A Semantically Disentangled Unified Model for Multi-category 3D Anomaly Detection

Supplementary Material

Contents

A. Implementation Details

Model setup. We adopt PointMAE [7], pretrained on ModelNet40 (8k), as the local feature extractor $E(\cdot)$. Each point cloud is partitioned into 1024 groups via FPS [8], and multi-resolution neighborhoods are formed with sizes $\{128, 256, 512\}$. Positional embeddings follow Point Transformer [10], where (x, y, z) coordinates are encoded by a two-layer MLP with ReLU. The shared transformer encoder in CFGT and the Geometry-Guided Decoder each contain 4 layers with 8 attention heads. Before entering the transformer encoder, feature jittering is applied with scale 20.0 and probability 1.0. The buffer in C3L is set to 64 entries.

Training and inference setup. We train the model using AdamW [5] with a base learning rate of 1×10^{-4} and a cosine annealing schedule. The batch size is 1 and training is performed for 1000 epochs. The loss weights in C3L are set to $\lambda_{\text{scl}} = 0.001$, $\lambda_{\text{cls}} = 0.001$, and $\lambda_{\text{cos}} = 0.01$. During inference, reconstruction discrepancies are smoothed with a 1D Gaussian filter with kernel size $k_g = 511$ and standard deviation $\sigma = 0.2$. All experiments are conducted with PyTorch 1.13.0 and CUDA 11.7 on a single NVIDIA GeForce RTX 3090 GPU.

B. Details of Geometry-Derived Bias

Given a point cloud $\mathbf{P} = \{\mathbf{p}_n\}_{n=1}^{N}$, Farthest Point Sampling [8] selects g group centers $\mathcal{S} = \{\mathbf{s}_i\}_{i=1}^{g}$. For each center $\mathbf{s}_i$, an adaptive radius r is estimated per instance, and its neighborhood is defined as $\mathcal{N}_i = \{\mathbf{p}_j \mid \|\mathbf{p}_j - \mathbf{s}_i\| \le r\}$.

Normal and curvature estimation. Let $\mathbf{C}_i$ denote the covariance matrix of centered neighbors:

$$\mathbf{C}_i = \text{Cov}\big(\mathcal{N}_i - \text{mean}(\mathcal{N}_i)\big). \tag{1}$$

Eigen-decomposition produces eigenvalues $\lambda_0 \le \lambda_1 \le \lambda_2$ and associated eigenvectors $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2$. The surface normal and curvature at the group center $\mathbf{s}_i$ are obtained as:

$$\mathbf{n}_i = \mathbf{e}_0, \qquad \kappa_i = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}. \tag{2}$$

Local geometric variations. Geometric irregularity is quantified through angular deviations of normals and curvature fluctuations within each neighborhood:

$$v_i^{\text{norm}} = \frac{1}{|\mathcal{N}_i|} \sum_{\mathbf{p}_j \in \mathcal{N}_i} \angle(\mathbf{n}_i, \mathbf{n}_j), \tag{3}$$

$$v_i^{\text{curv}} = \frac{1}{|\mathcal{N}_i|} \sum_{\mathbf{p}_j \in \mathcal{N}_i} |\kappa_i - \kappa_j|. \tag{4}$$

These two measures form a geometric descriptor:

$$\mathbf{v}_i = [v_i^{\text{norm}}, v_i^{\text{curv}}]. \tag{5}$$

After normalization, the descriptor is passed through a lightweight MLP to produce a scalar bias:

$$\mathbf{B}_{\text{geo}}(i) = \text{MLP}(\mathbf{v}_i). \tag{6}$$

The lightweight MLP is a two-layer MLP with ReLU.

C. More Quantitative Results

P-AUROC results. Tab. S9 and Tab. S10 present the full category-level results for Real3D-AD [4] and Anomaly-ShapeNet [2], respectively. While PO3AD [9] and ISMP [3] achieve strong performance on specific datasets—excelling on synthetic and real-world domains respectively—our method maintains consistently solid performance across both datasets under the unified multi-category setting. These results highlight the robustness and general applicability of our framework across diverse object categories and domains.

O-AUPR results on Anomaly-ShapeNet. We additionally report object-level AUPR on the Anomaly-ShapeNet dataset in Tab. S11. Our method achieves strong performance across categories, outperforming existing approaches. These results further confirm that the proposed semantically disentangled reconstruction contributes to reliable object-level anomaly detection under the unified multi-category setting.

Table S1. Comparison of computational cost.

Method	Parameters (M)	FLOPs (G)
MC3D-AD	34.16	439.21
SeDiR (Ours)	33.25	508.25

Computational cost. Tab. S1 compares the computational efficiency of our method with MC3D-AD [1]. Our model

uses fewer parameters (33.25M vs. 34.16M), indicating a slightly lighter architecture. Although the FLOPs are higher (508.25 vs. 439.21), this overhead mainly stems from the additional global tokenization and geometric bias modules. Overall, the model achieves improved performance while maintaining comparable computational complexity.

D. More Ablation Studies

To better understand how each design choice contributes to our framework, we conduct extensive ablation studies across several key components on the Real3D-AD dataset. These experiments provide a comprehensive analysis of how each factor influences semantic disentanglement, reconstruction quality, and overall anomaly detection performance.

Table S2. Ablation studies on the CFGT and C3L losses.

CFGT	$\mathcal{L}_{\text{scl}}$	$\mathcal{L}_{\text{cos}}$	$\mathcal{L}_{\text{cls}}$	O-AUROC (↑)	P-AUROC (↑)
	✓			78.8	79.0
		✓		77.9	78.7
	✓	✓		80.5	78.6
✓	✓			80.0	79.1
✓		✓		79.1	78.8
✓			✓	79.3	79.6
✓	✓	✓		80.5	79.8
✓		✓	✓	80.1	79.9
✓	✓		✓	80.0	79.3
✓	✓	✓	✓	**81.0**	**80.6**

Ablation on the CFGT and C3L losses. Tab. S2 reports the effect of enabling Coarse-to-Fine Global Tokenization (CFGT) module and the individual loss terms in C3L. Without CFGT, no global token is formed; the model instead applies supervised contrastive learning to a globally averaged feature, resulting in reasonable yet suboptimal performance. Enabling CFGT introduces a unified global token derived from multi-resolution features, which substantially strengthens category-conditioned contrastive learning. Adding the cosine alignment and auxiliary classification terms further improves performance, with the full configuration achieving the highest O-AUROC and P-AUROC. These results suggest that explicitly constructing a global token and supervising it with all C3L objectives yields the most discriminative and semantically disentangled representation.

Impact of coarse–fine neighborhood ratios. Tab. S3 reports the performance of different neighborhood ratio configurations used in the CFGT. We evaluate several coarse–fine combinations based on the base resolution k, including $\{k, 4k, k/4\}$, $\{k, 3k, k/3\}$, $\{k, 1.5k, 2k/3\}$, and $\{k, 2k, k/2\}$. Among these settings, the configuration $\{k, 2k, k/2\}$ achieves the best results on both O-AUROC and P-AUROC. Across the evaluated ratios, settings with a more balanced fine–base–coarse distribution—such as $\{k, 2k, k/2\}$—tend to show more stable performance. This observation indicates that the choice of multi-resolution neighborhood arrangement can influence the effectiveness of global representation learning within CFGT.

Table S3. Comparison of different coarse-to-fine neighborhood ratios.

Ratios	O-AUROC (↑)	P-AUROC (↑)
{k, 4k, k/4}	80.3	79.4
{k, 3k, k/3}	80.9	79.8
{k, 1.5k, 2k/3}	79.9	79.8
{k, 2k, k/2}	**81.0**	**80.6**

Table S4. Comparison of different base neighborhood sizes k.

Size	O-AUROC (↑)	P-AUROC (↑)
{256, 512, 1024}	77.5	78.5
{64, 128, 256}	79.9	79.6
{128, 256, 512}	**81.0**	**80.6**

Effect of neighborhood sizes. Tab. S4 evaluates different neighborhood sizes used in the CFGT module. Three configurations with various *base* sizes are compared, among which the setting with a base size of 256 ($\{128, 256, 512\}$) achieves the highest O-AUROC and P-AUROC. The results show that neighborhood size introduces performance gap: larger neighborhoods tend to smooth out fine geometric cues, whereas smaller ones may lack sufficient structural context. A base size of 256 provides the most balanced fine–base–coarse composition, enabling stable multi-resolution feature extraction and yielding the most reliable performance among the evaluated choices.

Table S5. Comparison of different numbers of groups.

N	O-AUROC (↑)	P-AUROC (↑)
512	75.7	77.4
2048	77.7	78.4
1024	**81.0**	**80.6**

Effect of the number of groups. Tab. S5 evaluates the impact of varying the number of groups used to partition the point cloud. Among the three tested settings—512, 1024, and 2048 groups—the configuration with 1024 groups yields the best performance. A clear performance gap emerges due to the trade-off in spatial granularity. Using too few groups (e.g., 512) produces large spatial regions, causing each token to aggregate overly coarse geometry and reducing sensitivity to fine anomalies. In contrast, using too many groups (e.g., 2048) results in highly fragmented neighborhoods that capture limited local structure and introduce noise, weakening multi-resolution feature extraction.

The configuration with 1024 groups achieves a balanced decomposition, providing sufficient geometric detail while maintaining stable local neighborhoods. This result highlights the importance of selecting an appropriate grouping resolution for robust multi-resolution representation learning.

Table S6. Comparison of different buffer sizes in C3L.

L	O-AUROC (↑)	P-AUROC (↑)
32	79.6	79.8
128	80.3	80.3
10,000	80.5	79.2
64	**81.0**	**80.6**

Effect of buffer size in C3L. Tab. S6 examines the influence of buffer size in C3L. Reducing the buffer from 64 to 32 slightly degrades both O-AUROC and P-AUROC, implying that too few contrastive samples hinder stable supervised contrastive learning. Increasing the buffer (e.g., to 128 or 10,000) yields performance similar to that of size 64, suggesting diminishing returns once a moderate number of samples is available; excessively large buffers may even introduce stale embeddings that are misaligned with the current representation space. Overall, a buffer size of 64 provides the best balance between stability and efficiency, yielding the highest performance among the evaluated configurations.

Table S7. Comparison of different loss weights in C3L.

λ_{scl}	λ_{cos}	λ_{cls}	O-AUROC (↑)	P-AUROC (↑)
0.01	0.001	0.01	79.3	79.6
0.001	0.01	0.01	78.6	78.3
0.001	0.001	0.1	77.0	78.9
0.001	**0.001**	**0.01**	**81.0**	**80.6**

Effect of loss weights in C3L. Tab. S7 examines the sensitivity of C3L to different loss-weight configurations, obtained by scaling each term by a factor of 10. Overweighting any single component disrupts the balance among the supervised contrastive, cosine alignment, and auxiliary classification losses, leading to noticeable drops in both O-AUROC and P-AUROC. The weights used in the main experiments (λ_{scl}=0.001, λ_{cos}=0.001, λ_{cls}=0.01) achieve the best performance, indicating that C3L benefits from a well-balanced combination of objectives.

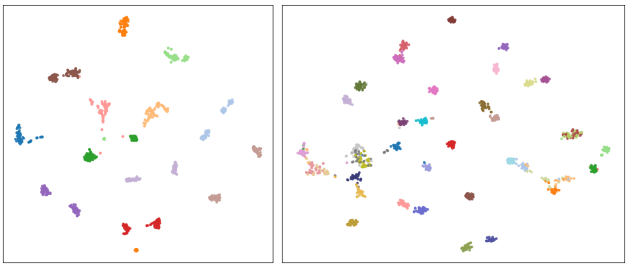
Table S8. Category prediction accuracy with different global representations.

Method	Accuracy (%)
Mean	78.1
Max	77.3
ACT	**97.3**

Effect of ACT for category representation. We evaluate the effectiveness of the proposed Adaptive Context Token (ACT) as a global summary of category information. After freezing the pretrained model, we train an additional category prediction head for 100 epochs using three different global representations: mean pooling, max pooling, and ACT. As shown in Tab. S8, ACT achieves substantially higher classification accuracy, indicating that it captures category-discriminative semantics far more effectively than conventional pooling-based global tokens. This validates the use of ACT as the primary representation for category-aware learning in our framework.

E. More Qualitative Results

Qualitative comparison with MC3D-AD on Anomaly-ShapeNet. Fig. S2 presents additional qualitative comparisons with MC3D-AD on the Anomaly-ShapeNet dataset. While MC3D-AD often yields incomplete or noisy anomaly responses, our method more distinctly highlights anomaly regions and produces more coherent anomaly maps. These results further verify the stability and effectiveness of our method across diverse geometric patterns.



(a) (b)

Figure S1. t-SNE visualization of the learned global tokens on (a) Real3D-AD (12 categories) and (b) Anomaly-ShapeNet (40 categories). The embeddings form well-structured clusters, indicating that the proposed method learns coherent and category-aware global representations across both datasets.

t-SNE visualization of global tokens. Figure S1 presents the t-SNE [6] embeddings of the global tokens learned on Real3D-AD and Anomaly-ShapeNet. Across both datasets, the embeddings exhibit well-formed clusters that reflect meaningful category structure, demonstrating that CFGT and C3L effectively promote category-aware global representations. The consistent clustering pattern, even as the number of categories increases, highlights the robustness of the proposed global tokenization strategy.

F. Limitation

While our model adopts a category-aware design, it does not outperform MC3D-AD on every category. This behavior is

expected, as our framework is built for class-aware generalization rather than class-specific tuning. Instead of learning highly specialized representations for individual categories, the model emphasizes semantic stability across categories, which inherently introduces trade-offs that may reduce performance in certain cases. Nevertheless, this design choice provides clear advantages in overall robustness and generalization. Analyzing the per-category AUROC variance reveals that MC3D-AD exhibits considerably larger fluctuations, whereas our method yields much lower variance and thus more uniform performance. On the Real3D-AD dataset, the O-AUROC variances of MC3D-AD and ours are 90.0 and 86.7, respectively; on the Anomaly-ShapeNet dataset, MC3D-AD reaches 102.9, while ours remains substantially lower at 43.4. These observations indicate that our approach is less sensitive to category-specific biases and achieves more stable reconstruction in the multi-category setting.

References

[1] Jiayi Cheng, Can Gao, Jie Zhou, Jiajun Wen, Tao Dai, and Jinbao Wang. Mc3d-ad: A unified geometry-aware reconstruction model for multi-category 3d anomaly detection. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pages 837–845. International Joint Conferences on Artificial Intelligence Organization, 2025. Main Track. 1

[2] 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*, pages 22207–22216, 2024. 1

[3] Hanzhe Liang, Guoyang Xie, Chengbin Hou, Bingshu Wang, Can Gao, and Jinbao Wang. Look inside for more: Internal spatial modality perception for 3d anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5146–5154, 2025. 1

[4] 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. 1

[5] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 1

[6] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9 (Nov):2579–2605, 2008. 3

[7] Yatian Pang, Eng Hock Francis Tay, Li Yuan, and Zhenghua Chen. Masked autoencoders for 3d point cloud self-supervised learning. *World Scientific Annual Review of Artificial Intelligence*, 1:2440001, 2023. 1

[8] 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, 2017. 1

[9] Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, and Kaizhu Huang. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1353–1362, 2025. 1

[10] 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*, pages 16259–16268, 2021. 1

Table S9. Comparison of point-level AUROC (%) of various methods on the Real3D-AD dataset.

P-AUROC(↑)													
Method	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Mean
Category-Specific Method													
BTF(Raw)	56.4	64.7	73.5	60.9	56.3	60.1	51.4	59.7	52.0	48.9	39.2	62.3	57.1
BTF(FPFH)	73.8	70.8	86.4	73.5	88.2	87.5	70.9	89.1	51.2	57.1	50.1	81.5	73.3
M3DM	54.7	60.2	67.9	67.8	60.8	66.7	60.6	67.4	56.0	73.8	53.2	68.2	62.0
PatchCore(FPFH)	56.2	75.4	78.0	42.9	82.8	26.4	82.9	**91.0**	73.9	73.9	60.6	74.7	68.2
PatchCore(PointMAE)	56.9	60.9	62.7	72.9	71.8	52.8	71.7	44.4	63.3	70.9	58.0	58.0	62.0
CPMF	61.8	**83.6**	73.4	55.9	75.3	71.9	**98.8**	44.9	**96.2**	72.5	**80.0**	**95.9**	75.8
Reg3D-AD	63.1	71.8	72.4	67.6	83.5	50.3	82.6	54.5	81.7	81.1	61.7	75.9	70.5
Group3AD	63.6	74.5	73.8	75.9	86.2	63.1	83.6	56.4	82.7	79.8	62.5	80.3	73.5
ISMP	**75.3**	**83.6**	**90.7**	**79.8**	**92.6**	**87.6**	88.6	85.7	81.3	**83.9**	64.1	89.5	**83.6**
PO3AD	71.5	57.4	76.3	56.3	61.4	64.4	87.7	53.0	58.5	54.2	65.9	72.9	65.0
Unified Method													
MC3D-AD	62.8	81.9	91.0	**64.0**	94.2	82.2	93.2	45.8	65.9	77.8	69.0	93.4	76.8
SeDiR (Ours)	**66.6**	**90.8**	**95.1**	63.2	**97.7**	**83.4**	**96.7**	**54.7**	**73.9**	**79.8**	**70.9**	**94.0**	**80.6**

Table S10. Comparison of point-level AUROC (%) of various methods on the Anomaly-ShapeNet dataset.

P-AUROC(↑)													
Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5	bucket0	bucket1
Category-Specific Method													
BTF(Raw)	51.2	43.0	55.1	49.1	72.0	52.4	46.4	42.6	68.5	56.3	51.7	61.7	68.6
BTF(FPFH)	62.4	74.6	64.1	54.9	62.2	71.0	76.8	51.8	59.0	67.9	69.9	40.1	63.3
M3DM	57.7	63.7	66.3	63.7	53.2	65.8	66.3	69.4	65.7	62.4	48.9	69.8	69.9
PatchCore(FPFH)	59.7	57.4	65.4	68.7	51.2	52.4	53.1	62.5	32.7	72.0	35.8	45.9	57.1
PatchCore(PointMAE)	49.5	67.4	55.3	60.6	65.3	52.7	52.4	51.5	58.1	50.1	56.2	58.6	57.4
CPMF	61.5	65.5	52.1	57.1	43.5	74.5	48.8	63.5	64.1	68.3	68.4	48.6	60.1
Reg3D-AD	69.8	71.5	88.6	69.6	52.5	77.5	61.5	59.3	65.4	80.0	69.1	61.9	75.2
IMRNet	67.1	66.8	55.6	70.2	64.1	78.1	70.5	68.4	59.9	57.6	71.5	58.5	77.4
PO3AD	**96.2**	**94.9**	**91.2**	**84.4**	**88.0**	**97.8**	**91.4**	**91.8**	**93.5**	**96.7**	**94.1**	**75.5**	**89.9**
Unified Method													
MC3D-AD	**80.1**	81.5	89.5	88.3	90.1	82.7	53.4	60.7	78.3	65.4	55.2	**79.0**	**89.5**
SeDiR (Ours)	76.4	**89.8**	**91.1**	**89.8**	**94.5**	**86.6**	**70.0**	**79.5**	**86.9**	**80.6**	**64.6**	71.7	88.4

Method	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1	helmet0	helmet1	helmet2	helmet3
Category-Specific Method													
BTF(Raw)	52.4	68.7	46.9	37.3	63.2	56.1	63.7	57.8	47.5	50.4	44.9	60.5	70.0
BTF(FPFH)	73.0	65.8	52.4	58.6	79.0	61.9	71.9	62.0	59.1	57.5	74.9	64.3	72.4
M3DM	53.1	60.5	71.8	65.5	71.5	55.6	71.0	58.1	58.5	59.9	42.7	62.3	65.5
PatchCore(FPFH)	47.2	65.3	59.5	79.5	65.5	59.6	81.0	58.3	46.4	54.8	48.9	45.5	73.7
PatchCore(PointMAE)	54.4	48.8	72.5	54.5	51.0	85.6	37.8	57.5	42.3	58.0	56.2	65.1	61.5
CPMF	60.1	55.1	55.3	55.1	49.7	50.9	68.9	69.9	45.8	55.5	54.2	51.5	52.0
Reg3D-AD	63.2	71.8	81.5	46.7	68.5	69.8	75.5	58.0	62.6	60.0	62.4	82.5	62.0
IMRNet	71.5	70.6	75.3	74.2	64.3	68.8	54.8	70.5	47.6	59.8	60.4	64.4	66.3
PO3AD	**95.7**	**94.8**	**94.0**	**86.4**	**90.9**	**93.2**	**97.4**	**82.3**	**90.7**	**87.8**	**94.8**	**93.2**	**84.6**
Unified Method													
MC3D-AD	86.7	92.2	88.0	88.5	82.9	71.5	**85.6**	66.6	**70.9**	74.4	58.2	81.7	60.2
SeDiR (Ours)	**90.0**	**97.2**	**94.0**	**95.0**	**87.1**	**84.2**	74.8	**70.8**	68.7	**81.0**	**59.6**	**88.5**	**70.1**

Method	jar	phone	shelf0	tap0	tap1	vase0	vase1	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Mean
Category-Specific Method															
BTF(Raw)	42.3	58.3	46.4	52.7	56.4	61.8	54.9	40.3	60.2	61.3	58.5	57.8	55.0	56.4	55.0
BTF(FPFH)	42.7	67.5	61.9	56.8	59.6	64.2	61.9	64.6	69.9	71.0	42.9	54.0	66.2	56.8	62.8
M3DM	54.1	35.8	55.4	65.4	71.2	60.8	60.2	73.7	65.8	65.5	64.2	51.7	55.1	66.3	61.6
PatchCore(FPFH)	47.8	48.8	61.3	73.3	**76.8**	65.5	45.3	72.1	43.0	50.5	44.7	69.3	57.5	66.3	58.0
PatchCore(PointMAE)	48.7	**88.6**	54.3	**85.8**	54.1	67.7	55.1	74.2	46.5	52.3	57.2	65.1	36.4	42.3	57.7
CPMF	61.1	54.5	**78.3**	45.8	65.7	45.8	48.6	58.2	58.2	51.4	65.1	50.4	52.9	54.5	57.3
Reg3D-AD	59.9	59.9	68.8	58.9	74.1	54.8	60.2	40.5	51.1	75.5	62.4	88.1	81.1	69.4	66.8
IMRNet	76.5	74.2	60.5	68.1	69.9	53.5	68.5	61.4	40.1	52.4	68.2	59.3	63.5	69.1	65.0
PO3AD	**87.1**	81.0	66.3	78.3	69.2	**95.5**	**88.2**	**97.8**	**88.4**	**90.2**	**93.7**	**98.2**	**95.0**	**95.2**	**89.8**
Unified Method															
MC3D-AD	84.8	85.4	65.8	52.4	53.7	**87.0**	**72.6**	82.2	79.7	78.6	64.7	61.6	88.5	77.6	75.9
SeDiR (Ours)	**91.1**	**89.1**	**67.9**	**59.7**	**63.0**	84.6	72.2	**85.0**	**83.0**	**88.4**	**76.6**	**77.7**	**90.0**	**83.5**	**81.0**

Table S11. Comparison of object-level AUPR (%) of various methods on the Anomaly-ShapeNet dataset.

O-AUPR(↑)													
Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5	bucket0	bucket1
Category-Specific Method													
BTF(Raw)	57.8	45.8	46.6	57.3	54.3	58.8	46.4	57.6	65.4	60.1	61.5	65.2	62.0
BTF(FPFH)	65.1	55.1	64.4	62.5	60.2	57.6	64.8	51.5	49.9	63.2	69.9	48.3	64.8
M3DM	63.2	64.2	76.3	67.4	45.1	52.5	51.5	63.0	63.5	57.1	60.1	60.9	50.7
PatchCore(FPFH)	44.5	60.8	61.5	67.7	57.9	54.8	54.5	61.1	62.0	57.5	54.1	60.4	56.5
PatchCore(PointMAE)	67.9	60.1	54.5	64.5	65.1	56.2	61.1	45.6	55.6	60.1	58.5	54.1	64.2
CPMF	45.3	65.5	58.8	59.2	50.5	77.5	62.1	60.1	41.8	68.3	68.5	66.2	50.1
Reg3D-AD	58.8	60.8	63.2	69.5	47.4	49.4	51.5	49.5	44.1	62.4	55.5	63.2	71.4
IMRNet	61.2	66.5	55.8	70.2	64.8	48.1	50.4	68.1	61.4	63.0	65.2	57.8	73.2
PO3AD	**99.9**	**80.9**	**92.7**	**95.9**	**96.2**	**94.6**	**90.5**	**88.8**	**92.7**	**98.5**	**90.4**	**92.3**	**88.2**
Unified Method													
SeDiR (Ours)	**97.5**	**88.0**	**93.4**	**94.3**	**99.0**	**97.5**	**96.1**	**93.6**	**97.0**	**98.9**	**98.5**	**94.1**	**94.4**

Method	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1	helmet0	helmet1	helmet2	helmet3
Category-Specific Method													
BTF(Raw)	65.9	61.2	51.5	65.3	60.1	70.1	42.5	37.9	51.5	55.9	38.8	61.5	52.6
BTF(FPFH)	61.8	57.9	54.5	59.3	58.5	65.1	71.9	53.1	52.3	56.8	72.1	58.8	56.4
M3DM	56.4	65.2	47.7	64.2	57.0	75.2	62.5	63.2	62.3	52.8	62.7	63.6	45.8
PatchCore(FPFH)	58.5	45.7	65.5	72.5	60.4	58.6	58.4	70.1	60.1	52.5	63.0	47.5	49.4
PatchCore(PointMAE)	56.1	58.3	72.1	54.2	64.2	71.0	80.1	51.5	42.3	63.3	57.1	49.6	61.1
CPMF	60.1	54.1	64.5	69.7	64.7	60.9	54.4	60.2	61.9	33.3	50.1	47.7	64.5
Reg3D-AD	69.3	71.1	62.3	77.0	53.1	63.8	42.4	53.8	61.7	60.0	38.1	61.8	46.8
IMRNet	71.1	70.2	65.8	50.2	45.5	62.7	59.9	70.1	65.6	69.7	61.5	60.2	57.5
PO3AD	**84.1**	**90.6**	**87.6**	**80.1**	**87.9**	**87.0**	**99.5**	**76.5**	**91.4**	**86.4**	**96.1**	**93.4**	**84.9**
Unified Method													
SeDiR (Ours)	**98.8**	**99.2**	**99.7**	**87.5**	**99.5**	**100.0**	**57.8**	**86.9**	**89.9**	**97.9**	**100.0**	**89.4**	**100.0**

Method	jar	phone	shelf0	tap0	tap1	vase0	vase1	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Mean
Category-Specific Method															
BTF(Raw)	42.8	61.3	62.4	53.5	59.4	56.2	44.1	41.3	71.7	42.8	61.5	54.7	41.6	48.2	54.9
BTF(FPFH)	47.9	66.2	61.1	61.0	57.5	64.1	65.5	56.9	65.2	58.7	47.2	59.2	62.4	63.8	59.8
M3DM	55.5	46.4	66.5	72.2	63.8	**78.8**	65.2	61.5	55.1	52.6	63.3	64.8	46.3	65.1	60.3
PatchCore(FPFH)	49.9	33.2	50.4	71.2	68.4	64.5	62.3	80.1	48.1	77.7	51.5	62.1	51.5	66.0	58.8
PatchCore(PointMAE)	46.3	65.2	54.3	71.2	54.2	54.8	57.2	71.1	45.5	58.6	58.5	65.2	65.5	63.4	59.5
CPMF	61.8	65.5	**68.1**	63.9	69.7	63.2	64.5	63.2	58.8	65.5	51.8	43.2	67.3	61.8	59.7
Reg3D-AD	60.1	61.4	67.5	67.6	59.9	61.5	46.8	64.1	65.1	50.5	58.8	45.5	62.9	57.4	58.4
IMRNet	76.0	55.2	62.5	40.1	**79.6**	57.3	72.5	65.5	70.8	52.8	65.4	60.1	63.9	46.2	62.1
PO3AD	**91.5**	**80.3**	68.0	**85.6**	70.9	75.3	**78.9**	**96.3**	**90.2**	**82.4**	**87.9**	**97.1**	**83.3**	**90.4**	**88.1**
Unified Method															
SeDiR (Ours)	**100.0**	**88.8**	**91.7**	**93.9**	**82.6**	**87.6**	**97.4**	**97.5**	**84.7**	**95.3**	**94.8**	**100.0**	**91.0**	**96.4**	**93.8**

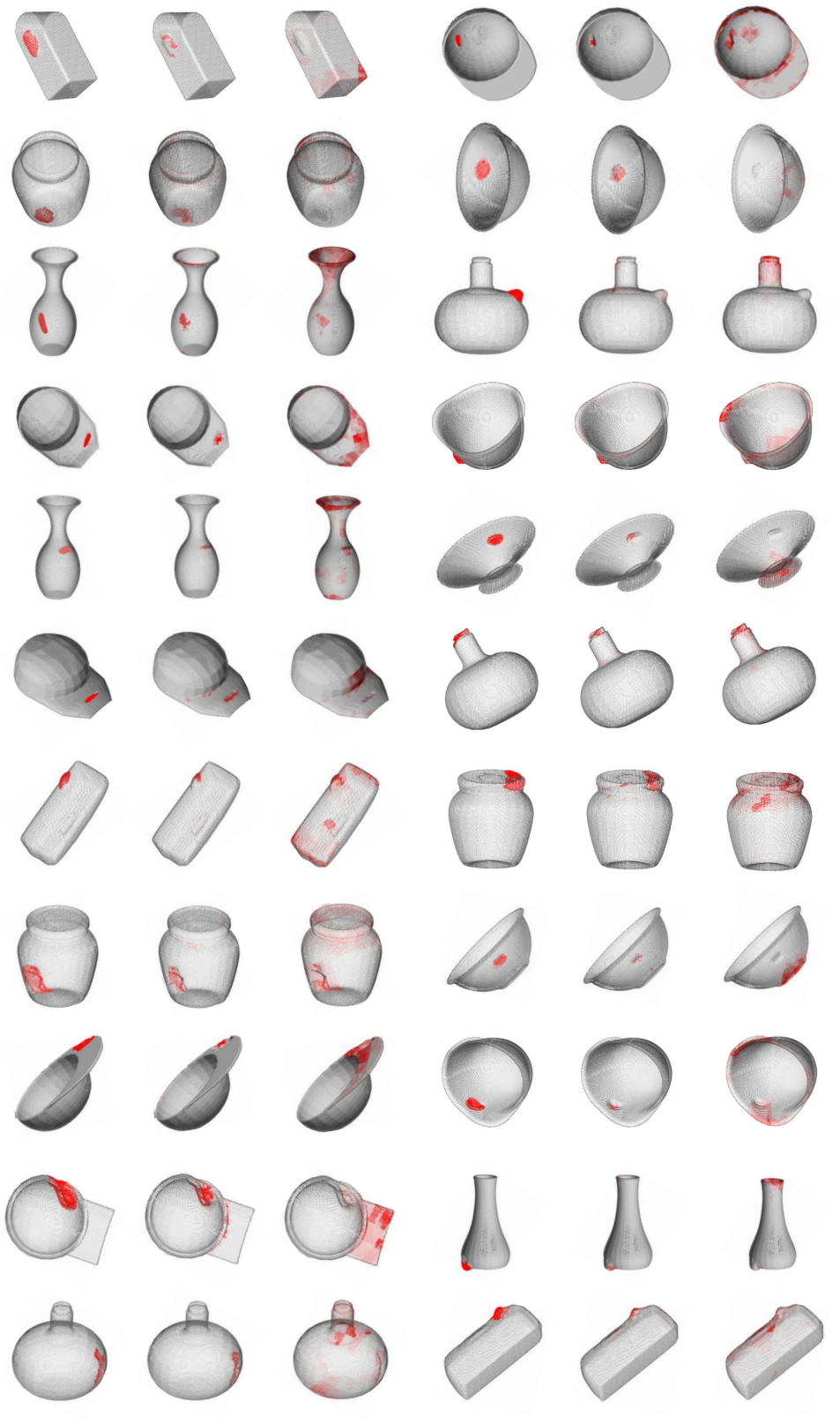


Figure S2. Additional qualitative comparison with MC3D-AD on the Anomaly-ShapeNet dataset.