POINTSIMAD: TRAINING-FREE ZERO-SHOT 3D ANOMALY DETECTION VIA INTRINSIC SELF-SIMILARITY

Anonymous authors
Paper under double-blind review

ABSTRACT

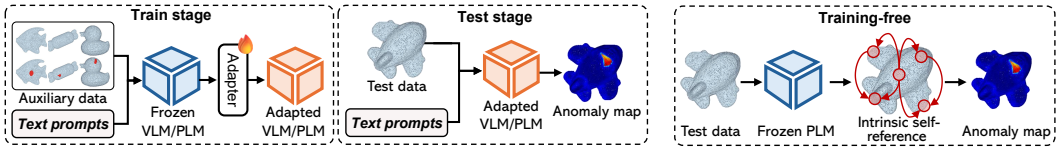
Zero-shot 3D anomaly detection (ZS-3DAD) aims to detect and localize anomalies in unseen object categories without requiring target-category training samples, making it particularly attractive for data-scarce industrial inspection. However, existing ZS-3DAD methods often rely on external auxiliary data to adapt pretrained models to the industrial domain, making their performance sensitive to the distribution of such data. In this paper, we propose PointSimAD, a training-free framework for ZS-3DAD that requires no fine-tuning with external industrial data and directly exploits the intrinsic structure of target point clouds. Specifically, we introduce a structural fingerprint that encodes local geometry, global shape, and spatial distribution into a sample-level descriptor. Based on this structural fingerprint, a structure-aware point grouping module adaptively determines the local structural granularity and partitions the point cloud into local point sets. A frozen 3D encoder then encodes the resulting local point sets into patch-level feature representations. By comparing each patch feature with similar features within the same object, we assess intra-object structural consistency and detect and localize anomalies based on the resulting local inconsistency. Experiments on three 3D anomaly detection benchmarks demonstrate that PointSimAD achieves strong and consistent performance, particularly for point-level anomaly localization, while requiring substantially lower memory and computation than existing adaptation-based approaches. The source code will be made publicly available upon acceptance.

1 INTRODUCTION

3D anomaly detection (3DAD) is a key technology for industrial quality inspection and is important for ensuring the safety, functionality, and reliability of products (Bergmann et al., 2019; Roth et al., 2022; Bergmann et al., 2021; Zhu et al., 2025). Most unsupervised methods (Chu et al., 2023; Horwitz & Hoshen, 2023; Zhao et al., 2025; Wang et al., 2025a; Cao et al., 2024; Li et al., 2024; Bergmann & Sattlegger, 2023) mainly rely on memory-bank retrieval or reconstruction paradigms, require a large number of training samples, and are difficult to generalize to unseen anomaly types. Zero-shot 3D anomaly detection (ZS-3DAD) performs anomaly detection and localization on unseen 3D objects without using training samples from the target category, thereby alleviating the difficulty of data acquisition in industrial scenarios as well as the generalization problems caused by unknown categories and anomaly patterns.

Existing ZS-3DAD methods can be broadly categorized into approaches based on vision-language model (VLM) (Zhou et al., 2024b; Deng et al., 2026a; Cheng et al., 2025; Xu et al., 2026) and point-language model (PLM) (Li et al., 2026). VLM-based methods render point clouds as two-dimensional images from multiple viewpoints and use a VLM (*e.g.*, CLIP (Radford et al., 2021)) to perform anomaly detection, whereas PLM-based methods directly use a pretrained PLM (*e.g.*, ULIP (Xue et al., 2023; 2024)) for anomaly detection. Although the two types of methods differ in their input forms and feature spaces, both rely on pretrained models to transfer anomaly detection capabilities to unseen target objects. Since these VLMs and PLMs are primarily pretrained on natural images or general-purpose 3D data, their representations exhibit a clear domain gap with respect to the fine-grained geometric defects and local anomalous structures in industrial point clouds.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Existing auxiliary-data-assisted approaches. (b) Our proposed training-free approach.

Figure 1: Comparison between auxiliary-data-assisted and training-free paradigms for ZS-3DAD. (a) Existing methods adapt pretrained VLM/PLM using external auxiliary data before inference on target objects. (b) Our training-free method employs a frozen pretrained PLM to extract point-cloud features and performs intrinsic self-reference within each target object, without external industrial data or task-specific fine-tuning.

To bridge this gap, existing VLM- and PLM-based methods (Zhou et al., 2024b; Deng et al., 2026a; Cheng et al., 2025; Li et al., 2026) generally need to use external auxiliary data for model adaptation, prompt learning, or anomaly-representation optimization, in order to transfer general representations to the industrial 3DAD scenario, as illustrated in Fig. 1(a). However, this adaptation process is strongly influenced by the distribution of the auxiliary data. The model can easily become biased toward structures and anomaly patterns similar to those in the auxiliary data, resulting in degraded detection performance when the target object or anomaly pattern differs substantially.

To address the above issues, we propose a new training-free ZS-3DAD framework, PointSimAD, which does not require fine-tuning on external industrial data and can directly perform anomaly detection on target point cloud samples, as shown in Fig. 1(b). Specifically, we introduce the concept of a "structural fingerprint," which encodes multiple dimensions of information, including the local geometry, global shape, and spatial distribution of a point cloud, into a sample-level structural descriptor, thereby comprehensively characterizing the intrinsic structural properties of the target point cloud. Based on this structural fingerprint, we further design a Structure-Aware Point Grouping Module, which maps the fingerprint to a local structural granularity adapted to the current object and uses this granularity to partition the point cloud into local point sets. We then use a 3D encoder to encode the resulting local point sets into a set of patch-level feature representations, thereby forming the local feature set of the current object. Finally, by establishing mutual references among patch features within the same object, we measure the consistency between each local feature and similar structural patterns and perform anomaly detection and localization according to the degree of structural deviation.

Our main contributions are summarized as follows:

- To the best of our knowledge, PointSimAD is the first training-free and reference-free ZS-3DAD framework to directly detect and localize anomalies from a single target point cloud via intra-object structural self-similarity.
- We introduce structural fingerprints that encode local geometry, global shape, and spatial distribution into sample-level descriptors, enabling the adaptive determination of object-specific local granularity for structure-aware local point-set partitioning.
- We propose an object-intrinsic patch modeling mechanism that encodes local point sets into patch-level features and compares each feature with its most similar intra-object features to characterize local structural consistency, enabling anomaly detection and localization.
- Extensive experiments on three 3D anomaly detection benchmarks demonstrate strong localization performance and high inference efficiency.

2 Related Work

2.1 Local Structural Modeling for Point Clouds

Point clouds are irregular and unordered, making local geometric structure modeling challenging. Early methods (Johnson & Hebert, 1999; Hanocka et al., 2019; Masci et al., 2015; Maturana & Scherer, 2015; Wu et al., 2015) typically transform point clouds into voxel grids or meshes, or use

handcrafted geometric descriptors to encode local shape information. PointNet (Qi et al., 2017a) directly extracts point-wise features using shared MLPs and symmetric aggregation, while PointNet++ (Qi et al., 2017b) captures hierarchical local structures through sampling, neighborhood grouping, and multi-scale aggregation. DGCNN (Wang et al., 2019) explicitly models point relations with dynamically constructed graphs, and Point Transformer (Zhao et al., 2021) employs attention mechanisms to aggregate neighboring features. More recent approaches, such as Point-BERT (Yu et al., 2022) and Point-MAE (Pang et al., 2022), learn transferable point-cloud representations from local patches through masked modeling, while Point Transformer V3 (Wu et al., 2024) improves the efficiency and receptive field of point-cloud encoding.

2.2 Zero-shot 3D Anomaly Detection

Zero-shot 3D anomaly detection aims to identify and localize anomalies in unseen object categories without using target-category training samples. Existing methods mainly exploit pretrained multimodal models and can be broadly divided into VLM-based and PLM-based approaches, while some point-cloud-native methods rely on auxiliary data for transferable representation learning. VLM-based methods, such as MVP-PCLIP (Cheng et al., 2025), PointAD (Zhou et al., 2024b), GS-CLIP (Deng et al., 2026b) and CoGeoAD (Xu et al., 2026), typically project point clouds into multi-view depth or RGB images and leverage CLIP-based visual and textual representations for anomaly detection. PLM-based methods were first explored by BTP (Li et al., 2026), which directly extracts point-cloud features from a pretrained PLM and aligns patch-level representations with textual anomaly semantics. In addition, 3DzAL (Wang et al., 2025b) learns transferable point-cloud representations by generating pseudo anomalies from task-irrelevant data and performing patch-level contrastive learning. Although these methods avoid target-category training, they generally depend on external semantic knowledge, auxiliary data, prompt learning, or task-specific representation training.

3 Method

3.1 Overview

In this paper, we propose PointSimAD, a training-free and reference-free framework for zero-shot 3D anomaly detection that exploits the intrinsic structural self-similarity within a single target point cloud. As illustrated in Fig. 2, PointSimAD first extracts a sample-level structural fingerprint from three complementary aspects, including local scale, global shape, and spatial distribution. The resulting structural fingerprint, together with the input point cloud, is fed into the Structure-Aware Point Grouping Module (SAPGM), which evaluates structural consistency across candidate granularities and determines an object-adaptive grouping granularity. A frozen 3D encoder then maps the resulting local point sets into patch-level features. To separate patch-specific variations from the shared structural content of the object, the Structural Consensus Decomposition Module suppresses the common components among patch features. Patch-wise cosine distances are subsequently computed on the decomposed features, and each patch is scored according to its average distance to the k most similar counterparts within the same object, yielding an intrinsic self-similarity anomaly score. Finally, these scores are propagated to the original points to produce the anomaly map, while their aggregation provides the object-level anomaly score. The entire pipeline operates without external industrial data, target-category normal references, or task-specific fine-tuning.

3.2 Structure-Aware Point Grouping

Point clouds from different objects may exhibit substantial variations in sampling density, geometric proportions, and spatial distributions. Consequently, a fixed grouping configuration can correspond to different effective receptive fields across objects, resulting in inconsistent characterization of local structures. To address this issue without introducing additional training, we propose the Structure-Aware Point Grouping Module (SAPGM), which characterizes the intrinsic geometry of each input point cloud and selects an object-adaptive grouping scale through deterministic structural consistency evaluation.

Multi-level Structural Fingerprint Extraction. Given an input point cloud $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$, we construct a compact structural fingerprint

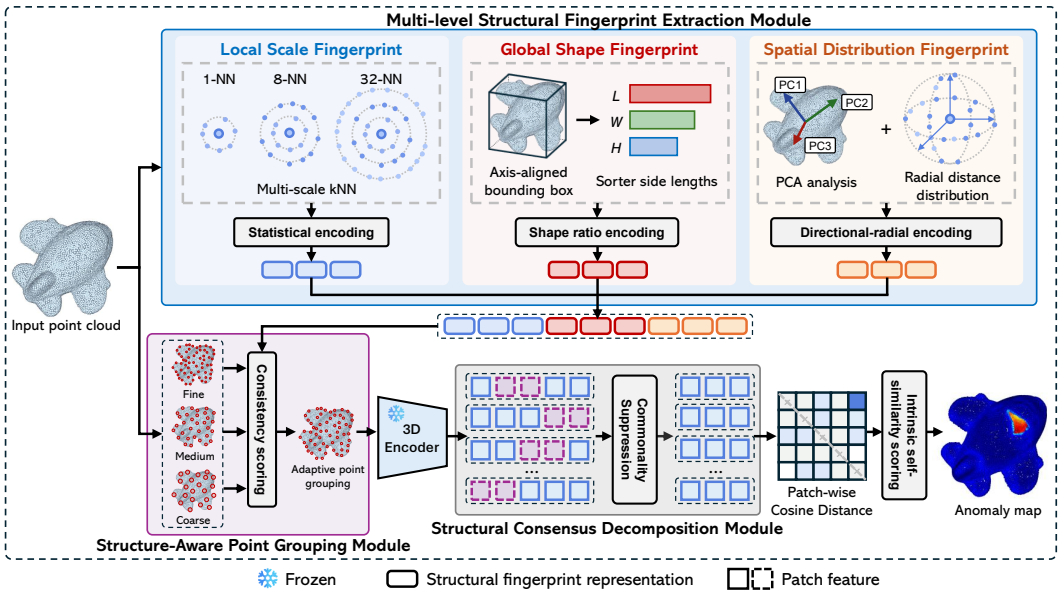


Figure 2: Overview of PointSimAD. MSFEM extracts multi-level structural fingerprints, SAPGM selects an object-adaptive grouping scale, and a frozen 3D encoder produces patch-level features. SCDM suppresses common structural components, followed by intrinsic self-similarity scoring for anomaly detection and localization.

$$\mathbf{s}^o = \Psi(\mathbf{X}_{\text{fp}}) = [\mathbf{s}_{\text{local}}, \mathbf{s}_{\text{shape}}, \mathbf{s}_{\text{dist}}]^\top \in \mathbb{R}^{12}, \tag{1}$$

where $\mathbf{X}_{\text{fp}}$ is a centered and unit-scale coordinate subset. The fingerprint contains three complementary groups: local-scale statistics describing sampling density and neighborhood support, global-shape statistics based on bounding-box and PCA geometry, and a spatial-distribution statistic describing radial variation. Their dimensions are 7, 4, and 1, respectively; the exact slot definitions are given in Appendix 7. All components are lower-bounded by a small positive constant before logarithmic comparison.

$$\begin{aligned} \mathbf{s}_{\text{local}} &= \Psi_{\text{local}}(\mathbf{X}_{\text{fp}}) \in \mathbb{R}^7, \\ \mathbf{s}_{\text{shape}} &= \Psi_{\text{shape}}(\mathbf{X}_{\text{fp}}) \in \mathbb{R}^4, \\ \mathbf{s}_{\text{dist}} &= \Psi_{\text{dist}}(\mathbf{X}_{\text{fp}}) \in \mathbb{R}^1. \end{aligned} \tag{2}$$

All three mappings are deterministic and use only point coordinates; no ULIP features, anomaly scores, or labels are involved in fingerprint construction.

Structure-Aware Configuration Selection. We use a small, predefined scale space $\mathcal{C}$ of candidate grouping configurations. For each candidate $c = (p_c, g_c) \in \mathcal{C}$, FPS selects the candidate center set $\mathcal{C}^c = \text{FPS}(\mathbf{X}_{16K}, p_c)$ and kNN forms the corresponding local groups. The candidate-induced structural response is summarized by

$$\mathbf{s}^c = \Phi(\mathbf{X}_{16K}; c) = [\mathbf{s}^c_{\text{local}}, \mathbf{s}^c_{\text{shape}}, \mathbf{s}^c_{\text{dist}}]^\top \in \mathbb{R}^{12}, \tag{3}$$

where the slots are aligned with $\mathbf{s}^o$ in Eq. 1; their exact coordinate-based definitions follow the same three groups and are provided in Appendix 7. We select the configuration with minimum logarithmic discrepancy:

$$\mathcal{L}_{\text{sc}}(c) = \sum_{d=1}^{D} w_d \left(\log \frac{s^c_d}{s^o_d} \right)^2, \qquad c^* = \arg\min_{c \in \mathcal{C}} \mathcal{L}_{\text{sc}}(c), \tag{4}$$

where $D = 12$ and $w_d = 1/12$ for all d. A lower cost indicates better agreement between the candidate-induced observation scale and the intrinsic geometry of the point cloud.

Finally, given the selected configuration c^*, we obtain the corresponding p^* patch centers using farthest point sampling (FPS) Eldar et al. (1997) and collect g^* neighboring points around each center via kNN. The resulting structure-adaptive local groups are subsequently fed into the frozen 3D encoder for patch-level feature extraction.

3.3 Intrinsic Self-Similarity Modeling

After structure-aware grouping, each local point group is independently fed into the frozen 3D encoder to obtain patch-level representations $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_p]^\top \in \mathbb{R}^{p \times d}$. Instead of relying on external normal samples, we exploit the structural redundancy within the test object itself to establish intrinsic references for anomaly detection. Specifically, we first suppress the common representation shared across local regions and then measure the remaining patch-wise self-similarity to identify structurally inconsistent regions.

Structural Consensus Decomposition. Patch representations from the same object generally contain a substantial amount of common information corresponding to its dominant structural characteristics, which may obscure subtle local deviations. We therefore estimate an object-level structural consensus using the feature-wise median,

$$\mathbf{m} = \mathrm{median}_{i=1,\ldots,p}(\mathbf{f}_i), \qquad \tilde{\mathbf{f}}_i = \mathbf{f}_i - \mathbf{m}. \tag{5}$$

This operation decomposes each patch representation into a shared object-level component and a local residual. Compared with directly using the original features, suppressing the common component emphasizes relative structural differences among local regions. Moreover, the median provides a robust estimate of the dominant representation and is less sensitive to a small number of potentially anomalous patches.

Intrinsic Self-Similarity Scoring. A normal local structure does not necessarily resemble all other regions of an object, but it is expected to find structurally similar counterparts within the same object. In contrast, an anomalous region is less likely to be supported by such intrinsic structural repetitions. Based on this observation, we compute the pairwise cosine distance between the residual patch features as

$$d_{ij} = 1 - \frac{\tilde{\mathbf{f}}_i^\top \tilde{\mathbf{f}}_j}{\|\tilde{\mathbf{f}}_i\|_2 \|\tilde{\mathbf{f}}_j\|_2}. \tag{6}$$

For each patch i, we select its k nearest counterparts with the smallest feature distances and define its anomaly score as the average distance to these intrinsic references,

$$r_i = \frac{1}{k} \sum_{j \in \mathcal{N}_k(i)} d_{ij}, \tag{7}$$

where $\mathcal{N}_k(i)$ denotes the k most similar patches excluding itself. A larger r_i indicates that the corresponding local region is difficult to explain by other structures within the same object and is therefore more likely to be anomalous.

Finally, the patch-level scores are spatially propagated to the original point cloud through distance-weighted interpolation to obtain the point-level anomaly map. For object-level detection, we average the highest-scoring 10% of patch responses, allowing localized structural inconsistencies to dominate the final anomaly score.

3.4 Anomaly Scoring and Localization

The patch-level anomaly responses are further converted into point-level and object-level predictions. For anomaly localization, we spatially propagate patch scores from their corresponding centers to the original point cloud using distance-aware interpolation, producing a dense point-level anomaly map. For object-level detection, high-response patches are aggregated to emphasize localized structural inconsistencies while suppressing the influence of predominantly normal regions. Detailed normalization, interpolation, and aggregation settings are provided in Sec. 4.2.

4 EXPERIMENTS

4.1 DATASETS AND EVALUATION METRICS

Real3D-AD. The Real3D-AD (Liu et al., 2023) dataset is a real-world industrial 3D anomaly detection dataset covering 12 object categories, with 1,254 samples including normal and defective instances.

Anomaly-ShapeNet. The Anomaly-ShapeNet (Li et al., 2024) dataset is a synthetic 3D anomaly detection dataset constructed from ShapeNetCoreV2 (Chang et al., 2015), covering 40 categories, with more than 1,600 samples including both normal and defective instances.

MulSen-AD. The MulSen-AD (Li et al., 2025) dataset is a high-resolution industrial anomaly detection dataset collected in a multi-sensor inspection scenario. In our experiments, we use its point-cloud modality only and do not use RGB or other sensor modalities to construct anomaly scores.

Evaluation Metrics. We evaluate both object-level detection and point-level localization. For object-level detection, we report Object-AUROC (O-AUROC) and Object-AP, which measure the ranking quality and average precision of the scalar anomaly score obtained by aggregating patch responses. For point-level localization, we report Point-AUROC (P-AUROC), which evaluates the ranking quality of the spatially propagated anomaly map against the point-wise anomaly mask.

4.2 IMPLEMENTATION DETAILS

Each input point cloud is downsampled to 16,384 points using farthest point sampling (FPS). For each grouping configuration, p patch centers are selected by FPS, with g neighboring points collected by kNN. We use the frozen point-cloud branch of ULIP-2 PointBERT (Xue et al., 2023; 2024) as the feature extractor. Following prior observations that shallow features better preserve fine-grained local geometry (Yue et al., 2026; Roth et al., 2022), we extract 384-dimensional patch features from the intermediate L2 layer. Patch features are median-centered, and anomaly scores are computed from cosine distances to the 4 most similar patches within the same object. The object-level score averages the top 10% patch responses, while point-level scores are obtained by Gaussian interpolation from the eight nearest patch centers. The structure-aware router computes the five-dimensional geometric fingerprint defined in Sec. 3.2 from a bounded coordinate subset and selects among three predefined grouping configurations: coarse $(p, g) = (512, 64)$, medium $(1024, 64)$, and fine $(2048, 64)$. All experiments are conducted on an NVIDIA RTX 6000 Ada GPU.

4.3 MAIN RESULTS

In this section, we evaluate PointSimAD on the Real3D-AD (Liu et al., 2023), Anomaly-ShapeNet (Li et al., 2024), and MulSen-AD (Li et al., 2025) datasets. We compare it with representative unsupervised methods, including BTF (Horwitz & Hoshen, 2023), M3DM (Wang et al., 2023), PatchCore (Roth et al., 2022), Reg3DAD (Liu et al., 2023), R3DAD (Zhou et al., 2024c), and IMR (Li et al., 2024), as well as VLM-based zero-shot methods, including PointCLIPV2 (Zhu et al., 2023), MVP-PCLIP (Cheng et al., 2025), AnomalyCLIP (Zhou et al., 2024a), PointAD (Zhou et al., 2024b), and GS-CLIP (Deng et al., 2026b). We also include the PLM-based zero-shot method BTP (Li et al., 2026).

On Real3D-AD, PointSimAD achieves 79.7% O-AUROC, 80.7% O-AP, and 86.4% P-AUROC, the best results among all compared methods. It exceeds the strongest zero-shot baselines by 3.3, 3.0, and 1.9 percentage points, respectively, and surpasses R3DAD by 6.3 percentage points in O-AUROC. On Anomaly-ShapeNet, PointSimAD obtains 79.4% O-AUROC, 84.5% O-AP, and 88.5% P-AUROC. GS-CLIP remains stronger at the object level (84.1% O-AUROC and 86.8% O-AP), whereas PointSimAD achieves the best P-AUROC, exceeding BTP and GS-CLIP by 1.2 and 13.3 percentage points, respectively. On MulSen-AD, PointSimAD reaches 72.4% O-AUROC and 81.5% P-AUROC. Although BTP obtains a higher O-AUROC of 74.5%, PointSimAD improves the P-AUROC by 7.0 percentage points and exceeds GS-CLIP by 18.6 percentage points. Overall, PointSimAD achieves the best point-level localization on all three datasets, while its strongest object-level result is obtained on Real3D-AD.

Table 1: Quantitative comparison with state-of-the-art methods on Real3D-AD, Anomaly-ShapeNet, and MulSen-AD. Methods are organized by type as unsupervised, VLM-based ZS, PLM-based ZS, and training-free ZS. Within the zero-shot category, the best results are highlighted in **bold**, and the second-best results are underlined. "–" indicates that the corresponding result is not available under the same benchmark or metric protocol. Results marked with * were reproduced using the official implementation.

Paradigm	Method	Venue	Real3D-AD			Anomaly-ShapeNet			MulSen-AD	
			O-AUROC	O-AP	P-AUROC	O-AUROC	O-AP	P-AUROC	O-AUROC	P-AUROC
Unsupervised	BTF (Horwitz & Hoshen, 2023)	CVPR'23	51.0	61.0	51.9	51.0	58.3	51.0	71.1	57.1
	$M3DM_{MAE}$ (Wang et al., 2023)	CVPR'23	54.6	54.7	59.8	55.1	62.0	65.1	62.8	58.7
	$M3DM_{BERT}$ (Wang et al., 2023)	CVPR'23	56.4	55.4	57.9	55.5	62.0	63.3	70.5	61.1
	$PatchCore_{MAE}$ (Roth et al., 2022)	CVPR'22	50.5	60.2	55.1	54.7	61.8	64.5	84.0	60.5
	$PatchCore_{BERT}$ (Roth et al., 2022)	CVPR'22	54.7	56.4	61.4	56.7	63.3	64.8	–	–
	Reg3DAD (Liu et al., 2023)	NeurIPS'23	69.0	70.9	69.5	52.7	60.4	63.2	74.9	64.1
	R3DAD (Zhou et al., 2024c)	ECCV'24	73.4	–	–	74.9	–	–	68.2	53.5
	IMR (Li et al., 2024)	CVPR'24	72.5	–	–	65.9	–	65.0	60.1	46.7
VLM-based ZS	PointCLIPV2 (Zhu et al., 2023)	ICCV'23	53.1	58.1	52.9	–	–	–	–	–
	MVP-PCLIP (Cheng et al., 2025)	T-SMC'25	74.9	75.3	75.9	78.7	83.8	70.4	–	–
	AnomalyCLIP (Zhou et al., 2024a)	ICLR'24	55.2	57.1	50.3	–	–	–	–	–
	PointAD* (Zhou et al., 2024b)	NeurIPS'24	74.8	76.9	73.5	82.6	85.6	74.1	64.6	60.5
	GS-CLIP* (Deng et al., 2026b)	CVPR'26	76.4	77.7	76.3	**84.1**	**86.8**	75.2	57.3	62.9
PLM-based ZS	BTP (Li et al., 2026)	CVPR'26	61.4	65.1	84.5	65.2	71.4	87.3	**74.5**	74.5
Training-free ZS	**PointSimAD (Ours)**	–	**79.7**	**80.7**	**86.4**	79.4	84.5	**88.5**	72.4	**81.5**

Table 2: Comparison of computational efficiency on Real3D-AD.

Method	Peak GPU memory (MB) ↓	GFLOPs ↓	FPS ↑
MVP-PCLIP (Cheng et al., 2025)	3,856	2,036.9	7.0
PointAD (Zhou et al., 2024b)	3,288	3,851.7	6.6
GS-CLIP (Deng et al., 2026b)	4,682	7,730.8	1.3
BTP (Li et al., 2026)	10,540	**52.5**	11.1
PointSimAD	**498**	122.6	**14.2**

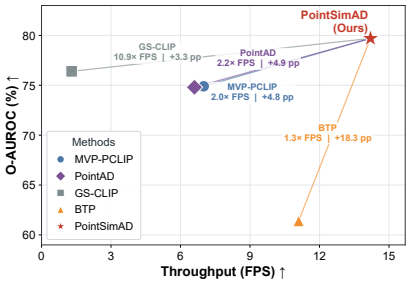


Figure 3: Efficiency–accuracy comparison on Real3D-AD.

Beyond detection accuracy, we further benchmark the computational efficiency of PointSimAD in Table 2 and visualize the accuracy–throughput trade-off in Fig. 3. Preprocessing time is excluded from all reported measurements, while the inference time of BTP and PointSimAD includes mapping anomaly scores from sampled points or patches back to the original point cloud. For a consistent comparison, both BTP and PointSimAD are evaluated with 16,384 input points. Under this setting, PointSimAD requires only 498 MB of peak GPU memory, with 122.6 GFLOPs and a throughput of 14.2 FPS. Compared with MVP-PCLIP, PointAD, and GS-CLIP, PointSimAD substantially reduces both memory consumption and computational cost while achieving higher inference throughput. Compared with BTP, PointSimAD reduces peak GPU memory from 10,540 MB to 498 MB and improves throughput from 11.1 to 14.2 FPS, although it incurs higher computational complexity in terms of GFLOPs.

As shown in Fig. 4, PointSimAD produces more compact and accurately localized anomaly responses across different object categories on both Real3D-AD and Anomaly-ShapeNet. Its responses are concentrated around the defective parts, with fewer scattered high-response points outside the anomalous regions. Compared with BTP, our method better aligns high-response regions with the ground-truth defects while suppressing irrelevant activations on normal structures. This reduced response dispersion is especially beneficial for identifying small or geometrically subtle defects, where diffuse activations can obscure the precise anomaly boundaries. These qualitative results further demonstrate the effectiveness of intrinsic self-similarity for fine-grained anomaly localization and help explain the stronger point-level performance of PointSimAD.

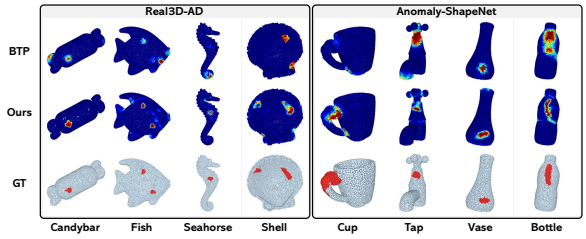


Figure 4: Qualitative anomaly localization results.

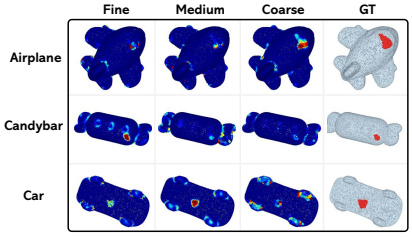


Figure 5: Visualization of grouping configurations.

Table 3: Ablation study of the main components of PointSimAD.

Variant	SAPGM	SCDM	Real3D-AD O-AUROC	Real3D-AD P-AUROC	Anomaly-ShapeNet O-AUROC	Anomaly-ShapeNet P-AUROC
Baseline			64.2	78.6	63.2	82.7
+SAPGM	✓		71.8	83.7	68.9	83.1
+SCDM		✓	75.0	83.8	74.7	87.4
Full	✓	✓	**79.7**	**86.4**	**79.4**	**88.5**

Table 4: Ablation study of different intrinsic reference strategies.

Reference Strategy	Real3D-AD O-AUROC	Real3D-AD P-AUROC	Anomaly-ShapeNet O-AUROC	Anomaly-ShapeNet P-AUROC
All patches	52.8	49.6	51.3	46.3
Random-k	40.8	50.1	46.6	49.5
Top-k similar (Ours)	**79.7**	**86.4**	**79.4**	**88.5**

4.4 Ablation Study

We analyze the impact of each module and various hyper parameters through a series of ablations.

Effectiveness of Key Components. We evaluate the contributions of SAPGM and SCDM by progressively incorporating them into the baseline framework. As shown in Table 3, introducing SAPGM consistently improves the baseline, increasing O-AUROC from 64.2% to 71.8% on Real3D-AD and from 63.2% to 68.9% on Anomaly-ShapeNet, demonstrating the benefit of adapting the local grouping granularity to the intrinsic structure of each object. SCDM brings more substantial gains by suppressing dominant common components and emphasizing local structural deviations, improving O-AUROC to 75.0% and 74.7% on the two datasets, respectively. Combining both modules further boosts the performance to 79.7%/86.4% O-AUROC/P-AUROC on Real3D-AD and 79.4%/88.5% on Anomaly-ShapeNet. These results demonstrate that structure-aware grouping and structural consensus decomposition provide complementary benefits for intrinsic self-similarity-based anomaly detection.

Table 5: Comparison between fixed and structure-adaptive grouping strategies.

Grouping Strategy	Real3D-AD O-AUROC	Real3D-AD P-AUROC	Anomaly-ShapeNet O-AUROC	Anomaly-ShapeNet P-AUROC
Fixed-Fine	76.5	82.1	72.3	84.6
Fixed-Medium	78.0	85.8	76.0	86.3
Fixed-Coarse	79.1	84.4	78.7	85.4
SAPGM	**79.7**	**86.4**	**79.4**	**88.5**

Table 6: Performance of the one-shot extension of PointSimAD.

Dataset	Setting	O-AUROC	P-AUROC
Real3D-AD	Zero-shot	79.7	86.4
	One-shot	**80.4 (+0.7)**	**88.6 (+2.2)**
Anomaly-ShapeNet	Zero-shot	79.4	88.5
	One-shot	**80.3 (+0.9)**	**90.2 (+1.7)**

Effect of Intrinsic Reference Strategy. We further investigate how the selection of intra-object references affects anomaly detection. As shown in Table 4, directly averaging the distances to all other patches yields only 52.82%/49.55% O-AUROC/P-AUROC on Real3D-AD, while randomly selecting k reference patches further degrades the object-level performance to 40.83%. Similar trends are observed on Anomaly-ShapeNet. In contrast, selecting the top-k most similar patches as intrinsic references substantially improves the performance to 79.69%/86.37% on Real3D-AD and 79.38%/88.53% on Anomaly-ShapeNet. These results indicate that effective intra-object self-reference should be established among structurally similar local regions rather than arbitrary patches, supporting our assumption that normal local structures can be explained by their most similar counterparts within the same object.

Effect of Structure-Adaptive Grouping. As shown in Table 5, the fine, medium, and coarse configurations correspond to $(p, g) = (2048, 64)$, $(1024, 64)$, and $(512, 64)$, respectively, and exhibit noticeable performance differences. Among the fixed strategies, the coarse configuration achieves the highest O-AUROC on both datasets, reaching 79.1% on Real3D-AD and 78.7% on Anomaly-ShapeNet, whereas the medium configuration yields the best P-AUROC, with 85.8% and 86.3%, respectively. In contrast, SAPGM achieves the best overall performance, reaching 79.7%/86.4% O-AUROC/P-AUROC on Real3D-AD and 79.4%/88.5% on Anomaly-ShapeNet. As illustrated in Fig. 5, different object categories favor different grouping configurations, demonstrating the benefit of object-adaptive grouping.

Effect of Feature Extraction Layer. We further investigate the influence of feature depth by extracting representations from different layers of the frozen ULIP-2 PointBERT encoder. As shown in Fig. 6, shallow features consistently exhibit stronger anomaly detection performance than deeper representations. In particular, L2 achieves the best O-AUROC and P-AUROC on Real3D-AD, while also obtaining the highest P-AUROC on Anomaly-ShapeNet. As the feature depth increases from L2 to L11, a clear performance degradation is observed on both datasets, especially for object-level detection. This trend supports our motivation that shallow representations better preserve fine-grained local geometric variations, whereas deeper layers increasingly emphasize high-level semantic abstraction and may suppress subtle structural differences that are critical for identifying local anomalies. Considering its overall performance across object- and point-level evaluation, we therefore adopt L2 as the default feature extraction layer.

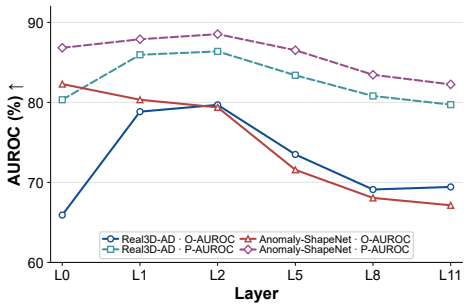


Figure 6: Layer-wise O-/P-AUROC on Real3D-AD and Anomaly-ShapeNet.

4.5 Extension to One-Shot Normal Reference

In practical industrial scenarios, zero-shot inference should not be treated as a strict constraint when limited normal data are available. We therefore extend PointSimAD to a one-shot setting by introducing a single normal sample as an additional reference, while preserving the training-free nature of the framework without any model training or fine-tuning. This provides a simple transition from fully reference-free inference to scenarios with limited normal data. As shown in Table 6, the one-shot extension consistently improves performance on both datasets. On Real3D-AD, O-AUROC and P-AUROC increase by 0.7 and 2.2 percentage points, respectively, while corresponding gains of 0.9 and 1.7 points are observed on Anomaly-ShapeNet. The larger improvements in point-level localization indicate that even a single normal reference can provide useful complementary structural information for identifying subtle local deviations. This further demonstrates the adaptability of PointSimAD to different levels of normal-data availability.

5 Conclusion

In this paper, we address ZS-3DAD under a training-free and reference-free setting. We propose PointSimAD, which exploits intrinsic structural self-similarity within each target point cloud without external industrial data or task-specific fine-tuning. Specifically, SAPGM adaptively determines object-specific grouping granularity, while SCDM suppresses common structural components to emphasize local deviations. Anomalies are then identified by comparing each patch with its most similar intra-object counterparts. Extensive experiments on Real3D-AD, Anomaly-ShapeNet, and MulSen-AD demonstrate the effectiveness, generalization, and efficiency of PointSimAD, particularly for point-level anomaly localization. Overall, PointSimAD provides a simple and practical alternative for ZS-3DAD when external references are unavailable, and offers a promising direction for exploiting intrinsic object structure in data-scarce scenarios, while reducing the reliance on task-specific adaptation and auxiliary industrial data.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, code refinement, and improving the readability of the manuscript. All research ideas, methodological designs, experiments, analyses, and conclusions were developed and verified by the authors. The authors reviewed all AI-assisted content and take full responsibility for the final manuscript.

REFERENCES

Paul Bergmann and David Sattlegger. Anomaly detection in 3d point clouds using deep geometric descriptors. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2612–2622. IEEE, 2023.

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9592–9600, 2019.

Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045*, 2021.

Yunkang Cao, Xiaohao Xu, and Weiming Shen. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition*, 156:110761, 2024.

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.

Yuqi Cheng, Yunkang Cao, Guoyang Xie, Zhichao Lu, and Weiming Shen. Toward zero-shot point cloud anomaly detection: A multiview projection framework. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025.

Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 6185–6194, 2023.

Zehao Deng, An Liu, and Yan Wang. Gs-clip: Zero-shot 3d anomaly detection by geometry-aware prompt and synergistic view representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35587–35596, June 2026a.

Zehao Deng, An Liu, and Yan Wang. Gs-clip: Zero-shot 3d anomaly detection by geometry-aware prompt and synergistic view representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 35587–35596, 2026b.

Yuval Eldar, Michael Lindenbaum, Moshe Porat, and Yehoshua Y Zeevi. The farthest point strategy for progressive image sampling. *IEEE transactions on image processing*, 6(9):1305–1315, 1997.

Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. Meshcnn: a network with an edge. *ACM Transactions on Graphics (ToG)*, 38(4):1–12, 2019.

Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2968–2977. IEEE, 2023.

Andrew E Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):433–449, 1999.

Kaiqiang Li, Gang Li, Mingle Zhou, Min Li, Delong Han, and Jin Wan. Back to point: Exploring point-language models for zero-shot 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14167–14177, 2026.

Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, and Yingna Wu. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22207–22216, 2024.

Wenqiao Li, Bozhong Zheng, Xiaohao Xu, Jinye Gan, Fading Lu, Xiang Li, Na Ni, Zheng Tian, Xiaonan Huang, Shenghua Gao, et al. Multi-sensor object anomaly detection: Unifying appearance, geometry, and internal properties. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 9984–9993, 2025.

Jiaqi Liu, Guoyang Xie, Ruitao Chen, Xinpeng Li, Jinbao Wang, Yong Liu, Chengjie Wang, and Feng Zheng. Real3d-ad: A dataset of point cloud anomaly detection. *Advances in Neural Information Processing Systems*, 36:30402–30415, 2023.

Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pp. 37–45, 2015.

Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 922–928. Ieee, 2015.

Yatian Pang, Wenxiao Wang, Francis EH Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pp. 604–621. Springer, 2022.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017a.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017b.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14318–14328, 2022.

Chengjie Wang, Haokun Zhu, Jinlong Peng, Yue Wang, Ran Yi, Yunsheng Wu, Lizhuang Ma, and Jiangning Zhang. M3dm-nr: Rgb-3d noisy-resistant industrial anomaly detection via multimodal denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025a.

Yizhou Wang, Kuan-Chuan Peng, and Yun Fu. Towards zero-shot 3d anomaly localization. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1447–1456. IEEE, 2025b.

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

Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8032–8041, 2023.

Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler, faster, stronger. In *CVPR*, 2024.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.

Ke Xu, Xinle Wang, Yanning Hou, Xueliang Ma, Juan Xie, and Jianfeng Qiu. Cogeoad: Hierarchical color-geometric fusion with multi-view attention for zero-shot 3d anomaly detection. *arXiv preprint arXiv:2606.25273*, 2026.

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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1179–1189, 2023.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27091–27101, 2024.

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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 19313–19322, 2022.

Yuanwen Yue, Damien Robert, Jianyuan Wang, Sunghwan Hong, Jan Dirk Wegner, Christian Rupprecht, and Konrad Schindler. Litept: Lighter yet stronger point transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24173–24184, June 2026.

Baozhu Zhao, Xiaohan Zhang, Jingfeng Guo, and Qi Liu. Pointcore: An efficient framework for unsupervised point cloud anomaly detection using joint local-global features. *Neural Networks*, pp. 108446, 2025.

Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 16259–16268, 2021.

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection. In *The Twelfth International Conference on Learning Representations*, 2024a. URL https://openreview.net/forum?id=buC4E91xZE.

Qihang Zhou, Jiangtao Yan, Shibo He, Wenchao Meng, and Jiming Chen. Pointad: Comprehending 3d anomalies from points and pixels for zero-shot 3d anomaly detection. *Advances in Neural Information Processing Systems*, 37:84866–84896, 2024b.

Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3d-ad: Reconstruction via diffusion for 3d anomaly detection. In *European conference on computer vision*, pp. 91–107. Springer, 2024c.

Wenbing Zhu, Lidong Wang, Ziqing Zhou, Chengjie Wang, Yurui Pan, Ruoyi Zhang, Zhuhao Chen, Linjie Cheng, Bin-Bin Gao, Jiangning Zhang, et al. Real-iad d3: A real-world 2d/pseudo-3d/3d dataset for industrial anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 15214–15223, 2025.

Xiangyang Zhu, Renrui Zhang, Bowei He, Ziyu Guo, Ziyao Zeng, Zipeng Qin, Shanghang Zhang, and Peng Gao. Pointclip v2: Prompting clip and gpt for powerful 3d open-world learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2639–2650, 2023.

A Experimental Details

We use the same training-free protocol for Real3D-AD, Anomaly-ShapeNet, and MulSen-AD. Each input point cloud is first downsampled to 16,384 points by farthest point sampling (FPS). The frozen point-cloud branch of ULIP-2 PointBERT extracts 384-dimensional patch features from its intermediate L2 layer. The structure-aware router uses the 12-dimensional geometric fingerprint defined in Eqs. 1–2, computed from at most 4,096 points, and selects one of three predefined grouping configurations.

Input and grouping. For each candidate configuration, FPS selects the patch centers and kNN collects the local neighborhoods. The router compares the candidate-induced structural response with the original geometric fingerprint and chooses the configuration with the minimum structural-consistency cost. The three candidates are coarse $(p, g) = (512, 64)$, medium $(1024, 64)$, and fine $(2048, 64)$, where p is the number of patch centers and g is the number of points in each local neighborhood.

Feature extraction and scoring. The ULIP-2 encoder remains frozen throughout inference. For each selected patch, we extract the intermediate L2 representation, median-center the patch features within the current object, and compute cosine distances between patch features. Each patch is scored by the average distance to its $k = 4$ most similar patches. The object-level score is the mean of the top 10% patch responses, and the point-level anomaly map is obtained by Gaussian interpolation from the eight nearest patch centers.

Evaluation protocol. All experiments use the point-cloud modality and do not train or fine-tune the feature extractor. No target-category normal samples, external industrial data, memory bank, text features, or anomaly labels are used to construct anomaly scores or select the grouping configuration. Object-level evaluation reports O-AUROC and O-AP, while point-level evaluation reports P-AUROC. Scores are computed over the official evaluation samples of each semantic category and then macro-averaged across categories; ground-truth labels are used only for the final metric computation. All experiments are conducted on an NVIDIA RTX 6000 Ada GPU.

Table 7: Definitions of the 12-dimensional geometric fingerprint. Here δ_r is the vector of r-th nearest-neighbor distances, $\ell_1 \geq \ell_2 \geq \ell_3$ are bounding-box side lengths, $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are PCA eigenvalues, and ρ_i is the distance to the object centroid. For a candidate, the corresponding center-set and local-group statistics are used.

Dim.	Component	Definition
1	Local scale	$\mathrm{med}(\delta_1)$
2	Local scale	$\mathrm{cv}(\delta_1)$
3	Local scale	$q_{0.9}(\delta_1)/\,\mathrm{med}(\delta_1)$
4	Local scale	$\mathrm{med}(\delta_8)$
5	Local scale	$\mathrm{med}(\delta_8)/\,\mathrm{med}(\delta_1)$
6	Local scale	$\mathrm{med}(\delta_{32})$
7	Local scale	$\mathrm{med}(\delta_{32})/\,\mathrm{med}(\delta_1)$
8	Global shape	ℓ_1/ℓ_2
9	Global shape	ℓ_1/ℓ_3
10	Global shape	$\lambda_2/(\lambda_1+\lambda_2+\lambda_3)$
11	Global shape	$\lambda_3/(\lambda_1+\lambda_2+\lambda_3)$
12	Spatial distribution	$\mathrm{cv}(\{\rho_i\})$

B Algorithm

Algorithm 1 summarizes the complete inference procedure. The structural fingerprint is used only to select the grouping scale. After the configuration is selected, anomaly evidence is obtained exclusively from the intrinsic self-similarity of patch features within the current object.

Algorithm 1 Training-free inference with PointSimAD.

Require: Input point cloud $\mathbf{X}$; frozen ULIP-2 PointBERT encoder E.
Ensure: Object score S_{obj} and point-level map S_{point}.

$\mathbf{X}_{16K} \leftarrow \text{FPS}(\mathbf{X}, 16384)$
$\mathbf{X}_{\text{fp}} \leftarrow \text{Subsample}(\mathbf{X}_{16K}, \min(4096, |\mathbf{X}_{16K}|))$
$\mathbf{s}^o \leftarrow \Psi(\mathbf{X}_{\text{fp}})$
$\mathcal{C} \leftarrow \{(512, 64), (1024, 64), (2048, 64)\}$
for all $c = (p_c, g_c) \in \mathcal{C}$ **do**
    $\mathbf{s}^c \leftarrow \Phi(\mathbf{X}_{16K}; c)$
    $L_{\text{sc}}(c) \leftarrow \sum_d w_d \left(\log \frac{s_d^c}{s_d^o}\right)^2$
end for
$c^* = (p^*, g^*) \leftarrow \arg\min_{c \in \mathcal{C}} L_{\text{sc}}(c)$
$\{\mathbf{c}_i\}_{i=1}^{p^*} \leftarrow \text{FPS}(\mathbf{X}_{16K}, p^*)$
$\mathbf{P}_i \leftarrow \text{kNN}(\mathbf{X}_{16K}, \mathbf{c}_i, g^*)$ for $i = 1, \ldots, p^*$
$\mathbf{f}_i \leftarrow E_{\text{L2}}(\mathbf{P}_i)$ for $i = 1, \ldots, p^*$
$\mathbf{m} \leftarrow \text{median}_{i=1,\ldots,p^*}(\mathbf{f}_i)$; $\widetilde{\mathbf{f}}_i \leftarrow \mathbf{f}_i - \mathbf{m}$
for $i = 1$ to p^* **do**
    $d_{ij} \leftarrow 1 - \dfrac{\widetilde{\mathbf{f}}_i^\top \widetilde{\mathbf{f}}_j}{\|\widetilde{\mathbf{f}}_i\|_2 \|\widetilde{\mathbf{f}}_j\|_2}$ for all $j \neq i$
    $\mathcal{N}_4(i) \leftarrow \text{TopKSmallest}(\{d_{ij}\}_{j \neq i}, 4)$
    $r_i \leftarrow \dfrac{1}{4} \sum_{j \in \mathcal{N}_4(i)} d_{ij}$
end for
$S_{\text{obj}} \leftarrow \text{mean}(\text{Top}_{10\%}\{r_i\}_{i=1}^{p^*})$
$S_{\text{point}} \leftarrow \text{GaussianInterpolate}(\{r_i, \mathbf{c}_i\}_{i=1}^{p^*}, \mathbf{X}, 8)$
return $S_{\text{obj}}, S_{\text{point}}$

C Per-category Results

We report the per-category results of PointSimAD on the three evaluation datasets. All metrics are shown as percentages. The final row reports the aggregate result provided in the corresponding experiment log. The Anomaly-ShapeNet results are arranged in two side-by-side blocks for compact presentation.

Table 8: Per-category results on Real3D-AD.

Category	O-AUROC (%)	O-AP (%)	P-AUROC (%)
airplane	77.4	81.1	82.7
candybar	82.0	84.9	95.8
car	73.2	78.4	85.7
chicken	80.4	76.7	83.5
diamond	99.8	99.8	96.2
duck	68.0	65.4	92.9
fish	93.1	95.2	90.5
gemstone	72.2	74.8	95.7
seahorse	67.5	71.8	65.2
shell	85.9	77.4	89.3
starfish	66.1	69.1	63.5
toffees	90.6	93.9	95.3
Mean	**79.7**	**80.7**	**86.4**

Table 9: Detailed per-category results on Anomaly-ShapeNet.

Object	P-AUROC	O-AUROC	O-AP
ashtray0	88.1	94.3	94.8
bag0	92.0	83.8	81.8
bottle0	92.2	92.9	94.5
bottle1	84.1	76.1	85.0
bottle3	91.7	88.9	93.8
bowl0	86.9	87.4	92.9
bowl1	92.7	76.3	83.5
bowl2	95.1	85.9	91.2
bowl3	91.4	82.6	88.6
bowl4	96.3	70.0	77.1
bowl5	90.6	75.1	81.9
bucket0	85.9	76.2	87.0
bucket1	90.8	90.5	94.1
cap0	86.4	96.7	98.0
cap3	96.7	83.5	90.1
cap4	97.1	82.8	89.1
cap5	95.0	92.3	95.0
cup0	91.8	85.2	86.1
cup1	90.1	77.6	78.1
eraser0	95.3	100.0	100.0
headset0	82.7	83.6	86.0
headset1	81.3	61.9	68.0
helmet0	90.4	74.2	86.7
helmet1	80.4	74.3	77.0
helmet2	96.0	89.6	94.6
helmet3	77.4	70.0	79.1
jar0	94.0	95.2	96.0
microphone0	90.8	98.1	98.0
shelf0	79.9	64.6	75.1
tap0	77.0	55.2	66.6
tap1	77.1	65.9	72.6
vase0	91.8	97.9	98.2
vase1	69.5	68.1	70.5
vase2	94.4	86.2	85.8
vase3	82.1	81.5	89.6
vase4	96.9	73.9	83.5
vase5	83.1	57.6	64.2
vase7	89.4	56.7	66.1
vase8	90.4	66.1	77.9
vase9	86.1	56.7	62.4
Mean	**88.5**	**79.4**	**84.5**

Table 10: Per-category results on MulSen-AD.

Category	O-AUROC (%)	O-AP (%)	P-AUROC (%)
button_cell	81.0	65.0	86.9
capsule	99.9	99.9	71.1
cotton	54.3	76.6	56.7
cube	68.4	82.4	91.0
flat_pad	94.6	97.0	90.1
light	62.7	60.7	84.5
nut	78.9	82.5	84.0
piggy	62.6	60.2	83.7
plastic_cylinder	61.6	72.6	95.5
screen	77.5	86.6	75.7
screw	80.2	82.1	66.8
solar_panel	46.2	23.2	92.2
spring_pad	61.8	74.4	73.9
toothbrush	74.0	80.7	86.8
zipper	81.9	87.9	84.1
Mean	**72.4**	**75.5**	**81.5**