L}_{\text{cont}}$	$\mathcal{L}_{\text{obj}}$	INF	AUROC ↑	FPR@95 ↓	AP ↑
(A)				ML	90.92	43.49	0.97
(B)	✓			ML	92.64	26.60	1.05
(C)	✓			C	93.61	**24.84**	1.89
(D)	✓	✓		ML	93.13	26.73	2.54
(E)	✓	✓		C	91.71	34.91	4.86
(F)	✓	✓		SH	95.04	26.71	12.88
(G)	✓	✓	✓	SH	92.68	36.57	14.82
(H)	✓	✓	✓	CH	88.16	100.0	16.67
(I)	✓	✓	✓	s_n	**95.05**	34.86	**27.53**

6. Conclusion

In this paper, we introduced a novel approach for LiDAR anomaly segmentation that operates directly in the feature space, modeling the distribution of inlier class representations to identify out-of-distribution objects. Our approach combines prototype, contrastive, and objectosphere losses to constrain anomaly features in the embedding space. In addition, we presented a new set of mixed real-synthetic out-of-distribution datasets, constructed from popular autonomous driving benchmarks with geometrically aligned synthetic anomalies, addressing the scarcity of datasets for this task. Extensive experiments demonstrated the effectiveness of our approach, which achieves state-of-the-art and competitive results on both real and mixed datasets.

Future Work. Despite the encouraging results, there is room for improvement. We plan to extend our approach to cross-domain tasks [33, 52] to enhance generalization and to further investigate the impact of uncertainty [28, 36] in LiDAR semantic segmentation to improve reliability and robustness of anomaly predictions.

References

[1] Eren Erdal Aksoy, Saimir Baci, and Selcuk Cavdar. Salsanet: Fast road and vehicle segmentation in lidar point clouds for autonomous driving. In *IV*, pages 926–932, 2020. 2

[2] Angelika Ando, Spyros Gidaris, Andrei Bursuc, Gilles Puy, Alexandre Boulch, and Renaud Marlet. Rangevit: Towards vision transformers for 3d semantic segmentation in autonomous driving. In *CVPR*, pages 5240–5250, 2023. 2

[3] Alberto Bacchin, Davide Allegro, Stefano Ghidoni, and Emanuele Menegatti. Sood-imagenet: a large-scale dataset for semantic out-of-distribution image classification and semantic segmentation. In *ECCV*, pages 80–97, 2024. 2

[4] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In *ICCV*, pages 9297–9307, 2019. 1, 5, 6

[5] Maxim Berman, Amal Rannen Triki, and Matthew B Blaschko. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *CVPR*, pages 4413–4421, 2018. 4

[6] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *IJCV*, 129 (11):3119–3135, 2021. 1, 2, 6, 4

[7] 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 *CVPR*, 2020. 1, 5, 6

[8] Anirudh S Chakravarthy, Meghana Reddy Ganesina, Peiyun Hu, Laura Leal-Taixé, Shu Kong, Deva Ramanan, and Aljosa Osep. Lidar panoptic segmentation in an open world. *IJCV*, 133(3):1153–1174, 2025. 1, 2

[9] Robin Chan, Krzysztof Lis, Svenja Uhlemeyer, Hermann Blum, Sina Honari, Roland Siegwart, Pascal Fua, Mathieu Salzmann, and Matthias Rottmann. Segmentmeifyoucan: A benchmark for anomaly segmentation. *NeurIPS*, 2021. 1, 2

[10] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *ICCV*, pages 5128–5137, 2021. 2

[11] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607. PmLR, 2020. 4

[12] Hui Xian Cheng, Xian Feng Han, and Guo Qiang Xiao. Cenet: Toward concise and efficient lidar semantic segmentation for autonomous driving. In *ICME*, pages 01–06, 2022. 2

[13] Ran Cheng, Ryan Razani, Ehsan Taghavi, Enxu Li, and Bingbing Liu. Af2-s3net: Attentive feature fusion with adaptive feature selection for sparse semantic segmentation network. In *CVPR*, pages 12547–12556, 2021. 2

[14] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *CVPR*, pages 3075–3084, 2019. 2, 3, 5

[15] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *International Symposium on Visual Computing (ISVC)*, pages 207–222, 2020. 2

[16] Anja Delić, Matej Grcic, and Siniša Šegvić. Outlier detection by ensembling uncertainty with negative objectness. In *BMVC*, 2024. 1, 2

[17] Akshay Raj Dhamija, Manuel Günther, and Terrance Boult. Reducing network agnostophobia. *NeurIPS*, 31, 2018. 4

[18] Giancarlo Di Biase, Hermann Blum, Roland Siegwart, and Cesar Cadena. Pixel-wise anomaly detection in complex driving scenes. In *CVPR*, pages 16918–16927, 2021. 2

[19] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *CoRL*, pages 1–16, 2017. 5

[20] Tuo Feng, Wenguan Wang, Fan Ma, and Yi Yang. Lsk3dnet: Towards effective and efficient 3d perception with large sparse kernels. In *CVPR*, pages 14916–14927, 2024. 1, 2

[21] Daniel Fusaro, Simone Mosco, Emanuele Menegatti, and Alberto Pretto. Exploiting local features and range images for small data real-time point cloud semantic segmentation. In *IROS*, pages 4980–4987, 2024. 2

[22] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, pages 1050–1059. PMLR, 2016. 2

[23] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE TPAMI*, 43(12):4338–4364, 2020. 1

[24] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *ICLR*, 2017. 2, 6, 7, 4

[25] Yuenan Hou, Xinge Zhu, Yuexin Ma, Chen Change Loy, and Yikang Li. Point-to-voxel knowledge distillation for lidar semantic segmentation. In *CVPR*, pages 8479–8488, 2022. 2

[26] Yen-Chang Hsu, Yilin Shen, Hongxia Jin, and Zsolt Kira. Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data. In *CVPR*, pages 10951–10960, 2020. 7

[27] Lingdong Kong, Youquan Liu, Runnan Chen, Yuexin Ma, Xinge Zhu, Yikang Li, Yuenan Hou, Yu Qiao, and Ziwei Liu. Rethinking range view representation for lidar segmentation. In *ICCV*, pages 228–240, 2023. 1, 2

[28] Lingdong Kong, Xiang Xu, Jun Cen, Wenwei Zhang, Liang Pan, Kai Chen, and Ziwei Liu. Calib3d: Calibrating model preferences for reliable 3d scene understanding. In *WACV*, pages 1965–1978, 2025. 7, 8, 5

[29] Xin Lai, Yukang Chen, Fanbin Lu, Jianhui Liu, and Jiaya Jia. Spherical transformer for lidar-based 3d recognition. In *CVPR*, pages 17545–17555, 2023. 2

[30] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *NeurIPS*, 30, 2017. 1, 2, 6, 7, 4

[31] Kaican Li, Kai Chen, Haoyu Wang, Lanqing Hong, Chaoqiang Ye, Jianhua Han, Yukuai Chen, Wei Zhang, Chunjing Xu, Dit-Yan Yeung, et al. Coda: A real-world road corner

case dataset for object detection in autonomous driving. In *ECCV*, pages 406–423. Springer, 2022. 1, 2, 5

[32] Li Li, Hubert PH Shum, and Toby P Breckon. Rapid-seg: Range-aware pointwise distance distribution networks for 3d lidar segmentation. In *ECCV*, pages 222–241, 2025. 2, 7

[33] Wanmeng Li, Simone Mosco, Daniel Fusaro, and Alberto Pretto. Dpgla: Bridging the gap between synthetic and real data for unsupervised domain adaptation in 3d lidar semantic segmentation. In *IROS*, pages 11553–11560, 2025. 8

[34] Wenrui Liu, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Diversity-measurable anomaly detection. In *CVPR*, pages 12147–12156, 2023. 2

[35] Youquan Liu, Runnan Chen, Xin Li, Lingdong Kong, Yuchen Yang, Zhaoyang Xia, Yeqi Bai, Xinge Zhu, Yuexin Ma, Yikang Li, et al. Uniseg: A unified multi-modal lidar segmentation network and the openpcseg codebase. In *ICCV*, pages 21662–21673, 2023. 2

[36] Hanieh Shojaei Miandashti, Qianqian Zou, and Claus Brenner. Calibrated and efficient sampling-free confidence estimation for lidar scene semantic segmentation. *arXiv*, 2024. 8

[37] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate lidar semantic segmentation. In *IROS*, pages 4213–4220, 2019. 2, 5

[38] Noam Mor and Lior Wolf. Confidence prediction for lexicon-free ocr. In *WACV*, pages 218–225, 2018. 2

[39] Simone Mosco, Daniel Fusaro, Wanmeng Li, Emanuele Menegatti, and Alberto Pretto. Point-plane projections for accurate lidar semantic segmentation in small data scenarios. *arXiv*, 2025. 2

[40] Simone Mosco, Daniel Fusaro, Wanmeng Li, and Alberto Pretto. Revisiting retentive networks for fast range-view 3d lidar semantic segmentation. In *WACV*, pages 2499–2509, 2026. 2

[41] Nazir Nayal, Misra Yavuz, Joao F Henriques, and Fatma Güney. Rba: Segmenting unknown regions rejected by all. In *ICCV*, pages 711–722, 2023. 1, 2, 6, 7, 4

[42] Nazir Nayal, Youssef Shoeb, and Fatma Güney. A likelihood ratio-based approach to segmenting unknown objects. *IJCV*, pages 1–13, 2025. 1, 2

[43] Alexey Nekrasov, Malcolm Burdorf, Stewart Worrall, Bastian Leibe, and Julie Stephany Berrio Perez. Spotting the unexpected (stu): A 3d lidar dataset for anomaly segmentation in autonomous driving. In *CVPR*, pages 11875–11885, 2025. 1, 2, 5, 6, 7, 3, 4

[44] Michael Oren and Shree K Nayar. Generalization of the lambertian model and implications for machine vision. *IJCV*, 14 (3):227–251, 1995. 5, 1

[45] Yancheng Pan, Biao Gao, Jilin Mei, Sibo Geng, Chengkun Li, and Huijing Zhao. Semanticposs: A point cloud dataset with large quantity of dynamic instances. In *IV*, pages 687–693. IEEE, 2020. 1, 5, 6

[46] Gilles Puy, Alexandre Boulch, and Renaud Marlet. Using a waffle iron for automotive point cloud semantic segmentation. In *ICCV*, pages 3379–3389, 2023. 2

[47] C. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *CVPR*, pages 77–85, 2017. 2

[48] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *NeurIPS*, 30, 2017. 2

[49] Shyam Nandan Rai, Fabio Cermelli, Dario Fontanel, Carlo Masone, and Barbara Caputo. Unmasking anomalies in road-scene segmentation. In *ICCV*, pages 4037–4046, 2023. 2

[50] Yongming Rao, Wenliang Zhao, Guangyi Chen, Yansong Tang, Zheng Zhu, Guan Huang, Jie Zhou, and Jiwen Lu. Denseclip: Language-guided dense prediction with context-aware prompting. In *CVPR*, pages 18082–18091, 2022. 2

[51] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, pages 14318–14328, 2022. 2

[52] Cristiano Saltori, Fabio Galasso, Giuseppe Fiameni, Nicu Sebe, Elisa Ricci, and Fabio Poiesi. Cosmix: Compositional semantic mix for domain adaptation in 3d lidar segmentation. In *ECCV*, pages 586–602. Springer, 2022. 8

[53] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. 4

[54] Aasheesh Singh, Aditya Kamireddypalli, Vineet Gandhi, and K Madhava Krishna. Lidar guided small obstacle segmentation. In *IROS*, pages 8513–8520, 2020. 2, 5

[55] Matteo Sodano, Federico Magistri, Lucas Nunes, Jens Behley, and Cyrill Stachniss. Open-World Semantic Segmentation Including Class Similarity. In *CVPR*, 2024. 1, 2, 3, 4

[56] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *JMLR*, 15 (56):1929–1958, 2014. 6, 4

[57] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. Searching efficient 3d architectures with sparse point-voxel convolution. In *ECCV*, pages 685–702, 2020. 2

[58] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *ICCV*, pages 6411–6420, 2019. 2

[59] Chin-Chia Tsai, Tsung-Hsuan Wu, and Shang-Hong Lai. Multi-scale patch-based representation learning for image anomaly detection and segmentation. In *WACV*, pages 3065–3073, 2022. 2

[60] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM TOG*, 38(5): 1–12, 2019. 2

[61] Kelvin Wong, Shenlong Wang, Mengye Ren, Ming Liang, and Raquel Urtasun. Identifying unknown instances for autonomous driving. In *CoRL*, pages 384–393, 2020. 1, 2, 5

[62] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *ICRA*, pages 1887–1893, 2018. 2

[63] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and

unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *ICRA*, pages 4376–4382, 2019. 2

[64] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *CVPR*, pages 9621–9630, 2019. 2

[65] Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, and Hengshuang Zhao. Point transformer v2: Grouped vector attention and partition-based pooling. *NeurIPS*, 35:33330–33342, 2022. 2

[66] Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler faster stronger. In *CVPR*, pages 4840–4851, 2024. 1, 2

[67] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *CVPR*, pages 1912–1920, 2015. 2, 5, 1

[68] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, and Alan L Yuille. Synthesize then compare: Detecting failures and anomalies for semantic segmentation. In *ECCV*, pages 145–161, 2020. 2

[69] Chenfeng Xu, Bichen Wu, Zining Wang, Wei Zhan, Peter Vajda, Kurt Keutzer, and Masayoshi Tomizuka. Squeezesegv3: Spatially-adaptive convolution for efficient point-cloud segmentation. In *ECCV*, pages 1–19, 2020. 2

[70] Jianyun Xu, Ruixiang Zhang, Jian Dou, Yushi Zhu, Jie Sun, and Shiliang Pu. Rpvnet: A deep and efficient range-point-voxel fusion network for lidar point cloud segmentation. In *ICCV*, pages 16024–16033, 2021. 2, 7

[71] Xu Yan, Jiantao Gao, Chaoda Zheng, Chao Zheng, Ruimao Zhang, Shuguang Cui, and Zhen Li. 2dpass: 2d priors assisted semantic segmentation on lidar point clouds. In *ECCV*, pages 677–695, 2022. 2

[72] Kadir Yilmaz, Jonas Schult, Alexey Nekrasov, and Bastian Leibe. Mask4former: Mask transformer for 4d panoptic segmentation. In *ICRA*, pages 9418–9425, 2024. 6, 7, 4, 5

[73] Karim Abou Zeid, Kadir Yilmaz, Daan de Geus, Alexander Hermans, David Adrian, Timm Linder, and Bastian Leibe. Dino in the room: Leveraging 2d foundation models for 3d segmentation. *arXiv*, 2025. 2

[74] Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zerong Xi, Boqing Gong, and Hassan Foroosh. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *CVPR*, pages 9601–9610, 2020. 2

[75] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *ICCV*, pages 16259–16268, 2021. 2

[76] Ying Zhao. Omnial: A unified cnn framework for unsupervised anomaly localization. In *CVPR*, pages 3924–3933, 2023. 2

[77] Yiwu Zhong, Jianwei Yang, Pengchuan Zhang, Chunyuan Li, Noel Codella, Liunian Harold Li, Luowei Zhou, Xiyang Dai, Lu Yuan, Yin Li, et al. Regionclip: Region-based language-image pretraining. In *CVPR*, pages 16793–16803, 2022. 2

[78] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *CVPR*, pages 9939–9948, 2021. 2