

DOODLE: Diffusion-based Out-of-Distribution Learning for Open-set LiDAR Semantic Segmentation

Changgyoon Oh*

KAIST

Daejeon, Korea

changgyoon@kaist.ac.kr

Hyeonseong Kim*

KAIST

Daejeon, Korea

brian617@kaist.ac.kr

Daehyun We*

KAIST

Daejeon, Korea

greathyun@kaist.ac.kr

Jongoh Jeong

KAIST

Daejeon, Korea

jeong2@kaist.ac.kr

Yujeong Chae

KAIST

Daejeon, Korea

yujeong@kaist.ac.kr

Kuk-Jin Yoon

KAIST

Daejeon, Korea

kjyoon@kaist.ac.kr

Abstract

Open-set driving in complex real-world environments requires reliable identification of out-of-distribution (OOD) objects to avoid overconfident predictions on unseen categories. However, the sparsity and limited semantic richness of LiDAR point clouds make separating known and unknown classes difficult. This work proposes **DOODLE**, a diffusion model-based OOD learning framework for open-set 3D semantic segmentation. DOODLE trains a diffusion model to reconstruct in-distribution semantic features; feature-level reconstruction discrepancies then serve as OOD evidence. The resulting OOD scores are used to enhance backbone semantic features, improving discrimination of unknown regions during segmentation. To further reduce false positives arising from nonuniform measurements, a density-aware post-processing (DAP) module incorporates spatial variation in LiDAR point density when refining OOD predictions. DOODLE integrates seamlessly with existing open-set models and does not constrain backbone design. Experiments on SemanticKITTI and nuScenes demonstrate state-of-the-art OOD performance. On SemanticKITTI, DOODLE improves area under the precision-recall curve (AUPR) by 1.85%p and area under the receiver operating characteristic (AUROC) by 1.29%p over prior methods. Ablation studies confirm complementary benefits from diffusion-based reconstruction and DAP. Code is available at <https://github.com/chang9711/DOODLE>.

1. Introduction

LiDAR sensors have become an integral component of modern perception and navigation systems, providing precise,

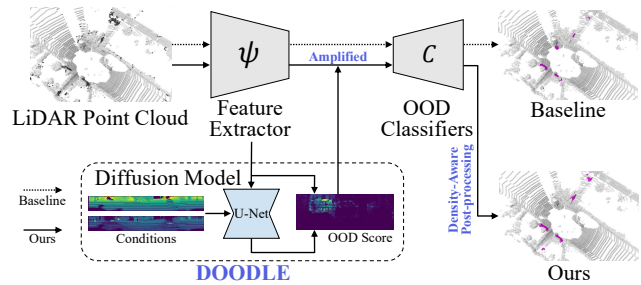


Figure 1. **Our diffusion-based out-of-distribution learning framework for open-set LiDAR semantic segmentation.** Compared to the traditional open-set semantic segmentation baselines that segments OOD using OOD classifiers, DOODLE further enhances open-set OOD segmentation based on the agreement between input and the reconstructed features from a diffusion model.

long-range distance measurements of complex, dynamic environments at high resolution. Leveraging these advantages, semantic segmentation of LiDAR point clouds has emerged as a fundamental 3D scene understanding task, playing a critical role in extracting semantically meaningful information for object localization. Current closed-set LiDAR semantic prediction methods [11, 18, 24, 26, 38, 49, 52, 53, 64–66] address this problem by classifying each point in 3D space into predefined categories such as vehicles, pedestrians, and road signs. While these methods excel at capturing the semantics of known classes, they struggle to detect unknown objects, posing safety risks. With an increasing number of accidents and near misses caused by such misclassifications [5], there is a pressing need for solutions that can robustly and reliably identify unknown objects in real-world environments.

To ensure safety and adaptability in dynamic envi-

ronments, recent research in out-of-distribution (OOD) localization has shifted its focus from closed-set methods [3, 16, 20, 32, 34, 55] to open-set recognition, which aims to classify known classes and identify unknown classes. Traditional closed-set networks are constrained by their training distributions and require substantial resources for retraining when adapting to new environments. To address these limitations, existing open set approaches to OOD detection in the 2D domain have explored confidence-based [9, 21, 27], synthesis-based [14, 31, 34, 51, 55], metric learning [7], and generalized clustering [23] techniques to assign scores for unknown classes.

However, naïve adaptations of 2D outlier detectors to 3D LiDAR point clouds have been found to perform poorly, often misclassifying outliers as inliers due to the high-dimensional nature of LiDAR data [8]. To overcome this, recent open-set 3D semantic segmentation methods employ redundancy classifiers that separately learn to discern unknown objects while normal classifiers handle known classes [8, 59]. However, existing methods still exhibit notable limitations. For instance, REAL [8] employs naive resizing operations that disrupt the natural LiDAR sampling pattern, leading to distorted geometric structures in the point cloud. LiON [59], on the other hand, generates synthetic outliers from ShapeNet [10] to compensate for the lack of semantic information in the LiDAR point clouds.

Although these discriminative methods with redundancy classifiers effectively distinguish unseen objects while classifying known ones, they still suffer from a high false positive rate. To address this, our objective is to further improve unknown class classification in open-set 3D semantic segmentation using generative models, which have demonstrated success in preserving semantic and geometric information in the 2D domain [41, 51]. Given the sparse and irregular nature of the LiDAR data, which inherently lacks semantic richness [8, 13, 59–61], we leverage the high semantic expressivity of diffusion models [22, 46] for effective segmentation of OOD classes from the learned distributions.

In this paper, we present **DOODLE**—a **D**iffusion-based **O**ut-Of-Distribution **L**earning framework for LiDAR point clouds that augments existing open-set 3D semantic segmentation methods with robust OOD detection. As illustrated in Fig. 1, DOODLE trains a diffusion model to learn the in-distribution semantic feature manifold and computes OOD scores from the discrepancy between input features and their reconstructions. To strengthen open-set segmentation using these scores, we introduce a feature amplification mechanism that steers representations along directions specified by redundancy-classifier weights, thereby increasing separability between known and unknown regions. In addition, we propose a density-aware post-processing (DAP) module that suppresses spurious OOD regions by accounting for spatial variation in LiDAR point density, reducing false positives

while preserving recall.

In summary, we pinpoint our key contributions as follows:

- We propose a novel diffusion model-based out-of-distribution learning framework for open-set LiDAR semantic segmentation, termed DOODLE. DOODLE enhances semantic features using OOD scores computed from the agreement between input and reconstructed features from a diffusion model. To the best of our knowledge, DOODLE is the first approach that utilizes diffusion models for open-set LiDAR semantic segmentation.
- We improve open-set segmentation with our density-aware post-processing (DAP), which refines OOD predictions based on LiDAR point density variations, reducing false positives.
- We achieve state-of-the-art performance on the SemanticKITTI and nuScenes open-set 3D semantic segmentation benchmark.

2. Related Work

In this section, we introduce some related literature such as 2D and 3D open-set semantic segmentation, and diffusion models for LiDAR point cloud.

2.1. 2D Open-set Semantic Segmentation

Following the pioneering work in open-set recognition [48] which emphasized the need for models to handle unseen classes, extensive research has been conducted in open-set image classification, primarily using uncertainty estimation and generative approaches. *Uncertainty-based* methods estimate prediction confidence to identify unknown samples. On the other hand, *generative model-based* methods use autoencoders and GANs to synthesize or reconstruct samples; e.g., [17, 40] generate synthetic unknown samples, while [42, 62] detect unknowns by comparing reconstructed and original inputs. In open-set semantic segmentation, similar techniques have been explored. *Metric learning-based* methods such as [7] classify unknown pixels based on similarity measures in feature space. In contrast, *generative model-based* approaches synthesize training features for unknown detection [31], or identify unknown regions by leveraging reconstruction errors from GANs [34, 55] or diffusion models [19, 35, 63].

2.2. 3D Open-set Semantic Segmentation

Despite extensive research in 2D, open-set recognition for 3D point clouds remains relatively under-explored. REAL [8] was the first to formulate open-set semantic segmentation for LiDAR by introducing a redundancy classifier framework. This framework employs additional classifiers to detect unknown objects by randomly resizing input objects, simulating unseen classes. However, this naive resizing operation disrupts the natural sampling pattern of

LiDAR sensors, leading to geometric distortions. To improve upon this, APF [33] integrates prototypical constraints and feature adversarial modules, leveraging class prototypes and adversarially generated features for open-set learning. LiON [59], the current state-of-the-art, enhances performance with point-wise abstaining penalties and synthetic outlier generation using ShapeNet [10].

Unlike in 2D, 3D open-set generative methods still remain relatively overlooked due to the complexity of handling LiDAR data [8]. In particular, diffusion models [22, 46], which have recently gained traction in 2D open-set semantic segmentation [19, 35, 63], have not yet been applied to open-set semantic segmentation in the 3D domain. In this light, we propose to leverage diffusion models with high semantic expressivity to enhance unknown class identification in open-set point cloud semantic segmentation.

2.3. Diffusion Models for LiDAR Point Cloud

Diffusion models, a class of generative models, have revolutionized image generation and have recently gained significant popularity in applications for their highly expressive feature representations. While primarily explored in 2D vision tasks, diffusion models have recently been adapted for 3D LiDAR processing [25, 39, 43, 57, 67]. For instance, R2DM [39] trains a DDPM-based diffusion model using depth and reflectance features derived from LiDAR range-view projections. LiDARDiffusion [43] extends the latent diffusion model (LDM) framework, enabling conditional multi-modal learning. Text2lidar [54] proposes a transformer-based architecture that enables text-guided LiDAR point cloud generation. Building on these advancements, we propose to leverage the high semantic expressivity of diffusion models to enhance unknown class identification in open-set 3D LiDAR semantic segmentation.

3. Method

Overview. In our diffusion model-based out-of-distribution learning framework, we propose to leverage a latent diffusion model [46], drawing inspiration from the success of reconstruction-based approaches in the 2D domain. In particular, we integrate a feature amplification mechanism into existing open-set 3D semantic segmentation classifiers [8, 59] using diffusion models, as illustrated in Fig. 2.

3.1. Method Formulation

Preliminaries. In the training pipeline of these baseline models, the LiDAR point cloud undergoes a feature extraction stage through the feature extractor $\psi(\cdot)$. The extracted point-wise semantic features are then passed through separate normal classifiers $g_{nm}(\cdot)$, and redundancy classifiers (*i.e.*, OOD classifiers), $g_{re}(\cdot)$, to output closed-set class logits and OOD scores, respectively. In our diffusion model-based enhancement stage, we first encode the semantic fea-

tures with a VQ-VAE encoder. Then, we feed them into the denoising U-Net conditioned with depth and reflectance maps, and sample at timestep t to output reconstructed range-projected images, containing in-distribution features. Finally, we obtain a score map of the out-of-distribution regions by measuring the agreement between the input and the reconstructed features.

Our Method. Given an input LiDAR point cloud $\mathcal{P} \in \mathbb{R}^{N \times 3}$, we first extract semantic point features $\mathcal{F}_{\text{point}} \in \mathbb{R}^{N \times C}$ from the feature extractor $\psi(\cdot)$ of a pre-trained open-set semantic segmentation model, *that is*, $\mathcal{F}_{\text{point}} = \psi(\mathcal{P})$, where N and C are the number of points and the dimension of the feature, respectively. The extracted features are then projected onto range view representations $\mathcal{F}_{\text{range}} \in \mathbb{R}^{H \times W \times C}$ for seamless integration with latent diffusion models, where H and W denote the vertical and horizontal resolutions of the range-view representation. Following previous work [25], we adopt the projected range view representation as a dense 2D LiDAR point cloud representation, since it is particularly well suited for generation tasks [1, 8, 12, 30, 38, 46, 58, 59]. Specifically, range-view representation is advantageous for several compelling reasons: compatibility with state-of-the-art generative models [67] and flexibility in conditional input compared to the volume representation approach [56]. By leveraging the range-view representation, we can harness the full potential of latent diffusion models for open-set segmentation models.

Following existing approaches to open-set 3D semantic segmentation, the initial semantic features $\mathcal{F}_{\text{range}}$ contain both in-distribution (ID) and out-of-distribution (OOD) features. To ensure our diffusion model learns to reconstruct exclusively the ID regions, we explicitly mask out OOD regions during training. Specifically, we apply a range-view masking to construct the input representation $\mathcal{F}_{\text{range}}^{\text{in}} = \mathcal{M} \odot \mathcal{F}_{\text{range}}$, using the binary mask $\mathcal{M} \in \{0, 1\}^{H \times W \times 1}$ for each range-view point $p \in \mathcal{F}_{\text{range}}$ defined as:

$$\mathcal{M}(p) = \begin{cases} 1, & \text{if } p \in \{\text{ID classes}\}, \\ 0, & \text{if } p \in \{\text{OOD classes}\}. \end{cases} \quad (1)$$

However, challenges arise from directly using $\mathcal{F}_{\text{range}}^{\text{in}}$ in its form due to high-frequency noise and high computational cost from its high feature dimension. To circumvent these issues, we employ a Vector-Quantized VAE (VQ-VAE) [50] to suppress noise and reduce the feature dimension, *i.e.*, $z = \mathcal{E}_{\text{VQ}}(\mathcal{F}_{\text{range}}^{\text{in}})$. Moreover, we further guide our semantic feature learning by conditionally using depth map $\mathcal{D}_{\text{range}} \in \mathbb{R}^{H \times W \times 1}$ and reflectance map $\mathcal{R}_{\text{range}} \in \mathbb{R}^{H \times W \times 1}$ from the LiDAR scan, projected to range-view in a similar manner. These maps are resized with bilinear interpolation to match the resolution of z . In essence, we concatenate the dimension-reduced latent feature z with the resized depth and reflectance maps, and feed the result to the denoising U-Net for training. At inference, the semantic features con-

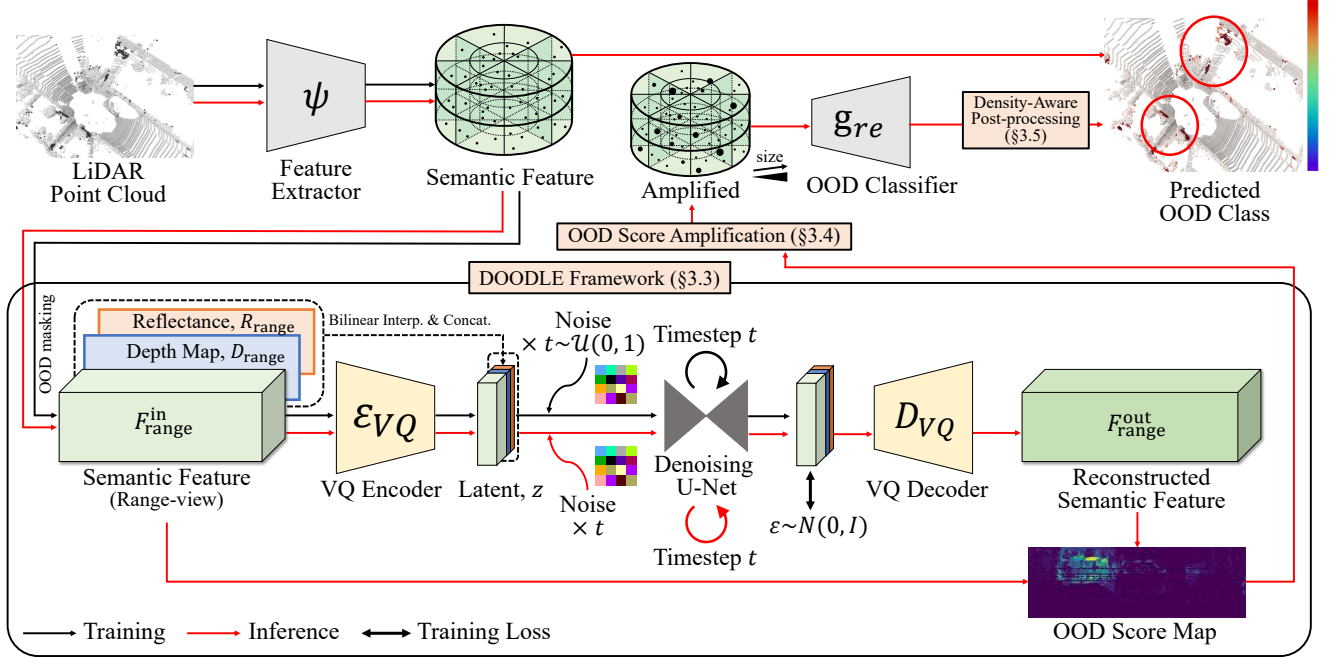


Figure 2. **Overview of the DOODLE framework.** Based on the semantic features obtained from the pre-trained open-set segmentation model, we project these features onto the range-view representation. The projected feature is then encoded into lower-dimensional features via the VQ-VAE encoder, which becomes the input to the latent diffusion model conditioned with depth and reflectance maps. For detecting out-of-distribution (OOD) objects, we first obtain OOD scores based on the agreement between input features and reconstructed features from a diffusion model. These scores are then used to amplify the semantic features indicating OOD objects. Subsequently, the depth-aware post-processing procedure mitigates false-positive predictions.

taining in-distribution regions are reconstructed. These reconstructed features are compared against the input features to measure the OOD score map, which locates potential OOD regions. These amplified regions are then fed to our post-processing, yielding the final OOD predictions.

3.2. Proposed DOODLE Framework

Diffusion models, a promising generative method, have shown considerable success in image generation tasks by progressively refining noisy data into high-quality samples. In the context of LiDAR point clouds, diffusion models offer several advantages, including their ability to capture intricate spatial relationships and generate realistic point cloud structures. These models can effectively help identify OOD regions by learning the underlying distribution of ID regions and distinguishing deviations that represent potential OODs. Moreover, diffusion models are well-suited for handling the continuous, unordered nature of LiDAR data, making them ideal for enriching 3D point cloud features in a generative framework. Motivated by related research in the 2D domain that has shown that training diffusion models on deep segmentation embedding spaces, rather than directly on the image distribution, is more effective for OOD detection [41, 51], we propose to train a latent diffusion model on

in-distribution point cloud segmentation embeddings.

Training VQ-VAE. To construct memory-efficient and smooth manifolds for compact latent representation, we employ VQ-VAE to compress the semantic features spatially with low feature dimensionality, based on the insights that vector quantized representations are robust to noise and easily compatible with generative models [15, 44, 45, 57]. For compression, we specifically train the implementation [46] with a smooth $L1$ objective [28], which is known to be effective as a reconstruction loss when the input features are noisy. We formulate the reconstruction loss as follows:

$$\mathcal{L}_{rec} = \begin{cases} \frac{1}{2}(\mathcal{F}_{range}^{in} - \mathcal{F}_{range}^{out})^2, & \text{if } |\mathcal{F}_{range}^{in} - \mathcal{F}_{range}^{out}| < 1, \\ |\mathcal{F}_{range}^{in} - \mathcal{F}_{range}^{out}| - 0.5, & \text{else.} \end{cases} \quad (2)$$

where $\mathcal{F}_{range}^{out} = \mathcal{D}_{VQ}(\mathcal{E}_{VQ}(\mathcal{F}_{range}^{in}))$ represents the reconstructed semantic feature with the VQ decoder, $\mathcal{D}_{VQ}$.

Training Phase. The pre-trained VQ encoder $\mathcal{E}_{VQ}$ encodes $\mathcal{F}_{range}^{in}$ to latent feature $z = \mathcal{E}_{VQ}(\mathcal{F}_{range}^{in})$. We then concatenate z with the resized $\mathcal{D}_{range}$ and $\mathcal{R}_{range}$, and use the result as input to conditionally train a denoising U-Net in our diffusion model. For implementation, we employ variational diffusion models [29] in our latent diffusion model framework to enable efficient optimization of the noise schedule

by utilizing the continuous time variational lower bound.

We formulate the forward transition process of the continuous-time diffusion model for the noisy latents from timestep $s \rightarrow t$ ($t, s \in [0, 1]$) as follows:

$$q(z_t|z_s) := \mathcal{N}(\alpha_{t|s}z_s, \sigma_{t|s}^2\mathbf{I}), \quad (3)$$

where α, σ are α -cosine schedule parameters, with $\alpha_{t|s} = \alpha_t/\alpha_s$ and $\sigma_{t|s} = \sigma_t - \alpha_{t|s}\sigma_s$. For the denoising process, we conditionally denoise the noisy feature z_t with $\mathcal{D}_{\text{range}}$ and $\mathcal{R}_{\text{range}}$ as follows:

$$p_\theta(z_{0:T}|\mathcal{D}_{\text{range}}, \mathcal{R}_{\text{range}}) = p(z_T) \prod_{t=1}^T p_\theta(z_{t-1}|z_t, \mathcal{D}_{\text{range}}, \mathcal{R}_{\text{range}}), \quad (4)$$

Altogether, we train a denoising U-Net with the following diffusion training objective:

$$\mathcal{L}_{ldm} = \mathbb{E}_{z, \epsilon \sim \mathcal{N}(0, \mathbf{I}), t \sim \mathcal{U}(0, 1)} [\|\epsilon - \hat{\epsilon}(z_t, \lambda_t, \mathcal{D}_{\text{range}}, \mathcal{R}_{\text{range}})\|_2^2], \quad (5)$$

where $\hat{\epsilon}$ is the output of denoising U-Net and $\lambda_t = \alpha_t^2/\sigma_t^2$.

Inference Phase. At inference, we utilize DOODLE as a reconstruction model that reconstructs the input semantic features, $\mathcal{F}_{\text{range}}^{\text{in}}$, *i.e.*, in-distribution semantic features are to be reconstructed. In the process, we first perturb the noisy latent, z , t times to yield z_t through the forward process. Then, the noisy latent z_t is denoised t times followed by decoding through the pre-trained VQ decoder to yield the final reconstructed output, $\mathcal{F}_{\text{range}}^{\text{out}}$. Finally, we can estimate the potential out-of-distribution feature locations by measuring the difference between the input, $\mathcal{F}_{\text{range}}^{\text{in}}$, and the reconstructed output, $\mathcal{F}_{\text{range}}^{\text{out}}$. Due to the randomness of Gaussian noise in diffusion sampling, reconstruction using diffusion models may result in noisy outputs. We thus utilize a diffusion noise ensemble strategy, in which the reconstruction process is repeated multiple times with stochastic noise variations and the outputs are averaged to obtain robust and stable range-view semantic features. These ensemble-averaged features are then compared against the input features to locate potential OOD regions.

However, since these features contain semantic information, direct pixel-wise comparison proved empirically ineffective for reliably detecting OOD regions. Instead, we resort to comparing distributional differences by the cross-entropy between the segmentation logits of the input features, $\mathcal{F}_{\text{range}}^{\text{in}}$, and the decoded output, $\mathcal{F}_{\text{range}}^{\text{out}}$, using normal classifiers from the pre-trained semantic segmentation model, g_{nm} . Here, the normal classifiers predict the closed-set prediction logits for known classes. The final OOD score map $\mathcal{S} \in \mathbb{R}^{H \times W}$ is computed as follows:

$$\mathcal{S} = - \sum g_{nm}(\mathcal{F}_{\text{range}}^{\text{in}}) \log(g_{nm}(\mathcal{F}_{\text{range}}^{\text{out}})). \quad (6)$$

3.3. OOD Score Amplification

From the predicted potential OOD regions in the score map, we aim to improve the OOD scores in the final semantic

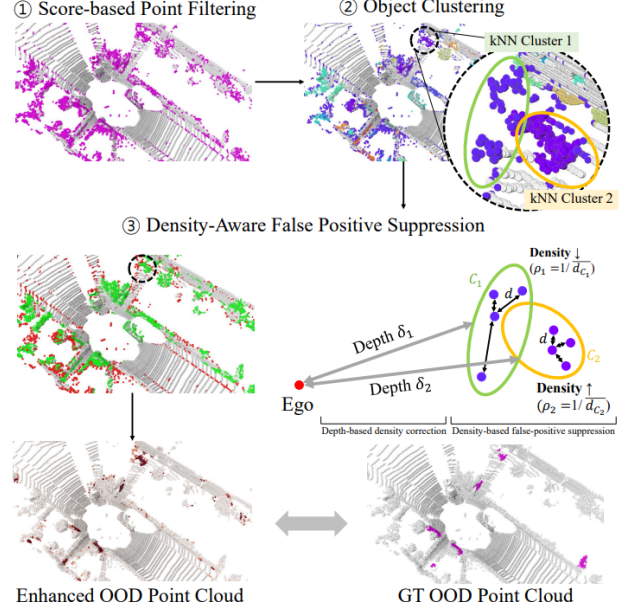


Figure 3. **Illustration of the proposed density-aware post-processing (DAP).** The amplified OOD scores undergo post-processing steps: (1) score-based point filtering, (2) object clustering, and (3) density-aware false positive suppression based on distance to KNN clusters and their depth.

segmentation output. To achieve this, we apply a series of refinement steps to the computed score map $\mathcal{S}$.

Firstly, we re-project the score map $\mathcal{S}$ onto the 3D space to derive point features. Since the input to the redundancy classifiers of the pre-trained open-set semantic segmentation model [8, 59] is voxelized, we reshape $\mathcal{S}$ into voxels, *i.e.*, $\mathcal{F}_{\text{voxel}} = \text{voxelize}(\mathcal{S})$, by aggregating the point features and assigning the mean score value in each voxel. Subsequently, we refine the voxel features by applying a threshold τ_{amp} to the scores to obtain $\mathcal{F}_{\text{voxel}}^{\text{mask}} = M_{\mathcal{S}} \odot \mathcal{F}_{\text{voxel}}$, using the binary mask defined as:

$$M_{\mathcal{S}} = \begin{cases} 1, & \mathcal{F}_{\text{voxel}} > \tau_{\text{amp}}, \\ 0, & \text{else.} \end{cases} \quad (7)$$

Finally, we select the classifier, $g_{re}^j(\cdot)$, from the redundancy classifiers $g_{re} = \{g_{re}^1(\cdot), g_{re}^2(\cdot), \dots, g_{re}^r(\cdot)\}$ of the pre-trained open-set segmentation model that yields the maximum OOD score given the refined score map. To this end, we project and normalize the refined score map features along the direction of the weight vector of the selected redundancy classifier $g_{re}^j(\cdot)$. The projected features are then amplified by applying a scaling factor λ , and fed into the classifier to output the final OOD scores $\hat{\mathcal{S}}$ as follows:

$$\hat{\mathcal{S}} = g_{re}^j \left(\lambda \cdot \frac{|\mathcal{F}_{\text{voxel}}^{\text{mask}} \odot \mathcal{W}_{g_{re}^j(\cdot)}|}{|\mathcal{F}_{\text{voxel}}^{\text{mask}}|} \right), \quad (8)$$

where $\circ$ denotes the dot product operator, $\mathcal{W}_{g_{re}^*}(\cdot)$ indicates the weight vector of the classifier $g_{re}^*(\cdot)$, and $j = \operatorname{argmax}_{i \in \{1, 2, \dots, r\}} \{|\mathcal{F}_{\text{voxel}}^{\text{mask}} \circ \mathcal{W}_{g_{re}^*}^i(\cdot)|\}$.

3.4. Density-aware Post-processing

The proposed reconstruction-based method may yield OOD scores with some degree of noise, which could potentially affect the false positive rate. To mitigate this, we introduce a density-aware post-processing (DAP) pipeline, which reduces noise in OOD predictions and improves accuracy. Our post-processing is illustrated in Fig. 3, with the corresponding pseudo-code algorithm in the *Supplementary Material*.

The process begins by filtering points based on a score threshold τ_{score} , retaining only those with scores above this threshold as potential OOD points for further processing. Fig. 3 illustrates this filtering process, where pink points denote retained candidates, while gray points represent points filtered out. Our approach is based on the assumption that sparse predictions are more likely to be noisy predictions, *i.e.*, false positives. This assumption follows from the observation that dense predictions are more likely to correspond to coherent object classes, while isolated points often stem from noise or spurious detections.

To this end, we apply HDBSCAN [37] to the remaining points to form OOD clusters $\{C_j\}_{j=1}^M$, as shown in Fig. 3. Points that fail to form clusters are treated as likely false positives, and their scores are reduced by a reduction factor γ . For each detected cluster, we compute its density ρ_j as the reciprocal of the mean distance to its k nearest neighbors. Since LiDAR point cloud density naturally decreases with distance from the LiDAR sensor, we apply a depth-based density correction to adjust density estimates. Distant clusters receive density amplification to account for point sparsity, while nearby clusters have their density reduced.

Clusters with corrected density below a pre-defined threshold τ_{density} are deemed likely false positives. For these clusters, we suppress the OOD scores by applying the reduction factor γ . In Step 3 of Fig. 3, the red points represent the suppressed points, while the green points correspond to those predicted as OOD objects. Detailed description of the DAP algorithm is provided in the *Supplementary Material*.

4. Experiments

4.1. Experimental Setup

In this section, we specify the detailed experimental setup used throughout our experiments. Further details in *Supp.*

Datasets. SemanticKITTI [4] and nuScenes [6] are large-scale 3D autonomous driving datasets widely used for LiDAR semantic segmentation. For fair comparisons with prior works, we designate the *other-vehicle* class in SemanticKITTI and the *barrier*, *construction-vehicle*, *traffic-cone*, *trailer* classes in nuScenes as OOD classes.

Implementation Details. We adopt the LDM [46] VQ-VAE and a denoising U-Net as our diffusion model. The VQ-VAE output channel dimension is set to 16 to balance performance and resource usage. For VQ-VAE training, we use AdamW [36] with $lr = 4.5 \times 10^{-6}$, $\beta_1 = 0.9$, and $\beta_2 = 0.5$; for the denoising U-Net, we use AdamW with $lr = 10^{-4}$, $\beta_1 = 0.9$, and $\beta_2 = 0.99$. We set the diffusion sampling steps to $T = 256$ and choose $t = 1$ for reconstruction. The denoising step is repeated five times and the results are averaged to form an ensemble. We use $\tau_{\text{amp}} = 0.8$ and $\lambda = 10$ for score amplification. In post-processing, we set the score threshold τ_{score} to retain the top 40% of points by score, fix the HDBSCAN minimum cluster size at 40 points with density computed using the three nearest neighbors, set $\tau_{\text{density}} = 3.0$, and reduce scores by a ratio $\gamma = 0.4$.

We utilize 3D semantic point features projected onto 2D range view as input to our DOODLE framework. The projected feature resolution is (64, 1024) for SemanticKITTI [4] acquired from 64-beam LiDAR sensors, and (32, 1070) for nuScenes [6] with 32-beam sensors.

Conventional U-Net [47]-based diffusion models require a significant amount of computational load. For ease of implementation and a cost-efficient framework, we employed U-Net with EfficientNet as the backbone. For diffusion model training, we set the total timesteps $T = 1,024$, and sampled with total timesteps $T = 256$ for cost efficiency.

Evaluation metrics. In accordance with existing works [8, 59], we assess the semantic segmentation performance on the OOD classes by the area under the ROC curve (AUROC) and the area under the precision-recall curve (AUPR) [2].

4.2. Main Results

Quantitative Comparison. Table 1 shows the performance of our diffusion-based method and compares it with other open-set point cloud semantic segmentation methods. Our framework is an add-on method that can be applied to existing open-set LiDAR semantic segmentation baselines. When applied to the baseline REAL [8], our method yields improvements of 2.09%p, 0.82%p on SemanticKITTI, and 1.33%p, 0.26%p on nuScenes in AUPR and AUROC, respectively. Similarly, applying our method to LiON [59] results in improvements of 1.85%p, 1.29%p on SemanticKITTI, and 2.37%p, 0.02%p on nuScenes in AUPR and AUROC, respectively. When our DOODLE framework is integrated as an add-on to existing baseline models, consistent performance improvements are observed across all cases. This demonstrates that our framework can seamlessly integrate. Furthermore, our method achieves state-of-the-art performance on both the SemanticKITTI and nuScenes datasets. On the SemanticKITTI dataset, DOODLE framework demonstrates significant improvements in both metrics, indicating an overall enhancement in the classification performance of the open-set semantic segmentation model.

Table 1. Benchmark comparisons of open-set semantic segmentation for LIDAR point clouds. Results are evaluated on the validation set. We achieve noticeable gains in both AUPR and AUROC on SemanticKITTI, and improvement in AUPR with comparable AUROC with the baseline on nuScenes. *Code is unavailable.

Method	SemanticKITTI [4]		nuScenes [6]	
	AUPR $\uparrow$	AUROC $\uparrow$	AUPR $\uparrow$	AUROC $\uparrow$
MSP [20]	6.70	74.00	4.30	76.70
MaxLogit [2]	7.60	70.50	8.30	79.40
MC-Dropout [16]	7.40	74.70	14.90	82.60
APF* [33]	36.10	85.60	-	-
REAL [8]	20.08	84.90	21.20	84.50
REAL + DOODLE	22.17 (+2.09 $\uparrow$)	85.72 (+0.82 $\uparrow$)	22.53 (+1.33 $\uparrow$)	84.76 (+0.26 $\uparrow$)
LiON [59]	44.68	92.69	31.58	95.24
LiON + DOODLE	46.53 (+1.85 $\uparrow$)	93.98 (+1.29 $\uparrow$)	33.95 (+2.37 $\uparrow$)	95.26 (+0.02 $\uparrow$)



Figure 4. **Qualitative comparisons** on SemanticKITTI (top) and nuScenes (bottom) datasets. OOD in vivid colors and ID in gray on SemanticKITTI; red circles on nuScenes mark OOD objects our method detects but the baseline misses. Our approach detects more OOD classes, demonstrating superior OOD identification.

In the nuScenes dataset, while the AUPR value showed a notable increase, the AUROC performance remained comparable. This suggests that the model is better focused on the positive class in the nuScenes dataset. The significant improvements on both datasets demonstrate that our outlier emphasis with diffusion guidance refines semantic features for better outlier detection. We remark that mIoU for the known classes remains the same since our DOODLE framework is designed to only amplify the scores of the outliers.

Qualitative Comparison. Figure 4 visualizes outputs from DOODLE and LiON [59] on SemanticKITTI and nuScenes. With diffusion-based reconstruction, DOODLE learns the geometrical distribution of OOD objects, enabling object-wise score estimation. By aggregating features from the pretrained model and amplifying scores at potential OOD points, DOODLE detects regions the pretrained model missed and reduces high false positives. Additional qualitative results are provided in the *Supplementary Material*.

4.3. Ablation Study

To demonstrate the advantages of DOODLE, we conducted ablation experiments on the SemanticKITTI.

Diffusion Noise Ensemble. Our DOODLE utilizes a dif-

fusion model to reconstruct semantic features. Due to the use of random noise in the diffusion reverse process, the sampling process tends to synthesize noisy features. To mitigate performance degradation caused by randomness, we employed a diffusion noise ensemble approach. When generating the score map using the score function, reconstructed features are used. To reduce randomness, we estimated the score map multiple times and averaged the results to produce the final score map. In Fig. 5, we report the AUPR and AUROC performance for ensemble counts of 1, 3, and 5. Both AUPR and AUROC showed improvements as the number of ensembles increased. Considering the trade-off between computational effort and performance gains, we selected five steps as the optimal configuration.

Diffusion Timestep Analysis. We present the performance variations with respect to the diffusion denoising timestep t used in the reconstruction and the diffusion training sampling timestep T in Table 2. Increasing the number of denoising steps t slightly reduces both AUPR and AUROC due to the accumulation of stochastic noise. Consequently, we chose to set $t = 1$ to mitigate the impact of stochastic noise accumulation. On the other hand, increasing T improves the sample quality and boosts both metrics. However, utilizing

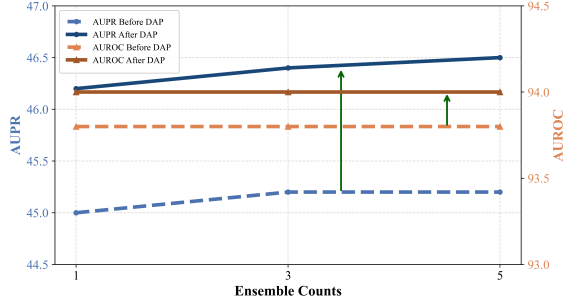


Figure 5. Ablation study on ensemble counts.

larger T diffusion models leads to an increase in computational cost during training. To balance computation and performance, we set $T = 256$.

Table 2. Ablation study on the diffusion timestep. The gray-shaded areas represent the variables chosen by our model.

# denoising t	AUPR $\uparrow$	AUROC $\uparrow$	T	# denoising t	AUPR $\uparrow$	AUROC $\uparrow$
$t = 56$	45.13	93.86	512	2	46.67	94.01
$t = 6$	46.38	93.96	256	1	46.50	94.00
$t = 1$	46.50	94.00				

Component Analysis. The DOODLE framework employs a latent diffusion model as a generative model for semantic feature reconstruction. To enhance performance, a post-processing pipeline (DAP) is incorporated to suppress false positives in the OOD predictions. In Table 3, the effectiveness of each component is validated through a series of ablation studies conducted on the SemanticKITTI dataset. Using the diffusion model for semantic feature reconstruction results in significant improvements in both AUPR and AUROC compared to using the VQ-VAE structure alone. This highlights the superior ability of diffusion models in capturing semantic information, making it more effective in computing OOD scores. Regardless of whether the diffusion model is used, the DAP pipeline consistently improves performance in both metrics, effectively reducing noise in OOD predictions. Even when only the diffusion model is used, performance surpasses the baseline LiON [59] in both AUPR and AUROC, underscoring the effectiveness of the OOD score amplification method. In conclusion, the combination of OOD score amplification, the diffusion model, and the DAP pipeline results in state-of-the-art performance in both metrics.

Depth-based Density Correction Analysis. Our DAP module leverages the proposed depth-based density correction to account for the natural decrease in the density of the point cloud with distance from the LiDAR center. To further evaluate the impact of this density correction method, we compare it with an approach that relies on a standard density. Keeping all other parameters constant, the results in Table 4 show a +0.32%p AUPR improvement on the Se-

Table 3. Ablation study on the components of DOODLE.

Method	Diffusion	DAP	AUPR $\uparrow$	AUROC $\uparrow$
Baseline [59]	-	-	44.68	92.69
Ours	$\times$	$\times$	43.66	93.74
	$\times$	$\checkmark$	45.59	93.91
	$\checkmark$	$\times$	45.20	93.79
	$\checkmark$	$\checkmark$	46.53	93.98

manticKITTI dataset when applying the density correction, together with a slight increase in AUROC.

Table 4. Ablation study on the depth-based density correction.

Method	Post-processing	Density Correction	AUPR $\uparrow$	AUROC $\uparrow$
Ours	-	-	45.20	93.80
	$\checkmark$	-	46.21	93.97
	$\checkmark$	$\checkmark$	46.53	93.98

Hyper-parameter Sensitivity. We conducted a sensitivity analysis on the post-processing hyper-parameters on SemanticKITTI in Tables 5. As the value of τ_{score} increases, the number of filtered points also increases, leading to better performance for both metrics, though at the expense of longer inference time. We selected the values given the trade-off between marginal increases in the two scores and non-trivial increases in the inference time. For τ_{density} , it is tuned to find the optimal value for determining OOD objects based on density, in order to achieve the best performance.

Table 5. Sensitivity and inference time analysis of τ_{score} and Sensitivity analysis of τ_{density} . We chose the τ_{score} and τ_{density} in gray.

τ_{score} (%)	10	20	30	40	50	60	
AUPR	0.4624	0.4634	0.4648	0.4653	0.4667	0.4661	
AUROC	0.9385	0.9390	0.9395	0.9398	0.9400	0.9402	
Inference time (s)	0.2017	0.3754	0.5487	0.7655	1.0827	1.3397	
τ_{density}	1	2	3	4	5	6	7
AUPR	0.4639	0.4640	0.4653	0.4590	0.4570	0.4331	0.4167
AUROC	0.9394	0.9395	0.9398	0.9392	0.9389	0.9374	0.9360

5. Conclusion

We introduce DOODLE, a novel diffusion model-based framework for out-of-distribution learning in open-set 3D LiDAR semantic segmentation. DOODLE enhances semantic features using OOD scores computed from a diffusion model and further refines OOD predictions through the density-aware post-processing (DAP). The overall pipeline is designed to integrate seamlessly into existing open-set 3D semantic segmentation frameworks. Our approach achieves state-of-the-art performance on both SemanticKITTI and nuScenes, establishing a new paradigm for applying diffusion models to open-set 3D semantic segmentation.

Acknowledgement. This work was supported by the Technology Innovation Program (1415187329,20024355, Development of autonomous driving connectivity technology based on sensor-infrastructure cooperation) funded by the Ministry of Trade, Industry & Energy(MOTIE, Korea) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF2022R1A2B5B03002636). We thank 42dot for funding this research. We also appreciate the support and valuable discussions from the members of the 42dot research team.

References

- [1] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5240–5250, 2023. 3
- [2] Steven Basart, Mazeika Mantas, Mostajabi Mohammadreza, Steinhardt Jacob, and Song Dawn. Scaling out-of-distribution detection for real-world settings. In *International Conference on Machine Learning*, 2022. 6, 7
- [3] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pages 161–169. Springer, 2019. 2
- [4] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9297–9307, 2019. 6, 7
- [5] Darko Bozhinoski, Davide Di Ruscio, Ivano Malavolta, Patrizio Pelliccione, and Ivica Crnkovic. Safety for mobile robotic systems: A systematic mapping study from a software engineering perspective. *Journal of Systems and Software*, 151:150–179, 2019. 1
- [6] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 6, 7
- [7] Jun Cen, Peng Yun, Junhao Cai, Michael Yu Wang, and Ming Liu. Deep metric learning for open world semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15333–15342, 2021. 2
- [8] Jun Cen, Peng Yun, Shiwei Zhang, Junhao Cai, Di Luan, Mingqian Tang, Ming Liu, and Michael Yu Wang. Open-world semantic segmentation for lidar point clouds. In *European Conference on Computer Vision*, pages 318–334. Springer, 2022. 2, 3, 5, 6, 7
- [9] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5128–5137, 2021. 2
- [10] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 2, 3
- [11] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3075–3084, 2019. 1
- [12] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *International Symposium on Visual Computing*, pages 207–222. Springer, 2020. 3
- [13] Tiago Cortinhal, Idriss Gouigah, and Eren Erdal Aksoy. Semantics-aware lidar-only pseudo point cloud generation for 3d object detection. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 3220–3226. IEEE, 2024. 2
- [14] Giancarlo Di Biase, Hermann Blum, Roland Siegwart, and Cesar Cadena. Pixel-wise anomaly detection in complex driving scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16918–16927, 2021. 2
- [15] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021. 4
- [16] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 2, 7
- [17] ZongYuan Ge, Sergey Demianov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*, 2017. 2
- [18] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9224–9232, 2018. 1
- [19] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, pages 8472–8480, 2024. 2, 3
- [20] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016. 2, 7
- [21] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joe Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*, 2019. 2

- [22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 3
- [23] Jie Hong, Weihao Li, Junlin Han, Jiyang Zheng, Pengfei Fang, Mehrtash Harandi, and Lars Petersson. Goss: Towards generalized open-set semantic segmentation. *The Visual Computer*, 40(4):2391–2404, 2024. 2
- [24] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randa-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11108–11117, 2020. 1
- [25] Qianjiang Hu, Zhimin Zhang, and Wei Hu. Rangeldm: Fast realistic lidar point cloud generation. In *European Conference on Computer Vision*, pages 115–135. Springer, 2025. 3
- [26] Hyun-Kurl Jang, Jihun Kim, Hyeokjun Kweon, and Kuk-Jin Yoon. Talos: Enhancing semantic scene completion via test-time adaptation on the line of sight. *Advances in Neural Information Processing Systems*, 37:74211–74232, 2024. 1
- [27] Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, and Jaegul Choo. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15425–15434, 2021. 2
- [28] Natalia Khanzhina, Alexey Lapenok, and Andrey Filchenkov. Towards robust object detection: Bayesian retinanet for homoscedastic aleatoric uncertainty modeling. *arXiv preprint arXiv:2108.00784*, 2021. 4
- [29] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021. 4
- [30] Deyvid Kochanov, Fatemeh Karimi Nejadasl, and Olaf Booij. Kprnet: Improving projection-based lidar semantic segmentation. *arXiv preprint arXiv:2007.12668*, 2020. 3
- [31] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 813–822, 2021. 2
- [32] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017. 2
- [33] Jianan Li and Qiulei Dong. Open-set semantic segmentation for point clouds via adversarial prototype framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9425–9434, 2023. 3, 7
- [34] Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2152–2161, 2019. 2
- [35] Victor Livernoche, Vineet Jain, Yashar Hezaveh, and Siamak Ravanbakhsh. On diffusion modeling for anomaly detection. *arXiv preprint arXiv:2305.18593*, 2023. 2, 3
- [36] I Loshchilov. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6
- [37] Leland McInnes, John Healy, Steve Astels, et al. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205, 2017. 6
- [38] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 4213–4220. IEEE, 2019. 1, 3
- [39] Kazuto Nakashima and Ryo Kurazume. Lidar data synthesis with denoising diffusion probabilistic models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14724–14731. IEEE, 2024. 3
- [40] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *Proceedings of the European conference on computer vision (ECCV)*, pages 613–628, 2018. 2
- [41] Julia Nitsch, Masha Itkina, Ransalu Senanayake, Juan Nieto, Max Schmidt, Roland Siegwart, Mykel J Kochenderfer, and Cesar Cadena. Out-of-distribution detection for automotive perception. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 2938–2943. IEEE, 2021. 2, 4
- [42] Poojan Oza and Vishal M Patel. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2307–2316, 2019. 2
- [43] Haoxi Ran, Vitor Guizilini, and Yue Wang. Towards realistic scene generation with lidar diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14738–14748, 2024. 3
- [44] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae-2. *Advances in neural information processing systems*, 32, 2019. 4
- [45] Christoph B Rist, David Emmerichs, Markus Enzweiler, and Darius M Gavrilu. Semantic scene completion using local deep implicit functions on lidar data. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7205–7218, 2021. 4
- [46] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 3, 4, 6
- [47] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 6
- [48] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boulton. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012. 2
- [49] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotequi, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6411–6420, 2019. 1

- [50] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 3
- [51] Tomas Vojir, Tomáš Šipka, Rahaf Aljundi, Nikolay Chumerin, Daniel Olmeda Reino, and Jiri Matas. Road anomaly detection by partial image reconstruction with segmentation coupling. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15651–15660, 2021. 2, 4
- [52] 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 *2018 IEEE international conference on robotics and automation (ICRA)*, pages 1887–1893. IEEE, 2018. 1
- [53] 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 *2019 international conference on robotics and automation (ICRA)*, pages 4376–4382. IEEE, 2019. 1
- [54] Yang Wu, Kaihua Zhang, Jianjun Qian, Jin Xie, and Jian Yang. Text2lidar: Text-guided lidar point cloud generation via equirectangular transformer. In *European Conference on Computer Vision*, pages 291–310. Springer, 2024. 3
- [55] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, and Alan L Yuille. Synthesize then compare: Detecting failures and anomalies for semantic segmentation. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 145–161. Springer, 2020. 2
- [56] Yuwen Xiong, Wei-Chiu Ma, Jingkang Wang, and Raquel Urtasun. Learning compact representations for lidar completion and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1074–1083, 2023. 3
- [57] Yuwen Xiong, Wei-Chiu Ma, Jingkang Wang, and Raquel Urtasun. Learning compact representations for lidar completion and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1074–1083, 2023. 3, 4
- [58] 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 *Computer vision—ECCV 2020: 16th European conference, glasgow, UK, August 23–28, 2020, proceedings, part XXVIII 16*, pages 1–19. Springer, 2020. 3
- [59] Shaocong Xu, Pengfei Li, Qianpu Sun, Xinyu Liu, Yang Li, Shihui Guo, Zhen Wang, Bo Jiang, Rui Wang, Kehua Sheng, et al. Lion: Learning point-wise abstaining penalty for lidar outlier detection using diverse synthetic data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8951–8959, 2025. 2, 3, 5, 6, 7, 8
- [60] Xu Yan, Jiantao Gao, Jie Li, Ruimao Zhang, Zhen Li, Rui Huang, and Shuguang Cui. Sparse single sweep lidar point cloud segmentation via learning contextual shape priors from scene completion. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3101–3109, 2021.
- [61] Hao Yang, Haiyang Wang, Di Dai, and Liwei Wang. Pred: pre-training via semantic rendering on lidar point clouds. *Advances in Neural Information Processing Systems*, 36:46041–46053, 2023. 2
- [62] Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. Classification-reconstruction learning for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4016–4025, 2019. 2
- [63] Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia. Unsupervised surface anomaly detection with diffusion probabilistic model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6782–6791, 2023. 2, 3
- [64] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9601–9610, 2020. 1
- [65] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4490–4499, 2018.
- [66] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9939–9948, 2021. 1
- [67] Vlas Zyrianov, Xiyue Zhu, and Shenlong Wang. Learning to generate realistic lidar point clouds. In *European Conference on Computer Vision*, pages 17–35. Springer, 2022. 3