

Exploiting Vision Language Model for Training-Free 3D Point Cloud OOD Detection via Graph Score Propagation

Tiankai Chen¹, Yushu Li^{2,3}, Adam Goodge³, Fei Teng¹, Xulei Yang³, Tianrui Li¹, Xun Xu^{†,3}

¹Southwest Jiaotong University, ²South China University of Technology

³Institute for infocomm research(I²R), A*STAR

Abstract

Out-of-distribution (OOD) detection in 3D point cloud data remains a challenge, particularly in applications where safe and robust perception is critical. While existing OOD detection methods have shown progress for 2D image data, extending these to 3D environments involves unique obstacles. This paper introduces a training-free framework that leverages Vision-Language Models (VLMs) for effective OOD detection in 3D point clouds. By constructing a graph based on class prototypes and testing data, we exploit the data manifold structure to enhancing the effectiveness of VLMs for 3D OOD detection. We propose a novel Graph Score Propagation (GSP) method that incorporates prompt clustering and self-training negative prompting to improve OOD scoring with VLM. Our method is also adaptable to few-shot scenarios, providing options for practical applications. We demonstrate that GSP consistently outperforms state-of-the-art methods across synthetic and real-world datasets for 3D point cloud OOD detection.

1. Introduction

The detection of out-of-distribution (OOD) samples in 3D point cloud data is critical for the safe and robust deployment of deep learning models in real-world environments. For example, identifying unexpected objects is crucial for ensuring the safety of personnel, vehicles, and robots by preventing collisions [18, 39, 41]. Existing approaches to OOD detection in 3D point clouds often focus on training separate models for specific tasks [54]. These methods typically rely on abundant in-distribution (ID) training data to build models based on reconstruction [34], distance metrics [10], or discriminative strategies [15]. While these methods achieve strong results, they depend on substantial ID training data, which is often challenging to collect in many real-world applications and generalization may be

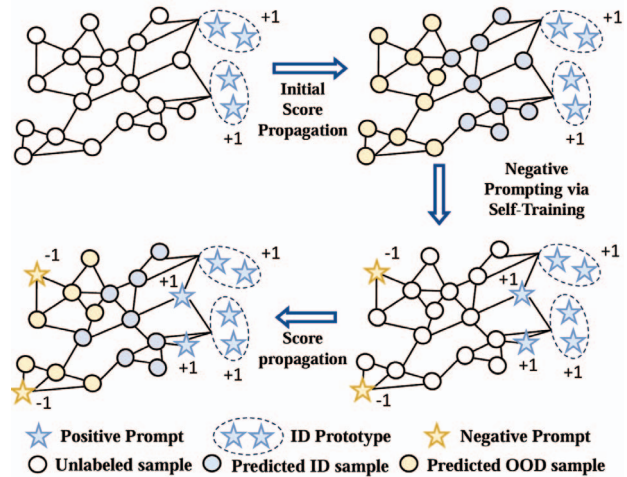


Figure 1. Exploiting the testing data manifold for zero-/few-shot OOD detection. ID prototypes propagate scores to test samples, with confident ones selected as pseudo negative prompts for refinement.

hindered by the limited diversity of ID training samples. These limitations highlight the need for alternative solutions that do not require extensive ID data collection.

Vision-language models (VLMs) have emerged as powerful foundation models that bridge visual features with textual understanding [23, 28, 42, 58]. VLMs are trained to associate the content of visual inputs with their textual labels and they demonstrate strong zero-shot generalization to unseen visual content [42]. This capacity makes VLMs appealing for OOD detection without retraining [5, 36, 37, 51]. A common VLM-based OOD detection pipeline maps ID class concepts, such as class names with various templates, into a shared feature space using the text encoder. The visual samples, similarly projected using the visual encoder, are then scored by comparing their distances to the ID class concepts in this feature space [36].

This VLM-based OOD detection pipeline has shown competitive performance for 2D image detection, primarily due to the vast understanding of 2D VLMs from the massive-scale image-text pre-training datasets (e.g. 400M

[†] Correspondence to Xun Xu: xu_xun@i2r.a-star.edu.sg, Code link: https://github.com/handsome999KK/GSP_OOD

pairs for pre-training CLIP [42]). In contrast, VLMs for 3D point clouds, such as the state-of-the-art ULIP2 [55], are typically pre-trained on much more restricted datasets (e.g., 800k point cloud samples). This restricts their ability to generalize across unseen categories and distributional shifts (e.g., style changes), which reduces the effectiveness of similarity-based methods for 3D point cloud OOD detection.

In this work, we propose to leverage the structure of the testing data manifold to compensate for this deficiency and improve VLM adaptation for downstream tasks [31, 44], as illustrated in Fig. 1. Specifically, we construct a graph based on class prototypes (e.g., text prompts for ID classes) and testing data features, where prototype nodes are assigned labels that propagate to neighboring testing nodes. Given the limited training data, we hypothesize that exploiting the testing data manifold enhances the generalization capabilities of 3D point cloud VLMs. Our method, Graph Score Propagation (GSP), assigns a positive OOD score (e.g., +1) to ID class prototypes while all testing samples start with a neutral score (0). As scores propagate along the graph, the resultant scores indicate OOD likelihood.

This straightforward adoption of score propagation significantly improves upon existing distance-based methods, and we introduce two additional enhancements to further boost OOD detection accuracy. First, while conventional training-free approaches typically use a single class prototype by averaging text prompts [44, 61], we argue that capturing the prompt distribution through clustering better leverages prompt diversity. Second, negative prompting has shown strong performance in VLM-based OOD detection [30, 38, 48]. However, simply negating the description (e.g. add “not”) is not very effective and learning negative prompts requires sophisticated optimization. Instead, we employ self-training to convert high-confidence OOD testing samples into pseudo-negative prompts and assign them a negative score (e.g., -1), for subsequent propagation iterations. Finally, our framework integrates seamlessly into traditional supervised OOD detection protocols by incorporating a few ID samples as positive prompts, further enhancing OOD detection under both zero-shot and few-shot protocols.

Our contributions are mainly summarized as follows.

- We employ a pre-trained vision-language model for 3D point cloud OOD detection and introduce Graph Score Propagation (GSP), a novel methodology that leverages the testing data manifold instead of solely relying on text-test feature distances.
- We enhance GSP with prompt clustering and self-training-based negative prompting. GSP is also compatible with both zero-shot and few-shot testing protocols.
- We validate the effectiveness of our method on synthetic and real-world 3D point cloud OOD detection bench-

marks, consistently achieving superior performance over state-of-the-art VLM-based OOD detection methods.

2. Related Work

Out-of-Distribution Detection: Traditional OOD detection methods focused on softmax-based probability scoring [15]. Recently, attention has shifted to novel evaluation metrics [26, 47, 57, 62], such as differential network channel processing [57, 62] and non-parametric nearest-neighbor distances [47]. Other strategies involve training specific OOD discriminators, like DAL [52], which constructs an OOD distribution set, and MOOD [29], which uses masked image modeling. NECO [2] utilizes neural collapse geometry for OOD identification. Vision-language models have introduced new methods using text prompts to distinguish between in-distribution (ID) and OOD data [5, 30, 37, 38, 51]. LoCoOP [37] leverages CLIP’s local features for OOD regularization, while Bai et al. [5] use CLIP to detect ID-like outliers, enhancing OOD detection and few-shot learning [59]. Training text encoders [51] and learnable prompt [30, 38] further increases the score gap between ID and OOD data. Recent research has expanded to include 3D datasets for point cloud OOD detection [1, 45]. Studies have shown that 3D OOD detection is particular important for autonomous driving scenarios [18, 27].

3D Vision-Language Models: Recent advancements in Vision-Language Models (VLMs) have facilitated innovative approaches for 3D understanding. Techniques utilizing CLIP [20, 35, 60] transform 3D point clouds into imagery data for extracting features aligned with text descriptions. More recently, methods like ULIP [54, 55] learn unified representations across images, text, and 3D point clouds, further improving zero-shot 3D classification. These approaches leverage the integration of text prompts and visual features which could better distinguish between in-distribution and out-of-distribution data in 3D contexts.

Label Propagation: Label propagation (LP) is a pivotal method in semi-supervised learning, leveraging limited labeled data to infer labels for unlabeled instances by exploiting data similarity across a graph [8, 14, 22, 40]. It assumes that labels vary smoothly across the graph, making it effective in transductive learning settings [22, 64]. Recent advancements integrate LP with graph neural networks to enhance its ability to separate ID from OOD data [6, 43, 53]. In the context of VLMs, LP employs text prompts as label representations, constructing graphs that propagate labels from text prototypes to test samples, thus improving zero-shot 2D classification [17, 31, 44]. Contrary to the existing attempts of combining LP with VLM, we tackle a one-class classification task. Only prototypes for in-distribution classes are available. We further propose a novel negative prompting technique via self-training and prompt clustering technique to better separate ID from OOD.

3. Methodology

3.1. Problem Formulation

We begin by formally defining the task of VLM-based out-of-distribution (OOD) detection in 3D point clouds. Let $z^f = f(x; \theta)$ and $z^t = g(t; \phi)$ denote the encoded features of point cloud data x and text t , where $f(\cdot; \theta)$ and $g(\cdot; \phi)$ are the VLM visual and text encoders respectively. The task involves the encoded features of unlabeled test data $\mathcal{D}_u = \{f(x_i)\}_{i=1 \dots N_u}$ and, optionally, the features and labels of few-shot labeled data $\mathcal{D}_l = \{f(x_j), y_j\}_{j=1 \dots N_l}$. The set of in-distribution (ID) classes $\mathcal{Y}_{in} = \{1, \dots, C_{in}\}$ is assumed to be known. Following established practices for VLM adaptation [36], we represent each ID class with N_t prompt templates $\mathcal{T} = \{t_j(y)\}_{N_t}$, e.g., “This is a point cloud of $\{y\}$.” or “A good point cloud of $\{y\}$ ”. These textual prompts are encoded by the text encoder, creating prototypes for ID classes $\mathcal{P}$. Existing works often create prototypes by averaging the encoded prompts for each class, resulting in a prototype z_c^p for the c -th class, and the full set of prototypes is denoted $\mathcal{P} = \{z_c^p\}_{c=1 \dots C_{in}}$. The OOD detection task is to determine whether a test sample $x_i \in \mathcal{D}_u$ belongs to an ID class, leveraging ID prototypes $\mathcal{P}$ and, optionally, the few-shot labeled data $\mathcal{D}_l$.

Preliminary of VLM: We briefly review the training objective of vision-language models (VLMs). The pre-training steps align the visual features with their associated text features [20, 54, 55], typically via a contrastive loss function:

$$\mathcal{L}_{vlm} = \sum_i \log \frac{\exp(z_i^f z_i^t / \tau)}{\sum_j \exp(z_i^f z_j^t / \tau)} \quad (1)$$

When a VLM is trained on a large amount of point cloud-text pairs, the cosine similarity of positive pairs (a point cloud sample and text prompt of its ground-truth class) is higher than that of negative pairs. Therefore, a pre-trained VLM provides a straightforward solution to 3D point cloud OOD detection by measuring the cosine similarity between textual descriptions of in-distribution classes and testing sample features. A high distance in the visual-text feature space suggests a greater likelihood of being OOD.

3.2. VLM for OOD Detection

Vanilla OOD Detection with VLM: Pre-trained VLMs offer an effective approach to OOD detection by measuring the distance between visual samples and text labels in the feature space. For example, [36] employs the maximum normalized cosine similarity as the OOD score. We define a normalized score S to indicate OOD-ness, where $S \rightarrow 0$ suggests OOD and $S \rightarrow 1$ suggests ID.

$$S = \max_c \frac{\exp(-d(z^f, z_c^p) / \tau)}{\sum_j \exp(-d(z^f, z_j^p) / \tau)} \quad (2)$$

It is assumed that in the feature space, OOD samples are further from the ID prototypes than ID samples. Subsequent research has further explored this approach by introducing techniques like negative prompts to better separate OOD from ID samples [24]. In this work, we are motivated by the challenge of measuring similarity/distance in the VLM feature space and propose to tackle the VLM OOD detection problem by exploiting the manifold of testing and, optionally, few-shot labeled data.

Manifold Distance for OOD Detection: The vanilla measurement overlooks the complex manifold of testing data. A straightforward approach to consider testing data manifold for OOD detection is to replace the standard distance metric with a manifold distance, represented via a graph.

Given a graph representing the data manifold $G = (E, V)$ with weight matrix $W \in \mathbb{R}^{(|\mathcal{P}|+|\mathcal{D}_l|+|\mathcal{D}_u|) \times (|\mathcal{P}|+|\mathcal{D}_l|+|\mathcal{D}_u|)}$, the manifold distance between two nodes on the graph is defined as the shortest path. The shortest path between arbitrary nodes can be obtained via Dijkstra’s algorithm [13]. More concretely, this can be formulated as finding a *path* between a prototype $z^p \in \mathcal{P}$ or a few-shot labeled sample $z_k^f \in \mathcal{D}_l$ and a testing sample $z_j^f \in \mathcal{D}_u$, consisting of edges $\tilde{E}_{ij} = \{e_k\}$ on the graph, such that the sum of these edge weights is minimized. The manifold distance between two nodes is thus defined as $d(z_j, z_i) = \min_{\tilde{E}_{ij}} \sum_{e_k \in \tilde{E}_{ij}} d(e_k)$, where $d(e_k)$ denotes the distance for edge e_k , e.g. the L2 distance between the encoded features.

Our empirical results in Tab. 3 reveal that incorporating the manifold improves the effectiveness of OOD detection. However, there are two main disadvantages associated with distance-based OOD detection. First, the manifold distance is unnormalised, if directly using the reciprocal of manifold distance as OOD score the result may be biased by the relative scale. Additionally, it is non-trivial to integrate negative prompts within the distance-based OOD detection paradigm.

3.3. VLM OOD Detection as Score Propagation

Instead of directly measuring the distance for OOD detection, we formulate OOD detection as a score propagation problem. Specifically, we assign a positive score (+1) to ID prompts $\mathcal{P}$, while all testing samples are initially assigned a neutral score (0). Labels are then propagated from the positive nodes to all testing samples in an iterative manner. We follow a standard propagation rule [63] as shown in Eq. 3,

$$\begin{aligned} S_t &= \tilde{W} S_{t-1} + \alpha S_0, \\ \text{s.t. } \tilde{W} &= D^{-\frac{1}{2}} W D^{-\frac{1}{2}}, \quad D = \text{diag}(\sum_j W_{ij}), \\ S_0 &= [\underbrace{1, \dots, 1}_{|\mathcal{P}|+|\mathcal{D}_l| \text{ elements}}, \underbrace{0, \dots, 0}_{|\mathcal{D}_u| \text{ elements}}], \end{aligned} \quad (3)$$

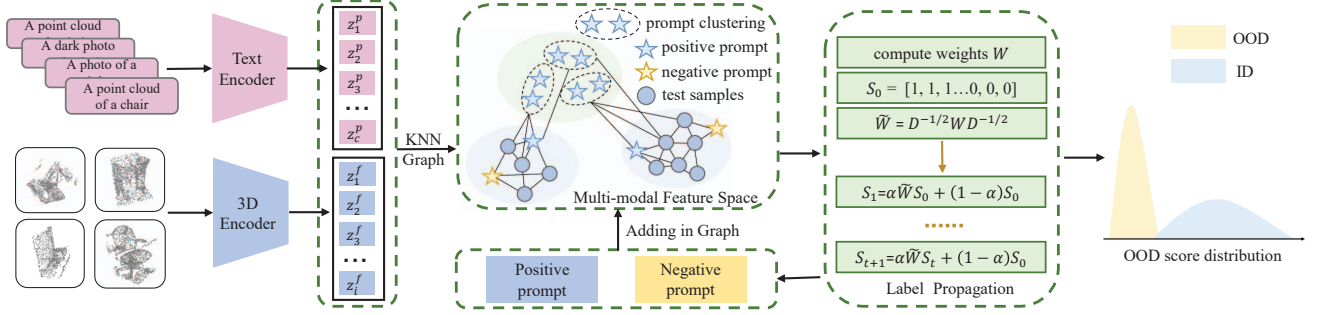


Figure 2. Overview of the proposed framework for zero-/few-shot OOD detection via graph score propagation. The framework encodes ID category prompts and dataset samples using frozen text and 3D encoders, constructing a KNN graph. Initial score propagation generates pseudo prompts, which refine the graph before final score propagation computes OOD scores.

where each step of propagation updates the score for unlabeled testing samples as a weighted average of scores from adjacent nodes and initial scores S_0 until T iterations are completed. Consequently, nodes that are closer to the prototypes on the manifold will obtain a higher score.

Prompt Clustering: The above propagation strategy propagates labels from ID prototypes to testing samples on the graph. Thus, it is critical to design effective ID prototypes that capture the diversity of textual descriptions for ID classes. Contrary to existing practices that average the text prompt features across all prompt templates, we propose modeling the distribution of prompt templates with multiple representative ones. Given the prompt templates $\mathcal{T}$, we first initialize a pool of text prompts $\mathcal{P}_c = \{z_{cj}^p = g(t_j(y_c))\}_{j=1 \dots |\mathcal{T}|}$ for the c -th class. We then introduce a clustering step to reduce this set to N_c representative prototypes $\tilde{\mathcal{P}}_c$. This is achieved by applying K-means [21] clustering on $\mathcal{P}_c$, and the cluster centers are adopted as the ID prototypes:

$$\begin{aligned} \tilde{\mathcal{P}}_c &= \arg \min_{\{\mu_k\}, \gamma} \sum_{N_c} \sum_{z_{cj}^p \in \mathcal{P}_c} \gamma_{jk} \|z_{cj}^p - \mu_k\|^2, \\ \text{s.t. } \gamma &\in \{0, 1\}^{|\mathcal{T}| \times N_c}, \sum_k \gamma_{jk} = 1, \end{aligned} \quad (4)$$

where μ_k are the cluster centers, and γ_{jk} indicates the assignment of each prompt feature z_{cj}^p to its closest center. The new prototypes, $\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_1 \cup \tilde{\mathcal{P}}_2 \cup \dots \cup \tilde{\mathcal{P}}_C$, are used for score propagation as positive nodes.

Negative Prompting via Self-Training: Negative prompts have been shown to substantially benefit OOD detection [30, 38, 48]. Incorporating negative prompts is particularly straightforward within the graph label propagation framework. We treat negative prompts as nodes $\mathcal{P}^n$ with a label of -1 on the graph. During propagation, the negative labels compete with the positive labels (ID prompts) to influence the score of testing samples. Existing approach exploits ID data to learn negative prompt [48] which prohibits zero-shot inference. To address this, we propose a

strategy that generates negative prompts via self-training, as illustrated in Fig. 1. Specifically, we first conduct initial score propagation for T iterations following Eq. 3, giving rise to initial propagated scores S_T . We select the top $m\%$ most confident scores from S_T following Eq. 5 where $\text{top}(S, m\%)$ indicates the top $m\%$ value of S .

$$\begin{aligned} \mathcal{I}_{pos} &= \{i \mid S_T[i] > \text{top}(S_T, m\%)\}, \\ \mathcal{I}_{neg} &= \{i \mid S_T[i] < \text{top}(S_T, 1 - m\%)\} \end{aligned} \quad (5)$$

We label the selected samples as positive and negative prompts respectively and re-initialize the score $\tilde{S}_0$. For negative prompts, we assign scores of -1 to further increase the gap between ID and OOD samples. The final score is obtained by running Eq. 3 with $\tilde{S}_0$ as initial scores.

$$\tilde{S}_0[i] = \begin{cases} S_0[i] & \text{if } i \notin \mathcal{I}_{pos} \text{ \& } i \notin \mathcal{I}_{neg}, \\ +1 & \text{if } i \in \mathcal{I}_{pos}, \\ -1 & \text{if } i \in \mathcal{I}_{neg}. \end{cases} \quad (6)$$

Graph Construction: We now introduce the approach for constructing a graph for score propagation. As observed by [31, 44], the distance between text prompts and visual samples is generally larger than the distance between visual samples. This observation also holds for 3D point cloud VLM because the 3D VLMs are often finetuned from 2D VLMs [54, 55]. Therefore, we adopt a similar approach to construct the graph. We write the weight matrix W in a blockwise form as shown in Eq. 7, where $W_u \in \mathbb{R}^{|\mathcal{D}_u| \times |\mathcal{D}_u|}$, $W_{up} \in \mathbb{R}^{|\mathcal{D}_u| \times |\tilde{\mathcal{P}}|}$ and $W_{ul} \in \mathbb{R}^{|\mathcal{D}_u| \times |\mathcal{D}_l|}$. No connections are created between prototypes and few-shot labeled samples. All blocks are further constructed as a k-nearest neighbor (KNN) graph following [31, 44]. It is also computationally efficient to adapt an existing graph to new testing samples by incrementally adding the testing samples to the graph [31].

$$W = \begin{bmatrix} \mathbf{I}^{|\tilde{\mathcal{P}}| \times |\tilde{\mathcal{P}}|} & \mathbf{0} & W_{up}^\top \\ \mathbf{0} & \mathbf{I}^{|\mathcal{D}_l| \times |\mathcal{D}_l|} & W_{ul}^\top \\ W_{up} & W_{ul} & W_u \end{bmatrix} \quad (7)$$

Overall Algorithm: We present the overall algorithm for unified iterative label propagation for vision-language models in Algorithm 1. In Fig. 2, we show an overview of the proposed framework for zero-/few-shot OOD detection via graph propagation.

Algorithm 1 Graph Score Propagation for OOD Detection

- 1: **Input:** training data $\mathcal{D}_l$, testing data $\mathcal{D}_u$ and ID class prompt templates $\mathcal{T}_c$
 - 2: **Output:** Final OOD scores $\tilde{S}_T$
 - 3: **for** $\mathcal{D}_l$ and $\mathcal{T}_c$, **do**
 - Encode: $z_l = f(x_l)$ for training data and $\mathcal{P}_c = \{g(t_j(y_c))\}$ for prompt templates.
 - Prompt clustering: positive prototypes $\tilde{\mathcal{P}}$ (Eq. 4)
 - 4: **end for**
 - 5: **for** $\mathcal{D}_u$, **do**
 - Encode: $z_i = f(x_i)$ for testing data
 - Build graph to compute W (Eq. 7)
 - Initialize score $S_0^{(i)}$ (Eq. 3)
 - Score propagate to compute S_T (Eq. 3)
 - Self-training to generate positive $\mathcal{I}_{pos}$ and negative prompts $\mathcal{I}_{neg}$ (Eq. 5)
 - Updated initial scores $\tilde{S}_0$ (Eq. 6)
 - Score propagate to compute $\tilde{S}_T$ (Eq. 3)
 - 6: **end for**
 - 7: **return** $\tilde{S}_T$
-

4. Experiments

4.1. Experimental setup

Dataset: In this work, we follow the 3DOS framework [1] to construct downstream datasets for evaluating our method. The 3DOS benchmark is based on two well-known 3D object datasets, **ShapeNetCore** [9] and **ScanObjectNN** [49]. ShapeNetCore includes 54 categories covering daily objects. ScanObjectNN is cropped out from ScanNet dataset featuring a real-world objects. It consists of 15 categories with 2,902 3D objects. To further validate our approach on real-world data, we introduce the **Sydney Urban Object Dataset** [11], **nuScenes** [7] and **S3DIS** [4]. Sydney Urban Object Dataset consists of LiDAR-scanned point clouds across 26 categories, totaling 631 unique instances. S3DIS dataset contains 6 large-scale indoor areas with 271 rooms annotated into 13 semantic categories. We cropped out individual objects using ground-truth segmentation masks, leading to 8931 objects. The nuScenes Dataset features point clouds captured by LiDAR sensor under autonomous driving scenario. We cropped out point clouds using ground-truth bounding boxes, leading to 2205

objects across 9 categories. Detailed visualization of each dataset is deferred to the supplementary.

Evaluation Protocol: Following the methodology outlined in 3DOS [1], we divide ShapeNetCore into three non-overlapping category sets, SN1, SN2, and SN3, referred to as the ShapeNetCore dataset. Each set contains 18 categories. For each OOD test scenario, one of these sets (SN1, SN2, or SN3) is treated as the known set of categories, while the remaining two sets are used as unknown categories, resulting in three distinct OOD test cases. Similarly, 3DOS defines three category sets, SR1, SR2, and SR3 on the ScanObjectNN dataset. Each SR-set includes 5 non-overlapping categories, with the other two sets acting as unknown categories in each scenario and call them as ScanObjectNN. As for Sydney Urban Object Dataset, we categorize these objects into two groups: movable objects, e.g. “truck” “pedestrian” and non-movable objects, e.g. “building” and “traffic sign”. Movable objects are treated as in-distribution (ID) classes, simulating a scenario in autonomous driving where object detection primarily focuses on movable objects. For consistency, we pre-process each point cloud sample by upsampling it to a uniform number of points. For S3DIS dataset, We retained instances containing more than 2048 original points, where each instance was randomly downsampled to retain exactly 2048 points. Then we set the fore-ground objects as ID and back-ground objects as OOD. For nuScenes dataset, we have conducted preliminary screening on the original point cloud and retained instances with more than 200 original points and upsampling it to 1024 points. Then we define Four-wheeled vehicles (including car, bus, truck, construction_vehicle, trailer) as ID and the rest as OOD. And further details provided in the supplementary material.

Implementation Details. We use the pre-trained ULIP/ULIP2 [54, 55] model to extract point cloud features and text features. For the 3D encoder, we select PointBERT [56]. In constructing the KNN graph, we set a universal K-value(N_k) of 10, connecting each text feature to its 10 nearest point cloud features. In Eq. 3, we set α to 0.5 and the number of iterations T to 5. The percentage of pseudo negative prompts, i.e. $m\%$ in Sect. 3.3, is also set to 5%. Additionally, we use K-means clustering(N_c) to group the prompts for each category into 3 clusters.

Competing Methods. We compare our method with several OOD detection approaches, including full-shot methods from 3DOS [1]. We also benchmark against recent zero-shot methods, such as DDCS [57], ZLaP[44], MCM [36], and NegLabel [24]. For zero-shot comparisons, all methods use the 3D encoder from [55]. Baseline results are also reported by directly using cosine similarity as the OOD score, with these methods referred to as ULIP and ULIP2, depending on the VLM used.

Evaluation Metrics. We evaluate OOD detection perfor-

	ScanObjNN								ShapeNetCore								
	SR3		SR2		SR1		Average		SN1		SN2		SN3		Average		
Method	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	
Requiring training dataset	MSP[15]	88.1	67.3	80.6	84.0	73.7	80.3	80.8	77.2	74.3	82.8	80.0	78.1	89.7	52.2	81.3	71.0
	MLS[50]	89.4	53.8	83.4	73.1	76.4	75.3	83.0	67.4	72.0	80.8	83.9	64.1	89.8	40.5	81.9	61.8
	ODIN[32]	90.2	47.9	83.3	71.7	76.3	76.8	83.3	65.5	74.2	79.4	79.4	71.7	87.8	41.8	80.5	64.3
	Energy[33]	89.5	50.6	81.6	75.8	76.6	75.5	82.6	67.3	72.1	81.2	83.0	64.7	89.8	39.4	82.0	61.8
	GradNorm[19]	88.5	50.7	77.4	75.3	75.2	76.8	80.4	67.6	72.1	81.8	57.7	88.9	57.8	79.0	62.6	83.3
	ReAct[46]	90.3	48.9	82.4	75.8	75.4	77.6	82.7	67.4	73.7	79.4	89.6	52.1	95.0	27.2	86.1	52.9
	NF	88.0	47.7	80.6	68.2	75.6	81.4	81.4	65.8	81.5	72.5	71.1	78.0	91.0	49.6	81.2	66.7
	OE+mixup[16]	72.6	83.5	72.0	88.5	62.5	87.8	69.0	86.6	72.7	78.9	80.3	68.8	87.3	62.2	80.1	69.9
	ARPL+CS[10]	-	-	-	-	-	-	-	-	74.8	80.3	80.7	72.4	85.4	50.8	80.3	67.8
	Cosine proto	91.0	41.0	82.1	78.2	77.6	75.6	83.6	64.9	80.3	68.3	88.7	60.8	91.9	38.0	86.9	55.7
	CE	85.1	64.4	78.9	83.9	73.2	79.1	79.1	75.8	83.4	66.8	89.5	37.7	92.9	28.1	88.6	44.2
	SupCon[25]	-	-	-	-	-	-	-	-	80.9	75.5	83.5	68.2	85.1	45.1	83.2	62.9
	SubArcface[12]	87.1	61.3	78.9	76.9	73.7	81.4	79.9	73.2	79.0	81.2	82.9	60.3	89.1	32.8	83.7	58.1
	GSP(1-shot)	79.3	61.3	80.9	84.1	88.6	51.7	82.9	65.7	91.3	44.0	86.4	43.1	83.8	91.4	87.2	59.5
	GSP(5-shot)	80.7	54.8	90.1	47.2	89.1	40.2	86.6	47.4	95.6	19.2	94.6	22.1	97.0	12.5	95.7	17.9
	GSP(full-shot)	90.1	33.0	93.0	30.5	92.2	34.4	91.8	22.6	97.9	10.1	97.9	7.9	98.8	6.0	98.2	8.0
Zero-shot	ULIP[54]	79.9	82.7	62.3	87.5	57.4	92.3	66.5	87.5	88.0	57.1	76.4	48.7	75.2	51.3	79.9	52.4
	ULIP2[55]	65.7	80.8	76.4	71.2	77.4	78.0	73.2	76.7	85.7	52.5	84.1	44.3	67.6	74.4	79.1	57.1
	DDCS[57]	63.3	90.5	62.4	80.0	64.4	96.2	63.4	88.9	61.1	86.7	66.8	86.5	77.2	76.2	68.4	83.1
	MCM[36]	52.0	92.3	62.3	91.3	81.3	80.4	65.2	88.0	85.1	51.6	83.2	46.1	66.4	75.8	78.2	57.8
	NegLabel[24]	65.7	85.5	66.8	91.3	67.4	88.2	66.6	88.3	60.6	87.8	80.6	78.6	88.0	48.3	76.4	71.6
	ZLaP[44]	70.1	81.7	67.0	84.7	67.8	83.6	68.3	83.3	88.2	52.8	72.3	66.4	77.4	90.0	79.3	69.7
	GSP	75.8	63.0	76.5	77.6	83.6	61.6	78.6	67.4	90.6	38.9	70.7	64.0	79.7	93.7	80.4	65.5

Table 1. Evaluation of OOD detection for ScanObjNN and ShapeNetCore datasets. Each column title for SR/SN indicates the chosen known class set and the rest sets of SR/SN are unknown.

mance using two widely adopted metrics: AUROC and FPR95. AUROC calculates the area under the receiver operating characteristic curve, with higher values indicating better performance. FPR95 measures the false positive rate at a true positive rate of 95%, with lower values reflecting improved performance.

4.2. Results on OOD Detection

Analysis of ScanObjNN Results: From the results presented in Tab. 1, we derive several key insights. **i)** Our proposed method demonstrates highly competitive performance under the zero-shot protocol, achieving an improvement of over 5% in AUROC compared to ULIP2, which uses cosine distance directly as an OOD score. This substantial enhancement suggests that employing a manifold-based approach significantly enhances the differentiation between OOD samples and ID classes. **ii)** Our few-shot results are equally impressive, rivaling even fully supervised methodologies. Notably, with 5-shot training samples, our method attains an average AUROC of 86.6%, surpassing the best fully supervised method, Cosine proto, which scores 83.6%, by a margin of 3%. Furthermore, when integrating all available training samples into the graph, our approach achieves an outstanding average AUROC of 91.8%. **iii)** A noticeable 5% gap in average AUROC is observed between our zero-shot and 1-shot methods. This gap underscores the critical role of incorporating real in-distribution samples in effectively distinguishing OOD samples. Consequently, practitioners are advised to introduce at least a few ID samples for OOD detection, in addition to leveraging Vision-Language Models (VLMs).

Analysis of ShapeNetCore Results: The results for ShapeNetCore, as shown in Tab. 1, lead us to several ob-

Method	Sydney		nuScenes		S3DIS	
	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓
ULIP[54]	63.4	81.7	53.2	93.3	53.1	95.1
ULIP2[55]	66.2	83.7	64.2	93.0	62.5	92.3
DDCS[57]	61.8	87.3	62.2	91.2	62.5	92.5
MCM[36]	65.7	82.4	57.1	82.3	65.2	88.1
NegLabel[24]	58.4	92.7	69.7	81.6	55.4	85.4
ZLaP[44]	58.5	80.6	70.0	87.9	71.3	88.4
GSP	72.5	68.9	80.8	80.7	77.6	94.0

Table 2. Zero-shot OOD detection results on the Sydney Urban Objects dataset, nuScenes dataset and S3DIS dataset.

servations. **i)** Unlike the findings from the ScanObjNN dataset, the performance gap between the leading baseline, ULIP, and our method under the zero-shot protocol is relatively narrow, approximately 0.5%. This modest improvement can be attributed to the substantial overlap between the downstream and pre-training datasets used in Vision-Language Models (VLMs). Specifically, ULIP/ULIP2 are pre-trained on ShapeNet, which shares significant commonalities with ShapeNetCore, thereby impacting OOD detection capabilities. Despite this, the slight enhancement achieved by our method underscores the utility of leveraging data manifolds to improve the differentiation of OOD samples. **ii)** A marked enhancement in performance is observed when few-shot samples are available. For instance, introducing 1-shot training samples results in nearly a 7% increase in average AUROC, while 5-shot samples yield an impressive 15% improvement. This highlights the critical role of few-shot learning in enhancing model performance.

Analysis of Additional Real-World Datasets: The zero-shot OOD detection results for the Sydney Urban Objects dataset [11] and nuScenes Dataset [7], presented in Tab. 2, reinforce our previous conclusions. Our method consistently outperforms all competing approaches in zero-shot OOD detection, demonstrating the efficacy of utilizing test-

ing data manifolds, particularly when there is a substantial shift between the downstream task and the pre-training data. This finding emphasizes the robustness of our approach in adapting to significant distributional shifts in real-world applications. For S3DIS [4], GSP achieves the highest AUROC, outperforming all baselines by a significant margin (e.g., +12.4% over MCM). This suggests superior ability to distinguish between classes across all decision thresholds. However, GSP’s FPR95 (94.0%) is slightly higher than ULIP (95.1%), indicating that while GSP excels in overall classification, it faces challenges in suppressing false positives when operating at a 95% true positive rate.

Further Insights Synergizing All Results: By comparing the outcomes from the ScanObjNN, Sydney Urban Objects, S3DIS, nuScenes and ShapeNetCore datasets, we derive several overarching insights. **i)** There is a notable correlation between the distribution shift from pre-training to downstream datasets and the efficacy of manifold exploitation for OOD detection. This can be attributed to several factors. Firstly, when pre-training is conducted on datasets that closely resemble the downstream datasets, the contrastive training loss effectively aligns text prompts with visual features, resulting in a well-defined separation between ID and OOD samples. Conversely, in scenarios where the distribution shift is more substantial, VLMs may struggle to accurately quantify the distance between ID text prompts and visual samples. In such cases, leveraging the testing data manifold becomes essential to enhance the calibration of prompt-to-visual feature distances, thereby improving OOD detection performance.

4.3. Ablation Study

We begin by analyzing the performance of three Out-of-Distribution (OOD) detection methods: Cosine Similarity, Shortest Path, and Label Propagation. As shown in Tab. 3, we employ the ULIP2 backbone for extracting text and point cloud features and evaluate these methods for OOD scoring. The approach of directly computing the maximal normalized cosine similarity between text and point cloud features yields a baseline performance. Substituting cosine similarity with the shortest path metric results in a performance increment, albeit modest. It is noteworthy that the shortest path is computed using Dijkstra’s algorithm, and the reciprocal of manifold distance serves as the OOD score. Lastly, we introduce a label propagation method to determine the OOD score. The experimental outcomes in Tab. 3 demonstrate that integrating graph-based information via label propagation significantly enhances the model’s ability to differentiate between In-Distribution (ID) and OOD categories. This suggests that graph-based methods provide a more nuanced understanding of the feature space, leading to improved OOD detection capabilities.

OOD Det. Mthd.	Prompt Clust.	Neg. Prompt	AUROC↑	FPR95↓
Cosine Dist.	-	-	73.2	76.7
Manifold Dist.	-	-	74.1	73.0
Score Prop.	-	-	76.4	72.6
Score Prop.	✓	-	77.2	73.6
Score Prop.	-	✓	77.3	71.4
Score Prop.	✓	✓	78.6	67.4

Table 3. Ablation study on ScanObjNN dataset.

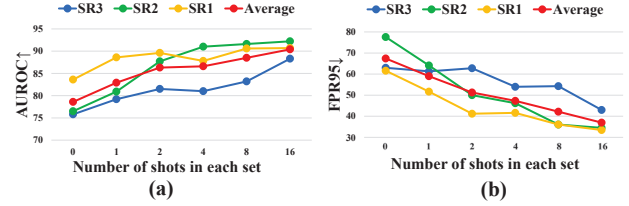


Figure 3. Few-shot OOD detection on the ScanObjNN dataset. (a) Results on AUROC metric. (b) Results on FPR95 metric.

4.4. Further Analysis

Few-shot OOD Detection: To evaluate the performance of label propagation in few-shot OOD detection, we incorporated a subset of training data from the 3DOS training dataset, treating them as in-distribution categories. The outcomes, presented in Fig. 3, show the effects of varying the number of training samples used in label propagation, with all samples randomly selected from the training dataset. The results indicate that as the number of training samples increases, OOD detection performance improves and eventually stabilizes.

Qualitative Analysis: To validate our method, we visualize sample predictions and scores from baseline method and our full GSP. As shown in Fig. 5, for easy-to-predict samples from the Real-world dataset, conventional distance-based methods perform well. However, in challenging cases (e.g., occlusions or missing parts), they often fail. In contrast, our method utilizes spatially proximate samples in feature space for score propagation, ensuring accurate OOD detection even in difficult scenarios. These qualitative results confirm the effectiveness and robustness of our approach.

Impact of Prompt Clustering: To illustrate the benefits of prompt expansion via K-means clustering, we examined how varying the number of prompts affects OOD (Out-of-Distribution) detection. As depicted in Fig. 4 (a), we experimented with different values of N_c in the K-means algorithm, where each N_c value corresponds to a distinct number of clusters, and consequently, a different number of prompts. An increase in N_c results in a greater number of distinct prompts.

The results presented in the figure reveal that the quantity of prompts considerably impacts OOD detection performance. Specifically, when $N_c = 6$, the AUROC surpasses 78.7%. However, further increasing N_c can negatively af-

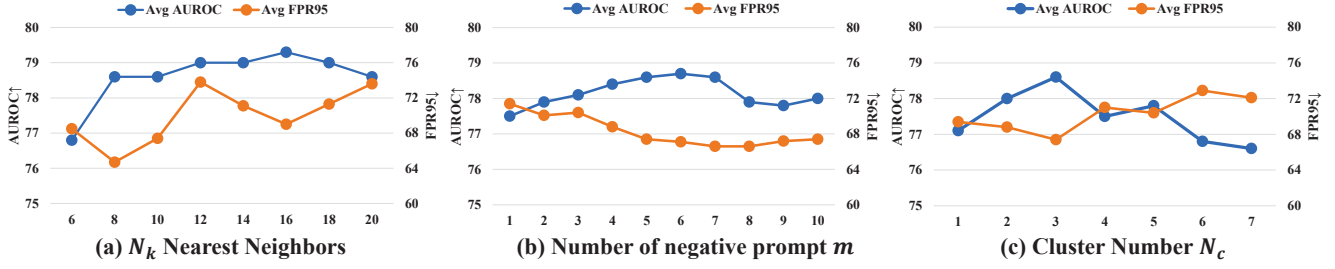


Figure 4. Results from ScanObjNN Benchmark Experiments: (a) Evaluation of prompt clustering with different K-means cluster numbers. (b) Impact of varying negative prompts on label propagation. (c) Results of different number of K Nearest Neighbors in KNN.

	ScanObjectNN Dataset		nuScenes Dataset	
	Easy-to-predict	Hard-to-predict	Easy-to-predict	Hard-to-predict
ID category				
Distance-based score:	0.618	0.093	0.689	0.192
GSP score:	0.465	0.357	0.374	0.278
OOD category				
Distance-based score:	0.174	0.396	0.148	0.431
GSP score:	0.072	0.054	0.074	0.127

Figure 5. Visualization of some samples for ScanObjectNN and nuScenes Dataset with two methods predict scores. All scores are normalized to 0-1.

fect the spatial alignment between text features and point cloud features. This phenomenon is likely due to the fact that typical prompts in Vision-Language Models (VLMs) are generally standardized, with only occasional outliers. Excessive clustering may isolate these outliers, leading to a decline in performance.

In conventional VLM approaches, the mean text feature vector is typically calculated for each category (equivalent to setting $N_c = 1$), achieving an average AUROC of less than 78%. The findings in Fig. 4 (c) suggest that selecting an appropriate N_c enhances the model’s ability to differentiate OOD categories, underscoring the advantages of prompt clustering.

Effectiveness of Negative Prompting via Self-Training: We conducted experiments to evaluate the effectiveness of the proposed negative prompting. Using the ShapeNetCore dataset, we explored the impact of self-training for negative prompting by varying the percentage of samples tagged with pseudo labels, denoted as $m\%$ as introduced in Section 3.3. In Fig. 4 (b), We observe a performance increase from $m = 1$ up to $m = 6$, indicating that incorporating negative prompts through self-training is an effective strategy to leverage the manifold for score propagation. However, performance declines beyond $m = 6$, likely due to the introduction of incorrect pseudo-negative prompts. This highlights the importance of mitigating the effects of confirmation bias [3] to prevent the model from being adversely

	Modelload + Datasetload	Text Feature Extract	Prompt Clustering	Feature Extract	Build Graph	Score Propagation	Total time
SR	1552ms	2212ms	1128ms	2780ms	106ms	8ms	7786ms
SN	1641ms	2459ms	1657ms	25031ms	671ms	34ms	31493ms

Table 4. The inference time of the GSP on SR and SN.

affected.

Impact of K-Nearest-Neighbor. We compare performance with different values of N_k Nearest Neighbors. As we can see in Fig 4 (a), the results demonstrate that as the number of N_k values increases, the OOD detection performance is tend to be stable. While $N_k = 16$ achieve the best performance and Less than 8 or larger than 18 get the low results.

Computational Cost: We profiled each step and test the inference time of GSP using a full GPU implementation in Tab. 4. Experiments are carried out on a single RTX3090 GPU and a i7-12700k CPU. As observed, the graph-based operations (Build Graph + Score Propagation) account for only around 2% of the total inference time. The additional computation cost introduced by graph operations is negligible compared with point cloud feature extraction.

5. Conclusion

In this work, we presented a novel training-free approach to 3D point cloud OOD detection that effectively leverages Vision-Language Models through Graph Score Propagation. By constructing a graph that incorporates both positive prototypes and pseudo-negative samples via self-training, our method captures the underlying structure of the testing data manifold. This approach allows for the propagation of informative OOD scores, thus surpassing existing distance-based techniques. Empirical results across multiple 3D point cloud datasets demonstrate the robustness and adaptability of our framework, particularly in zero- and few-shot settings. The GSP method addresses key limitations in prior OOD approaches by reducing dependency on extensive in-distribution training data while achieving high accuracy. Future work could explore further enhancements through dynamic graph updates, enabling real-time OOD detection in continuously evolving environments.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. U2468207), Sichuan Science and Technology Program (No. 2024NSFTD0036), and the Agency for Science, Technology and Research (A*STAR) under its MTC Programmatic Funds (No. M23L7b0021).

References

- [1] Antonio Alliegro, Francesco Cappio Borlino, and Tatiana Tommasi. 3dos: Towards 3d open set learning-benchmarking and understanding semantic novelty detection on point clouds. *Advances in Neural Information Processing Systems*, 2022. 2, 5
- [2] Mouin Ben Ammar, Nacim Belkhir, Sebastian Popescu, Antoine Manzanera, and Gianni Franchi. Neco: Neural collapse based out-of-distribution detection. *arXiv preprint arXiv:2310.06823*, 2023. 2
- [3] Eric Arazo, Diego Ortego, Paul Albert, Noel E O'Connor, and Kevin McGuinness. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *International joint conference on neural networks*, 2020. 8
- [4] I Armeni, O Sener, Ar Zamir, H Jiang, I Brilakis, M Fischer, S Savarese, and IEEE. 3d semantic parsing of large-scale indoor spaces. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016. 5, 7
- [5] Yichen Bai, Zongbo Han, Bing Cao, Xiaoheng Jiang, Qinghua Hu, and Changqing Zhang. Id-like prompt learning for few-shot out-of-distribution detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 1, 2
- [6] Tianyi Bao, Qitian Wu, Zetian Jiang, Yiting Chen, Jiawei Sun, and Junchi Yan. Graph out-of-distribution detection goes neighborhood shaping. In *International Conference on Machine Learning*. 2
- [7] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multi-modal dataset for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 5, 6
- [8] Tianle Cai, Ruiqi Gao, Jason Lee, and Qi Lei. A theory of label propagation for subpopulation shift. In *International Conference on Machine Learning*, 2021. 2
- [9] 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. 5
- [10] Guangyao Chen, Peixi Peng, Xiangqian Wang, and Yonghong Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 2021. 1, 6
- [11] Mark De Deuge, Alastair Quadros, Calvin Hung, and Bertrand Douillard. Unsupervised feature learning for classification of outdoor 3d scans. In *Australasian Conference on Robotics and Automation*, 2013. 5, 6
- [12] Jiankang Deng, Jia Guo, Tongliang Liu, Mingming Gong, and Stefanos Zafeiriou. Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In *European Conference on Computer Vision*, 2020. 6
- [13] Edsger W Dijkstra. A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: his life, work, and legacy*. 2022. 3
- [14] Shreyas Havaladar, Navodita Sharma, Shubhi Sareen, Karthikeyan Shanmugam, and Aravindan Raghuvier. Learning from label proportions: Bootstrapping supervised learners via belief propagation. *arXiv preprint arXiv:2310.08056*, 2023. 2
- [15] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2022. 1, 2, 6
- [16] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *International Conference on Learning Representations*, 2019. 6
- [17] Xuefeng Hu, Ke Zhang, Lu Xia, Albert Chen, Jiajia Luo, Yuyin Sun, Ken Wang, Nan Qiao, Xiao Zeng, Min Sun, et al. Reclip: Refine contrastive language image pre-training with source free domain adaptation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024. 2
- [18] Chengjie Huang, Vahdat Abdelzad, Christopher Gus Mannes, Luke Rowe, Benjamin Therien, Rick Salay, Krzysztof Czarnecki, et al. Out-of-distribution detection for lidar-based 3d object detection label propagation for zero-shot classification with visionion. In *IEEE International Conference on Intelligent Transportation Systems*, 2022. 1, 2
- [19] Rui Huang, Andrew Geng, and Yixuan Li. On the importance of gradients for detecting distributional shifts in the wild. *Advances in Neural Information Processing Systems*, 2021. 6
- [20] Tianyu Huang, Bowen Dong, Yunhan Yang, Xiaoshui Huang, Rynson WH Lau, Wanli Ouyang, and Wangmeng Zuo. Clip2point: Transfer clip to point cloud classification with image-depth pre-training. In *IEEE/CVF International Conference on Computer Vision*, 2023. 2, 3
- [21] Abiodun M. Ikotun, Absalom E. Ezugwu, Laith Abualigah, Belal Abuhaija, and Jia Heming. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information science*, 2023. 4
- [22] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. Label propagation for deep semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [23] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International Conference on Machine Learning*, 2021. 1
- [24] Xue Jiang, Feng Liu, Zhen Fang, Hong Chen, Tongliang Liu, Feng Zheng, and Bo Han. Negative label guided ood detection with pretrained vision-language models. In *International Conference on Learning Representations*, 2024. 3, 5, 6

- [25] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 2020. 6
- [26] Konstantin Kirchheim, Tim Gonschorek, and Frank Ortmeier. Out-of-distribution detection with logical reasoning. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024. 2
- [27] Michael Kösel, Marcel Schreiber, Michael Ulrich, Claudius Gläser, and Klaus C. J. Dietmayer. Revisiting out-of-distribution detection in lidar-based 3d object detection. *IEEE Intelligent Vehicles Symposium*, 2024. 2
- [28] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning*, 2022. 1
- [29] Jingyao Li, Pengguang Chen, Zexin He, Shaozuo Yu, Shu Liu, and Jiaya Jia. Rethinking out-of-distribution (ood) detection: Masked image modeling is all you need. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2
- [30] Tianqi Li, Guansong Pang, Xiao Bai, Wenjun Miao, and Jin Zheng. Learning transferable negative prompts for out-of-distribution detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2, 4
- [31] Yushu Li, Yongyi Su, Adam Goodge, Kui Jia, and Xun Xu. Efficient and context-aware label propagation for zero-/few-shot training-free adaptation of vision-language model. In *International Conference on Learning Representations*, 2025. 2, 4
- [32] Shiyu Liang, Yixuan Li, and R Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *International Conference on Learning Representations*, 2018. 6
- [33] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 2020. 6
- [34] Mana Masuda, Ryo Hachiuma, Ryo Fujii, Hideo Saito, and Yusuke Sekikawa. Toward unsupervised 3d point cloud anomaly detection using variational autoencoder. In *IEEE International Conference on Image Processing*, 2021. 1
- [35] Guofeng Mei, Luigi Riz, Yiming Wang, and Fabio Poiesi. Geometrically-driven aggregation for zero-shot 3d point cloud understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2
- [36] Yifei Ming, Ziyang Cai, Jiuxiang Gu, Yiyao Sun, Wei Li, and Yixuan Li. Delving into out-of-distribution detection with vision-language representations. In *Advances in Neural Information Processing Systems*, 2022. 1, 3, 5, 6
- [37] Atsuyuki Miyai, Qing Yu, Go Irie, and Kiyoharu Aizawa. Locoop: Few-shot out-of-distribution detection via prompt learning. *Advances in Neural Information Processing Systems*, 2024. 1, 2
- [38] Jun Nie, Yonggang Zhang, Zhen Fang, Tongliang Liu, Bo Han, and Xinmei Tian. Out-of-distribution detection with negative prompts. In *International Conference on Learning Representations*, 2024. 2, 4
- [39] Julia Nitsch, Masha Itkina, Ransalu Senanayake, Juan Nieto, Max Schmidt, Roland Siegwart, Mykel J Kochenderfer, and Cesar Cadena. Out-of-distribution detection for automotive perception. In *IEEE International Intelligent Transportation Systems Conference*, 2021. 1
- [40] Rattana Pukdee, Dylan Sam, Maria-Florina Balcan, and Pradeep Ravikumar. Label propagation with weak supervision. *arXiv preprint arXiv:2210.03594*, 2022. 2
- [41] Rui Qian, Xin Lai, and Xirong Li. 3d object detection for autonomous driving: A survey. *Pattern Recognition*, 2022. 1
- [42] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021. 1, 2
- [43] Maximilian Stadler, Bertrand Charpentier, Simon Geisler, Daniel Zügner, and Stephan Günnemann. Graph posterior network: Bayesian predictive uncertainty for node classification. *Advances in Neural Information Processing Systems*, 2021. 2
- [44] Vladan Stojnić, Yannis Kalantidis, and Giorgos Tolias. Label propagation for zero-shot classification with vision-language models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2, 4, 5, 6
- [45] Jiachen Sun, Qingzhao Zhang, Bhavya Kailkhura, Zhiding Yu, Chaowei Xiao, and Z Morley Mao. Benchmarking robustness of 3d point cloud recognition against common corruptions. *arXiv preprint arXiv:2201.12296*, 2022. 2
- [46] Yiyao Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *Advances in Neural Information Processing Systems*, 2021. 6
- [47] Yiyao Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In *International Conference on Machine Learning*, 2022. 2
- [48] Xinyu Tian, Shu Zou, Zhaoyuan Yang, and Jing Zhang. Argue: Attribute-guided prompt tuning for vision-language models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2, 4
- [49] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *IEEE/CVF International Conference on Computer Vision*, 2019. 5
- [50] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. In *International Conference on Learning Representations*, 2022. 6
- [51] Hualiang Wang, Yi Li, Huifeng Yao, and Xiaomeng Li. Clipn for zero-shot ood detection: Teaching clip to say no. In *IEEE/CVF International Conference on Computer Vision*, 2023. 1, 2
- [52] Qizhou Wang, Zhen Fang, Yonggang Zhang, Feng Liu, Yixuan Li, and Bo Han. Learning to augment distributions for out-of-distribution detection. *Advances in Neural Information Processing Systems*, 2023. 2

- [53] Qitian Wu, Yiting Chen, Chenxiao Yang, and Junchi Yan. Energy-based out-of-distribution detection for graph neural networks. *ArXiv*, 2023. [2](#)
- [54] Le Xue, Mingfei Gao, Chen Xing, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, and Silvio Savarese. Ulip: Learning a unified representation of language, images, and point clouds for 3d understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#)
- [55] Le Xue, Ning Yu, Shu Zhang, Artemis Panagopoulou, Junnan Li, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, et al. Ulip-2: Towards scalable multimodal pre-training for 3d understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. [2](#), [3](#), [4](#), [5](#), [6](#)
- [56] Xumin Yu, Lulu Tang, Yongming Rao, Tiejun Huang, Jie Zhou, and Jiwen Lu. Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. [5](#)
- [57] Yue Yuan, Rundong He, Yicong Dong, Zhongyi Han, and Yilong Yin. Discriminability-driven channel selection for out-of-distribution detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. [2](#), [5](#), [6](#)
- [58] Ji Zhang, Shihan Wu, Lianli Gao, Heng Tao Shen, and Jingkuan Song. Dept: Decoupled prompt tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12924–12933, 2024. [1](#)
- [59] Ji Zhang, Jingkuan Song, Lianli Gao, Nicu Sebe, and Heng Tao Shen. Reliable few-shot learning under dual noises. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. [2](#)
- [60] Renrui Zhang, Ziyu Guo, Wei Zhang, Kunchang Li, Xupeng Miao, Bin Cui, Yu Qiao, Peng Gao, and Hongsheng Li. Pointclip: Point cloud understanding by clip. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. [2](#)
- [61] Renrui Zhang, Wei Zhang, Rongyao Fang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-adapter: Training-free adaption of clip for few-shot classification. In *European Conference on Computer Vision*. Springer, 2022. [2](#)
- [62] Zihan Zhang and Xiang Xiang. Decoupling maxlogit for out-of-distribution detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. [2](#)
- [63] Xiaojin Zhu and Zoubin Ghahramani. Learning from labeled and unlabeled data with label propagation. *ProQuest number: information to all users*, 2002. [3](#)
- [64] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *International Conference on Machine Learning*, 2003. [2](#)