LEARNING CANONICAL REPRESENTATIONS FOR UNIFIED 3D ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

3D anomaly detection is critical in industrial manufacturing because defects are defined by surface geometry that 2D inspection often misses. In practice, a single inspection pipeline covers many object categories, and the objects arrive in arbitrary poses. Recent unified methods train a single model for many categories, but they assume a fixed input pose and build on backbones pre-trained on external shape data. Methods that address rotation align each test sample to a reference, which is costly and does not carry over to the unified setting. Our key idea is to canonicalize before scoring, so that local geometry no longer carries the object pose. The canonicalizer must therefore keep the canonical frame unchanged when a defect is present and consistent across samples of the same category. We present a framework that learns **C**anonical representations for **R**otation-invariant and **U**nified 3D anomaly detection (**CRU3D-AD**), which meets both conditions with an SO(3)-equivariant canonicalizer trained under a pairing loss and a prototype loss. The canonicalizer is then frozen while a feature extractor, a geometry-aware local attention module, and a point-wise scorer are trained on synthetically generated defects, using no external pre-training and no memory bank. Experiments on Real3D-AD and Anomaly-ShapeNet show the best point-level localization among unified methods on both datasets, with competitive object-level detection. The code will be made available upon acceptance.

1 INTRODUCTION

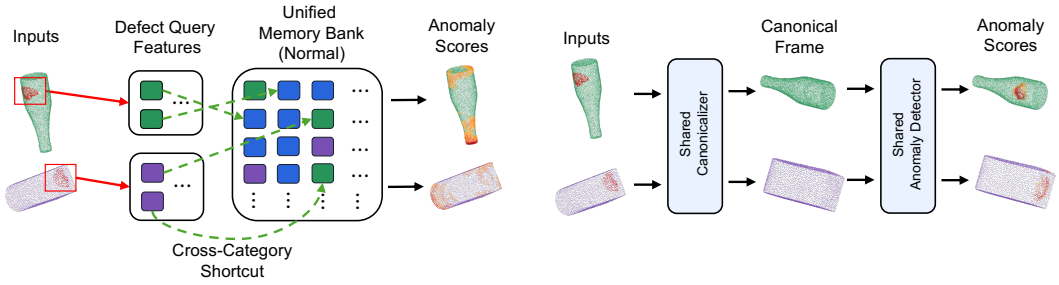
Detecting anomalies in 3D point clouds is a critical capability for automated industrial inspection, where defects frequently appear as localized geometric deviations such as dents, bulges, cracks, or missing parts (Wang et al., 2023; Horwitz & Hoshen, 2023; Liu et al., 2023a; Li et al., 2024). While 2D visual inspection is widely deployed (Roth et al., 2022; Liu et al., 2023b; You et al., 2022; Zhou et al., 2024; Lee et al., 2026), it can miss subtle structural failures because planar projection discards depth cues and reduces rich surface geometry to appearance. Point clouds preserve explicit 3D structure and enable direct reasoning about local shape and neighborhood relations, which is essential for precise defect localization.

One inspection pipeline is expected to cover many object categories, and the objects arrive in arbitrary poses. *Unified* 3D anomaly detection addresses the first with a single model (Cheng et al., 2025; Kim et al., 2026), replacing a separate detector per category and scoring every input under a common rule. Rotation has been treated in a separate line of work (Zheng et al., 2025; Yu et al., 2025), and no method so far handles both.

Most 3D anomaly detection methods are single-category models that build a memory bank of normal features and score a test point by its distance to the nearest normal neighbor (Horwitz & Hoshen, 2023; Wang et al., 2023; Cao et al., 2024b; Yu et al., 2025). That distance is informative only while the nearest normal neighbor is a comparable structure; this no longer holds once all categories are pooled (Figure 1a). Categories differ in their local geometric statistics, so anomalous regions often match normal features from unrelated categories. Rotation-invariant features do not prevent this. In the unified setting, Reg2Inv (Yu et al., 2025) matches 1,216 defect query points of bowl0 to normal features of helmet3 and 1,021 points of bucket0 to vase1 (Figure B in Appendix).

To handle pose variation, PASDF (Zheng et al., 2025) and Reg2Inv (Yu et al., 2025) register each test scan to a reference of its own category. The category must therefore be known at test time, and each

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) Memory bank-based unified detection.

(b) CRU3D-AD (ours).

Figure 1: Cross-category retrieval versus scoring in a canonical frame. (a) A unified memory bank holds normal features from every category, so a defect query can match a normal feature of an unrelated category and receive a low anomaly score. (b) CRU3D-AD maps every input into a canonical frame with a shared canonicalizer, and a shared anomaly detector scores each point in that frame without retrieval. Ground-truth anomalous points are shown in red. The boxes in (a) mark the defect region used as the query.

model is trained and deployed for that category alone. Registration is also iterative, so its latency depends on the input, as in the FPFH, RANSAC (Bolles & Fischler, 1981), and ICP (Besl & McKay, 1992) stack used in PASDF. In contrast, MC3D-AD (Cheng et al., 2025) and SeDiR (Kim et al., 2026) share one reconstruction model across all categories. Both, however, build on a backbone pre-trained on external shape data and include no mechanism for arbitrary input poses.

To address these limitations, we propose CRU3D-AD, a framework that learns **C**anonical representations for **R**otation-invariant and **U**nified 3D anomaly detection across diverse object categories. Our key idea is to canonicalize before scoring, so that local geometry no longer carries the object pose (Figure 1b). Such canonicalization has proven effective for point cloud classification and segmentation (Kaba et al., 2023). Anomaly detection, however, puts two conditions on the frame that recognition does not. The frame must not move when the input contains a defect, because a defect-dependent frame displaces every point and hides the true defect among those shifts. It also needs to be consistent across samples of the same category, because the scorer compares new samples against normal geometry learned in that frame. We design a canonicalizer with SO(3)-equivariant operations (Deng et al., 2021), so that its features rotate with the input and every orientation of a shape maps to the same canonical output. We then enforce the two conditions with a pairing loss that holds the frame fixed between a normal shape and its defective copy, and a prototype loss that pulls samples of one category together.

We then freeze the canonicalizer so that the frame cannot drift while the detector is trained on canonical geometry. Canonicalization removes the pose, but a defect is still a local deviation. Global self-attention aggregates over all points regardless of distance, which flattens the response of such a deviation. We therefore use geometry-aware local attention over k-nearest neighbors with relative positional embeddings (Zhao et al., 2021), and a lightweight MLP predicts a score per point. Because no real defects are available for training, we synthesize bulges, dents, and scratches on normal shapes and train the scorer to separate perturbed points from the rest. On Real3D-AD and Anomaly-ShapeNet, CRU3D-AD achieves the best point-level localization among unified methods while remaining competitive at object level. It is also lightweight in practice, and we report efficiency measurements in the supplementary material.

To summarize, our contribution is fourfold:

- We present CRU3D-AD, a unified framework that couples learned canonicalization with shared point-wise scoring, handling arbitrary input poses and multiple categories with a single network that needs no memory-bank retrieval or external pre-training.
- We train the canonicalizer with a pairing loss and a prototype loss so that the estimated canonical frame remains stable when inputs contain defects and consistent across samples of the same category.
- We introduce a geometry-aware local attention scoring head with relative positional embeddings that refines canonical features to preserve sharp, localized anomaly evidence.

- We develop a self-supervised training scheme that dynamically synthesizes structured defect operators from normal point clouds, enabling point-wise scoring without real anomalous samples.

2 Related Work

2.1 2D Anomaly Detection

Anomaly detection in 2D established the paradigms that carry over to 3D, namely embedding distance to a nominal distribution (Roth et al., 2022; Defard et al., 2021), pixel-level reconstruction error (Zavrtanik et al., 2021a;b), and discriminators trained on synthetic defects (Liu et al., 2023b). UniAD (You et al., 2022) relaxed the single-category assumption and CLIP-based methods (Jeong et al., 2023; Zhou et al., 2024; Cao et al., 2024a; Huang et al., 2024) extended it to zero- and few-shot detection, but these pipelines operate on a fixed image grid rather than on unaligned 3D geometry.

2.2 3D Anomaly Detection

M3DM (Wang et al., 2023) builds a multimodal memory bank from PointMAE (Pang et al., 2023) and Vision Transformer (Caron et al., 2021; Dosovitskiy et al., 2021) features, and BTF (Horwitz & Hoshen, 2023) pairs handcrafted FPFH descriptors with raw coordinates. Group3AD (Zhu et al., 2024) structures feature clusters by group-level contrastive learning, ISMP (Liang et al., 2025) adds internal projection slices, and CPMF (Cao et al., 2024b) renders multi-view images for 2D features. A further challenge is rotation, which methods address either by invariant features or by explicit alignment. Reg3D-AD (Liu et al., 2023a) registers each test sample to a nominal reference before memory-bank scoring, Reg2Inv (Yu et al., 2025) folds registration objectives into feature learning, and PASDF (Zheng et al., 2025) pairs pose alignment with a pose-aware signed distance field (Park et al., 2019). Iterative alignment is costly at multi-category scale, and memory-bank scoring collides across categories. Unified reconstruction replaces the bank: MC3D-AD (Cheng et al., 2025) reconstructs group tokens with a shared encoder–decoder, and SeDiR (Kim et al., 2026) conditions reconstruction on a disentangled category identity. Neither aligns the input, however, so both remain tied to the pose of the test scan. We therefore enforce a shared canonical frame so that unified scoring becomes independent of the input pose.

2.3 Rotation-Equivariant Representation

Vector Neurons (Deng et al., 2021) represent each feature as a set of 3D vectors instead of scalars. Standard operations such as linear layers, nonlinearities, and pooling are redefined so that a rotation applied to the input propagates to the features by the same rotation. As a result, the network maps rotated inputs to consistently rotated representations, which removes the need for external pre-training or explicit pose labels. A canonical rotation frame can then be regressed from these equivariant features and used to align the input. We adopt this equivariant formulation to build our canonicalizer.

3 Methodology

3.1 Geometric Setup and Overview

We are given an unaligned point cloud $P = \{p_i\}_{i=1}^{N} \in \mathbb{R}^{N\times 3}$ (optionally with normals n_i for training-time synthesis). The core challenge in *unified* 3D anomaly detection is to produce anomaly scores that remain comparable across categories while being stable under arbitrary object rotations. We address this by predicting a *canonical chart* in $SO(3)$ for each input shape and then performing anomaly scoring in the aligned canonical space. Figure 2 summarizes CRU3D-AD. Given an unaligned input point cloud, we first apply an equivariant canonicalizer that maps the input to a shared canonical space. A feature extractor then produces dense canonical features, which a geometry-aware *local* attention module (Zhao et al., 2021) with relative positional embeddings refines into point-wise representations. Finally, a lightweight point-wise MLP predicts anomalies. We train the canonicalizer on normal shapes and their synthetically defective copies, then freeze it to keep

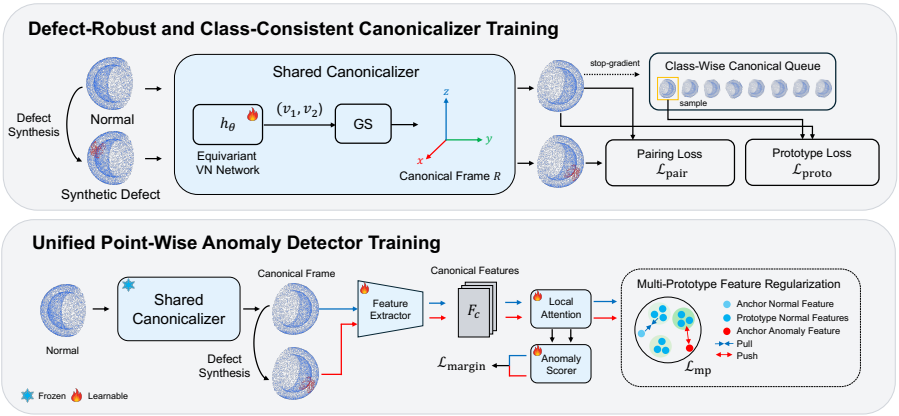


Figure 2: Overview architecture of the proposed CRU3D-AD.

the canonical space fixed while the feature extractor, the local attention module, and the scorer are trained on the same synthetic defects.

3.2 Canonical Alignment

Extracting consistent features across diverse object categories requires representing unaligned inputs in a shared canonical space. We learn an equivariant canonicalizer mapping, and train it to remain reliable under defects and consistent across samples of the same category.

3.2.1 Equivariant canonicalization

We define a canonicalizer $R(\cdot)$ that takes a point cloud and outputs the rotation aligning it to the canonical frame. An encoder h_θ built on SO(3)-equivariant Vector Neuron layers (Deng et al., 2021) maps the input P to two equivariant vectors $v_1, v_2 \in \mathbb{R}^3$, a Gram–Schmidt step $\mathrm{GS}(\cdot)$ turns into a rotation (Zhou et al., 2019):

$$(v_1, v_2) = h_\theta(P), \qquad R(P) = \mathrm{GS}(v_1, v_2) \in SO(3). \tag{1}$$

The canonical point cloud is obtained by centering and rotating the input:

$$\tilde{P} = (P - c)\, R(P), \qquad c = \frac{1}{N} \sum_i p_i. \tag{2}$$

Since h_θ is equivariant, the canonicalizer satisfies $R(PQ) = Q^\top R(P)$ for any $Q \in SO(3)$. Together with centering, this maps every orientation of a shape to the same canonical frame, making $\tilde{P}$ invariant to the input orientation.

3.2.2 Defect-robust and class-consistent alignment

Equivariance fixes a canonical frame for each shape, but two gaps remain. A defect can shift the predicted frame, and two shapes of the same category need not share a common one. We close both gaps with two objectives, applied under random SO(3) augmentation. Both are Chamfer distances (Fan et al., 2017) between canonicalized point sets, a standard measure of point-set overlap that we minimize to drive the two clouds into the same pose. For defect robustness, we pair each normal shape P with a defective copy P' and its defect mask m, both produced by our synthetic anomaly generation. We describe this procedure in detail later, as part of our self-supervised training strategy. After canonicalizing both, we match the normal shape to the non-defect points of the copy:

$$\mathcal{L}_{\text{pair}} = \mathrm{Chamfer}(\tilde{P},\, \{\tilde{p}'_j : m_j = 0\}). \tag{3}$$

Aligning only the shared geometry keeps the injected defect from perturbing the frame. For class consistency, we keep a canonical prototype $\bar{P}$ per category in memory and pull each canonicalized shape toward this stop-gradient target:

$$\mathcal{L}_{\text{proto}} = \mathrm{Chamfer}(\tilde{P},\, \bar{P}). \tag{4}$$

This anchors the shapes of a category to a common frame. The canonicalizer is trained with the combined objective

$$\mathcal{L} = \lambda_{\text{pair}}\mathcal{L}_{\text{pair}} + \lambda_{\text{proto}}\mathcal{L}_{\text{proto}}, \tag{5}$$

and is then frozen for the detection stage.

3.2.3 CANONICAL FEATURE EXTRACTION

With the canonicalizer frozen, we train a DGCNN (Wang et al., 2019) feature extractor g_ϕ on the canonical point cloud $\tilde{P}$. It aggregates features over each point's local k-nearest-neighbor graph, yielding a dense feature map:

$$F_c = g_\phi(\tilde{P}) \in \mathbb{R}^{N\times C}. \tag{6}$$

These canonical features serve as the input to the subsequent topology-aware aggregation.

3.3 GEOMETRY-AWARE LOCAL ATTENTION

Many anomaly patterns in 3D inspection are spatially localized, so global self-attention (Guo et al., 2021; Vaswani et al., 2017) can be inefficient and may dilute localized evidence, as the attention maps in the supplementary material show. We therefore perform *local* attention over k-nearest-neighbor graphs and inject explicit 3D geometry via relative positional embeddings (Zhao et al., 2021). Given $\tilde{P} \in \mathbb{R}^{N\times 3}$ and $F_c \in \mathbb{R}^{N\times C}$, we build a local neighborhood graph. For each query point p_i with feature f_i, we find its k-NN set $\mathcal{N}(i)$ in Euclidean space. For each neighbor $p_j \in \mathcal{N}(i)$, we compute a relative positional encoding:

$$e_{ij} = \Phi_{\text{pos}}(p_j - p_i), \tag{7}$$

where Φ_{pos} is an MLP mapping 3D relative coordinates to an embedding. We then compute geometry-aware attention by injecting e_{ij} into both key and value paths:

$$\alpha_{ij} = \text{Softmax}_{j\in\mathcal{N}(i)}\left(\frac{(W_q f_i)\cdot(W_k f_j + e_{ij})^T}{\sqrt{d}}\right), \tag{8}$$

$$h_i = \sum_{j\in\mathcal{N}(i)} \alpha_{ij}(W_v f_j + e_{ij}), \tag{9}$$

where d is the head dimension. The outputs are passed through a standard FFN with residual connections to obtain the refined anomaly-aware representation for each point. This design yields $O(Nk)$ aggregation cost while explicitly preserving canonical-space local topology.

3.4 UNIFIED ANOMALY SCORING

Given refined features $\{h_i\}_{i=1}^N$, we predict point-wise anomaly scores using a shared MLP Φ_{score}:

$$s_{\text{pt}}(i) = \Phi_{\text{score}}(h_i) \in \mathbb{R}. \tag{10}$$

For object-level scoring, we aggregate the K_{top} highest point scores to avoid diluting localized defects, given as:

$$I_K \triangleq \text{TopK}\big(\{s_{\text{pt}}(j)\}_{j=1}^N\big), \qquad S_{\text{obj}} = \frac{1}{K}\sum_{i\in I_K} s_{\text{pt}}(i), \tag{11}$$

where K here represents K_{top}.

3.5 SELF-SUPERVISED TRAINING STRATEGY

3.5.1 SYNTHETIC ANOMALY GENERATION

We train the attention and scoring modules using structured pseudo-anomalies synthesized on normal point clouds. Given a normal point cloud with normals $P = \{(p_i, n_i)\}_{i=1}^N$, we randomly choose a center point p_c to define a local perturbation region and apply one of three defect operators. The smooth deformation operator applies a Gaussian-shaped displacement along the normal direction, centered at p_c, where a sign $\delta \in \{-1, 1\}$ selects a bulge or a dent. The linear incision operator applies an inward displacement that decays with the distance to a line segment ℓ through p_c. Normals

inside the perturbed region are updated to follow the displaced surface. The operator equations, the normal updates, and the sampling range of every parameter are given in Appendix A.1.2. We validate the synthesis by comparing local normal variation and a curvature proxy computed from neighborhood PCA between real and synthesized defect regions, and report this dataset-level check in the same appendix.

3.5.2 TRAINING LOSS OBJECTIVE

Binary pseudo-labels $y_i \in \{0, 1\}$ indicate synthetically perturbed anomaly points ($y_i = 1$) and normal points ($y_i = 0$). We enforce separation between score distributions using a hinge-style margin loss:

$$\begin{aligned}\mathcal{L}_{\text{margin}} = &\frac{1}{|\mathcal{P}_{\text{norm}}|} \sum_{i \in \mathcal{P}_{\text{norm}}} \max(0, s_{\text{pt}}(i) + m_{\text{margin}}) \\ &+ \frac{1}{|\mathcal{P}_{\text{anom}}|} \sum_{j \in \mathcal{P}_{\text{anom}}} \max(0, -s_{\text{pt}}(j) + m_{\text{margin}}),\end{aligned} \tag{12}$$

where $\mathcal{P}_{\text{norm}}$ and $\mathcal{P}_{\text{anom}}$ are the sets of normal and anomalous points, respectively. The constant m_{margin} is the margin applied to both hinge terms. This objective explicitly pushes the two score distributions apart, supporting robust defect localization at inference time.

Multi-prototype feature regularization. The score-level margin objective does not explicitly organize the underlying feature space. Moreover, normal point features are inherently multi-modal in unified anomaly detection because object categories and local geometries vary substantially. Representing them with a single center may therefore suppress useful geometric structure. We instead maintain a class-agnostic dictionary $\mathcal{U} = \{u_r\}_{r=1}^{M}$ of M unit-norm normal feature prototypes. These feature prototypes are distinct from the category-specific shape prototypes used during canonicalizer pretraining and require no category labels.

The normal feature set $\mathcal{F}_N$ comprises features of all points from clean normal samples and unchanged points from synthetic samples, whereas $\mathcal{F}_A$ comprises features of synthetic anomaly points. We define the similarity between a feature and its closest normal mode as:

$$s_N(f; \mathcal{U}) = \max_{1 \le r \le M} f^\top u_r. \tag{13}$$

The regularizer pulls each normal feature toward its closest prototype while forcing each synthetic anomaly feature away from even its most similar normal mode:

$$\begin{aligned}\mathcal{L}_{\text{pull}} &= \frac{1}{|\mathcal{F}_N|} \sum_{f \in \mathcal{F}_N} \left[1 - s_N(f; \mathcal{U})\right], \\ \mathcal{L}_{\text{push}} &= \frac{1}{|\mathcal{F}_A|} \sum_{f \in \mathcal{F}_A} \left[s_N(f; \mathcal{U}) - \gamma_a\right]_+, \\ \mathcal{L}_{\text{mp}} &= \mathcal{L}_{\text{pull}} + \lambda_{\text{push}} \mathcal{L}_{\text{push}}.\end{aligned} \tag{14}$$

Here, $[x]_+ = \max(0, x)$ and γ_a is the maximum allowed similarity to a normal prototype for an anomalous feature. Each normal feature is assigned to $a(f) = \arg\max_r f^\top u_r$. For the non-empty assignment set $\mathcal{A}_r = \{f \in \mathcal{F}_N \mid a(f) = r\}$, we update the corresponding prototype without gradients:

$$u_r \leftarrow \text{Norm}\left(\mu u_r + (1 - \mu) \frac{1}{|\mathcal{A}_r|} \sum_{f \in \mathcal{A}_r} f\right), \tag{15}$$

where μ is the exponential-moving-average momentum and $\text{Norm}(\cdot)$ denotes ℓ_2 normalization. The complete detection objective is

$$\mathcal{L}_{\text{det}} = \mathcal{L}_{\text{margin}} + \lambda_{\text{mp}} \mathcal{L}_{\text{mp}}. \tag{16}$$

The dictionary is used only during training and discarded at inference, adding no inference-time computation.

Table 1: Average object-level (O) and point-level (P) AUROC results (%) on Real3D-AD and Anomaly-ShapeNet. Methods marked with * indicate that they were trained and evaluated in a unified setting. **Bold** values indicate the best performance in the unified setting.

		Real3D-AD		Anomaly-ShapeNet	
Type	Method	O-AUROC	P-AUROC	O-AUROC	P-AUROC
Single	Reg3D-AD Liu et al. (2023a)	70.4	70.5	57.2	66.8
	Group3AD Zhu et al. (2024)	75.1	73.5	–	–
	ISMP Liang et al. (2025)	75.7	83.6	–	–
	Reg2Inv Yu et al. (2025)	78.0	87.8	86.1	88.2
	PASDF Zheng et al. (2025)	80.2	74.5	90.0	89.7
Unified	ISMP* Liang et al. (2025)	48.4	46.7	–	–
	Reg2Inv* Yu et al. (2025)	57.1	61.8	54.2	69.8
	PASDF* Zheng et al. (2025)	51.2	49.5	54.0	51.5
	MC3D-AD Cheng et al. (2025)	78.2	76.8	84.2	74.8
	SeDiR Kim et al. (2026)	**81.0**	80.6	**93.3**	81.0
	CRU3D-AD (Ours)	75.5	85.5	87.0	92.1
	CRU3D-AD w/ TTA (Ours)	78.9	**87.0**	89.2	**93.9**

4 EXPERIMENTS

4.1 DATASETS AND METRICS

Datasets. We evaluate CRU3D-AD on Real3D-AD (Liu et al., 2023a) and Anomaly-ShapeNet (Li et al., 2024). Real3D-AD contains 12 categories of real industrial components with authentic defects, acquired as partial single-view scans. Each category provides only four normal training samples and 100 test samples. Anomaly-ShapeNet is a synthetic benchmark of 1,600 point clouds across 40 categories. Following the unified protocol (Cheng et al., 2025), we train one model on the pooled normal samples of all categories and evaluate it on the combined test set.

Metrics. Following the conventional protocol (Bergmann & Sattlegger, 2023), we report AUROC at two granularities: object-level AUROC (O-AUROC) for detection, i.e., how well the model separates anomalous point clouds from normal ones, and point-level AUROC (P-AUROC) for localization, computed by comparing per-point anomaly scores with the corresponding ground-truth labels. Both are macro-averaged over categories so that larger categories do not dominate.

4.2 EXPERIMENTAL SETTINGS

Baseline methods. To comprehensively evaluate the performance of our proposed CRU3D-AD, we compare it against various anomaly detection methods. The main text focuses on quantitative comparisons with recent and highly competitive baselines, specifically Reg3D-AD (Liu et al., 2023a), Group3AD (Zhu et al., 2024), ISMP (Liang et al., 2025), Reg2Inv (Yu et al., 2025), PASDF (Zheng et al., 2025), MC3D-AD (Cheng et al., 2025), and SeDiR (Kim et al., 2026). The complete evaluation results, which also include earlier approaches such as BTF (Horwitz & Hoshen, 2023), M3DM (Wang et al., 2023), PatchCore (Roth et al., 2022), CPMF (Cao et al., 2024b), are detailed in the supplementary material.

Implementation details. For each benchmark, we trained the canonicalizer from scratch using only its normal training split and kept it frozen during subsequent training. Both benchmarks use 4,096-point inputs obtained by Farthest Point Sampling (FPS). The DGCNN feature extractor, geometry-aware local attention module, and point-wise scorer were jointly trained for 400 epochs using Adam with a learning rate of 10^{-4} and cosine annealing. Remaining training details and the hardware setup are provided in the supplementary material.

Inference protocol. Unless stated otherwise, each test sample was evaluated with a single forward pass on a deterministic FPS subsample of the same size used during training. We additionally report a test-time augmentation (TTA) variant that averages predictions over $T = 8$ independent random subsamples of the same object. For each subsample, the object score is the mean of the K_{top} highest point scores, and the reported object score is the average over the T subsamples. Point scores are propagated from each subsample to the full-resolution cloud by nearest-neighbor assignment and then averaged before computing point-level AUROC. The network weights, the neighborhood size

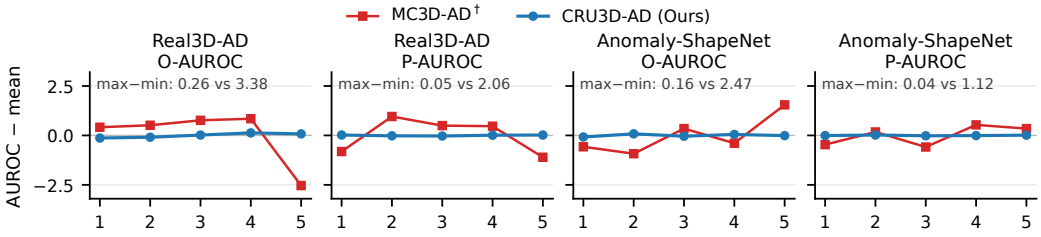


Figure 3: Stability under test-set reorientation. Each of the five runs applies one random SO(3) rotation to the whole test set, the same five rotations for both methods, without retraining. The vertical axis is the AUROC of a run minus that method's mean AUROC over the five runs, so each curve is centered on its own mean and the four panels share one scale. The number in each panel is the max−min spread of CRU3D-AD against MC3D-AD. †MC3D-AD is our reproduction. Absolute values are listed in Table B in the appendix.

k, and K_{top} are identical to the single-pass setting, and no retraining is involved. Rows marked "w/ TTA" in Table 1 follow this protocol at a cost of T forward passes per sample.

4.3 Main Evaluation Results

Table 1 compares CRU3D-AD with recent baselines under the unified setting on both benchmarks. CRU3D-AD gives the strongest point-level localization on Real3D-AD and Anomaly-ShapeNet, while its object-level detection stays within the range of the unified baselines. Notably, our empirical evaluations reveal that prior rotation-invariant and alignment-based paradigms, namely Reg2Inv (Yu et al., 2025) and PASDF (Zheng et al., 2025), suffer severe performance degradation in the unified setting. Reg2Inv relies on a memory bank of normal features, where mixing diverse categories leads to frequent cross-category retrievals (Figure B). For PASDF, the limitation is two-fold. First, modeling highly diverse geometric topologies within a single shared Signed Distance Field (SDF) overwhelms the network capacity, resulting in over-smoothed shape reconstructions. Second, the pose alignment mechanism in PASDF strictly requires pairing the input point cloud with a category-matched template to function correctly. Since the category of an incoming test sample is unknown in the unified setting, evaluating PASDF inherently necessitates an oracle to manually provide the corresponding template. Relying on such oracle template pairing fundamentally violates the strict class-agnostic assumption of the unified setting.

Comparisons on Real3D-AD. Table 1 shows that CRU3D-AD attains the highest point-level accuracy among unified methods on Real3D-AD, reaching 85.5% P-AUROC and 87.0% with test-time augmentation. Under TTA, CRU3D-AD outperforms SeDiR (Kim et al., 2026) by 6.4% and MC3D-AD (Cheng et al., 2025) by 10.2%. Object-level detection is more constrained, at 75.5% and 78.9% under TTA, which is comparable to MC3D-AD but 2.1% behind SeDiR. Real3D-AD provides only four normal scans per category, and its single-view acquisition leaves sensor noise and incomplete surfaces that raise the scores of normal regions, so the object-level margin stays narrow even when the defect itself is localized correctly. Qualitative results on Real3D-AD are shown in Figures E and F in the appendix.

Comparisons on Anomaly-ShapeNet. The margin over unified baselines is wider on Anomaly-ShapeNet. As presented in Table 1, CRU3D-AD achieves 92.1% P-AUROC and 87.0% O-AUROC, which test-time augmentation raises to 93.9% and 89.2%. Under TTA, CRU3D-AD outperforms SeDiR by 12.9% and MC3D-AD by 19.1% at point level. Object-level detection exceeds MC3D-AD by 5.0% in the same setting but remains 4.1% behind SeDiR.

The two unified baselines score reconstruction discrepancy on group tokens. MC3D-AD pools the difference between reconstructed and input tokens, and SeDiR applies the same token-level construction over 1,024 groups. CRU3D-AD instead predicts a score for every input point and trains it against synthetic defects with a margin objective. Our method is above both baselines at point level

Table 2: Ablation study on both benchmarks (AUROC %). *Align* is canonical alignment, *Local* is geometry-aware local attention, and *Synthetic* is synthetic anomaly generation. ∘ marks a canonicalizer with random, untrained weights, which is still SO(3)-equivariant by construction.

Components			Real3D-AD		Anomaly-ShapeNet	
Align	*Local*	*Synthetic*	O-AUROC	P-AUROC	O-AUROC	P-AUROC
	✓	✓	57.7	75.1	74.0	83.2
∘	✓	✓	70.0	85.4	79.1	86.6
✓		✓	44.5	81.1	68.3	91.0
✓	✓		49.0	51.1	51.4	46.7
✓	✓	✓	**75.5**	**85.5**	**87.0**	**92.1**

but within their range at object level, which may reflect this difference in scoring. Qualitative results on Anomaly-ShapeNet are shown in Figures E and F in the appendix.

4.4 Robustness to Test-Set Rotation

Both benchmarks deliver each test sample in the pose recorded in the dataset, so the results reported so far do not show what happens when an object arrives turned some other way. We draw five random SO(3) rotations and re-evaluate the trained models, applying one rotation to the whole test set in each run. Figure 3 shows the deviation of each run from the method's own mean, and the largest movement of CRU3D-AD is 0.26 points at object level on Real3D-AD, against 3.38 for MC3D-AD. Each draw is a uniformly random rotation from the full rotation group rather than a small tilt, so the five runs span the range that a sweep from small to large angles would cover.

4.5 Ablation Study

Table 2 removes each module of CRU3D-AD in turn on both benchmarks under the unified setting, and adds a control in which the canonicalizer keeps its random initial weights. Removing *Align* degrades O-AUROC by 13.0 points and P-AUROC by 8.9 points on Anomaly-ShapeNet, and by 17.8 and 10.4 points on Real3D-AD, since an instance-varying frame shifts the top-K_{top} scores that the object score compares across samples. A canonicalizer with random, untrained weights is still equivariant by construction and recovers part of this loss. It stays 7.9 and 5.5 points below the trained canonicalizer at object level but within 5.5 and 0.1 points at point level, so the learned frame benefits object-level scoring more than localization. Global attention in place of *Local* costs 18.7 and 31.0 O-AUROC points but only 1.1 and 4.4 P-AUROC points, because aggregating over every point flattens the score peaks that the object score averages while largely preserving their ordering. Training without *Synthetic* is the most damaging ablation: both metrics fall to chance on both benchmarks, since the margin objective then sees only normal points and nothing in the training signal separates a defect from an unusual but normal region.

5 Conclusion

We introduced CRU3D-AD, a unified 3D anomaly detection framework that maps unaligned point clouds into a shared canonical space and detects defects from the resulting geometry. On Real3D-AD and Anomaly-ShapeNet, CRU3D-AD achieves the best point-level localization among unified methods while remaining competitive at the object level. Learned canonicalization provides a practical route to scalable inspection without per-category retraining. A shape with a nontrivial symmetry group admits no unique canonical frame, and neither of our losses resolves that ambiguity. Object-level detection also trails the strongest unified baseline, by 2.1 and 4.1 points under test-time augmentation, while point-level localization leads on both benchmarks. The gap follows false positives on locally unusual normal geometry, where sensor noise and incomplete surfaces raise point scores that the object score averages. Suppressing these responses and resolving symmetry explicitly are one promising future work.

AI USE STATEMENT

In this work, we used generative AI tools for proofreading and for editing the writing of this paper. We have not used generative AI tools for research ideation, for the design of the method, or for the design of the experiments, and the remaining disclosure categories are not applicable to this work. We have reviewed all AI-assisted work. Every suggested edit was checked against our intended meaning and was accepted or rewritten by the authors, and the technical content of each sentence is our own. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies anomaly detection on two public point-cloud benchmarks, Real3D-AD (Liu et al., 2023a) and Anomaly-ShapeNet (Li et al., 2024). Both consist of scans and synthetic models of manufactured objects and contain no human subjects, no personally identifiable information, and no sensitive content. We used them under their released terms and did not collect any new data. The intended application is industrial quality inspection. As with any inspection system, a missed defect can carry a safety cost, so the point-level scores our method produces are meant to support human inspection rather than replace it, and the limitations we report in Section 5 bear on where that support is weaker. We declare no conflicts of interest.

REPRODUCIBILITY STATEMENT

Both benchmarks are public, and the preprocessing we apply to them, including the voxel downsampling of Real3D-AD and the normal re-estimation that follows canonicalization and each synthetic defect, is described in Appendix A.1. That subsection also lists every optimizer setting, loss weight, and architectural hyperparameter for both stages, and Appendix A.1.2 gives the defect operator equations together with the sampling range of each of their parameters. The hardware and software used for all runs, including the efficiency measurements, are stated in Appendix A.1.1. Appendix A.1.3 describes how we adapted the single-category baselines to the unified setting, since those numbers are not taken from published tables. Table C reports the variation across five random seeds. The code will be made available upon acceptance.

REFERENCES

Paul Bergmann and David Sattlegger. Anomaly detection in 3d point clouds using deep geometric descriptors. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2613–2623, 2023.

Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pp. 586–606. Spie, 1992.

Robert C Bolles and Martin A Fischler. A ransac-based approach to model fitting and its application to finding cylinders in range data. In *Ijcai*, volume 1981, pp. 637–643, 1981.

Y. Cao, J. Zhang, L. Frittoli, Y. Cheng, W. Shen, and G. Boracchi. Adaclip: Adapting clip with hybrid learnable prompts for zero-shot anomaly detection. In *Proceedings of the European Conference on Computer Vision*, July 2024a.

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

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

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 International Joint Conference on Artificial Intelligence*, 2025.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International conference on pattern recognition*, pp. 475–489. Springer, 2021.

Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J Guibas. Vector neurons: A general framework for SO(3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12200–12209, 2021.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 605–613, 2017.

Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R Martin, and Shi-Min Hu. Pct: Point cloud transformer. *Computational visual media*, 7(2):187–199, 2021.

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

C. Huang, A. Jiang, J. Feng, Y. Zhang, X. Wang, and Y. Wang. Adapting visual-language models for generalizable anomaly detection in medical images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11375–11385, 2024.

Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 19606–19616, 2023.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pp. 15546–15566. PMLR, 2023.

SuYeon Kim, Wongyu Lee, and MyeongAh Cho. A semantically disentangled unified model for multi-category 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 33036–33045, 2026.

Geonu Lee, Yujeong Oh, Geonhui Jang, Soyoung Lee, Jeonghyo Song, Sungmin Cha, and YoungJoon Yoo. Continual-mega: A large-scale benchmark for generalizable continual anomaly detection. *Neurocomputing*, pp. 134460, 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.

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*, volume 39, pp. 5146–5154, 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, 2023a.

Z. Liu, Y. Zhou, Y. Xu, and Z. Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20402–20411, 2023b.

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.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 165–174, 2019.

K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. 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.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

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.

Z. You, L. Cui, Y. Shen, K. Yang, X. Lu, Y. Zheng, and X. Le. A unified model for multi-class anomaly detection. In *Advances in Neural Information Processing Systems*, volume 35, pp. 4571–4584, 2022.

Yuyang Yu, Zhengwei Chen, Xuemiao Xu, Lei Zhang, Haoxin Yang, Yongwei Nie, and Shengfeng He. Registration is a powerful rotation-invariance learner for 3d anomaly detection. In *Advances in Neural Information Processing Systems*, 2025.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8330–8339, 2021a.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021b.

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.

Bozhong Zheng, Jinye Gan, Xiaohao Xu, Xintao Chen, Wenqiao Li, Xiaonan Huang, Na Ni, and Yingna Wu. Bridging 3d anomaly localization and repair via high-quality continuous geometric representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27063–27072, 2025.

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. In *International Conference on Learning Representations*, volume 2024, pp. 49705–49737, 2024.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5745–5753, 2019.

Hongze Zhu, Guoyang Xie, Chengbin Hou, Tao Dai, Can Gao, Jinbao Wang, and Linlin Shen. Towards high-resolution 3d anomaly detection via group-level feature contrastive learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 4680–4689, 2024.

APPENDIX

A ADDITIONAL IMPLEMENTATION DETAILS

A.1 TRAINING CONFIGURATION.

Real3D-AD is far denser than Anomaly-ShapeNet, so we voxel-downsample each point cloud to 30,000 points before applying FPS. Normals are re-estimated from geometry with a radius of 0.1 and at most 30 neighbors, once after canonicalization and again after each synthetic defect is applied. For both Anomaly-ShapeNet and Real3D-AD, we trained the canonicalizer for 200 epochs with a batch size of 16. We used AdamW with an initial learning rate of 10^{-3} and a weight decay of 10^{-2}. We set $\lambda_{\text{pair}} = \lambda_{\text{proto}} = 1$, and each category queue stored eight canonical point clouds. Because the prototype loss keeps one queue per category, the canonicalizer needs a category label for every training sample. That category label is never used again downstream, and a test sample is canonicalized and scored without any category information. The registration-based baselines instead need a category label at test time, since the template they align to has to come from the test sample's own category. We trained the feature extractor, geometry-aware local attention module, and point-wise scorer with a batch size of six and 100 updates per epoch. We set the local neighborhood size to $k = 64$ with $K_{\text{top}} = 100$ on Anomaly-ShapeNet and $k = 32$ with $K_{\text{top}} = 200$ on Real3D-AD, since the single-view scans of Real3D-AD favor a tighter neighborhood and a broader score pool. The normal feature prototypes were initialized from the first training batch by farthest-first sampling under cosine distance and updated with momentum $\mu = 0.9$. We used $M = 8$ normal feature prototypes with $\lambda_{\text{mp}} = 0.25$, and set $\gamma_a = 0$, $\lambda_{\text{push}} = 1$, and $m_{\text{margin}} = 0.5$ on both benchmarks. For the multi-prototype loss, we sampled at most 512 point features per object from each of the following sets: the original normal point cloud, the synthetic anomaly region, and the normal region of the synthetic anomaly sample.

A.1.1 HARDWARE AND SOFTWARE.

We ran all experiments on a single NVIDIA RTX A6000 48 GB with dual Intel Xeon Gold 6248R CPUs at 3.00 GHz, using PyTorch 2.1.2 with CUDA 12.1 on Ubuntu 22.04. The efficiency measurements in Table H were taken on the same machine.

A.1.2 SYNTHETIC ANOMALY CONFIGURATION.

Smooth Deformation Operator. We formalize smooth surface deformations by applying a Gaussian-shaped displacement along the normal direction, centered at p_c:

$$p_i' = p_i + \delta\,\eta\,\exp\!\left(-\frac{\|p_i - p_c\|_2^2}{2\sigma_s^2}\right) n_i, \tag{17}$$

where $\delta \in \{-1, 1\}$ selects bulge or dent, η sets the peak displacement, and σ_s controls the spatial extent. Here $S(p_i)$ denotes the scalar displacement magnitude in the above equation. To keep the perturbed geometry consistent, we update normals using a tangent-plane gradient step:

$$g_i = \left(\mathbf{I} - n_i n_i^\top\right)(p_i - p_c), \qquad \nabla S(p_i) = -\frac{S(p_i)}{\sigma_s^2}\, g_i, \tag{18}$$

followed by renormalization:

$$n_i' = \frac{n_i - \nabla S(p_i)}{\|n_i - \nabla S(p_i)\|_2}. \tag{19}$$

Linear Incision Operator. We create thin linear incisions by sampling a 3D line segment ℓ through p_c and applying an inward normal displacement that decays with the point-to-line distance:

$$p_i' = p_i - \eta\,\exp\!\left(-\frac{d(p_i, \ell)^2}{2\sigma_l^2}\right) n_i, \tag{20}$$

where $d(p_i, \ell)$ is the Euclidean distance from p_i to the line segment, η controls incision depth, and σ_l controls width. To emulate surface roughness, we perturb normals with $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2\mathbf{I})$ and renormalize:

$$n_i' = \frac{n_i + \epsilon_i}{\|n_i + \epsilon_i\|_2}. \tag{21}$$

Table A: Per-category P-AUROC (%) on Anomaly-ShapeNet in the unified setting. **Bold** values indicate the best performance.

P-AUROC(↑)											
Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5
Reg2Inv*	80.5	82.0	76.8	81.5	93.6	81.0	57.4	61.6	74.9	68.1	54.4
PASDF*	56.7	45.6	60.2	50.0	52.1	54.1	39.3	46.8	39.9	49.7	37.9
MC3D-AD	80.7	85.7	90.2	86.7	90.2	77.5	56.2	59.7	77.9	67.0	56.2
SeDiR	76.4	89.8	91.1	89.8	94.5	86.6	70.0	79.5	86.9	80.6	64.6
CRU3D-AD (Ours)	83.1	93.2	99.2	98.3	95.5	96.2	89.4	96.9	**97.5**	97.6	86.3
CRU3D-AD w/ TTA (Ours)	**85.2**	**96.4**	**99.7**	**98.8**	**97.0**	**98.1**	**92.1**	**98.7**	97.4	**98.0**	**89.3**
Method	bucket0	bucket1	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1
Reg2Inv*	54.7	77.0	75.6	80.7	75.7	76.7	71.8	64.7	84.1	41.1	62.4
PASDF*	46.3	46.5	46.9	58.9	40.6	56.7	61.6	44.3	61.2	57.5	57.1
MC3D-AD	90.2	86.8	85.4	90.3	85.8	88.2	76.3	69.4	82.0	66.6	59.2
SeDiR	71.7	88.4	90.0	**97.2**	**94.0**	95.0	87.1	84.2	74.8	70.8	68.7
CRU3D-AD (Ours)	97.0	**96.8**	96.3	95.0	90.9	92.7	95.8	94.5	98.9	83.7	**83.8**
CRU3D-AD w/ TTA (Ours)	**97.9**	96.5	**97.3**	95.7	92.5	**95.8**	**97.3**	**97.3**	**99.3**	**87.7**	83.2
Method	helmet0	helmet1	helmet2	helmet3	jar0	phone	shelf0	tap0	tap1	vase0	vase1
Reg2Inv*	53.4	52.0	75.2	66.9	83.8	83.4	55.9	60.8	48.7	68.6	70.6
PASDF*	48.2	41.3	49.4	57.1	53.5	62.4	64.0	60.7	59.9	65.9	44.5
MC3D-AD	74.9	59.1	81.8	58.5	84.7	89.1	62.5	50.2	58.4	89.7	60.8
SeDiR	81.0	59.6	88.5	70.1	91.1	89.1	67.9	59.7	63.0	84.6	72.2
CRU3D-AD (Ours)	90.6	83.9	91.2	95.9	98.5	99.0	92.9	74.2	77.1	94.3	87.6
CRU3D-AD w/ TTA (Ours)	**90.8**	**89.9**	**93.1**	**96.9**	**99.1**	**99.5**	**94.0**	**75.2**	**79.4**	**96.8**	**91.5**
Method	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Average			
Reg2Inv*	82.6	65.4	54.2	62.6	67.8	93.5	68.9	69.8			
PASDF*	56.4	52.9	41.2	40.1	56.6	52.6	42.9	51.5			
MC3D-AD	78.1	80.0	77.2	58.8	57.6	87.4	76.2	74.8			
SeDiR	85.0	83.0	88.4	76.6	77.7	90.0	83.5	81.0			
CRU3D-AD (Ours)	97.9	94.7	84.6	85.1	86.8	98.5	92.4	92.1			
CRU3D-AD w/ TTA (Ours)	**98.5**	**96.7**	**90.1**	**86.3**	**90.8**	**99.0**	**96.4**	**93.9**			

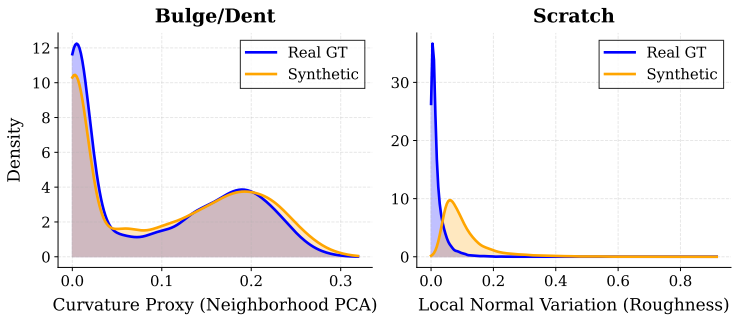


Figure A: **Synthetic-to-real similarity of local geometry.** We compare distributions of local normal variation and a curvature proxy (neighborhood PCA) between real defect regions and synthesized defect regions, aggregated across categories.

The operator-specific hyperparameters are sampled from uniform distributions, denoted by $\mathcal{U}(a, b)$, as follows:

Smooth Deformation Operator (Bulges and Dents).

- Deformation direction (δ): Randomly chosen from $\{-1, 1\}$ with equal probability to create either a bulge or a dent.
- Peak displacement (η): Sampled from $\mathcal{U}(0.01, 0.06)$, taken at the centre of the region and decaying outwards with the Gaussian below.
- Spatial extent (σ_s): We first define a local region radius $\rho \sim \mathcal{U}(0.05, 0.17)$. The spatial extent of the Gaussian is then set to $\sigma_s = 0.4\rho$, resulting in an effective range of $\sigma_s \in [0.02, 0.068]$.

Linear Incision Operator (Scratches).

- Incision length (L): The length of the 3D line segment ℓ is sampled from $\mathcal{U}(0.05, 0.30)$, with a random 3D orientation.
- Incision depth (η): Sampled from $\mathcal{U}(0.01, 0.05)$. Note that scratches are strictly applied as inward displacements.
- Incision width (σ_l): We define a maximum thickness boundary $w_{\max}$ sampled from $\mathcal{U}(0.025, 0.050)$. The Gaussian decay width is set to $\sigma_l = w_{\max}/2$, yielding $\sigma_l \in [0.0125, 0.025]$.
- Surface roughness (σ_n): The operator perturbs the normals inside the incision with a standard deviation of $\sigma_n = 0.2$. Because normals are re-estimated afterwards, only the displaced points carry the incision into training.

Synthetic-to-Real Similarity of Local Geometry. We check how closely the operators above reproduce the local geometry of real defects. For the smooth deformation operator we measure a curvature proxy computed from neighborhood PCA, and for the linear incision operator we measure local normal variation, since each descriptor captures the property that the corresponding operator manipulates. Figure A shows the two distributions aggregated across categories. Bulges and dents follow the real distribution closely, including its bimodal shape. Scratches span the same range but are shifted toward higher normal variation, so the synthesized incisions are rougher than real ones.

A.1.3 Unified Adaptation of Single-Category Baselines.

Rows marked with an asterisk report methods whose released implementations fit one model per category, so we adapted them to the unified setting ourselves.

Reg2Inv. We retrained the feature extractor on the normal samples of all categories pooled into one training set and used the resulting single checkpoint for every category. The memory bank is pooled in the same way. We extracted features for every training sample of every category and normalized the position and shape components by their maxima over that pooled set. The bank was then reduced to 10,000 entries with the approximate greedy coreset sampler at a rate of 0.1. Scoring uses five nearest neighbors, and scores are min-max normalized within each category. We left the registration stage untouched, so a test sample is still aligned to a reference drawn from its own category. Reg2Inv* therefore keeps the category information that its registration needs, and only the memory bank becomes category-agnostic.

PASDF. Since the released method fits one signed distance field per category, the unified variant trains a single field over the pooled normal samples of all categories, for 2,000 epochs with a latent size of 64. Its registration stage is unchanged and still runs FPFH, RANSAC and ICP against a template of the test sample's own category. The reported numbers are therefore obtained with that template available. What the unified setting removes from PASDF* is the per-category field, not the reference it registers to.

ISMP. Two memory banks are used, one over learned point features and one over raw coordinates, and we pooled both across all categories. Both banks are built and subsampled as above, and scoring uses three nearest neighbors. Its registration is unchanged as well, so every test sample is aligned by FPFH and RANSAC to a template taken from the training set of its own category.

B Cross-Category Retrieval Analysis

We examine which normal features the defect query points of Reg2Inv* retrieve from the pooled memory bank described above. For every defect query point in the unified setting, we take its nearest normal feature in the bank and record the category that feature comes from. Figure B lists the most frequent pairs in which the retrieved feature belongs to a different category than the query. Rotation-invariant features do not keep the retrieval within the query's own category, and defect regions of bowl0, bucket0, and cap0 repeatedly match normal geometry of helmet3 and several vase categories. Such a match returns a small distance, so the defect receives a low anomaly score even though the retrieved structure is unrelated to the query object.

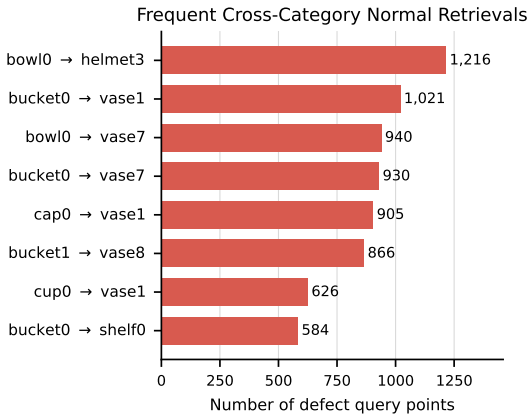


Figure B: Cross-category retrievals of Reg2Inv* in the unified setting. Each bar counts the defect query points whose nearest normal feature in the pooled memory bank comes from the category on the right of the arrow.

Table B: Stability of anomaly detection under test-set reorientation (AUROC %). We apply five random SO(3) rotations, each to the whole test set, and re-evaluate without retraining. Within each dataset block the two methods are paired on each metric, and **bold** marks the smaller range in each pair. †MC3D-AD numbers are our own reproduction and therefore differ from the values quoted in the other tables.

Dataset	Method	Mean ± Std	Max − Min
Anomaly-ShapeNet	MC3D-AD† O-AUROC	83.16 ± 0.981	2.47
	CRU3D-AD O-AUROC	86.79 ± 0.065	**0.16**
	MC3D-AD† P-AUROC	66.58 ± 0.499	1.12
	CRU3D-AD P-AUROC	92.10 ± 0.016	**0.04**
Real3D-AD	MC3D-AD† O-AUROC	73.65 ± 1.429	3.38
	CRU3D-AD O-AUROC	75.35 ± 0.110	**0.26**
	MC3D-AD† P-AUROC	73.20 ± 0.903	2.06
	CRU3D-AD P-AUROC	85.53 ± 0.023	**0.05**

C Extended Evaluation Results

We extend the comparison with four earlier point cloud anomaly detection baselines: BTF (Horwitz & Hoshen, 2023), M3DM (Wang et al., 2023), PatchCore (Roth et al., 2022), and CPMF (Cao et al., 2024b). Table I details the full O-AUROC and P-AUROC scores across all individual categories for Real3D-AD. For Anomaly-ShapeNet we report the two settings separately, since the 40 categories make a single combined table hard to read. Tables G and A cover the unified setting, where every method is trained once on the pooled normal samples of all categories; this is the comparison our claims rest on. Tables J and K list the single-category results, where each method is trained separately per category and therefore solves an easier problem. The four added baselines stay below MC3D-AD and SeDiR on both benchmarks. Among the unified methods, CRU3D-AD reaches the best point-level score on each. With the eight-pass test-time augmentation it scores 93.9% on Anomaly-ShapeNet and 87.0% on Real3D-AD, against 81.0% and 80.6% for SeDiR. A single pass lowers these to 92.1% and 85.5%, still ahead of SeDiR on both. At object level it trails SeDiR, at 89.2% and 78.9% against 93.3% and 81.0%.

D More Ablation Studies

In this section, we provide extended ablation studies to further validate the robustness and the architectural design choices of our proposed CRU3D-AD.

Table C: Performance robustness across 5 random seeds (Mean $\pm$ Std). Evaluated under the unified setting.

Dataset	O-AUROC	P-AUROC
Real3D-AD	73.96 $\pm$ 1.26	86.58 $\pm$ 0.62
Anomaly-ShapeNet	86.48 $\pm$ 0.51	92.50 $\pm$ 0.36

Table D: Comparison of synthetic anomaly generation methods on Anomaly-ShapeNet (AUROC %). Evaluated under the unified setting.

Method	O-AUROC	P-AUROC
Gaussian Noise	80.0	74.8
Structured (Ours)	**87.0**	**92.1**

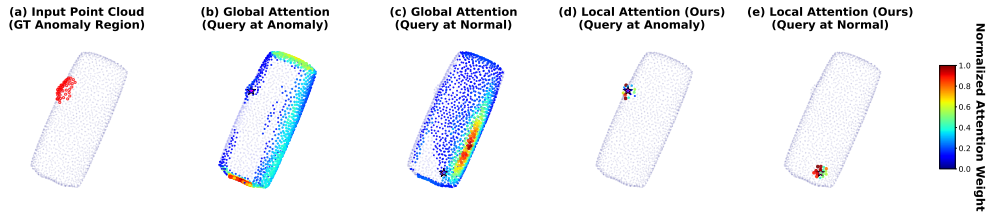


Figure C: **Effectiveness of the Local Attention module.** The attention weight distributions are visualized for both anomalous and normal query points. While the standard global attention (b, c) is easily distracted by distant, irrelevant points and fails to highlight the defect, our local attention module (d, e) confines high attention weights to the actual anomaly boundary (a).

Robustness to Random Seeds. We keep the canonicalizer fixed and retrain the detector in the unified setting with five random seeds. Table C reports the resulting mean and standard deviation. On both benchmarks the standard deviation stays below 1.3 at object level and below 0.7 at point level. The detector therefore does not depend on a particular weight initialization.

Loss Terms. Each row of Table F sets one loss weight to zero. $\mathcal{L}_{\text{proto}}$ matters most: without it AUROC falls by 7.6 points at object level on Real3D-AD and 5.0 on Anomaly-ShapeNet, since nothing then ties separate instances of a category to one frame. Dropping $\mathcal{L}_{\text{pair}}$ costs 4.4 points at object level on Real3D-AD and 1.6 on Anomaly-ShapeNet, while point level rises by 1.2 on the first and stays within 0.1 on the second. The loss that holds the frame steady under defects buys detection accuracy rather than localization. On Anomaly-ShapeNet, removing $\mathcal{L}_{\text{mp}}$ costs 1.7 points at both levels, but on Real3D-AD it moves object level by 0.5, inside the seed standard deviation of 1.26 in Table C, and leaves point level unchanged. The canonicalizer losses carry most of the object-level contribution, and the detector-side term adds to it on Anomaly-ShapeNet alone.

Stability under Test-Set Reorientation. Both benchmarks deliver each test sample in the pose recorded in the dataset, so the results reported so far do not show what happens when an object arrives turned some other way. We draw five random SO(3) rotations and re-evaluate the trained models, applying one rotation to the whole test set in each run. No model is retrained, and the only difference between runs is the rotation. The voxel downsampling that Real3D-AD receives happens once when the cache is built, so it precedes every rotation. Table B reports the five runs, and for both methods the largest movement occurs on Real3D-AD, where CRU3D-AD changes by 0.26 at object level and 0.05 at point level, against 3.38 and 2.06 for MC3D-AD. On Anomaly-ShapeNet CRU3D-AD changes by 0.16 and 0.04, against 2.47 and 1.12 for MC3D-AD. This stability follows from scoring in a shared canonical frame, which makes the result independent of the input pose.

Effectiveness of Structured Synthetic Anomalies. We replace our structured operators with a Gaussian noise perturbation and retrain the detector. As presented in Table D, Gaussian noise degrades O-AUROC by 7.0 and P-AUROC by 17.3. Isotropic noise displaces points in random di-

Table E: Effect of the local neighborhood size k on both benchmarks. Evaluated under the unified setting. **Bold** values indicate the best performance in each column.

	Real3D-AD		Anomaly-ShapeNet	
k	O-AUROC	P-AUROC	O-AUROC	P-AUROC
16	72.4	83.4	**87.0**	90.2
32	**75.5**	85.5	85.8	90.8
64	71.9	**86.7**	**87.0**	**92.1**

Table F: Loss-term ablation on both benchmarks (AUROC %). Each row sets one weight to zero. λ_{pair} and λ_{proto} belong to the canonicalizer, so removing either retrains it and rebuilds the canonical cache, whereas λ_{mp} belongs to the detector. The reference row repeats the full model of Table 2. The ablated rows were validated once, at the final epoch.

	Real3D-AD		Anomaly-ShapeNet	
Configuration	O-AUROC	P-AUROC	O-AUROC	P-AUROC
CRU3D-AD	**75.5**	85.5	**87.0**	**92.1**
w/o $\mathcal{L}_{\text{pair}}$	71.1	86.7	85.4	92.0
w/o $\mathcal{L}_{\text{proto}}$	67.9	84.8	82.0	89.1
w/o $\mathcal{L}_{\text{mp}}$	75.0	85.5	85.3	90.4

Table G: Per-category O-AUROC (%) on Anomaly-ShapeNet in the unified setting. **Bold** values indicate the best performance.

O-AUROC(↑)											
Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5
Reg2Inv*	91.4	31.4	54.3	73.7	64.8	57.0	64.8	67.0	45.6	51.1	42.5
PASDF*	62.4	66.7	61.4	75.1	56.8	54.8	39.3	44.4	54.1	65.2	61.4
MC3D-AD	96.2	80.5	79.5	70.9	75.6	93.0	**97.8**	71.9	88.5	91.1	75.4
SeDiR	97.6	88.6	92.9	91.6	**98.4**	95.9	95.6	92.6	96.7	**98.5**	**97.9**
CRU3D-AD (Ours)	94.8	89.0	99.5	**99.3**	93.3	99.3	59.3	88.1	95.9	95.2	78.2
CRU3D-AD w/ TTA (Ours)	**99.0**	**92.4**	**100.0**	93.0	95.9	**100.0**	75.2	**95.9**	**97.0**	93.0	94.4
Method	bucket0	bucket1	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1
Reg2Inv*	58.1	52.1	47.0	56.8	58.6	65.6	53.8	26.7	68.6	45.3	42.4
PASDF*	50.3	60.0	36.3	48.1	40.7	54.7	50.5	56.7	58.1	80.4	56.7
MC3D-AD	89.8	78.4	79.3	70.1	83.5	76.1	74.3	95.2	77.6	**86.2**	88.6
SeDiR	91.1	**93.3**	**98.5**	**98.9**	**99.6**	**87.4**	99.5	**100.0**	66.7	84.9	89.0
CRU3D-AD (Ours)	**98.7**	82.9	94.8	78.6	74.4	77.2	**100.0**	88.6	99.1	83.6	**94.3**
CRU3D-AD w/ TTA (Ours)	97.1	84.8	88.9	77.5	88.4	82.5	99.5	91.9	**100.0**	82.7	93.3
Method	helmet0	helmet1	helmet2	helmet3	jar0	phone	shelf0	tap0	tap1	vase0	vase1
Reg2Inv*	50.7	63.8	64.1	45.2	50.0	45.2	42.3	64.2	53.0	41.7	38.1
PASDF*	46.4	51.4	50.4	62.7	57.6	77.1	37.2	60.0	52.2	78.8	46.7
MC3D-AD	67.2	**100.0**	60.9	97.9	97.1	91.9	84.1	**94.5**	**97.0**	82.1	85.7
SeDiR	**96.5**	**100.0**	**87.0**	**100.0**	**100.0**	90.5	**88.7**	92.7	89.3	88.8	**97.1**
CRU3D-AD (Ours)	77.4	91.4	83.5	82.7	99.5	**99.1**	82.0	59.4	60.4	96.2	85.7
CRU3D-AD w/ TTA (Ours)	89.0	92.4	79.4	86.4	99.5	98.1	76.2	66.7	75.9	**97.5**	90.5
Method	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Average			
Reg2Inv*	45.2	53.3	50.9	47.1	65.2	77.3	51.8	54.2			
PASDF*	49.0	49.7	37.9	40.7	39.5	39.1	50.3	54.0			
MC3D-AD	92.9	76.1	87.6	**97.6**	93.8	67.0	73.6	84.2			
SeDiR	97.6	78.8	**92.4**	94.8	**100.0**	87.3	**94.2**	**93.3**			
CRU3D-AD (Ours)	**100.0**	**84.5**	71.2	78.6	93.3	**90.6**	82.1	87.0			
CRU3D-AD w/ TTA (Ours)	99.0	79.4	61.5	80.5	91.9	88.5	91.2	89.2			

rections, whereas our operators displace them along the surface normal and reproduce the bulges, dents, and scratches that the benchmark contains.

Local Attention versus Global Attention. Figure C contrasts the attention weights of the two designs at an anomalous query point and at a normal one. Global attention spreads weight over distant points that share no local structure with the query, so the defect boundary receives no more weight than the rest of the surface. Local attention instead restricts the receptive field to the k-nearest neighbors and concentrates weight on that boundary. This is consistent with the ablation

Table H: Detailed efficiency comparison on the Real3D-AD dataset. To ensure a fair comparison, we report the maximum inference latency and peak memory footprint recorded over 100 iterations. The total latency represents the true end-to-end time encompassing necessary pre-processing steps and the model forward pass.

Method	Latency (ms)			Peak Memory (MB)
	Pre-processing	Inference	Total	
PASDF	2,405	579,807	582,212	2695.28
Reg2Inv	146	97,626	97,772	3145.26
MC3D-AD	46	1466	1,512	4011.95
CRU3D-AD (Ours)	415	111	526	1197.98

in Table 2: the diluted global response preserves the ordering of point scores, which point-level AUROC measures, but flattens the peaks that the object score averages.

Local Neighborhood Size. Table E varies k while every other setting is held fixed. The two values follow from how each benchmark is acquired rather than from this ablation. Anomaly-ShapeNet supplies complete synthetic shapes, while Real3D-AD supplies single-view scans whose partial surfaces favor a tighter neighborhood. The ablation is consistent with both settings. On Anomaly-ShapeNet $k = 64$ is the only value at the top on both metrics, tied with $k = 16$ at object level and alone at point level. On Real3D-AD the two metrics disagree, and the object-level gap favoring $k = 32$ is three times the point-level gap favoring $k = 64$.

E Qualitative Results

Canonicalization Results. Figure D shows what the canonicalizer does to the input pose. The benchmark delivers the samples of a category in unrelated orientations, and after canonicalization the six samples of each category face the same way. The three defective samples land in that same frame, and the red region sits at a different place on each of them, so the frame is set by the object rather than by the defect. These are the two conditions that the prototype loss and the pairing loss are meant to enforce.

Anomaly Localization. Figure E compares anomaly localization on both benchmarks. No anomaly detection checkpoint has been released for MC3D-AD, so we trained it with the authors' code, and part of the gap visible in Figure E may come from that reproduction. Across both benchmarks the response of CRU3D-AD concentrates on the annotated region and follows its shape. MC3D-AD instead predicts anomalies across wide portions of the object, which is consistent with scoring group tokens rather than individual points. Figure F adds Reg2Inv (Yu et al., 2025) as a third method on six categories, each in its original test pose. Reg2Inv misses the ground-truth region on Real3D-AD, and its prediction breaks into fragments on Anomaly-ShapeNet. CRU3D-AD predicts a single compact region that overlaps the ground truth in every column. Its shape also follows the ground-truth region, recovering the two separate spots on the gemstone and the striped patch on the jar.

The remaining errors are false positives on normal geometry that is locally unusual. On Real3D-AD, duck and shell show some response outside the annotated region. The single-view acquisition of this benchmark leaves sensor noise and incomplete surfaces that raise the scores of normal regions. On Anomaly-ShapeNet, our method localizes helmet0 and tap0 more tightly than MC3D-AD, but scores still rise along thin flanges and protruding parts, where a local neighborhood covers a sharp change in surface orientation.

F Computational Efficiency

We compare the computational efficiency of CRU3D-AD against the baselines on Real3D-AD, whose scans are the denser of the two benchmarks and therefore the more demanding to process. All latencies and peak memory are measured on a single NVIDIA RTX A6000, reporting the maximum over 100 iterations after 10 warm-up runs. Table H separates pre-processing from the model forward pass. PASDF (Zheng et al., 2025) and Reg2Inv (Yu et al., 2025) spend 580 s and 98 s per sample

Table I: Comparative analysis of AUROC results (%) for localization and detection tasks on Real3D-AD. Methods marked with * indicate that they were trained and evaluated in a unified setting. **Bold** denotes the best score within the unified setting.

O-AUROC(↑) / P-AUROC(↑)								
Type	Method	Airplane	Car	Candy	Chicken	Diamond	Duck	Fish
Single	BTF(RAW)	73.0/56.4	64.7/64.7	53.9/73.5	78.9/60.9	70.7/56.3	69.1/60.1	60.2/51.4
	BTF(FPFH)	52.0/73.8	56.0/70.8	63.0/86.4	43.2/73.5	54.5/88.2	78.4/87.5	54.9/70.9
	M3DM	43.4/54.7	54.1/60.2	55.2/67.9	68.3/67.8	60.2/60.8	43.3/66.7	54.0/60.6
	M3DM(PointBERT)	40.7/52.3	50.6/59.3	44.2/68.2	67.3/79.0	62.7/59.4	46.6/66.8	55.6/58.9
	PatchCore(FPFH)	88.2/56.2	59.0/75.4	54.1/78.0	83.7/42.9	57.4/82.8	54.6/26.4	67.5/82.9
	PatchCore(FPFH+Raw)	84.8/55.6	77.7/74.0	62.6/74.9	85.3/55.8	78.4/85.4	62.8/65.8	83.7/78.1
	PatchCore(PMAE)	72.6/56.9	49.8/60.9	66.3/62.7	82.7/72.9	78.3/71.8	48.9/52.8	63.0/71.7
	CPMF	70.1/61.8	55.1/83.6	55.2/73.4	50.4/55.9	52.3/75.3	58.2/71.9	55.8/98.8
	Reg3D-AD	71.6/63.1	69.7/71.8	68.5/72.4	85.2/67.6	90.0/83.5	58.4/50.3	91.5/82.6
	Group3AD	74.4/63.6	72.8/74.5	84.7/73.8	78.6/75.9	93.2/86.2	67.9/63.1	97.6/83.6
	ISMP	85.8/75.3	73.1/83.6	85.2/90.7	71.4/79.8	94.8/92.6	71.2/87.6	94.5/88.6
	Reg2Inv	81.8/92.3	75.8/94.4	100.0/96.9	94.4/91.0	100.0/97.9	75.0/93.7	67.2/84.6
	PASDF	62.8/77.7	95.9/80.2	78.8/54.6	73.9/76.8	89.4/69.9	65.8/78.2	98.9/83.7
Unified	ISMP*	62.9/45.5	35.5/46.7	39.3/43.5	43.7/51.8	35.0/42.3	56.8/43.3	51.2/51.2
	Reg2Inv*	**95.3/97.0**	46.8/70.4	48.4/57.6	46.2/57.2	61.8/53.0	47.7/68.7	52.1/61.9
	PASDF*	50.5/45.3	49.7/46.4	52.4/50.3	40.4/45.7	55.9/55.8	57.4/51.7	40.5/49.8
	MC3D-AD	85.0/62.8	74.9/81.9	83.0/91.0	71.5/64.0	95.5/94.2	83.1/82.2	86.5/93.2
	SeDiR	86.0/66.6	**78.3**/90.8	81.9/95.1	72.9/63.2	94.8/**97.7**	**86.2**/83.4	**93.8/96.7**
	CRU3D-AD (Ours)	68.4/83.5	74.5/90.4	85.2/95.6	**74.0**/81.5	97.6/92.7	73.1/87.8	76.6/87.5
	CRU3D-AD w/ TTA (Ours)	70.3/84.0	76.5/**92.2**	**87.2/96.8**	73.2/**83.1**	**98.6**/95.7	79.9/**88.8**	80.2/90.6

Type	Method	Gemstone	Seahorse	Shell	Starfish	Toffees	Average
Single	BTF(RAW)	68.6/59.7	59.6/52.0	39.6/48.9	53.0/39.2	70.3/62.3	63.5/57.1
	BTF(FPFH)	64.8/89.1	77.9/51.2	75.4/57.1	57.5/50.1	46.2/81.5	60.3/73.3
	M3DM	64.4/67.4	49.5/56.0	69.4/73.8	55.1/53.2	45.0/68.2	55.2/63.1
	M3DM(PointBERT)	61.7/64.6	49.4/57.4	57.7/73.2	52.8/56.3	56.2/67.7	53.8/63.6
	PatchCore(FPFH)	37.0/91.0	50.5/73.9	58.9/73.9	44.1/60.6	56.5/74.7	59.3/68.2
	PatchCore(FPFH+Raw)	35.9/53.9	76.7/80.8	66.3/75.3	47.1/61.3	57.0/54.9	68.2/68.0
	PatchCore(PMAE)	37.4/44.4	53.9/63.3	50.1/70.9	51.9/58.0	58.5/58.0	59.4/62.0
	CPMF	58.9/44.9	72.9/96.2	65.3/72.5	70.0/80.0	39.0/95.9	58.6/75.8
	Reg3D-AD	41.7/54.5	76.2/81.7	58.3/81.1	50.6/61.7	82.7/75.9	70.4/70.5
	Group3AD	53.9/56.4	84.1/82.7	58.5/79.8	56.2/62.5	79.6/80.3	75.1/73.5
	ISMP	46.8/85.7	72.9/81.3	62.3/83.9	66.0/64.1	84.2/89.5	75.7/83.6
	Reg2Inv	73.5/90.7	53.2/64.5	69.2/90.6	84.1/84.0	62.6/73.7	78.0/87.8
	PASDF	63.4/65.4	100.0/88.7	85.0/64.8	61.7/70.3	86.6/83.8	80.2/74.5
Unified	ISMP*	42.4/46.4	56.3/50.2	55.0/45.6	45.9/43.0	56.8/50.9	48.4/46.7
	Reg2Inv*	48.9/69.7	52.5/47.1	73.0/39.1	68.5/63.6	44.2/56.1	57.1/61.8
	PASDF*	53.0/47.4	48.5/48.3	56.2/50.7	55.2/52.9	54.3/50.1	51.2/49.5
	MC3D-AD	56.0/45.8	71.6/65.9	**80.3**/77.8	76.6/69.0	73.8/93.4	78.2/76.8
	SeDiR	62.7/54.7	67.4/**73.9**	77.9/79.8	**85.4/70.9**	84.5/94.0	**81.0**/80.6
	CRU3D-AD (Ours)	65.5/89.7	72.8/66.0	75.3/89.1	60.3/67.2	82.0/95.2	75.5/85.5
	CRU3D-AD w/ TTA (Ours)	**69.2/91.2**	**83.0**/66.2	78.1/**92.4**	61.4/67.3	**88.7/96.1**	78.9/**87.0**

on iterative optimization and registration. MC3D-AD (Cheng et al., 2025) is far faster at 1512 ms in total, but peaks at 4012 MB of memory. CRU3D-AD requires 415 ms for pre-processing and 111 ms for the forward pass, 526 ms end to end, at 1198 MB of peak memory, the lowest among the compared methods. The remaining cost is therefore dominated by point-cloud preparation rather than by the network.

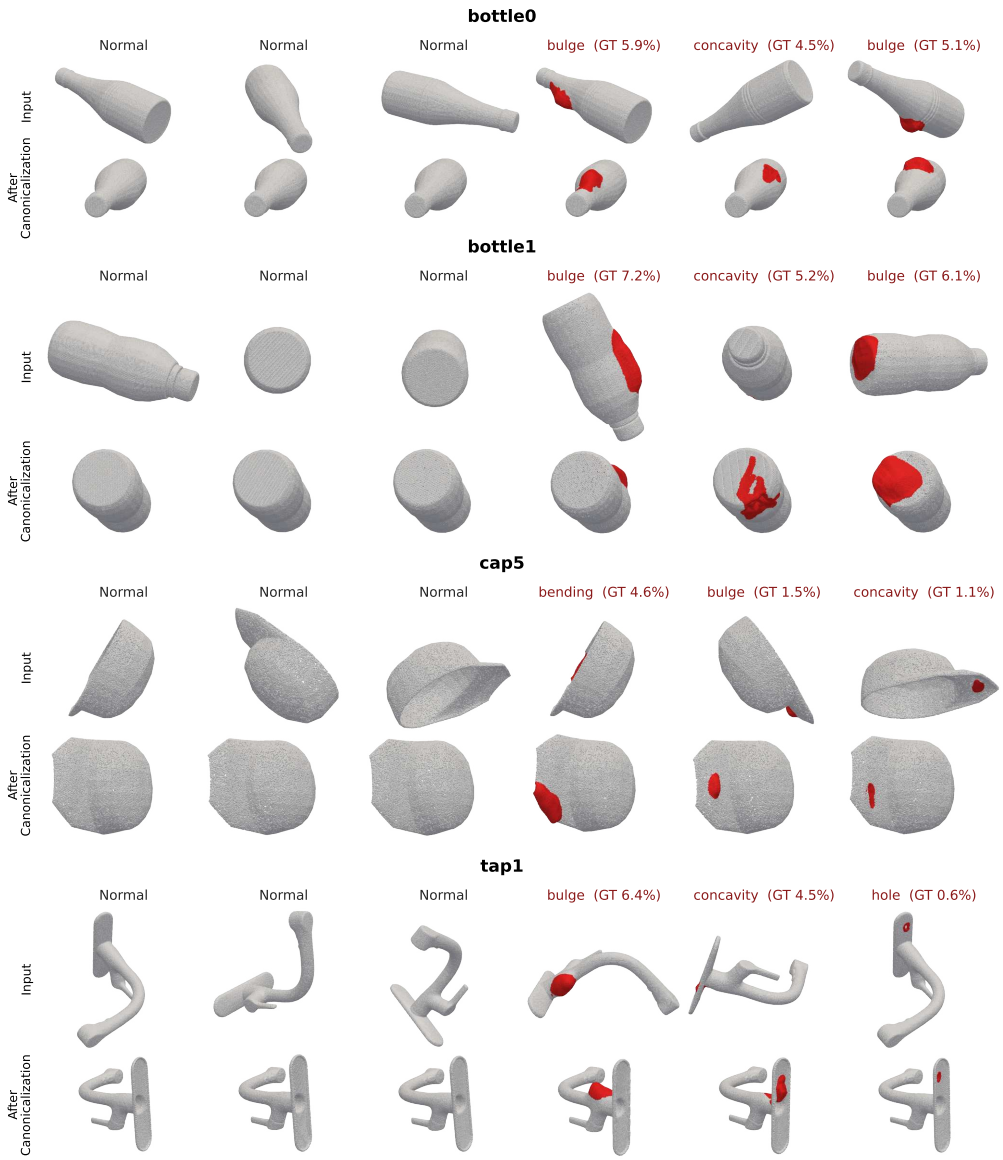


Figure D: Effect of the canonicalizer on Anomaly-ShapeNet categories. For each category the top row shows six test samples in the pose provided by the benchmark, and the bottom row shows the same samples after our canonicalizer. All six panels within a row share one camera, so any difference in appearance along a row is a difference in frame. The first three samples of each category are normal and the last three carry a real defect. Ground-truth anomalous points are shown in red, and the percentage beside each defect type is the fraction of points that the ground truth labels anomalous. The canonical frame stays the same whether or not a defect is present.

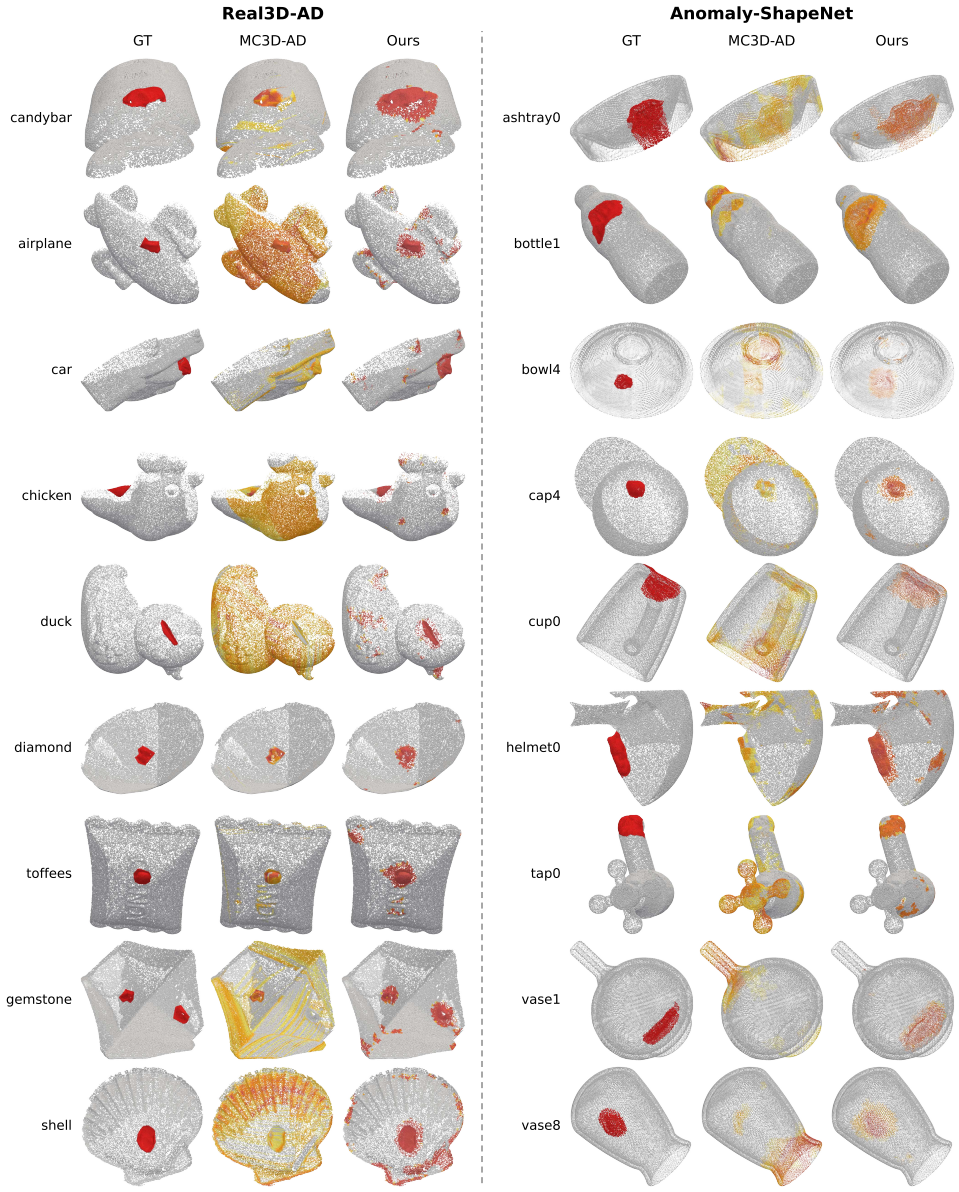


Figure E: Qualitative comparison of anomaly localization in the unified setting. Each row is one category, with Real3D-AD on the left and Anomaly-ShapeNet on the right, and each half shows the ground truth, MC3D-AD (Cheng et al., 2025), and our method (CRU3D-AD). Within a row the three panels share one camera, chosen per sample so that the annotated region faces the viewer. Ground-truth anomalous points are shown in red. For the two prediction panels, only points scoring above 0.5 after per-sample min-max normalization are highlighted, from orange to red with increasing score.

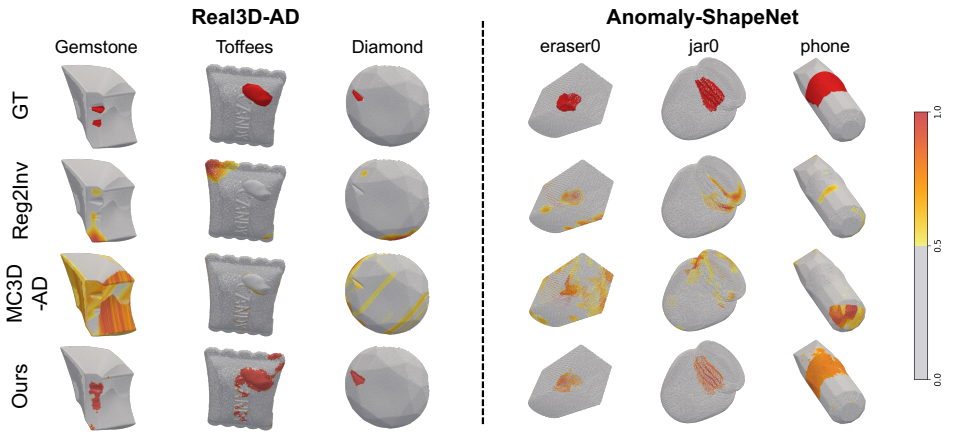


Figure F: Qualitative comparison of anomaly localization in the unified setting. Rows show the ground truth and the predictions of Reg2Inv (Yu et al., 2025), MC3D-AD (Cheng et al., 2025), and our method (CRU3D-AD) on Real3D-AD (left) and Anomaly-ShapeNet (right). Only points scoring above 0.5 after per-sample min-max normalization are highlighted.

Table J: Per-category O-AUROC (%) on Anomaly-ShapeNet in the single-category setting.

	O-AUROC(↑)										
Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5
BTF(RAW)	57.8	41.0	59.7	51.0	56.8	56.4	26.4	52.5	38.5	66.4	41.7
BTF(FPFH)	42.0	54.6	34.4	54.6	32.2	50.9	66.8	51.0	49.0	60.9	69.9
M3DM	57.7	53.7	57.4	63.7	54.1	63.4	66.3	68.4	61.7	46.4	40.9
PatchCore(FPFH)	58.7	57.1	60.4	66.7	57.2	50.4	63.9	61.5	53.7	49.4	55.8
PatchCore(PMAE)	59.1	60.1	51.3	60.1	65.0	52.3	62.9	45.8	57.9	50.1	59.3
CPMF	35.3	64.3	52.0	48.2	40.5	78.3	63.9	62.5	65.8	68.3	68.5
Reg3D-AD	59.7	70.6	48.6	69.5	52.5	67.1	52.5	49.0	34.8	66.3	59.3
Reg2Inv	90.0	100.0	100.0	100.0	100.0	100.0	80.7	65.6	58.5	85.2	81.8
PASDF	100.0	99.5	100.0	100.0	100.0	100.0	94.8	100.0	100.0	93.3	91.2
Method	bucket0	bucket1	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1
BTF(RAW)	61.7	32.1	66.8	52.7	46.8	37.3	40.3	52.1	52.5	37.8	51.5
BTF(FPFH)	40.1	63.3	61.8	52.2	52.0	58.6	58.6	61.0	71.9	52.0	49.0
M3DM	30.9	50.1	55.7	42.3	77.7	63.9	53.9	55.6	62.7	57.7	61.7
PatchCore(FPFH)	46.9	55.1	58.0	45.3	75.7	79.0	60.0	58.6	65.7	58.3	63.7
PatchCore(PMAE)	59.3	56.1	58.9	47.6	72.7	53.8	61.0	55.6	67.7	59.1	62.7
CPMF	48.2	60.1	60.1	55.1	55.3	69.7	49.7	49.9	68.9	64.3	45.8
Reg3D-AD	61.0	75.2	69.3	72.5	64.3	46.7	51.0	53.8	34.3	53.7	61.0
Reg2Inv	81.3	90.2	65.9	86.3	68.1	90.2	73.3	93.3	100.0	100.0	84.3
PASDF	96.8	77.5	85.2	64.9	64.6	85.3	97.1	85.7	95.2	100.0	79.5
Method	helmet0	helmet1	helmet2	helmet3	jar0	phone	shelf0	tap0	tap1	vase0	vase1
BTF(RAW)	55.3	34.9	60.2	52.6	42.0	56.3	16.4	52.5	57.3	53.1	54.9
BTF(FPFH)	57.1	71.9	54.2	44.4	42.4	67.1	60.9	56.0	54.6	34.2	21.9
M3DM	52.6	42.7	62.3	37.4	44.1	35.7	56.4	75.4	73.9	42.3	42.7
PatchCore(FPFH)	54.6	48.4	42.5	40.4	47.2	38.8	49.4	75.3	76.6	45.5	42.3
PatchCore(PMAE)	55.6	55.2	44.7	42.4	48.3	48.8	52.3	45.8	53.8	44.7	55.2
CPMF	55.5	58.9	46.2	52.0	61.0	50.9	68.5	35.9	69.7	45.1	34.5
Reg3D-AD	60.0	38.1	61.4	36.7	59.2	41.4	68.8	67.6	64.1	53.3	70.2
Reg2Inv	81.7	98.6	87.5	87.6	100.0	100.0	57.7	94.8	80.4	99.6	60.5
PASDF	81.2	93.8	76.5	84.6	100.0	100.0	71.3	88.2	79.3	100.0	92.9
Method	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Average			
BTF(RAW)	41.0	71.7	42.5	58.5	44.8	42.4	56.4	49.3			
BTF(FPFH)	54.6	69.9	51.0	40.9	51.8	66.8	26.8	52.8			
M3DM	73.7	43.9	47.6	31.7	65.7	66.3	66.3	55.2			
PatchCore(FPFH)	72.1	44.9	50.6	41.7	69.3	66.2	66.0	56.8			
PatchCore(PMAE)	74.1	46.0	51.6	57.9	65.0	66.3	62.9	56.2			
CPMF	58.2	58.2	51.4	61.8	39.7	52.9	60.9	55.9			
Reg3D-AD	60.5	65.0	50.0	52.0	46.2	62.0	59.4	57.2			
Reg2Inv	100.0	84.5	81.8	100.0	64.3	81.8	87.3	86.1			
PASDF	100.0	80.6	91.2	100.0	100.0	92.4	83.6	90.0			

Table K: Per-category P-AUROC (%) on Anomaly-ShapeNet in the single-category setting.

P-AUROC(↑)											
Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5
BTF(RAW)	51.2	43.0	55.1	49.1	72.0	52.4	46.4	42.6	68.5	56.3	51.7
BTF(FPFH)	62.4	74.6	64.1	54.9	62.2	71.0	76.8	51.8	59.0	67.9	69.9
M3DM	57.7	63.7	66.3	63.7	53.2	65.8	66.3	69.4	65.7	62.4	48.9
PatchCore(FPFH)	59.7	57.4	65.4	68.7	51.2	52.4	53.1	62.5	32.7	72.0	35.8
PatchCore(PMAE)	49.5	67.4	55.3	60.6	65.3	52.7	52.4	51.5	58.1	50.1	56.2
CPMF	61.5	65.5	52.1	57.1	43.5	74.5	48.8	63.5	64.1	68.3	68.4
Reg3D-AD	69.8	71.5	88.6	69.6	52.5	77.5	61.5	59.3	65.4	80.0	69.1
Reg2Inv	78.5	99.1	99.5	84.9	81.7	98.3	82.8	82.2	76.1	78.8	82.4
PASDF	91.9	95.8	95.1	92.6	94.8	96.3	90.0	81.6	93.9	86.5	90.9
Method	bucket0	bucket1	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1
BTF(RAW)	61.7	68.6	52.4	68.7	46.9	37.3	63.2	56.1	63.7	57.8	47.5
BTF(FPFH)	40.1	63.3	73.0	65.8	52.4	58.6	79.0	61.9	71.9	62.0	59.1
M3DM	69.8	69.9	53.1	60.5	71.8	65.5	71.5	55.6	71.0	58.1	58.5
PatchCore(FPFH)	45.9	57.1	47.2	65.3	59.5	79.5	65.5	59.6	81.0	58.3	46.4
PatchCore(PMAE)	58.6	57.4	54.4	48.8	72.5	54.5	51.0	85.6	37.8	57.5	42.3
CPMF	48.6	60.1	60.1	55.1	55.3	55.1	49.7	50.9	68.9	69.9	45.8
Reg3D-AD	61.9	75.2	63.2	71.8	81.5	46.7	68.5	69.8	75.5	58.0	62.6
Reg2Inv	61.0	85.5	86.1	94.5	86.4	97.0	79.8	88.1	98.0	94.6	97.0
PASDF	87.5	82.4	94.8	86.1	89.4	92.0	94.8	88.4	94.5	86.3	89.1
Method	helmet0	helmet1	helmet2	helmet3	jar0	phone	shelf0	tap0	tap1	vase0	vase1
BTF(RAW)	50.4	44.9	60.5	70.0	42.3	58.3	46.4	52.7	56.4	61.8	54.9
BTF(FPFH)	57.5	74.9	64.3	72.4	42.7	67.5	61.9	56.8	59.6	64.2	61.9
M3DM	59.9	42.7	62.3	65.5	54.1	35.8	55.4	65.4	71.2	60.8	60.2
PatchCore(FPFH)	54.8	48.9	45.5	73.7	47.8	48.8	61.3	73.3	76.8	65.5	45.3
PatchCore(PMAE)	58.0	56.2	65.1	61.5	48.7	88.6	54.3	85.8	54.1	67.7	55.1
CPMF	55.5	54.2	51.5	52.0	61.1	54.5	78.3	45.8	65.7	45.8	48.6
Reg3D-AD	60.0	62.4	82.5	62.0	59.9	59.9	68.8	58.9	74.1	54.8	60.2
Reg2Inv	92.5	90.6	89.1	95.6	98.2	99.2	63.2	91.8	86.9	98.0	70.5
PASDF	81.6	64.6	80.9	95.8	95.9	95.1	86.5	88.4	90.2	94.4	79.7
Method	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Average			
BTF(RAW)	40.3	60.2	61.3	58.5	57.8	55.0	56.4	55.0			
BTF(FPFH)	64.6	69.9	71.0	42.9	54.0	66.2	56.8	62.8			
M3DM	73.7	65.8	65.5	64.2	51.7	55.1	66.3	61.6			
PatchCore(FPFH)	72.1	43.0	50.5	44.7	69.3	57.5	66.3	58.0			
PatchCore(PMAE)	74.2	46.5	52.3	57.2	65.1	36.4	42.3	57.7			
CPMF	58.2	58.2	51.4	65.1	50.4	52.9	54.5	57.3			
Reg3D-AD	40.5	51.1	75.5	62.4	88.1	81.1	69.4	66.8			
Reg2Inv	99.7	84.4	92.7	87.9	86.3	93.4	97.1	88.2			
PASDF	95.6	86.8	89.9	91.5	95.9	90.9	86.3	89.7			