LiON: Learning Point-wise Abstaining Penalty for LiDAR Outlier DetectioN Using Diverse Synthetic Data

Shaocong Xu[1,2*], Pengfei Li[1*], Qianpu Sun[1], Xinyu Liu[1], Yang Li[1], Shihui Guo[2†], Zhen Wang[3], Bo Jiang[3], Rui Wang[3], Kehua Sheng[3], Bo Zhang[3], Li Jiang[4], Hao Zhao[1†], Yilun Chen[1]

[1]Tsinghua University
[2]Xiamen University
[3]Didi Chuxing
[4]Chinese University of Hong Kong, ShenZhen

Abstract

LiDAR-based semantic scene understanding is an important module in the modern autonomous driving perception stack. However, identifying outlier points in a LiDAR point cloud is challenging as LiDAR point clouds lack semantically-rich information. While former SOTA methods adopt heuristic architectures, we revisit this problem from the perspective of Selective Classification, which introduces a selective function into the standard closed-set classification setup. Our solution is built upon the basic idea of abstaining from choosing any inlier categories but learns a point-wise abstaining penalty with a margin-based loss. Apart from learning paradigms, synthesizing outliers to approximate unlimited real outliers is also critical, so we propose a strong synthesis pipeline that generates outliers originated from various factors: object categories, sampling patterns and sizes. We demonstrate that learning different abstaining penalties, apart from point-wise penalty, for different types of (synthesized) outliers can further improve the performance. We benchmark our method on SemanticKITTI and nuScenes and achieve SOTA results.

Code — https://github.com/Daniellli/LiON/

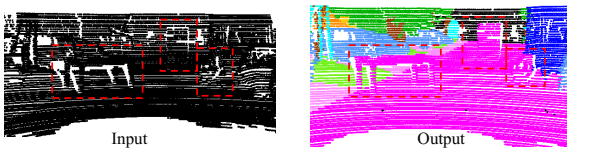
Introduction

LiDAR outlier detection (Li and Dong 2023) complements LiDAR semantic segmentation (Wang et al. 2024), aiming to enhance the model's ability to recognize outliers without compromising its inlier segmentation performance.

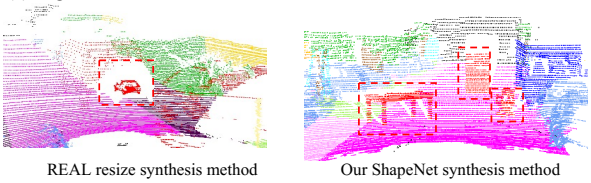
This task is important and practical. Traditional segmentation methods (Wang et al. 2024, Li, Shum, and Breckon 2024) assume that the samples in the training and test sets belong to the same set of categories. Thus, these models are trained on inlier categories and tend to classify all inputs into one of the inlier categories. However, this assumption fails in real-world scenarios where outliers are present. For example, as shown in Fig. 1-(a), the model may randomly classify furniture that has not been seen in training, leading to disastrous consequences in the downstream planning stage.

While 2D outlier detection has made significant strides, including optimizing inlier prediction (Tian et al. 2022, Liu et al. 2023, Miao et al. 2024), addressing outlier class imbalance (Choi, Jeong, and Choi 2023), evolving from pixel-wise to mask-based outlier detection methods (Zhang et al. 2024, Rai et al. 2023, Nayal et al. 2023, Zhang et al. 2024), developing promptable outlier detectors (Zhao et al. 2024, Li et al. 2024, Zhou et al. 2022), and utilizing model ensembling (Liu et al. 2024), the field of LiDAR outlier detection is still in its early stages (Li and Dong 2023). Seminal work REAL (Cen et al. 2022) proposes randomly choosing and resizing objects existing in the scene to synthesize outliers to approximate unlimited real outliers. For example, as shown in the left panel of Fig. 1-(b), the car is chosen and shrunk. This is viable but fail to represent the long-tail distribution of real outliers, in two regards. Firstly, objects from existing road scene understanding datasets are limited in category. Secondly, naive resizing violates the sampling pattern of real LiDAR sensors: points on enlarged objects get sparser while those on shrunk objects get denser. This leads to a shortcut problem: the model may find a trivial solution to distinguish outliers from inliers solely using point sparsity.

Besides, REAL empirically finds 2D outlier detection methods (Hendrycks and Gimpel 2018, Hendrycks et al. 2019, Gal and Ghahramani 2016) perform poorly in the 3D



(a) Segmentation model's random classification of furniture in the road



(b) Synthesis Method Comparison

Figure 1: (a). The semantic segmentation model fails to identify furniture because the training set does not include such objects; (b). Comparison of our ShapeNet outlier synthesis method and the former resize outlier synthesis method.

*These authors contributed equally.
†Corresponding author.

domain, as a large number of outliers are predicted as inliers with high confidence scores . To alleviate this phenomenon, in addition to the Cross Entropy (CE) loss, REAL designs a Calibration Cross Entropy(CCE) loss to calibrate the outlier probabilities in inlier prediction. However, we empirically find that CCE alleviates this at the cost of accuracy in predicting inliers. As shown in the right panel of Fig. 4, with CCE, the outlier probabilities of inlier points, which account for the vast majority of the total points, become very high (higher than 0.1), which is undesirable.

To address these issues, in this work, we propose a novel method LiON, aiming to mitigate the lack of semantically-rich information in LiDAR point clouds for outlier detection. We contribute from two perspectives: learning and data.

Learning. We reformulate the LiDAR outlier detection problem by applying Selective Classification (SC) principles (Feng et al. 2023, Chow 1970) and introduce a point-wise abstaining penalty learning paradigm to address the problem of unclear distinction in point clouds. While inspired by SC, our method differs significantly from SC in our point-wise design. Specifically, a diverse calibration factor is learned in a point-wise manner to more effectively capture subtle differences within a point cloud and calibrate the relationship between inlier and outlier classifiers. As a result, we mitigate the unclear distinction in point clouds caused by their lack of semantically-rich information by learning a diverse factor in a point-wise manner.

Data. Inspired by outlier exposure (Hendrycks, Mazeika, and Dietterich 2019), we propose introducing objects from an external dataset, ShapeNet (Chang et al. 2015), into existing scenes to synthesize outliers. ShapeNet, with its wide spectrum of categories and diverse geometries, can compensate for the long-tail distribution of real outliers. To ensure the realism of synthesized outliers, we take the LiDAR sampling pattern into consideration when merging randomly selected ShapeNet objects into road scenes. As illustrated in the right of Fig. 1-(b), our synthesized outliers are precisely aligned with objects in the scene with respect to point sparsity and occlusion. In this way, we mitigate the lack of semantically-rich information in LiDAR point clouds by utilizing realistic, ShapeNet outliers with diverse geometries.

Finally, the risk-coverage evaluation metrics associated with SC are also adapted for this task to serve as supplementary metrics of the holistic metrics AUPR/AUROC/mIoU$_{\text{old}}$, allowing us to gain a deeper understanding of the performance gains. These metrics are also a key to narrow the gap between academic and industrial communities. This is because these metrics allow us to identify the rejection threshold that incurs the least cost but yields the highest gain (coverage), which is very important for industrial applications.

Our contributions can be summarized as follows:

- We propose a point-wise abstaining penalty learning paradigm using the principle of SC to calibrate the relationship between inlier and outlier classifiers in a point-wise manner. Additionally, the risk-coverage evaluation metrics associated with SC are adapted for this problem, serving as supplementary metrics to the holistic metrics AUPR/AUROC/mIoU$_{\text{old}}$.
- We utilize ShapeNet objects to synthesize realistic and diverse outliers to approximate unlimited real outliers.
- Our method has achieved new SOTA performance for LiDAR outlier detection not only in previously established outlier detection metrics, but also in the risk-coverage curve metric, on SemanticKITTI and NuScenes.

Related Works

Outlier Detection in Autonomous Driving. Outlier detection is vital for ensuring the safety of ego-cars in open-world environments by identifying outlier objects. Extensive research has been conducted in 2D perception, specifically with semantic segmentation models. Unsupervised methods (Jung et al. 2021, Bevandić et al. 2021) involve post-processing predicted logits from frozen segmentation models to detect outliers. Supervised methods (Tian et al. 2022, Grcić, Bevandić, and Šegvić 2022, Chan, Rottmann, and Gottschalk 2021) utilize auxiliary datasets like COCO (Lin et al. 2014) to synthesize outlier objects in training images (e.g., Cityscapes (Cordts et al. 2016)) and retrain the segmentation model using outlier exposure (Hendrycks, Mazeika, and Dietterich 2019, Zhou et al. 2024). While significant advancements have been made in 2D outlier detection, the exploration in the context of LiDAR point clouds remains limited. Cen et al. (2022) use resize synthesis pipeline and calibration loss to achieve the discrimination of outlier points. However, their focus primarily revolves around the open-world segmentation setting without extensively analyzing calibration effectiveness. Li and Dong (2023) propose an adversarial prototype framework that improves performance but involves complex network design and computationally expensive training. Considering these limitations, we present our method and substantiate its effectiveness through extensive experiments.

Selective Classification. SC can be broadly classified into two groups: 1) the first group focuses on addressing SC through the use of additional heads/logits (Geifman and El-Yaniv 2017, 2019, Liu et al. 2019, Feng et al. 2023, Gal and Ghahramani 2016, Chow 1970); 2) the second group tackles SC through cost-sensitive classification techniques (Charoenphakdee et al. 2021, Mozannar and Sontag 2020, Su et al. 2015, Movshovitz-Attias, Kanade, and Sheikh 2016, Zhang et al. 2017b, Handa et al. 2016, McCormac et al. 2017, Zhang et al. 2017a, Song et al. 2017, Gao et al. 2021, 2024a,b, 2023, Xu et al. 2024, Ding et al. 2024). Motivated by the extra head/logits design, resembling outlier detection, we design a new outlier detector from the perspective of SC and utilize SC's evaluation metrics as alternative performance measures for our outlier detector.

Preliminaries: Selective Classification

We first formalize the definition of SC and put different methods under a unified lens. Experimentally, the risk-coverage analysis that comes along with this framework, allows us to reveal in-depth differences between methods.

Definition. In SC, our goal is to learn predictive models that know what they do not know or when they should ab-

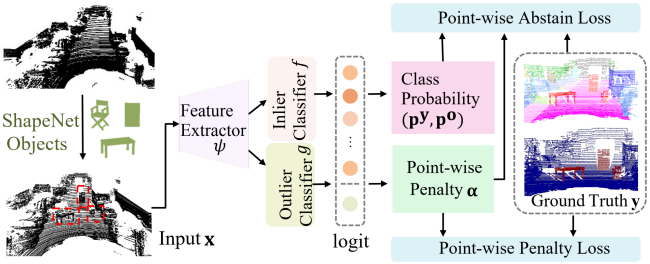


Figure 2: Method pipeline: a point cloud containing outliers synthesized by ShapeNet is processed by a feature extractor to obtain features. These features are then used by inlier and outlier classifiers to predict class logits.

stain[1] from making decisions. Here we consider a generic SC definition, which is agnostic of network and application. A supervised classification task is formulated as follows. Let $\mathcal{X}$ be any feature space and $\mathcal{Y}$ a label space. In LiDAR outlier detection, $\mathcal{X}$ could be point clouds, and $\mathcal{Y}$ could be class labels[2] of each point cloud. Let $P(\mathcal{X}, \mathcal{Y})$ be a distribution over $\mathcal{X} \times \mathcal{Y}$. A model $f : \mathcal{X} \to \mathcal{Y}$ is called a prediction function and its true risk used to evaluate the performance of f w.r.t. P is $R(f) := E_{P(\mathcal{X},\mathcal{Y})}[\ell(f(\mathbf{x}), \mathbf{y})]$, where $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+$ is a given loss function, for example, the Cross Entropy (CE) loss. Given a labeled set $S_m = \{(\mathbf{x_i}, \mathbf{y_i})\}_{i=1}^m$ sampled i.i.d. from $P(\mathcal{X}, \mathcal{Y})$, the empirical risk of the classifier f is $\hat{r}(f|S_m) := \frac{1}{m}\sum_{i=1}^m \ell(f(\mathbf{x_i}), \mathbf{y_i})$.

Apart from risk, another important concept in the SC formulation is coverage. A selective model (El-Yaniv et al. 2010) is a pair (f, g), where f is a prediction function, and $g : \mathcal{X} \to \{0, 1\}$ is a selective function, which serves as a binary qualifier for f as follows:

$$(f, g)(\mathbf{x}) := \begin{cases} f(\mathbf{x}), & \text{if } g(\mathbf{x}) = 1 \\ \text{ABSTAIN}, & \text{if } g(\mathbf{x}) = 0 \end{cases} \tag{1}$$

Thus, the selective model abstains from prediction at $\mathbf{x}$ iff $g(\mathbf{x}) = 0$. A soft selection function can also be considered, where $g : \mathcal{X} \to [0, 1]$, and decisions can be taken probabilistically or deterministically (e.g., using a threshold). The introduction of a selective function g allows us to define coverage. Specifically, coverage is defined to be the ratio of the non-abstained subset within set P to the entirety of P, which can be formulated as:

$$\phi(g) := E_P[g(\mathbf{x})] \tag{2}$$

Accordingly, the standard risk for a classifier f can be augmented into the selective risk of (f, g) as

$$R(f, g) := \frac{E_P[\ell(f(\mathbf{x}), \mathbf{y})g(\mathbf{x})]}{\phi(g)} \tag{3}$$

Clearly, the risk of a selective model can be traded-off for coverage. The performance profile of such a model can be specified by its risk–coverage curve, defined to be the risk as a function of coverage.

[1]The abstaining penalty will be defined later.

[2]These class labels include an outlier label.

Finally, we clarify that the continuous risk and coverage defined above are calculated using a fixed set in practice. For any given labeled set S_m, the empirical selective risk is

$$\hat{r}(f, g|S_m) := \frac{\frac{1}{m}\sum_{i=1}^m \ell(f(\mathbf{x_i}), \mathbf{y_i})g(\mathbf{x_i})}{\phi(g|S_m)} \tag{4}$$

and the empirical coverage is

$$\hat{\phi}(g|S_m) := \frac{1}{m}\sum_{i}^{m} g(\mathbf{x_i}) \tag{5}$$

Method

LiDAR semantic segmentation is the task of assigning a class label from a predefined class label set to each point in a given point cloud. LiDAR outlier detection, on the other hand, is an extension to semantic segmentation, which aims to identify points that do not belong to the predefined inlier label set. While the seminal work proposes a viable solution REAL, we revisit this issue through the unified lens of SC and propose a solution that effectively mitigates the problem of unclear distinctions in point clouds. Apart from the learning paradigm, we design a novel outlier synthesis pipeline that leverages the richness of the ShapeNet and adheres to the realistic LiDAR distribution to compensate for the lack of semantically-rich information in point clouds.

Network Architecture Overview

As shown in Fig. 2, the input point cloud $\mathbf{x} \in \mathbb{R}^{n\times 3}$, sampled from S_m, is denoted on the left with ShapeNet objects integrated into it. Then, $\mathbf{x}$ is fed into the feature extractor ψ followed by an inlier classifier f to predict the inlier logit $\hat{\mathbf{y}} \in \mathbb{R}^{n\times c}$, where c is the number of inlier classes. An outlier classifier g is used to predict the outlier logit $\hat{\mathbf{o}} \in \mathbb{R}^{n\times 1}$. As such, (f, g) instantiates a selective model mentioned above. These operations can be expressed as follows:

$$\begin{gathered} \hat{\mathbf{y}} := f(\psi(\mathbf{x})) \quad \hat{\mathbf{o}} := g(\psi(\mathbf{x})) \\ \tilde{\mathbf{y}} := [\hat{\mathbf{y}}, \hat{\mathbf{o}}] := \left\{ \tilde{\mathbf{y}}_\mathbf{i} := [\hat{y}_i, \hat{o}_i] \middle| i = 1, \ldots, n \right\} \\ \mathbf{p} := \left\{ p_{i,j} = \frac{e^{\tilde{y}_{i,j}}}{\sum_{k=1}^{c+1} e^{\tilde{y}_{i,k}}} \middle| i = 1, \ldots, n; j = 1, \ldots, c+1 \right\} \\ \mathbf{p^y}, \mathbf{p^o} := \mathbf{p} \end{gathered} \tag{6}$$

Here, the operation $[\cdot]$ denotes concatenation. Prediction probability $\mathbf{p} \in [0, 1]^{n\times(c+1)}$ consists of inlier probability $\mathbf{p^y} \in [0, 1]^{n\times c}$ and outlier probability $\mathbf{p^o} \in [0, 1]^{n\times 1}$.

Revisiting the REAL Formulation

Cen et al. (2022) introduces the first LiDAR outlier detector, REAL. Their added dummy classifiers can also be thought of as g, under the SC framework, but there is a key difference. They observe that numerous real outliers are wrongly classified as inlier classes with high probabilities. To address this issue, they propose a Calibration Cross Entropy (CCE) loss function to drive the outlier logit of the inlier sample to

CE	CCE	AUPR	AUROC	mIoU_{old}
✓	✓	20.00	84.90	57.80
✓		**26.68**	**87.60**	**58.28**

Table 1: Additional ablation of baseline REAL.

the second largest. We formalize this loss as follows:

$$\ell := \frac{1}{m}\sum_{S_m}\frac{1}{n}\sum_{i=1}^{n}\Bigg\{\underbrace{-\log p_{i,y_i}}_{\text{CE}} \underbrace{-\lambda\mathbb{I}(y_i \neq c+1)\log\frac{e^{\tilde{y}_{i,c+1}}}{\sum_{k=1\&k\neq y_i}^{c+1} e^{\tilde{y}_{i,k}}}}_{\text{CCE}}\Bigg\} \tag{7}$$

Here, $y_i \in \{1,\ldots,c,c+1\}$ signifies the ground truth corresponding to $x_i \in \mathbf{x}$, where $\{1,\ldots,c\}$ represents inlier class labels while $\{c+1\}$ indicates the outlier class label. $\mathbf{y}$ and $\mathbf{x}$ are sampled i.i.d. from S_m. $\mathbb{I}(\cdot)$ is the indicator function. λ is a hyperparameter. A notable fact is that REAL does not provide an ablation study for this CCE loss and as shown in Tab. 1, removing this CCE loss can indeed improve standard outlier detection metrics. But why this happens cannot be understood through metrics like AUPR or AUROC, highlighting the need to revisit REAL under the lens of SC.

We use Fig. 4 as an intuitive case to analyze the negative impact of the CCE loss. It illustrates sample numbers within different $\mathbf{p^o}$ intervals for both inliers and outliers. By removing the CCE loss, sample number in the extremely low interval $[0, 0.1]$ grows significantly for both inliers and outliers. This is desirable for inliers but not for outliers. As shown by the red increase number, 9.8 million samples change to a desirable state while 1.7 million samples change to a undesirable state, and this large difference explains why the collective metrics in Tab. 1 become better. The reason why this in-depth statistics (Fig. 4) can reveal the negative impact of CCE is that different $\mathbf{p^o}$ values are investigated separately. And, the SC framework provides the principled tool risk-coverage curve to conduct this kind of analysis, because different coverage is achieved through selecting different thresholds on $\mathbf{p^o}$. More principled analyses that reveal reasons behind phenomena like Tab. 1 can be found in the experiments section.

Point-wise Abstaining Penalty Learning

Point-wise upgrade. The reason why the disadvantages of CCE exceed its benefits is its sub-optimal calibration for the inlier probabilities $\mathbf{p^y}$ and the outlier probabilities $\mathbf{p^o}$. Inspired by 'learning to abstain' (Liu et al. 2019), where the calibration between the reject and non-reject options is adeptly handled, we propose our new learning paradigm. Specifically, Liu et al. (2019) employs a fixed calibrating factor to achieve this. While this may suffice for image-level SC, it is inadequate for point-wise outlier detection. This is because, in LiDAR outlier detection, we need a different calibrating factor for each point to capture the subtle differences between them, such as between remote (sparse) and near (dense) points, as well as between inlier and outlier points. Therefore, we upgrade this fixed calibrating factor to a point-wise one, defined as abstaining penalty α, and introduce a point-wise penalty loss to supervise the network to learn the subtle differences between various points, which can be expressed as follows:

$$\alpha := \left\{\alpha_i = -\log(\sum_{j=1}^{c} e^{\hat{y}_{i,j}}) | i = 1,\ldots,n\right\}$$
$$\ell^{\text{penalty}} := \frac{1}{m}\sum_{S_m}\frac{1}{n}\sum_{i=1}^{n}\Big\{\mathbb{I}(y_i \neq c+1)\max(\alpha_i - m_{\text{in}}, 0) + \mathbb{I}(y_i = c+1)\max(m_{\text{out}} - \alpha_i, 0)\Big\} \tag{8}$$

As shown in Fig. 5-(c), the hyperparameters m_{in} and m_{out} ensure that the inliers are associated with penalties lower than m_{in}, while the outliers exhibit penalties higher than m_{out}. Note that in our experiments, the penalties are negative and we set the value of m_{in} to -12, and m_{out} to -6.

Moreover, our new 'learning to abstain' formulation for this task with this point-wise penalty is defined as follows:

$$\ell^{\text{abstain}} := \frac{1}{m}\sum_{S_m}\frac{1}{n}\sum_{i=1}^{n}\Bigg\{\underbrace{-\mathbb{I}(y_i \neq c+1)\log\left\{p^y_{i,y_i} + \underbrace{\frac{p^o_i}{(-\alpha_i)^2}}_{\text{abstaining term}}\right\}}_{\text{for inlier samples}} \underbrace{-\mathbb{I}(y_i = c+1)\sum_{j=1}^{c}\log\left\{p^y_{i,j} + \underbrace{\frac{p^o_i}{(-\alpha_i)^2}}_{\text{abstaining term}}\right\}}_{\text{for outlier samples}}\Bigg\} \tag{9}$$

Intuition. Minimizing the point-wise penalty loss Eq. (8) results in assigning lower α_i to inliers, thereby leading to higher values of $(-\alpha_i)^2$. This effectively suppresses the contribution of p^o_i and allows p^y_{i,y_i} to play a dominant role in the point-wise abstain loss Eq. (9). Likewise, higher α_i are allocated to outliers, resulting in lower values of $(-\alpha_i)^2$. This allows p^o_i to play a dominant role and, consequently, suppresses the contribution of $\{p^y_{i,j} | j = 1,\ldots,c\}$ in the point-wise abstain loss Eq. (9). Since this learning paradigm is defined in a point-wise manner, it has the potential to capture the subtle difference between inliers and outliers despite LiDAR point clouds lack semantically-rich information. The total loss can be expressed as follows:

$$\ell^{\text{total}} := \lambda^{\text{abstain}}\ell^{\text{abstain}} + \lambda^{\text{penalty}}\ell^{\text{penalty}} \tag{10}$$

Outlier Synthesis Pipeline

REAL synthesizes outliers by resizing the objects presenting in existing scene, as depicted in the left panel of Fig. 1-(b). However, we observe that this synthesis pipeline fails to

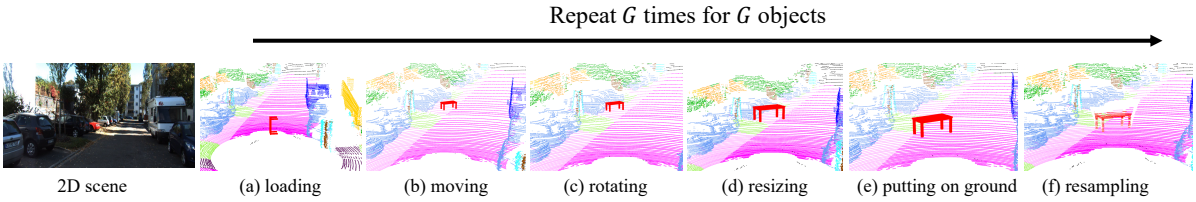


Figure 3: Our outlier synthesis pipeline.

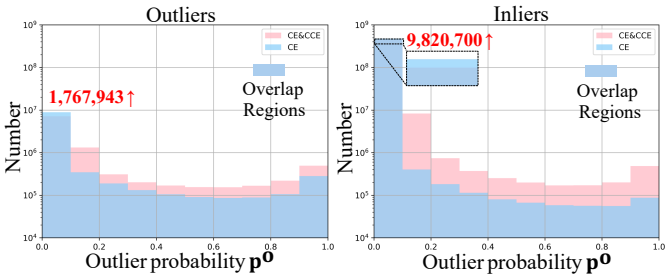


Figure 4: Additional ablation of REAL from the perspective of outlier probability $\mathbf{p^o}$.

represent the real outlier in two aspects: 1) the limited variety of objects in the existing scene makes it challenging to compensate for the lack of semantically-rich information through resizing; 2) learning from these synthesized outliers may lead to a model that classifies real outliers solely based on point sparsity. As such, we resort to an additional dataset, ShapeNet, which consists of 220,000 models classified into 3,135 categories. As shown in Fig. 3, we repeat the synthesis pipeline for G times to insert G outlier objects. For each object, there are six steps as follows:

(a) To synthesize diverse outliers, we first randomly decide the number of outlier objects G according to a Binomial distribution[3]. The specific distribution used is Binomial$(20, 0.3)$. We then **load** G objects from ShapeNet into the given scene $\mathbf{x}$, where the probability of not adding any object into $\mathbf{x}$ is also considered.

(b) Then, for each object $\mathbf{s} \in \mathbb{R}^{l\times 3}$, we **move** $\mathbf{s} \in \mathbb{R}^{l\times 3}$ away from scene center x^c, alone the x-axis by $d^x \sim$ Uniform$(r^{\text{min}}, 0.8 * r^{\text{max}})$[4] that r^{min} is the distance of the closest point from x^c and r^{max} is the furthest point from x^c.

(c) Next, we **rotate** $\mathbf{s}$ around x^c on the xy-plane (around the gravity direction) for $d^{\text{lon}} \sim$ Uniform$(0, 360)$ degrees and denote the resulting object as $\mathbf{s} = (\mathbf{u}, \mathbf{v}, \mathbf{w})$. There is a probability that $\mathbf{s}$ does not overlap with $\mathbf{x}$, after moving and rotating. Therefore, if $\mathbf{s}$ is positioned outside of $\mathbf{x}$, the subsequent steps are not carried out and we move on the next object. Specifically, we stop synthesis process if:

$$\min\{|\bar{u} - i| + |\bar{v} - j|\} > \Delta, (i, j, k) \in \mathbf{x} \quad (11)$$

[3]The Binomial(a, b) generates a random value from a binomial distribution with 'a' trials and 'b' probability of success per trial.

[4]The Uniform(a, b) generates a random value from a uniform distribution with lower bound 'a' and upper bound 'b'.

Here, $\bar{u}$ and $\bar{v}$ represent the mean u and mean v of s, respectively. We set Δ to 1 in our experiments.

(d) Then, since the objects from ShapeNet tend to be smaller in size compared to those in the existing scene, we proceed to **resize** $\mathbf{s}$ by a factor of $k \sim$ Uniform$(1, 7)$.

(e) Following this, we **put** $\mathbf{s}$ **on ground** by setting its last axis to $\tilde{\mathbf{w}} = \mathbf{w} - \Delta_w$, where Δ_w represents the distance between the bottom of $\mathbf{s}$ and the point on $\mathbf{x}$ that are the closest to $\mathbf{s}$ along the gravity direction and falls into the x-y plane projection of s. The resulting object is $\mathbf{s} = (\mathbf{u}, \mathbf{v}, \tilde{\mathbf{w}})$.

(f) Finally, to consider the realistic LiDAR's sampling pattern, we merge $\mathbf{s}$ into $\mathbf{x}$ by adjusting the radii of $\mathbf{x}$. Specifically, we represent $\mathbf{s}$ and $\mathbf{x}$ using spherical coordinates:

$$\begin{aligned} \mathbf{s} &:= \{s_j = (\text{lon}_j, \text{lat}_j, \text{r}_j) | j = 1, \ldots, l\} \\ \mathbf{x} &:= \{x_k = (\text{lon}_k, \text{lat}_k, \text{r}_k) | k = 1, \ldots, n\} \end{aligned} \quad (12)$$

Here, the lon represents longitude, lat represents latitude, and r represents radius. For each x_k, we replace r_k with r_j if s_j satisfy $|\text{lon}_k - \text{lon}_j| < \Delta_{\text{lon}}$ and $|\text{lat}_k - \text{lat}_j| < \Delta_{\text{lat}}$. During our experiment, we set Δ_{lon} to 0.02 and Δ_{lat} to 0.2. When multiple s_j satisfy the above criterion, we use their smallest r to replace r_k.

Dynamic Penalty

As shown in Fig. 1-(b), the outliers synthesized through resizing are further from the inlier data distribution in terms of point sparsity compared to those synthesized by our pipeline. Hence, to maximize the benefits of the point-wise learning paradigm, we introduce a dynamic penalty loss that handles points in a customized manner:

$$\begin{aligned} &\ell^{\text{dynamic penalty}} := \\ &\frac{1}{m}\sum_{S_m}\frac{1}{n}\sum_{i=1}^{n}\Big\{\mathbb{I}(y_i \neq c+1 \;\&\; y_i \neq c+2))\max(\alpha_i - \beta_{\text{in}} m_{\text{in}}, 0) \\ &\quad + \mathbb{I}(y_i = c+1)\max(\beta_{\text{rout}} m_{\text{rout}} - \alpha_i, 0) \\ &\quad + \mathbb{I}(y_i = c+2)\max(\beta_{\text{sout}} m_{\text{sout}} - \alpha_i, 0)\Big\} \end{aligned} \quad (13)$$

The $\{c+2\}$ denotes the outlier class label generated by ShapeNet. The weight parameter β associated with the threshold m is initialized as 1 and is learnable. In our experimental setting, we set the value of m_{sout} to -7, and m_{rout} to -6. Consequently, the total loss becomes:

$$\ell^{\text{total}} := \lambda^{\text{abstain}}\ell^{\text{abstain}} + \lambda^{\text{dynamic penalty}}\ell^{\text{dynamic penalty}} \quad (14)$$

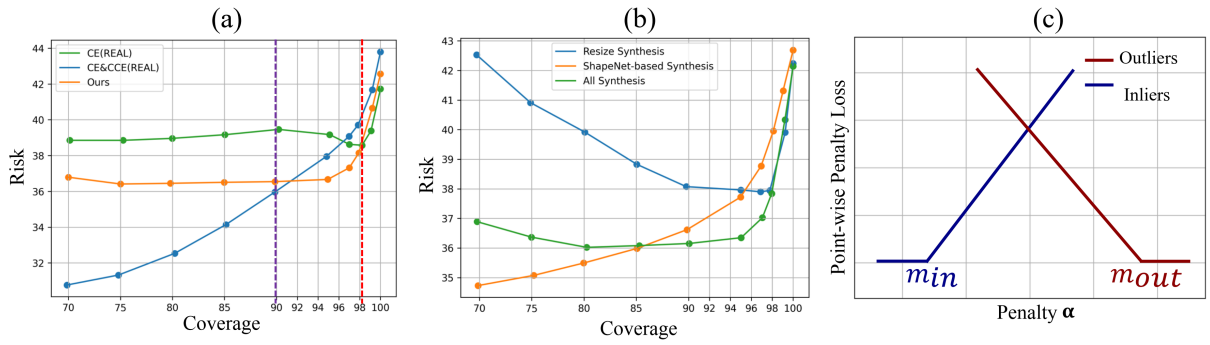


Figure 5: (a). Comparison with the SOTA using the Risk-Coverage curves; (b). Comparison with different outlier synthesis pipeline using the risk-coverage curve; (c). Relationship between point-wise penalty loss and penalty α.

Method	SemanticKITTI (Behley et al. 2019)			NuScenes (Caesar et al. 2020)		
	AUPR	AUROC	mIoU_{old}	AUPR	AUROC	mIoU_{old}
Closed-set C3D	-	-	58.00	-	-	58.70
C3D + MSP (Hendrycks and Gimpel 2018)	6.70	74.00	58.00	4.30	76.70	58.70
C3D + MaxLogit (Hendrycks et al. 2019)	7.60	70.50	58.00	8.30	79.40	58.70
C3D + MC-Dropout (Gal and Ghahramani 2016)	7.40	74.70	58.00	14.90	82.60	58.70
C3D + REAL (Cen et al. 2022)	20.08	84.90	57.80 (0.20 ↓)	21.20	84.50	56.80
C3D + APF (Li and Dong 2023)	36.10	85.60	57.30 (0.70 ↓)	-	-	-
C3D + LiON (ours)	44.68 (8.58 ↑)	92.69 (7.09 ↑)	57.56 (0.44 ↓)	31.58 (10.38 ↑)	95.24 (10.74 ↑)	59.11 (0.41 ↑)

Table 2: Comparisons with previous methods. C3D refers to the base segmentation model, Cylinder3D (Zhu et al. 2021).

Experiments

Dataset

SemanticKITTI is a driving-scene dataset designed for point cloud segmentation. The point clouds are collected using the Velodyne-HDLE64 LiDAR in Germany. The dataset consists of 22 sequences, with sequences 00 to 10 utilized as the training set, sequence 08 serves as the validation set, and sequences 11 to 21 used as the test set. After merging classes with different moving statuses and ignoring classes with a small number of points, 19 classes remain for training and evaluation. Consistent with prior work, we designate {other-vehicle} as outlier class. **NuScenes** consists of 1000 scenes, each lasting 20 seconds, captured using a 32-beam LiDAR sensor, which leads to its challenging nature (a sparser LiDAR point cloud makes the classification task more difficult). It contains 40,000 frames sampled at 20Hz and has official training and validation splits. After merging similar classes and removing rare/useless classes including 'ego-car', there are 16 remaining classes for training and evaluation. The classes designated as outliers include {barrier, constructive-vehicle, traffic-cone, trailer}.

Evaluation Metric

Traditional evaluation metrics: Consistent with previous work (Cen et al. 2022), we employ inlier mean intersection over union (mIoU_{old}) metric to evaluate the performance of inlier classification, while the AUPR and AUROC are utilized to assess the performance of outlier classification.

New evaluation metrics: The loss-based selective risk (Eq. (4)) is sub-optimal for serving as an evaluation metric for point-wise classification task because it is not sensitive to the class imbalance which is crucial in this task. Thus, we upgrade it into mIoU-based selective risk, as shown below:

$$\hat{r}(f, g|S_m) := \frac{100 - \text{mIoU}_{\text{old}}^{S_m|g}}{\phi(g|S_m)} \tag{15}$$

$\text{mIoU}_{\text{old}}^{S_m|g}$ represents the mIoU_{old} calculated for the sub-dataset S_m under the selective model condition g. With these new evaluation metrics, we can draw risk-coverage curves to obtain deeper understanding of model performance.

Comparisons with State-of-the-art Methods

We use Cylinder3D (Zhu et al. 2021) as the baseline segmentation model to ensure a fair comparison. The computational cost remains low with the addition of an outlier detection head, achieving 7 fps on a single NVIDIA 3090 GPU.

Quantitative comparison: As illustrated in Tab. 2, our method achieves a new SOTA in outlier class segmentation for SemanticKITTI and NuScenes, surpassing the previous SOTA by a significant margin. Specifically, our method achieves an AUPR of 44.68 and an AUROC of 92.69, which exceeds the previous SOTA scores by 8.58 and 7.09 in SemanticKITTI. Moreover, our method achieves an AUPR of 31.58 and an AUROC of 95.24 in NuScenes, which exceeds the previous SOTA scores by 10.38 and 10.74, respectively.

Qualitative comparison: Qualitative results, as illustrated in Fig. 6, demonstrate that our method not only locates the outliers more accurately but also does so with greater confidence compared to REAL. Furthermore, in NuScenes (Fig. 6-right), our method accurately identifies the 'ego-car' as outliers, although this category is not incorporated into the training and evaluation phases. These findings provide further evidence of the superiority of our method.

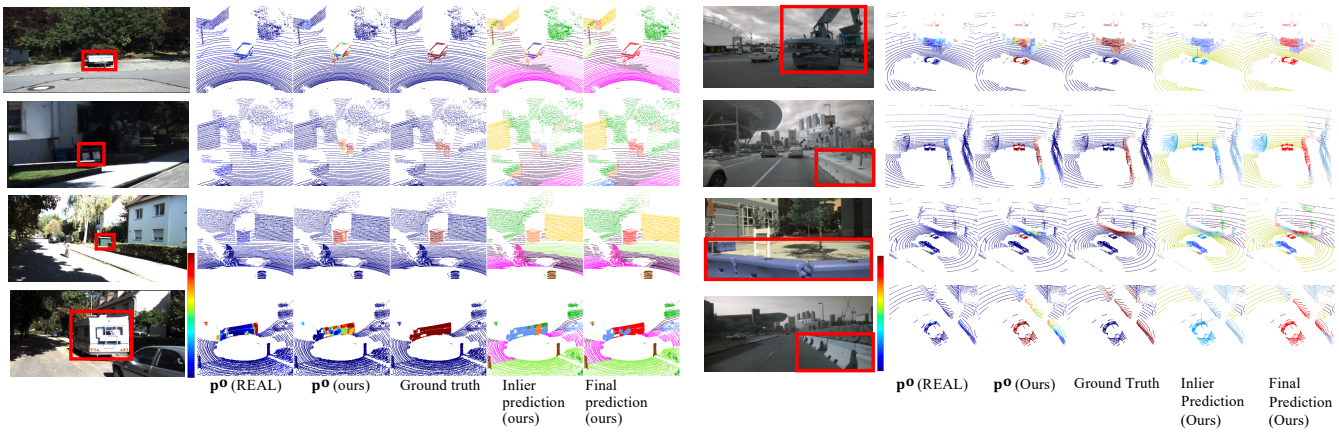


Figure 6: Qualitative comparison results for SemanticKITTI (left) and NuScenes (right); Inlier prediciton indicates semantic segmentation; The final prediction is obtained by integrating the inlier prediction with $\mathbf{p^o}$ using a threshold of 0.5.

Comparison between LiON and the previous SOTA (APF) from another perspective: In addition to better quantitative results, LiON is easier to implement and requires less computational cost compared to APF. LiON can be implemented by simply adding an extra classifier to an arbitrary LiDAR-based segmentation network and training the network in a single stage using our novel learning paradigm and two different outlier synthesis pipelines. In contrast, APF requires not only an arbitrary segmentation network but also several learnable prototypes, a prototypical constraint module, a generator, a discriminator, and an adversarial mapper, which significantly increase computational costs. Further, APF relies on a two-stage training process, complicating the reproducibility of its results.

Risk-Coverage curve comparison: As shown in Fig. 5-(a), the problem of the CE&CCE, analyzed above, is reflected as a high risk at high coverages, while the CE exhibits an unstable trend when coverage decreases. In contrast, our method achieves a highly competitive risk compared to them at high coverages. With decreasing coverage, our method shows a consistent decline to a plateau in risk.

Why our method achieve a higher risk compared to REAL when coverage is below 90%? (purple dotted line) This is because our method rejects most real outlier samples/points when coverage exceeds 90%. This is proved by that the threshold set to achieve 90% coverage is 0.0051, indicating that all samples with a predicted outlier probability higher than 0.0051 are rejected. This means that when coverage is around 90%, the outlier probabilities of the remaining samples are relatively small, and there are few real outliers left. Therefore, to further decrease the coverage, our method tends to reject more true inliers than true outliers. However, we believe that robust performance in high coverage is more critical than in low coverage due to the fact that the outlier objects are rare in the real world.

Ablation Study

As demonstrated in Tab. 3, the dynamic penalty setting achieves the best performance. Switching from the dynamic penalty mode to the penalty mode results in a decrease of 0.99 in AUPR and 0.18 in AUROC. Further excluding the ShapeNet synthesis pipeline results in a decrease in AUPR by 2.86 and an increase in AUROC by 0.35. The AUPR and AUROC drop significantly by 15.54 and 3.13, respectively, without the resize synthesis pipeline. This raises the question of **whether ShapeNet synthesis pipeline is trivial**.

Penalty	ShapeNet	Resize	Dynamic Penalty	AUPR	AUROC	mIoU_{old}
✓	✓			29.14 (15.54 ↓)	89.56 (3.13 ↓)	57.31 (0.25 ↓)
✓		✓		41.82 (2.86 ↓)	93.04 (0.35 ↑)	57.30 (0.26 ↓)
✓	✓	✓		43.69 (0.99 ↓)	92.51 (0.18 ↓)	57.47 (0.09 ↓)
✓	✓	✓	✓	**44.68**	**92.69**	**57.56**

Table 3: Ablation study in SemanticKITTI.

There are two types of real outliers: those that are distant from the inlier distribution (referred to as far real outliers) and those that lie closer to the inlier distribution (referred to as near real outliers). As shown in Fig. 5-(b), the resize synthesis pipeline shows an initial decrease in risk followed by an increase. This occurs because this pipeline can approximate the far real outlier distribution but struggles with approximating the near real outlier distribution.

On the other hand, our ShapeNet synthesis pipeline exhibits a consistent reduction in risk as the coverage decreases, as it can synthesize outliers that approximate both near and far real outlier distributions. The former result from considering the LiDAR sampling pattern, while the latter stem from considering the diversity of real outliers.

Conclusion

In this work, we first revisit previous methods using the unified lens of selective classification and propose a new formulation which effectively captures the subtle differences between inliers and outliers. Then, we design an outlier synthesis pipeline, compensating for the lack of semantically-rich information in point clouds. Experimental results demonstrate the superiority of LiON across both traditional metrics and newly introduced metrics.

References

Behley, J.; Garbade, M.; Milioto, A.; Quenzel, J.; Behnke, S.; Stachniss, C.; and Gall, J. 2019. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9297–9307.

Bevandić, P.; Krešo, I.; Oršić, M.; and Šegvić, S. 2021. Dense outlier detection and open-set recognition based on training with noisy negative images. *arXiv preprint arXiv:2101.09193*.

Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631.

Cen, J.; Yun, P.; Zhang, S.; Cai, J.; Luan, D.; Tang, M.; Liu, M.; and Yu Wang, M. 2022. Open-world semantic segmentation for lidar point clouds. In *ECCV*.

Chan, R.; Rottmann, M.; and Gottschalk, H. 2021. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proceedings of the ieee/cvf international conference on computer vision*, 5128–5137.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

Charoenphakdee, N.; Cui, Z.; Zhang, Y.; and Sugiyama, M. 2021. Classification with rejection based on cost-sensitive classification. In *International Conference on Machine Learning*, 1507–1517. PMLR.

Choi, H.; Jeong, H.; and Choi, J. Y. 2023. Balanced energy regularization loss for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15691–15700.

Chow, C. 1970. On optimum recognition error and reject tradeoff. *IEEE Transactions on information theory*, 16(1): 41–46.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3213–3223.

Ding, K.; Chen, B.; Wu, R.; Li, Y.; Zhang, Z.; Gao, H.-a.; Li, S.; Zhou, G.; Zhu, Y.; Dong, H.; et al. 2024. Preafford: Universal affordance-based pre-grasping for diverse objects and environments. *arXiv preprint arXiv:2404.03634*.

El-Yaniv, R.; et al. 2010. On the Foundations of Noise-free Selective Classification. *Journal of Machine Learning Research*, 11(5).

Feng, L.; Ahmed, M. O.; Hajimirsadeghi, H.; and Abdi, A. H. 2023. Towards Better Selective Classification. In *The Eleventh International Conference on Learning Representations*.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, 1050–1059. PMLR.

Gao, W.; Liu, Q.; Huang, Z.; Yin, Y.; Bi, H.; Wang, M.-C.; Ma, J.; Wang, S.; and Su, Y. 2021. RCD: Relation map driven cognitive diagnosis for intelligent education systems. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 501–510.

Gao, W.; Liu, Q.; Wang, H.; Yue, L.; Bi, H.; Gu, Y.; Yao, F.; Zhang, Z.; Li, X.; and He, Y. 2024a. Zero-1-to-3: Domain-Level Zero-Shot Cognitive Diagnosis via One Batch of Early-Bird Students towards Three Diagnostic Objectives. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8417–8426.

Gao, W.; Liu, Q.; Yue, L.; Yao, F.; Wang, H.; Gu, Y.; and Zhang, Z. 2024b. Collaborative Cognitive Diagnosis with Disentangled Representation Learning for Learner Modeling. *arXiv preprint arXiv:2411.02066*.

Gao, W.; Wang, H.; Liu, Q.; Wang, F.; Lin, X.; Yue, L.; Zhang, Z.; Lv, R.; and Wang, S. 2023. Leveraging transferable knowledge concept graph embedding for cold-start cognitive diagnosis. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, 983–992.

Geifman, Y.; and El-Yaniv, R. 2017. Selective classification for deep neural networks. *Advances in neural information processing systems*, 30.

Geifman, Y.; and El-Yaniv, R. 2019. Selectivenet: A deep neural network with an integrated reject option. In *International conference on machine learning*, 2151–2159. PMLR.

Grcić, M.; Bevandić, P.; and Šegvić, S. 2022. Densehybrid: Hybrid anomaly detection for dense open-set recognition. In *European Conference on Computer Vision*, 500–517. Springer.

Handa, A.; Patraucean, V.; Badrinarayanan, V.; Stent, S.; and Cipolla, R. 2016. Understanding real world indoor scenes with synthetic data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4077–4085.

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

Hendrycks, D.; and Gimpel, K. 2018. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In *ICLR 2017*. arXiv.

Hendrycks, D.; Mazeika, M.; and Dietterich, T. 2019. Deep Anomaly Detection with Outlier Exposure. *Proceedings of the International Conference on Learning Representations*.

Jung, S.; Lee, J.; Gwak, D.; Choi, S.; and Choo, J. 2021. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 15425–15434.

Li, J.; and Dong, Q. 2023. Open-Set Semantic Segmentation for Point Clouds via Adversarial Prototype Framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9425–9434.

Li, L.; Shum, H. P.; and Breckon, T. P. 2024. RAPiD-Seg: Range-Aware Pointwise Distance Distribution Networks for 3D LiDAR Segmentation. *arXiv preprint arXiv:2407.10159*.

Li, T.; Pang, G.; Bai, X.; Miao, W.; and Zheng, J. 2024. Learning transferable negative prompts for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17584–17594.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755. Springer.

Liu, Y.; Ding, C.; Tian, Y.; Pang, G.; Belagiannis, V.; Reid, I.; and Carneiro, G. 2023. Residual pattern learning for pixel-wise out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1151–1161.

Liu, Y.; Wei, X.; Lasang, P.; Pranata, S.; Subramanian, K.; and Seow, H. 2024. Ensemble Uncertainty Guided Road Scene Anomaly Detection: A Simple Meta-Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*.

Liu, Z.; Wang, Z.; Liang, P. P.; Salakhutdinov, R. R.; Morency, L.-P.; and Ueda, M. 2019. Deep gamblers: Learning to abstain with portfolio theory. *Advances in Neural Information Processing Systems*, 32.

McCormac, J.; Handa, A.; Leutenegger, S.; and Davison, A. J. 2017. Scenenet rgb-d: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation? In *Proceedings of the IEEE International Conference on Computer Vision*, 2678–2687.

Miao, W.; Pang, G.; Bai, X.; Li, T.; and Zheng, J. 2024. Out-of-distribution detection in long-tailed recognition with calibrated outlier class learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 4216–4224.

Movshovitz-Attias, Y.; Kanade, T.; and Sheikh, Y. 2016. How Useful Is Photo-Realistic Rendering for Visual Learning? In *ECCV*. arXiv.

Mozannar, H.; and Sontag, D. 2020. Consistent Estimators for Learning to Defer to an Expert. In *ICML2020*.

Nayal, N.; Yavuz, M.; Henriques, J. F.; and Güney, F. 2023. Rba: Segmenting unknown regions rejected by all. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 711–722.

Rai, S. N.; Cermelli, F.; Fontanel, D.; Masone, C.; and Caputo, B. 2023. Unmasking anomalies in road-scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4037–4046.

Song, S.; Yu, F.; Zeng, A.; Chang, A. X.; Savva, M.; and Funkhouser, T. 2017. Semantic scene completion from a single depth image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1746–1754.

Su, H.; Qi, C. R.; Li, Y.; and Guibas, L. J. 2015. Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 2686–2694. Santiago, Chile: IEEE. ISBN 978-1-4673-8391-2.

Tian, Y.; Liu, Y.; Pang, G.; Liu, F.; Chen, Y.; and Carneiro, G. 2022. Pixel-wise energy-biased abstention learning for anomaly segmentation on complex urban driving scenes. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, 246–263. Springer.

Wang, Y.; Zhao, W.; Cao, C.; Deng, T.; Wang, J.; and Chen, W. 2024. SFPNet: Sparse Focal Point Network for Semantic Segmentation on General LiDAR Point Clouds. *arXiv preprint arXiv:2407.11569*.

Xu, S.; Chen, X.; Zheng, Y.; Zhou, G.; Chen, Y.; Zha, H.; and Zhao, H. 2024. ECT: Fine-grained edge detection with learned cause tokens. *Image and Vision Computing*, 143: 104947.

Zhang, H.; Li, F.; Qi, L.; Yang, M.-H.; and Ahuja, N. 2024. CSL: Class-Agnostic Structure-Constrained Learning for Segmentation Including the Unseen. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7078–7086.

Zhang, Y.; Bai, M.; Kohli, P.; Izadi, S.; and Xiao, J. 2017a. Deepcontext: Context-encoding neural pathways for 3d holistic scene understanding. In *Proceedings of the IEEE international conference on computer vision*, 1192–1201.

Zhang, Y.; Song, S.; Yumer, E.; Savva, M.; Lee, J.-Y.; Jin, H.; and Funkhouser, T. 2017b. Physically-Based Rendering for Indoor Scene Understanding Using Convolutional Neural Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5057–5065. Honolulu, HI: IEEE. ISBN 978-1-5386-0457-1.

Zhao, W.; Li, J.; Dong, X.; Xiang, Y.; and Guo, Y. 2024. Segment Every Out-of-Distribution Object. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3910–3920.

Zhou, K.; Yang, J.; Loy, C. C.; and Liu, Z. 2022. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9): 2337–2348.

Zhou, Q.; Li, W.; Jiang, L.; Wang, G.; Zhou, G.; Zhang, S.; and Zhao, H. 2024. Pad: A dataset and benchmark for pose-agnostic anomaly detection. *Advances in Neural Information Processing Systems*, 36.

Zhu, X.; Zhou, H.; Wang, T.; Hong, F.; Ma, Y.; Li, W.; Li, H.; and Lin, D. 2021. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9939–9948.