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

In this supplementary material, we provide additional materials not included in the main paper due to the page limitation. The following sections are the implementation details and additional experimental results.

A. Implementation Details

Feature projection. 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 [1] acquired from 64-beam LiDAR sensors, and (32, 1070) for nuScenes [2] with 32-beam sensors.

Diffusion model training. Conventional U-Net [3]-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.

Detailed algorithm of DAP. The Density-Aware Post-processing (DAP) algorithm is designed to refine 3D point cloud predictions by eliminating low-confidence points and adjusting scores based on spatial density and depth cues. It operates in three main stages: semantic filtering, clustering with score adjustment, and density-aware suppression.

Stage 1: Semantic Score-Based Point Filtering. Given a set of points $\mathcal{P} = \{(x_i, y_i, z_i)\}_{i=1}^{n} \in \mathbb{R}^3$ with corresponding semantic scores $\mathcal{S} = \{s_i\}_{i=1}^{n}$, DAP first filters out points with scores below a threshold τ_{score}. This results in a reduced set $\mathcal{P}_{\text{filtered}}$ that contains only high-confidence predictions.

Stage 2: Object Clustering and Score Adjustment. The filtered point set $\mathcal{P}_{\text{filtered}}$ is clustered using HDBSCAN, producing M object clusters $\{C_j\}_{j=1}^{M}$ and a set of outliers $\mathcal{P}_{\text{outlier}}$. Points in the outlier set are assumed to be false negatives and have their semantic scores boosted by a factor $\gamma > 1$, forming the adjusted score set $\mathcal{S}'$.

Stage 3: Density-Aware False Positive Suppression. Each cluster C_j is evaluated for potential false positives based on spatial density and depth consistency. The local density ρ_j of a cluster is estimated using the average k-nearest neighbor (kNN) distance of its points. Depth consistency is then measured by comparing the mean depth of the cluster to that of the entire filtered point set, resulting in a relative deviation term $\delta_{\text{rel},j}$. If $\delta_{\text{rel},j}$ falls below a suppression threshold τ_{density}, the cluster is considered anomalous and its scores are scaled down by the same factor γ.

The algorithm then outputs a refined set of points $\mathcal{P}'$ and corresponding semantic scores $\mathcal{S}'$, where outliers are enhanced and noisy or inconsistent predictions are suppressed. This improves the overall reliability of point-level predictions in noisy or cluttered environments.

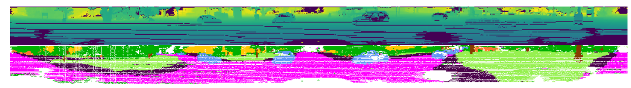

Figure 1. **Diffusion-generated semantic features.** Conditioned on the depth map shown on *Top* (as well as reflectance), our diffusion model generates the semantic features as shown on *Bottom*, demonstrating strong correspondence with the depth map.

Algorithm 1 Density-Aware Post-processing (DAP)

Input: Points $\mathcal{P} = \{(x_i, y_i, z_i) \mid i = 1, 2, \dots, n\} \in \mathbb{R}^3$, corresponding semantic scores $\mathcal{S} = \{s_i \mid i = 1, 2, \dots, n\}$
Output: Post-processed points $\mathcal{P}'$ w/ corresp. semantic scores $\mathcal{S}'$

```
// Stage 1. Semantic score-based point filtering
for (p, s) in zip(P, S) do
    P_filtered ← p  if  s > Percentile(S, 100 − τ_score)
end for
// Stage 2. Object clustering and score adjustment
{C_j}_{j=1}^M, P_outlier ← HDBSCAN(P_filtered)
S' ← γ · S,   for ∀p in P_outlier
// Stage 3. Density-aware false positive suppression
for j = 1 to M do
    // Compute density of kNN clusters
    d̄_{C_j} = (1/|C_j|) Σ_{p∈C_j} (1/k) Σ_{i=1}^k d(p, p^(i))
    ρ_j ← 1/d̄_{C_j}                                          ▷ ρ: density
    // Depth-based correction
    ρ_{rel,j} ← ρ_j · (Σ_{p∈C_j} δ_p / Σ_{p∈P_filtered} δ_p)^2     ▷ δ: depth
    if ρ_{rel,j} < τ_density then
        S'_j ← γ · S_j,  ∀p ∈ C_j
    end if
end for
```

B. Additional Experimental Results

In Fig. 2, 3, We present qualitative comparisons, showing that DOODLE effectively identifies out-of-distribution objects on the SemanticKITTI and nuScenes datasets.

C. Discussion

Our DOODLE utilizes a diffusion model to learn in-distribution semantic features. If the model has truly captured the underlying semantics, it should generate coherent semantic features under appropriate conditioning. In Fig. 1, we visualize features produced by our diffusion model conditioned on depth and reflectance; the generated features exhibit a high correspondence with the conditioning depth map, indicating that the model has effectively internalized the semantic distribution. As an add-on module, our approach incurs additional inference time; relative to the LiON baseline [4], runtime increases by 21.30%. To improve applicability to autonomous driving, where low latency is critical, future work will explore more efficient diffusion sampling to lower inference time while preserving the observed accuracy gains.

References

[1] 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. 1

[2] 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. 1

[3] 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. 1

[4] 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, 4

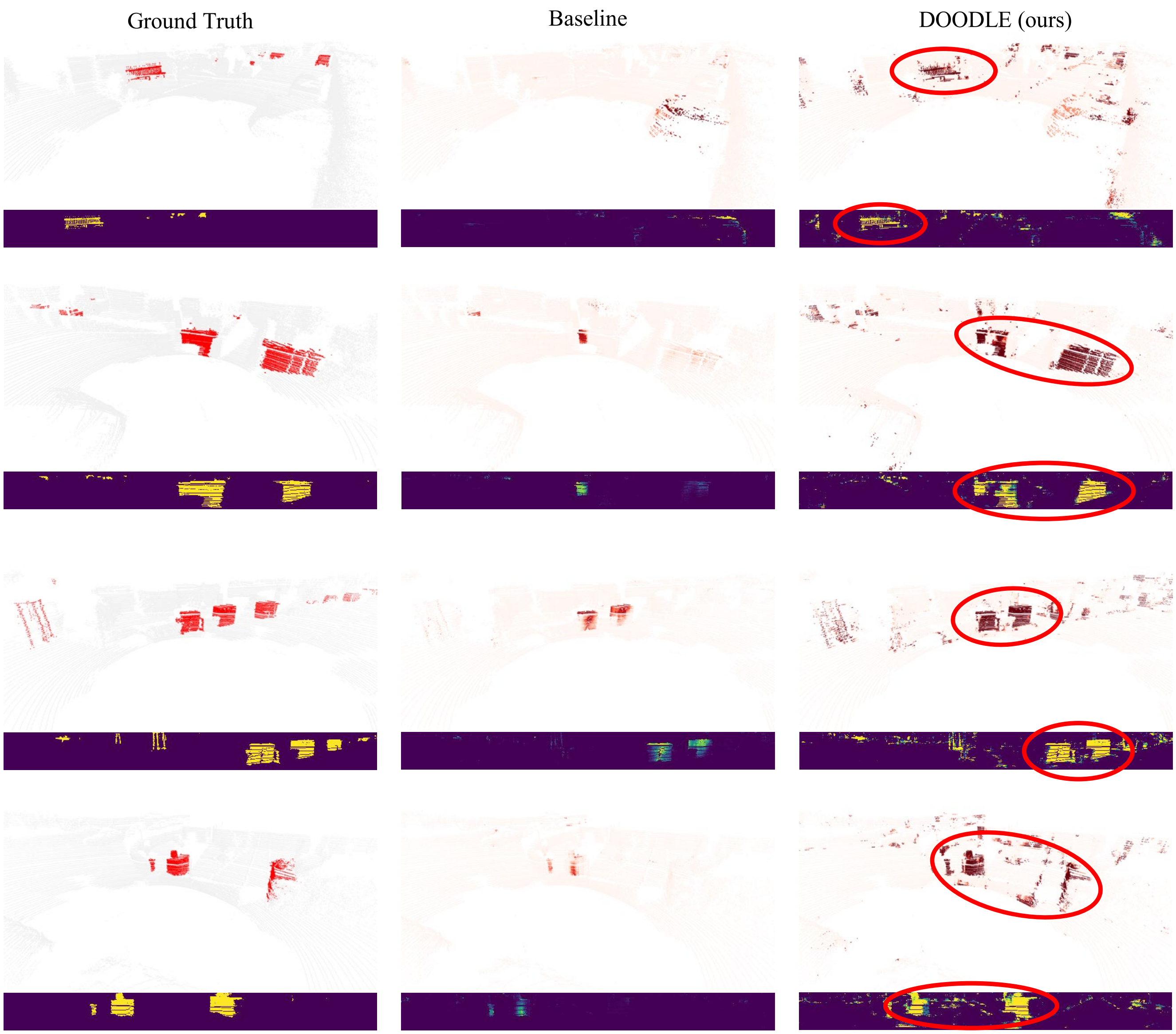

Figure 2. **Additional qualitative comparison.** We compare the OOD detection results on SemanticKITTI base model. Ours detect OOD class regions remarkably outperforms the baseline. *From top to bottom rows*: 3D LiDAR point clouds and 2D range-view projections. *From left to right columns*: Ground Truth (GT), LiON [4], and DOODLE.

Figure 3. **Additional qualitative comparison.** We compare the OOD detection results on nuScenes. Ours detect OOD class regions remarkably outperforms the baseline. *From top to bottom rows*: 3D LiDAR point clouds and 2D range-view projections. *From left to right columns*: Ground Truth (GT), LiON [4], and DOODLE.