TRAINING-FREE 3D ANOMALY DETECTION VIA VIEWPOINT-PRESERVING ALIGNMENT AND RESIDUAL CALIBRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

3D anomaly detection aims to identify defective objects and localize structural deviations from a small set of normal observations. Existing methods often detect anomalies indirectly through learned or engineered representations, or through discrepancies from reconstructed normal geometry. For structural anomalies, however, the anomaly signal is inherently geometric, suggesting a simpler and direct alternative: align a query point cloud with normal references and directly measure their geometric residuals. The challenge is that large residuals are not unique to defects; they can also arise from alignment errors, partial observations, and sampling noise. In this work, we argue that geometric residuals become meaningful anomaly cues when pose-induced discrepancies are controlled and residual evidence is evaluated consistently across spatial scales. Motivated by this principle, we introduce **MARS**, a training-free 3D anomaly-detection method that addresses the residual ambiguity issues through viewpoint-preserving **M**ulti-hypothesis **A**lignment and normal-calibrated **R**esidual **S**coring. MARS first generates candidate poses from precomputed reference views and preserves alternative viewing directions during hypothesis screening before selecting the final pose. Then, query-to-normal residuals are summarized across multiple spatial scales and evaluated against the corresponding residual statistics observed when normal references are compared with one another. With only a few normal references per category and no pretrained models, training, or synthesized anomalies, MARS achieves 97.0%/91.2% point-/object-level AUROC on Real3D-AD and 98.2%/95.7% on Anomaly-ShapeNet, outperforming recent methods on average.

1 INTRODUCTION

3D anomaly detection (3D-AD) aims to identify structural irregularities in 3D observations (e.g., point cloud) and localize the regions in which they occur (Bergmann & Sattlegger, 2023; Horwitz & Hoshen, 2023; Wang et al., 2023; Liu et al., 2023; Li et al., 2024; Ye et al., 2025). This problem is particularly relevant to visual inspection, where anomalous examples are scarce and the possible forms of defects are difficult to enumerate (Liu et al., 2023). In many practical settings, acquiring high-quality 3D scans is itself costly, leaving only a small number of anomaly-free examples for a given object category (Liu et al., 2023). Therefore, in this work, we focus on few-reference 3D anomaly detection, where a query point cloud must be assessed using only a handful of category-matched normal references and no anomalous examples.

Existing 3D-AD methods often infer anomalies through learned or engineered representations of normal geometry. Feature-based approaches compare query descriptors or learned features against normal patterns (Bergmann & Sattlegger, 2023; Horwitz & Hoshen, 2023; Chu et al., 2023; Chen et al., 2023; Zhao et al., 2026; Zhu et al., 2024), while reconstruction-based approaches identify discrepancies from predicted normal geometry (Li et al., 2024; Zhou et al., 2024b; Wu et al., 2026). Other approaches learn geometry-sensitive representations through curvature-guided reconstruction (Zha et al., 2026) or predict point offsets as anomaly evidence, optionally together with normal-codebook similarity (Ye et al., 2025; Kang et al., 2026). Their anomaly signal consequently depends on the chosen representation or on how faithfully normal geometry can be reconstructed. For structural anomalies, however, the evidence we seek is geometric: a defective region should

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.

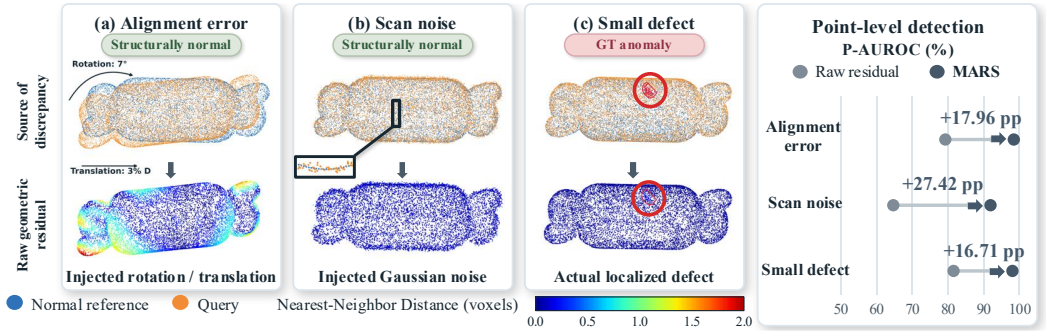

Figure 1: **Limitations of raw geometric residuals.** Alignment error and scan noise can induce residual responses on structurally normal geometry, whereas small defects can produce weak residuals. MARS improves point-level detection across these cases (see Appendix D.3 for details).

deviate from the geometry observed in normal objects (Liu et al., 2023; Yang et al., 2026). This motivates a simpler question and more direct approach: *Can a few normal objects themselves provide enough information to detect anomalies directly from geometric differences?* A natural approach is to align the query with the normal references and use the remaining geometric distances between them, or geometric residuals, as anomaly evidence, since these residuals can be computed directly from observed normal geometry without learning a category-specific representation or synthesizing anomalous examples. Registration-assisted 3D-AD methods (Liu et al., 2023; Liang et al., 2025; Yu et al., 2025; Zheng et al., 2025) already use geometric alignment before subsequent anomaly analysis, and recent work has also begun to consider post-registration residuals themselves as anomaly cues (Yang et al., 2026).

A large geometric residual is not necessarily anomalous, as inaccurate pose estimation, incomplete overlap, and sampling or scan noise can produce similar residuals (Liu et al., 2023; Yu et al., 2025; Yang et al., 2026); conversely, small localized defects can produce only weak pointwise residuals (Figure 1). This ambiguity creates two coupled requirements. First, the query must be aligned reliably despite uncertain pose and partial observations, because remaining alignment error directly enters the anomaly signal. Second, residual evidence must be aggregated consistently across spatial scales, because isolated pointwise discrepancies can arise from sampling noise or remaining local mismatch rather than structural defects. A direct geometric detector must therefore answer two questions: *Is the query aligned correctly?* and *does the remaining discrepancy form spatially supported anomaly evidence?*

We address these challenges with **MARS**, a training-free framework coupling **M**ulti-hypothesis **A**lignment with normal-calibrated **R**esidual **S**coring. MARS maintains diverse pose hypotheses under uncertain pose and partial visibility, then converts multi-scale query-to-normal residuals into comparable anomaly evidence using normal-only empirical calibration. MARS requires only a few normal point clouds and no pretrained models, training, or synthesized anomalies.

On Real3D-AD (Liu et al., 2023) and Anomaly-ShapeNet (Li et al., 2024), MARS achieves the highest average point- and object-level AUROC among the evaluated methods. These results indicate that carefully aligned and spatially aggregated geometric residuals can provide a strong training-free basis for 3D anomaly detection across both partial observations and complete shapes.

Our contributions are summarized as follows:

- We formulate few-reference 3D anomaly detection around post-registration geometric residuals, without learned anomaly representations or reconstruction models.
- We develop a viewpoint-preserving multi-hypothesis alignment strategy that maintains pose diversity during screening and yields tighter geometric alignment before residual scoring.
- We develop normal-calibrated multi-scale residual scoring that converts heterogeneous residual statistics into comparable anomaly evidence and reinforces spatially supported geometric deviations.

2 Related Work

3D anomaly detection. 3D anomaly detection aims to identify defective objects and localize anomalous regions in 3D observations (Liu et al., 2023; Li et al., 2024). Feature-based approaches compare query descriptors with normal patterns (Horwitz & Hoshen, 2023; Zhu et al., 2024), optionally combining geometric features with RGB or projected-view information (Wang et al., 2023; Cao et al., 2024). Reconstruction-based methods recover normal geometry and measure discrepancies from the reconstruction (Li et al., 2024; Zhou et al., 2024b; Wu et al., 2026), whereas offset-prediction methods learn the displacements toward normal geometry (Ye et al., 2025; Kang et al., 2026). Further studies explore geometric representation learning (Zha et al., 2026; Xiao et al., 2026) and feature reconstruction across multiple categories (Cheng et al., 2025; Kim et al., 2026). Zero-shot methods transfer pretrained point, image, and language representations to unseen categories (Zhou et al., 2024a; Li et al., 2026; Deng et al., 2026). In parallel, direct comparison with observed normal geometry can provide a complementary approach, but requires separating defects from geometric discrepancies caused by pose error, partial observations, and sampling.

Registration for anomaly detection. Pose differences complicate geometry-based comparison because misalignment can introduce discrepancies unrelated to defects (Liu et al., 2023; Yu et al., 2025). Registration-assisted methods align observations before feature comparison (Liu et al., 2023; Zhao et al., 2026), internal spatial analysis (Liang et al., 2025), template matching (Liu et al., 2025), or continuous surface modeling (Zheng et al., 2025). Registration can also guide the learning of rotation-invariant anomaly features (Yu et al., 2025). Classical geometric registration (Rusu et al., 2009; Fischler & Bolles, 1981; Besl & McKay, 1992) and learned correspondence models (Wang & Solomon, 2019; Yew & Lee, 2020; Qin et al., 2022) provide tools for estimating alignment, but partial visibility can lead to pose ambiguity, leaving several poses plausible. For anomaly detection, selecting an incorrect pose can make normal regions appear defective.

Geometric residuals for anomaly detection. Yang et al. (2026) combine geometric and feature-space residuals after multi-view registration, using feature differences to complement unreliable pointwise geometric correspondence. However, raw geometric residuals remain difficult to combine across locations and spatial scales because their magnitudes can be affected by sampling density, observed coverage, and residual misalignment (Liu et al., 2023; Yang et al., 2026). MARS instead maps multi-scale geometric residuals to a common empirical evidence scale using normal-reference comparisons, without an additional feature-space residual.

3 Method

Problem statement. Following the standard problem settings of prior work (Liu et al., 2023; Li et al., 2024; Yu et al., 2025), we consider 3D anomaly detection using a small set of K normal reference point clouds from the same object category. Let $\mathcal{P}_j^{\mathrm{r}} = \{\mathbf{p}_{j,i}^{\mathrm{r}}\}_{i=1}^{N_j} \subset \mathbb{R}^3$, $j = 1, \ldots, K$, denote the j-th normal reference point cloud, where $\mathbf{p}_{j,i}^{\mathrm{r}}$ is its i-th 3D point and N_j is the number of points in that reference. Further, we denote an unlabeled query point cloud by $\mathcal{P}^{\mathrm{q}} = \{\mathbf{p}_i^{\mathrm{q}}\}_{i=1}^{N_q} \subset \mathbb{R}^3$, where $\mathbf{p}_i^{\mathrm{q}}$ denotes the i-th query point and N_q is the number of query points. The reference point clouds may differ in rigid pose and are not assumed to be pre-aligned. Given these references, our goal is to detect geometric anomalies in the query and localize them at the point level.

Overview. Our proposed framework **MARS** (viewpoint-preserving **M**ulti-hypothesis **A**lignment and normal-calibrated **R**esidual **S**coring) aims to detect anomalies by measuring how an aligned query differs from normal reference geometry. Two factors make this direct geometric comparison difficult: (1) an incorrect pose can create large geometric residuals, and (2) isolated residuals can also arise from sampling, partial coverage, or small remaining local mismatches. We address the first through viewpoint-preserving multi-hypothesis alignment and the second through spatially aggregated, normal-calibrated residual scoring. As illustrated in Figure 2, we first align the normal references to a common coordinate system (Section 3.1). We then estimate the query pose by screening hypotheses within each viewpoint before refining and comparing candidates across viewpoints (Section 3.2). Finally, we measure query residuals at multiple spatial scales and compare them with residual statistics obtained among normal references (Section 3.3). Reference alignment, the view-conditioned reference bank, and normal calibration are prepared offline (i.e., computed

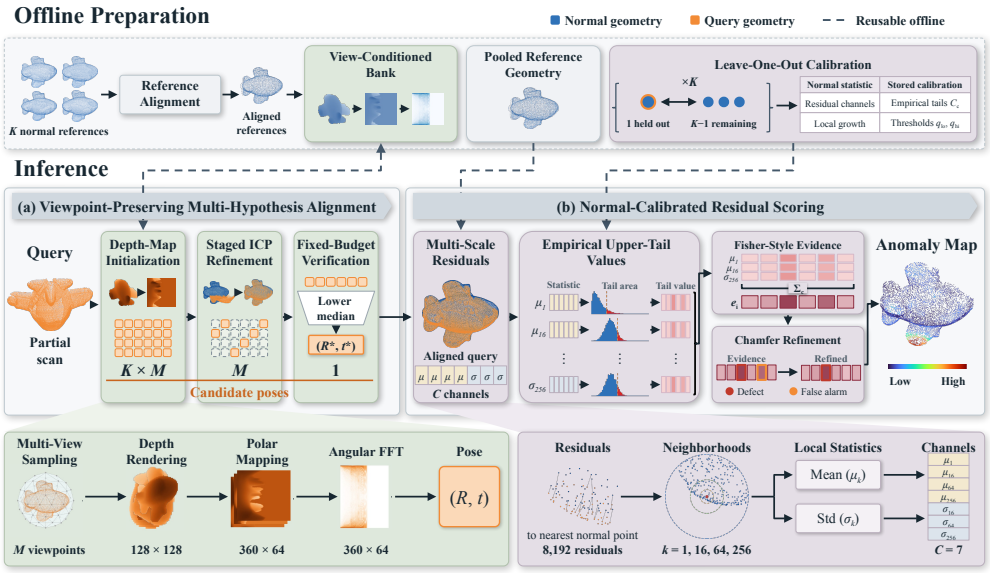

Figure 2: **Framework of MARS.** Offline, normal references are aligned into a shared coordinate system, rendered into a reusable view-conditioned bank, and used to construct leave-one-out calibration statistics. At inference, (a) depth-map initialization generates multiple pose hypotheses, which are progressively refined and verified, and (b) multi-scale query-to-normal residuals are evaluated against the stored normal calibration to produce anomaly scores.

once per category and reused for all queries). MARS requires no model training, pretrained models, or synthesized anomalies.

3.1 Common Reference Alignment

Arbitrary pose differences among the normal references would contaminate subsequent geometric comparisons with discrepancies unrelated to object shape. We center the first normal reference and orient its principal axes using skewness to define a canonical anchor. Using only this anchor to build the view-conditioned target bank, we independently align each of the remaining $K-1$ normal references to it with the procedure described in Section 3.2. We denote the resulting aligned references by $\widehat{\mathcal{P}}_j^{\mathrm{r}}$, which provide the shared geometric basis for both query alignment and residual scoring. Implementation details are provided in Appendix B.1.

3.2 Viewpoint-Preserving Multi-Hypothesis Alignment

A query may have an unknown orientation and expose only part of an object, making a single pose initialization unreliable. We therefore generate multiple candidate alignments from different reference viewpoints and progressively screen and refine them while retaining one candidate for each viewpoint. This preserves alternative viewing directions during screening, allowing them to undergo further refinement before final pose selection.

View-conditioned reference bank. A partial query observes only the object surface visible from its viewing direction, so directly using the full reference shape can introduce geometry that is absent from the query. We therefore render each aligned reference orthographically from M geodesic viewpoints distributed over a sphere, obtaining a viewpoint-specific representation of its visible surface. At each pixel, we retain the frontmost point, producing a depth map of the surface visible from that direction. For reference j and viewpoint m, we store the depth map $\mathbf{D}_{j,m}^{\mathrm{r}}$ and its polar Fourier descriptor $F_{j,m}^{r}$ for in-plane rotation estimation through angular correlation. The resulting bank is

$$\mathcal{B} = \left\{ (\mathbf{D}_{j,m}^{\mathrm{r}}, \mathbf{F}_{j,m}^{\mathrm{r}}) : j = 1, \ldots, K,\ m = 1, \ldots, M \right\}. \tag{1}$$

The viewpoint specifies a candidate viewing direction, while the descriptor is used to estimate rotation within the image plane. The bank is precomputed and shared across queries. Further details of the polar representation are provided in Appendix A.1.

Viewpoint-and-roll pose initialization. A candidate viewpoint determines which side of the object is observed, but not how the object is rotated within the projected image; this remaining in-plane rotation is the *roll*. We estimate this remaining rotation by comparing the query depth map with the corresponding reference depth map through angular correlation. At inference, we render the query using the same orthographic settings and compute its polar Fourier descriptor. In polar coordinates, an in-plane rotation becomes a circular shift along the angular axis. We therefore estimate the relative in-plane rotation, or roll, for each bank entry using normalized angular Fourier correlation.

The bank viewpoint and estimated roll together determine a candidate rotation relative to the query projection frame. Translation aligns the query projection origin with the stored reference-view center. This gives one initial rigid transformation for each reference–viewpoint pair, forming a hypothesis set $\mathcal{H}_0$ with KM candidates. These candidates cover different orientations across references and viewpoints without query-specific local descriptor extraction or correspondence matching.

Viewpoint-preserving screening and refinement. The initial bank favors pose coverage over compactness: each viewpoint contributes K reference-specific candidates, while the coarse viewpoint–roll estimate can retain alignment error that would later contaminate anomaly residuals. Candidates from different normal references can be redundant when they represent the same viewing direction, whereas different viewpoints correspond to distinct pose possibilities. We therefore first retain one reference-specific candidate within each viewpoint and then refine the surviving viewpoint candidates before final verification.

Let j_h denote the reference associated with hypothesis h. We apply batched point-to-plane ICP (Chen & Medioni, 1991) to all KM hypotheses using a downsampled query. Within each viewpoint, we retain the reference-specific hypothesis with the smallest median directed query-to-reference residual. This reduces $\mathcal{H}_0$ to $\mathcal{H}_1 \subset \mathcal{H}_0$ with one candidate per viewpoint, so $|\mathcal{H}_1| = M$, retaining one candidate for every bank viewpoint. Each retained candidate then undergoes further ICP refinement on the voxel-downsampled query, yielding a transformation $(\widehat{\mathbf{R}}_h, \widehat{\mathbf{t}}_h)$ for $h \in \mathcal{H}_1$.

This step corrects residual alignment error left by the discrete viewpoint and roll initialization. Details of the point-to-plane objective and staged refinement are provided in Appendix A.2, and the corresponding implementation settings are given in Appendix B.2.

Pose selection. After refinement, we compare the candidates across viewpoints on the same fixed-size query-point subset $\mathcal{I}_v$, so that all refined hypotheses are compared on the same query points. For hypothesis h and verification point i, we define its verification residual as

$$d_{h,i} = \min_{\mathbf{p}^{\mathrm{r}} \in \widehat{\mathcal{P}}^{\mathrm{r}}_{j_h}} \left\| \widehat{\mathbf{R}}_h \mathbf{p}^{\mathrm{q}}_i + \widehat{\mathbf{t}}_h - \mathbf{p}^{\mathrm{r}} \right\|_2 . \tag{2}$$

To reduce the influence of localized defects and isolated outliers during pose selection, we select the final hypothesis with the smallest lower-median residual:

$$h^* = \arg\min_{h \in \mathcal{H}_1} \mathrm{lmed}_{i \in \mathcal{I}_{\mathrm{v}}}\, d_{h,i}, \tag{3}$$

Here, lmed denotes the lower median, while the shared verification subset provides a common comparison criterion across hypotheses. The selected transformation $(\mathbf{R}^*, \mathbf{t}^*)$ associated with h^* gives the aligned query $\widehat{\mathcal{P}}^{\mathrm{q}} = \{\widehat{\mathbf{p}}^{\mathrm{q}}_i = \mathbf{R}^* \mathbf{p}^{\mathrm{q}}_i + \mathbf{t}^*\}_{i=1}^{N_q}$.

3.3 Normal-calibrated residual scoring

Even after alignment, a large query-to-reference residual need not indicate a defect because sampling differences, partial coverage, and residual alignment error can produce similar responses. We therefore summarize residuals spatially and map the resulting statistics to a common empirical scale using normal-reference comparisons. Concretely, we first measure how far each query point lies from normal geometry, then summarize whether this discrepancy persists across its neighborhood, and finally convert the resulting multi-scale statistics into comparable anomaly evidence. We then refine and aggregate the resulting anomaly evidence to obtain point- and object-level scores.

Residuals to pooled normal-reference geometry. Although the selected pose is obtained using one reference, scoring against only that instance could make residuals overly dependent on its particular sampling and surface coverage. We therefore use all aligned normal references as the scoring target.

From each reference, we sample $\mathcal{S}_j^{\mathrm{r}} \subseteq \widehat{\mathcal{P}}_j^{\mathrm{r}}$ and construct a pooled reference point cloud:

$$\mathcal{P}_{\cup}^{\mathrm{r}} = \bigcup_{j=1}^{K} \mathcal{S}_j^{\mathrm{r}}. \tag{4}$$

This pooled reference geometry serves as the comparison target for computing query-to-normal residuals. Let $\mathcal{I}_{\mathrm{s}} \subseteq \{1, \ldots, N_q\}$ denote the query points used for residual scoring. For each $i \in \mathcal{I}_{\mathrm{s}}$, we compute the nearest-neighbor residual to the pooled normal geometry and normalize it by the median query residual:

$$r_i = \min_{\mathbf{p}^{\mathrm{r}} \in \mathcal{P}_{\cup}^{\mathrm{r}}} \left\| \widehat{\mathbf{p}}_i^{\mathrm{q}} - \mathbf{p}^{\mathrm{r}} \right\|_2, \qquad \widetilde{r}_i = \frac{r_i}{\max(\mathrm{median}_{l \in \mathcal{I}_{\mathrm{s}}} r_l, \epsilon)}. \tag{5}$$

Here, $\epsilon > 0$ prevents division by zero. The directed query-to-reference distance measures deviations at observed query points without directly penalizing reference surfaces absent from the query. Median normalization expresses residuals relative to the overall residual level of the query while limiting the influence of a few localized large responses on the normalization factor.

Multi-scale residual statistics. A single large residual may reflect isolated noise, whereas a defect can affect a region of varying size. Thus, we summarize the normalized residuals using local means and standard deviations: the mean captures whether neighboring points consistently have large residuals, while the standard deviation captures how much those residuals vary within the neighborhood. Let $\mathcal{N}_k(i)$ contain the k nearest scored query points to $\widehat{\mathbf{p}}_i^{\mathrm{q}}$, including the point i itself. Then, for neighborhood-size sets for local means $\mathcal{K}_\mu$ and standard deviations $\mathcal{K}_\sigma$, we compute

$$\mu_k(i) = \mathrm{mean}_{l \in \mathcal{N}_k(i)} \widetilde{r}_l, \ k \in \mathcal{K}_\mu, \qquad \sigma_k(i) = \mathrm{std}_{l \in \mathcal{N}_k(i)} \widetilde{r}_l, \ k \in \mathcal{K}_\sigma. \tag{6}$$

These statistics form $C = |\mathcal{K}_\mu| + |\mathcal{K}_\sigma|$ residual channels, denoted by $\{s_{i,c}\}_{c=1}^{C}$.

Leave-one-out normal calibration. Because residual channels have different normal ranges, their raw magnitudes are not directly comparable. We therefore estimate a separate normal baseline for each channel through leave-one-out comparisons, treating each aligned normal reference as a query against the remaining references. Holding out the source reference prevents trivial self-matches. We apply the same residual normalization and neighborhood statistics as above, then pool the resulting values into an empirical calibration collection $\mathcal{C}_c$ for each channel.

For a query statistic $s_{i,c}$, we compute the fraction of calibration values that are at least as large, using add-one smoothing. The resulting empirical upper-tail value is then combined into the channel-wise evidence e_i using a Fisher-style log aggregation (Fisher, 1925):

$$p_{i,c} = \frac{1 + \sum_{z \in \mathcal{C}_c} \mathbf{1}[z \geq s_{i,c}]}{1 + |\mathcal{C}_c|}, \qquad e_i = -2 \sum_{c=1}^{C} \log p_{i,c}. \tag{7}$$

Add-one smoothing keeps the logarithm finite. A small $p_{i,c}$ indicates that the query statistic exceeds most values observed among normal references. Such a response lies in the upper tail of its channel-specific normal baseline and therefore provides stronger anomaly evidence. Because the channels share residuals and overlapping neighborhoods, they are dependent. Accordingly, we use e_i only as an anomaly-ranking score rather than as a calibrated probability or chi-square test statistic. Additional calibration details are provided in Appendix A.3, B.1 and D.4.

Chamfer-based evidence refinement. Residual pose error or incomplete coverage can still produce strong responses. As an auxiliary geometric check, we characterize how the local mismatch around a high-scoring point grows as its neighborhood expands. We compare this growth across two spatial scales with the range observed in leave-one-out normal comparisons and partially attenuate unusually large responses with a bounded correction. This refinement is conservative by design and cannot dominate the primary anomaly evidence. The exact formulation is provided in Appendix A.4 and D.5.

Point- and object-level aggregation. Pointwise evidence may contain isolated peaks. We first apply median smoothing to suppress unsupported responses while preserving spatially supported evidence, yielding the pointwise anomaly score a_i.

For object-level detection, let $\mathcal{T}_o \subseteq \mathcal{I}_s$ index the n_o largest smoothed scores. We aggregate them using the harmonic mean:

$$A(\mathcal{P}^{\mathrm{q}}) = \left(\frac{1}{n_o} \sum_{l \in \mathcal{T}_o} a_l^{-1} \right)^{-1}. \tag{8}$$

We define the harmonic mean as zero, i.e., $A(\mathcal{P}^{\mathrm{q}}) = 0$, if any selected score is zero. The harmonic mean reduces the influence of a single extreme score while retaining sensitivity to localized defects. Further details of the aggregation procedure are provided in Appendix A.5 and D.5.

4 Experiments

4.1 Experimental setup

Datasets and evaluation. We evaluate MARS on two geometry-only 3D anomaly-detection benchmarks: Real3D-AD (Liu et al., 2023) and Anomaly-ShapeNet (Li et al., 2024). Real3D-AD contains 12 real-world categories with partial single-view point clouds, whereas Anomaly-ShapeNet contains 40 categories of complete, object-centered point clouds with synthetic shape anomalies. For both benchmarks, we use four anomaly-free references per category to construct the reference bank and calibrate the scoring channels. Following prior work (Liu et al., 2023; Li et al., 2024; Yu et al., 2025; Zhao et al., 2026), we evaluate performance using category-macro point- and object-level AUROC (P/O-AUROC). Following R3D-AD (Zhou et al., 2024b), we randomly sample query points for scoring. Mean rank averages category-wise ranks over methods with complete results, with ties averaged.

Implementation details. MARS uses no pretrained models or model training. Unless stated otherwise, we use four normal references per category, $M = 42$ geodesic viewpoints, and 8,192 query points for residual scoring. All settings are shared across categories and both benchmarks, with full implementation details provided in Appendix B.

Baselines. We compare MARS with eleven baselines spanning memory-bank matching (PatchCore (Roth et al., 2022), BTF (Horwitz & Hoshen, 2023)), learned anomaly modeling (IMRNet (Li et al., 2024), CASL (Zha et al., 2026), AARD (Wu et al., 2026)), and registration-assisted scoring (Reg3D-AD (Liu et al., 2023), ISMP (Liang et al., 2025), Reg2Inv (Yu et al., 2025), PointCore (Zhao et al., 2026), PASDF (Zheng et al., 2025), and MVR-DR (Yang et al., 2026)).

4.2 Main results

4.2.1 Quantitative results

As shown in Table 1, MARS achieves the highest average P/O-AUROC on both benchmarks in the four-shot setting without model training or synthetic anomaly generation, and retains the highest reported averages in the two-shot comparison. Together with the component analyses in Section 4.3, the consistent results across partial scans, complete shapes, and reduced reference counts support our central premise that carefully aligned and spatially aggregated geometric residuals provide a strong basis for anomaly detection. Full category-wise results for both benchmarks are provided in Appendix C.

4.2.2 Qualitative results

Figure 3 compares registration and anomaly localization with Reg2Inv (Yu et al., 2025). MARS shows tighter alignment with fewer pose-induced responses, concentrating anomaly evidence more closely on the annotated defects.

4.3 Analysis and ablation studies

We analyze MARS from three perspectives: (1) viewpoint robustness, registration efficiency, and geometric fit; (2) the contributions of roll search, ICP refinement, and candidate pruning; and (3)

Table 1: **Overall AUROC comparison on Real3D-AD and Anomaly-ShapeNet.** MR denotes mean rank. For 2-shot, O-AUROC uses common top-16 harmonic pooling, and MR ranks only the four methods. ✗/✓denotes without/with method-specific training. **Bold**/underlined values denote the **best**/second-best results within the same shot setting.

Shot	Method	Training	Real3D-AD				Anomaly-ShapeNet			
			P-AUROC		O-AUROC		P-AUROC		O-AUROC	
			Avg.↑	MR↓	Avg.↑	MR↓	Avg.↑	MR↓	Avg.↑	MR↓
Non-registration-based										
4	IMRNet*	✓	-	-	72.5	8.42	65.0	8.89	66.1	6.69
	CASL*	✓	88.2	3.25	82.3	5.17	89.9	3.95	88.7	3.93
	AARD*	✓	86.0	4.08	82.0	4.58	93.6	2.95	-	-
	PatchCore*	✗	69.2	8.67	68.2	8.88	58.0	9.75	56.8	8.25
	BTF*	✗	72.2	7.58	63.5	9.92	62.8	9.30	52.8	8.78
Registration-based										
2	Reg2Inv†	✓	90.6	1.67	80.2	2.17	90.3	1.93	84.5	1.96
	Reg3D-AD†	✗	69.5	3.83	64.3	3.75	58.7	3.78	51.9	3.73
	PointCore†	✗	76.7	3.00	80.0	2.83	65.9	3.23	61.9	3.10
	MARS (Ours)	✗	**95.9**	**1.50**	**88.9**	**1.25**	**98.1**	**1.08**	**95.5**	**1.21**
4	PASDF*	✓	74.5	6.42	80.2	6.38	89.7	4.23	90.0	3.50
	Reg2Inv*	✓	87.8	3.46	78.0	5.63	88.2	4.10	86.1	3.91
	ISMP*	✗	83.6	4.92	76.7	6.83	69.1	8.05	-	-
	MVR-DR*	✗	-	-	82.0	5.75	85.6	5.25	95.0	2.38
	Reg3D-AD*	✗	70.0	7.75	70.4	9.08	66.8	8.31	57.2	8.34
	PointCore*	✗	73.1	7.50	82.9	5.13	-	-	67.7	7.00
	MARS (Ours)	✗	**97.0**	**1.38**	**91.2**	**2.25**	**98.2**	**1.23**	**95.7**	**2.24**

* Reported from the original paper or benchmark table when direct reproduction was not reliably available.
† Reproduced with publicly released code.

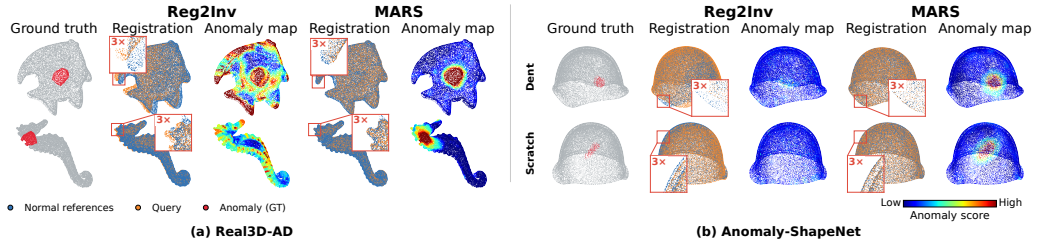

Figure 3: **Qualitative comparison across categories and defect types.** Boxes labeled 3× show alignment differences at threefold magnification; blue-to-red indicates increasing anomaly scores.

multi-scale residual statistics, channel fusion, and Chamfer refinement. We report Real3D-AD results here, with the corresponding component ablations on Anomaly-ShapeNet in Appendix D.1.

4.3.1 Registration robustness and efficiency

Viewpoint robustness and registration efficiency. Figure 4 shows that denser viewpoint coverage improves accuracy until saturation while retaining favorable latency relative to registration-based baselines (Liu et al., 2023; Yu et al., 2025; Zheng et al., 2025; Zhao et al., 2026). The FPFH+RANSAC variant achieves comparable accuracy but is slower, supporting our use of reusable view-conditioned depth maps to generate diverse pose hypotheses without query-specific correspondence search.

Registration residuals on normal queries. Figure 4(d) shows that MARS has the lowest median and narrowest interquartile range of nearest-neighbor residuals on normal queries, indicating more consistent geometric fit before anomaly scoring. These residuals compare query-to-reference fit under a common evaluation, not ground-truth pose error.

4.3.2 Component ablations

Alignment. Table 2 shows that ICP refinement is critical to the effectiveness of the alignment stage, while candidate pruning retains the resulting detection performance with substantially lower refinement cost. Together with the gain from roll search, these results support preserving diverse coarse pose hypotheses while reserving full refinement for promising candidates.

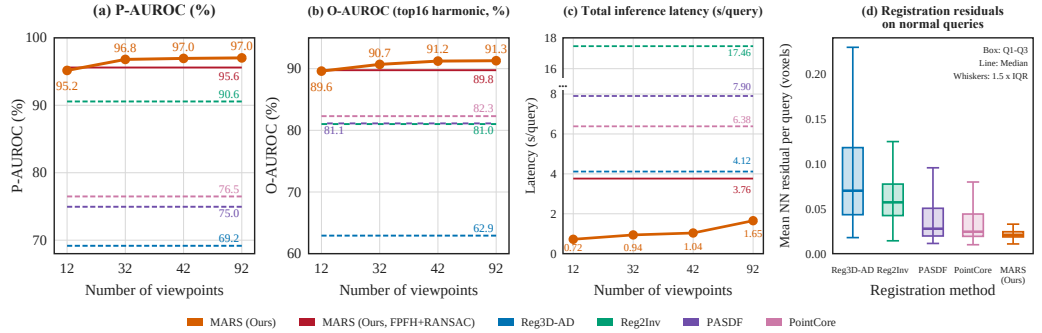

Figure 4: **Accuracy, latency, and normal-query registration residuals on Real3D-AD.** In (a)-(c), solid lines denote MARS and its FPFH+RANSAC variant; dashed lines denote registration-based baselines. MARS trades accuracy for speed across viewpoints while remaining more accurate and faster than the baselines; (d) shows its lower median of query-mean residuals on normal queries. All results are reproduced under the same evaluation conditions.

Table 2: **Alignment-component ablation on Real3D-AD.** Latency is relative to MARS.

Variant	P/O-AUROC (%)	Rel. latency ↓
No ICP refinement	53.1 / 49.9	0.97×
No roll search	95.0 / 89.0	1.02×
No candidate pruning	**97.0 / 91.3**	2.10×
MARS (Ours)	**97.0** / 91.2	**1.00**×

Table 3: **Residual-scoring ablation on Real3D-AD.**

Residual statistics	Fusion	P/O-AUROC (%)
Pointwise	–	83.7 / 91.4
Multi-scale means	Average	92.6 / 87.5
Multi-scale means + std. devs.	Average	94.9 / 89.6
Multi-scale means + std. devs.	Fisher-style	**97.0** / 90.8
+ Chamfer refinement (MARS)	**Fisher-style**	**97.0 / 91.2**

Residual scoring. Table 3 shows that multi-scale residual statistics substantially strengthen point-level localization, with local dispersion providing complementary information beyond neighborhood means. Fisher-style aggregation further improves the joint use of these channels, while Chamfer refinement provides a targeted correction for residual geometric mismatch. These trends support spatially aggregated residual evidence with channel-wise fusion.

5 Limitations

MARS focuses on geometric anomalies under rigid alignment; appearance-only defects and substantial non-rigid variation are outside its current scope. Its query-to-reference residual assigns anomaly evidence only to observed points, so completely missing regions are not scored directly. Although this avoids penalizing surfaces absent solely due to partial visibility, it can also leave missing defects undetected when the remaining geometry shows little deviation. Future work could incorporate visibility-aware reference-to-query comparison to distinguish missing defects from unobserved surfaces.

6 Conclusion

In this work, we introduce MARS, a training-free method built on the idea that geometric deviations should be interpreted relative to the variation observed among normal objects. To make this comparison reliable, MARS couples viewpoint-preserving multi-hypothesis alignment with normal-calibrated residual scoring. The alignment stage preserves alternative viewing directions during pose search, while the scoring stage evaluates geometric residuals at multiple spatial scales against statistics obtained among normal references. Experiments on Real3D-AD and Anomaly-ShapeNet demonstrate strong detection and localization performance using only four normal references per category, without pretrained models or synthesized anomalies. These results support using normal observations both as comparison geometry and as a baseline for interpreting deviations, enabling anomaly detection directly from observed geometry.

7 AI USE

Generative AI tools were used throughout this project as interactive assistants for research ideation and execution. They supported the exploration and refinement of successive formulations of MARS, including its multi-hypothesis alignment, normal-relative residual calibration, geometric refinement, and score aggregation. They also assisted in planning and executing prototype, ablation, robustness, and efficiency experiments; diagnosing unsuccessful variants; checking experimental configurations; and examining consistency among the implementation, reported settings, and experimental outputs.

Generative AI tools were also used to polish author-written text for clarity, grammar, organization, and readability. The authors reviewed all AI-assisted work, independently verified the reported results and technical claims, and made all final research and editorial decisions. The authors take full responsibility for the content of this work.

REFERENCES

Paul Bergmann and David Sattlegger. Anomaly detection in 3d point clouds using deep geometric descriptors. In *WACV*, 2023.

P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992.

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

Ruitao Chen, Guoyang Xie, Jiaqi Liu, Jinbao Wang, Ziqi Luo, Jinfan Wang, and Feng Zheng. Easynet: An easy network for 3d industrial anomaly detection. In *ACM MM*, 2023.

Y. Chen and G. Medioni. Object modeling by registration of multiple range images. In *ICRA*, 1991.

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 *IJCAI*, 2025.

Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. In *ICML*, 2023.

Zehao Deng, An Liu, and Yan Wang. Gs-clip: Zero-shot 3d anomaly detection by geometry-aware prompt and synergistic view representation learning. In *CVPR*, 2026.

Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981.

Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh, 1925.

Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *CVPR Workshops*, 2023.

Xueyang Kang, Zizhao Li, Tian Lan, Dong Gong, Kourosh Khoshelham, and Liangliang Nan. Hierarchical point-patch fusion with adaptive patch codebook for 3d shape anomaly detection. In *CVPR*, 2026.

SuYeon Kim, Wongyu Lee, and MyeongAh Cho. A semantically disentangled unified model for multi-category 3d anomaly detection. In *CVPR*, 2026.

Kaiqiang Li, Gang Li, Mingle Zhou, Min Li, Delong Han, and Jin Wan. Back to point: Exploring point-language models for zero-shot 3d anomaly detection. In *CVPR*, 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 *CVPR*, 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 *AAAI*, 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. In *NeurIPS*, 2023.

Yi Liu, Changsheng Zhang, and Yufei Yang. Template3d-ad: Point cloud template matching method based on center points for 3d anomaly detection. In *IJCAI*, 2025.

Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *CVPR*, 2022.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, 2022.

Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *ICRA*, 2009.

Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *ICCV*, 2019.

Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *CVPR*, 2023.

Linchun Wu, Qin Zou, Yuanhao Yue, and Zhongyuan Wang. Geometry-aligned and anomaly-aware reconstruction for 3d anomaly detection. In *CVPR*, 2026.

Haibo Xiao, Hanzhe Liang, Jie Zhou, Jinbao Wang, and Can Gao. Learning discriminative signed distance functions from multi-scale level-of-detail features for 3d anomaly detection. *arXiv preprint arXiv:2605.03437*, 2026.

Yuxing Yang, Zeyu Fu, Yang Li, Liewei Wang, Siyue Yu, and Jimin Xiao. Probing 3d anomalies via multi-view registration and dual-residual analysis. *Neurocomputing*, 2026.

Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, and Kaizhu Huang. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection. In *CVPR*, 2025.

Zi Jian Yew and Gim Hee Lee. Rpm-net: Robust point matching using learned features. In *CVPR*, 2020.

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 *NeurIPS*, 2025.

Yaohua Zha, Xue Yuerong, Chunlin Fan, Yuansong Wang, Tao Dai, Ke Chen, and Shu-Tao Xia. Casl: Curvature-augmented self-supervised learning for 3d anomaly detection. In *AAAI*, 2026.

Baozhu Zhao, Xiaohan Zhang, Jingfeng Guo, and Qi Liu. Pointcore: An efficient framework for unsupervised point cloud anomaly detection using joint local-global features. *Neural Networks*, 2026.

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 *ICCV*, 2025.

Qihang Zhou, Jiangtao Yan, Shibo He, Wenchao Meng, and Jiming Chen. Pointad: Comprehending 3d anomalies from points and pixels for zero-shot 3d anomaly detection. In *NeurIPS*, 2024a.

Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3d-ad: Reconstruction via diffusion for 3d anomaly detection. In *ECCV*, 2024b.

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 *ACM MM*, 2024.

A Additional Method Details

This section specifies the standard mathematical operations used by MARS and clarifies how they are organized within the proposed alignment and scoring pipeline. The individual operations–circular correlation, point-to-plane ICP, empirical tail ranking, directed Chamfer discrepancy, and robust aggregation–are established tools. The contribution of MARS lies in combining them into a viewpoint-preserving multi-hypothesis alignment procedure and a normal-calibrated residual-scoring framework. Complete parameter values and backend assignments are provided in Appendix B.

A.1 Polar-Fourier pose initialization

Let $D(\rho, \theta)$ denote a depth map resampled in polar coordinates around its image center. An in-plane rotation by $\Delta\theta$ induces a circular shift along the angular dimension:

$$D_{\Delta\theta}(\rho, \theta) = D(\rho, \theta - \Delta\theta). \tag{A.1}$$

Consequently, the relative roll between two projected depth maps can be estimated through circular correlation along their angular dimension.

For a query descriptor $\mathbf{F}^{\mathrm{q}}$ and a bank descriptor $\mathbf{F}^{\mathrm{r}}_{j,m}$, we evaluate normalized correlation over the discrete angular shifts and select

$$\widehat{\ell}_{j,m} = \arg\max_{\ell} \operatorname{corr}\left(\mathbf{F}^{\mathrm{q}}, \operatorname{shift}\left(\mathbf{F}^{\mathrm{r}}_{j,m}, \ell\right)\right). \tag{A.2}$$

The shift index $\widehat{\ell}_{j,m}$ determines the corresponding coarse roll angle $\rho_{j,m}$, while viewpoint m determines the candidate viewing direction. Let $\mathbf{E} = [\mathbf{e}_u, \mathbf{e}_v, \mathbf{d}]$ denote the query projection frame and $\mathbf{B}_{j,m} = [\mathbf{a}_{j,m}, \mathbf{b}_{j,m}, \mathbf{v}_{j,m}]$ the corresponding bank frame. For each reference–viewpoint pair, the coarse rigid transformation is recovered as $\mathbf{R}_{j,m} = \mathbf{B}_{j,m}\mathbf{R}_z(-\rho_{j,m})\mathbf{E}^\top$ and $\mathbf{t}_{j,m} = \mathbf{c}_{j,m} - \mathbf{R}_{j,m}\mathbf{p}_0$, where $\rho_{j,m}$ is the estimated roll, $\mathbf{c}_{j,m}$ is the stored view center, and $\mathbf{p}_0$ is the query projection origin. The additional image-plane and depth offsets are initialized to zero. Combining every reference j with every viewpoint m therefore produces KM initial rigid-pose hypotheses.

This construction differs from correspondence-based initialization in that the query is not required to generate and match local geometric descriptors. The reference projections and their polar-Fourier descriptors are computed once and reused across all queries. The resulting poses are coarse because of viewpoint discretization, depth-map resolution, and imperfect centering; the subsequent ICP stages correct these residual alignment errors.

A.2 Staged point-to-plane refinement

For a source point $\mathbf{x}_i$, its target correspondence $\mathbf{y}_{\pi(i)}$, and the corresponding target normal $\mathbf{n}_{\pi(i)}$, point-to-plane ICP minimizes

$$\sum_i \left[\mathbf{n}^\top_{\pi(i)}\left(\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_{\pi(i)}\right)\right]^2. \tag{A.3}$$

We use this standard objective for both hypothesis screening and subsequent pose refinement.

The distinguishing aspect of our procedure is the allocation of refinement effort. All KM reference–viewpoint hypotheses first receive a short batched ICP update. Within each viewpoint, the reference-specific hypothesis with the smallest median directed residual is retained. This structured screening reduces the candidate set from KM to M without prematurely discarding viewpoint diversity. Only these M candidates undergo further ICP refinement.

After refinement, every retained candidate is evaluated on the same verification subset $\mathcal{I}_{\mathrm{v}}$. Let $\{d_{h,(1)}, \ldots, d_{h,(|\mathcal{I}_{\mathrm{v}}|)}\}$ denote the sorted directed residuals for hypothesis h. We define the lower median as

$$\operatorname{lmed}_{i\in\mathcal{I}_{\mathrm{v}}} d_{h,i} = d_{h,(\lfloor(|\mathcal{I}_{\mathrm{v}}|+1)/2\rfloor)}. \tag{A.4}$$

Using a shared verification subset makes candidate scores directly comparable, while the lower median limits the influence of localized defects and isolated outliers. The final pose is selected according to Equation 3.

A.3 EMPIRICAL TAIL EVIDENCE

The residual channels have different normal ranges, so their raw magnitudes do not provide a common channel-wise evidence scale. For each channel c, the leave-one-out normal-reference comparisons form an empirical calibration collection $\mathcal{C}_c$. Equation 7 converts a query statistic $s_{i,c}$ into an add-one-smoothed empirical upper-tail rank.

The add-one term has two practical consequences. First, it prevents a zero tail value when the query statistic exceeds every calibration value. Second, it bounds the smallest attainable value by

$$p_{i,c} \geq \frac{1}{1 + |\mathcal{C}_c|}, \tag{A.5}$$

preventing an individual channel from contributing unbounded evidence. The non-strict comparison $\mathbf{1}[z \geq s_{i,c}]$ also handles ties conservatively by including calibration values equal to the query statistic.

We aggregate the channel-wise tail values using the Fisher-style log combination $e_i = -2\sum_c \log p_{i,c}$, which has the same algebraic form as Fisher's combination statistic. However, the channels share the same underlying residuals and use overlapping spatial neighborhoods, so they are not independent. We therefore use e_i only as a monotone evidence-ranking score. It is not interpreted as a calibrated probability, and no chi-square significance claim is made.

A.4 CHAMFER-BASED EVIDENCE REFINEMENT

Let $\mathcal{X}^{\mathrm{q}} = \{x_i\}_{i \in \mathcal{I}_{\mathrm{s}}}$ denote the aligned scored query points, let $\mathcal{X}^{\mathrm{r}}$ denote the pooled aligned normal-reference geometry, and let e_i be the unrefined anomaly evidence of x_i. We apply geometric refinement to the n_{g} points with the largest evidence values:

$$\mathcal{I}_{\mathrm{g}} = \mathrm{Top}_{n_{\mathrm{g}}}\left(\{e_i\}_{i \in \mathcal{I}_{\mathrm{s}}}\right). \tag{A.6}$$

For each candidate x_i, we locate its nearest point y_i in the pooled aligned normal-reference geometry $\mathcal{X}^{\mathrm{r}}$. For each scale $K \in \{K_{\mathrm{s}}, K_{\mathrm{l}}\}$, let $\mathcal{Q}_i^{(K)}$ contain the K nearest query points to x_i, and let $r_i^{(K)}$ denote the distance to its farthest member. We use this radius to construct the coverage-matched reference neighborhood

$$\mathcal{R}_i^{(K)} = \left\{ y \in \mathcal{X}^{\mathrm{r}} \,\middle|\, \|y - y_i\|_2 \leq r_i^{(K)} \right\}. \tag{A.7}$$

The directed reference-to-query discrepancy at scale K is

$$c_i^{(K)} = \frac{1}{|\mathcal{R}_i^{(K)}|} \sum_{y \in \mathcal{R}_i^{(K)}} \min_{x \in \mathcal{Q}_i^{(K)}} \|y - x\|_2. \tag{A.8}$$

The scale-dependent increase is defined as

$$g_i = c_i^{(K_{\mathrm{l}})} - c_i^{(K_{\mathrm{s}})}. \tag{A.9}$$

A large g_i indicates that the local discrepancy expands as the neighborhood scale increases, as can occur under residual alignment error or incomplete query coverage.

We calibrate the normal range of g_i through leave-one-out comparisons among the normal references. For each held-out source j, let $\mathcal{G}_j$ denote its set of scale-dependent increases and define

$$q_{\mathrm{lo}}^{(j)} = \mathrm{Quantile}_{\alpha_{\mathrm{lo}}}(\mathcal{G}_j), \qquad q_{\mathrm{hi}}^{(j)} = \mathrm{Quantile}_{\alpha_{\mathrm{hi}}}(\mathcal{G}_j). \tag{A.10}$$

We aggregate the source-specific statistics using their median:

$$q_{\mathrm{lo}} = \mathrm{median}_j\, q_{\mathrm{lo}}^{(j)}, \qquad q_{\mathrm{hi}} = \mathrm{median}_j\, q_{\mathrm{hi}}^{(j)}. \tag{A.11}$$

Here, q_{lo} determines the onset of attenuation, whereas q_{hi} controls its exponential scale.

The attenuation strength is

$$\psi_i = 1 - \exp\left(-\frac{[g_i - q_{\mathrm{lo}}]_+}{q_{\mathrm{hi}}}\right), \qquad [x]_+ = \max(x, 0). \tag{A.12}$$

We obtain the refined evidence as

$$\bar{e}_i = \begin{cases} e_i \left(1 - \lambda \psi_i\right), & i \in \mathcal{I}_{\mathrm{g}}, \\ e_i, & i \notin \mathcal{I}_{\mathrm{g}}. \end{cases} \tag{A.13}$$

Candidates within the calibrated normal range remain unchanged, while larger scale-dependent increases receive progressively stronger attenuation. The bound λ prevents the refinement from removing evidence produced by broad geometric defects that exhibit a similar discrepancy pattern. The values of n_{g}, K_{s}, K_{l}, α_{lo}, α_{hi}, and λ are specified in Appendix B.

A.5 Robust residual aggregation

Chamfer refinement is followed by median smoothing on the scored query graph. Let $\mathcal{N}_8(i)$ denote the self-inclusive eight-neighbor set of point i. The final pointwise score is

$$a_i = \mathrm{median}_{l \in \mathcal{N}_8(i)} \, \bar{e}_l.$$

This operation preserves locally supported responses while suppressing isolated score fluctuations.

For object-level detection, let $\mathcal{T}_{\mathrm{o}}$ contain the n_{o} largest smoothed point scores. The object score is their harmonic mean,

$$A(\mathcal{P}^{\mathrm{q}}) = \frac{n_{\mathrm{o}}}{\sum_{l \in \mathcal{T}_{\mathrm{o}}} a_l^{-1}}, \tag{A.14}$$

which we define as zero if any score in $\mathcal{T}_{\mathrm{o}}$ is zero (cf. Equation 8 in the main paper). For positive scores, we normalize the selected scores by their maximum before reciprocal evaluation and rescale the result afterward; this computation is algebraically equivalent to the harmonic mean in Equation 8. Compared with maximum pooling, harmonic aggregation reduces the influence of a single unusually high point score and instead requires support from multiple top-ranked responses. Restricting the aggregation to $\mathcal{T}_{\mathrm{o}}$ prevents a spatially localized defect from being diluted by the predominantly normal object surface.

B Reproducibility Details

This section provides the implementation settings omitted from the main paper, including offline reference preparation, query registration, residual scoring, backend assignments, and computational resources. Unless otherwise specified, the same configuration is used for every category on both Real3D-AD and Anomaly-ShapeNet.

B.1 Reference preparation and calibration

For each category, we use four anomaly-free reference point clouds. We center the first raw reference, align its principal axes, and resolve their signs using skewness to define the canonical anchor. Each of the remaining three references is independently aligned to this anchor using the same depth-map initialization, ICP refinement, and final pose-selection procedure as at inference time. Reference alignment is performed once per category.

Each aligned reference is orthographically rendered from $M = 42$ geodesic viewpoints on a 128×128 depth grid. The corresponding polar Fourier descriptors use 360 angular and 64 radial bins. The resulting depth maps and descriptors constitute the reusable view-conditioned reference bank.

For leave-one-out calibration, each aligned normal reference is treated in turn as a normal query and compared with the pooled remaining references. We apply the same residual normalization and multi-scale statistics used at test time, producing an empirical calibration collection $\mathcal{C}_c$ for every residual channel c. The normal-only neighborhood-growth statistics required by Chamfer refinement are constructed from the same leave-one-out comparisons. The view-conditioned bank, calibration collections, and Chamfer-refinement statistics are computed once per category and reused for all queries.

B.2 Query registration

Depth-map initialization produces one pose hypothesis for each reference–viewpoint pair. With four references and 42 viewpoints, this gives $4 \times 42 = 168$ initial hypotheses.

The screening stage applies three batched point-to-plane ICP iterations to at most 256 voxel-downsampled query points. For each viewpoint, we retain the reference-specific hypothesis with the smallest median directed residual, reducing the candidate set from 168 to 42. Starting from these screened poses, the retained hypotheses undergo up to 27 additional point-to-plane ICP iterations.

ICP uses a correspondence radius of $12d_{\mathrm{ref}}/40$ and a source-voxel size of $1.2d_{\mathrm{ref}}/40$, where d_{ref} is the maximum bounding-box diagonal across the aligned normal references and is shared by all pose hypotheses. The initial screening iterations are executed on the GPU, whereas the subsequent refinement is executed on the CPU.

Final pose verification evaluates all retained hypotheses on the same set of 512 query points. For each retained hypothesis, we compute directed query-to-reference nearest-neighbor residuals using exact float64 CPU search and rank the hypotheses by their lower-median residuals. Using a shared subset prevents differences in sampled verification points from affecting candidate comparison.

B.3 Residual scoring and Chamfer refinement

We sample 8,192 query points and compute their directed residuals to the pooled sampled normal references. Local residual means use

$$\mathcal{K}_{\mu} = \{1, 16, 64, 256\},$$

and local residual dispersions use

$$\mathcal{K}_{\sigma} = \{16, 64, 256\}.$$

These settings produce seven residual channels. Nearest-neighbor search, residual computation, and neighborhood statistics are evaluated on the GPU. Empirical calibration and Fisher-style evidence aggregation are evaluated on the CPU.

Chamfer refinement is applied to the $n_{\mathrm{g}} = 32$ points with the largest uncorrected evidence scores. We use $K_{\mathrm{s}} = 64$ and $K_{\mathrm{l}} = 512$ for the small and large query neighborhoods, respectively, and set $\alpha_{\mathrm{lo}} = 0.95$, $\alpha_{\mathrm{hi}} = 0.99$, and $\lambda = 0.25$. Their spatial radii determine coverage-matched reference neighborhoods centered at the nearest point in the pooled normal-reference geometry. The resulting neighborhood-growth statistic is compared with its normal-reference baseline. The bounded attenuation reduces evidence associated with broad geometric mismatch without completely removing responses to possible large defects.

The corrected point scores are median-smoothed over a self-inclusive eight-neighbor graph. The object-level anomaly score is the harmonic mean of the 16 largest smoothed point scores.

B.4 Computational environment

All experiments were conducted on a server running Ubuntu 20.04.6 LTS. The server contains two Intel Xeon Gold 6226R CPUs at 2.90 GHz, providing 32 physical cores and 64 hardware threads. The server has 251 GiB of system memory and eight NVIDIA GeForce RTX 3090 GPUs with 24 GiB of memory each. Each individual run is restricted to one GPU and eight logical CPU threads.

The software environment uses Python 3.8.20 and PyTorch 1.13.0 with CUDA 11.7. We use NumPy 1.24.4, SciPy 1.10.1, scikit-learn 1.3.2, and Open3D 0.19.0. All paired latency comparisons use the same server configuration and per-run resource allocation.

C Full Benchmark Results

C.1 Category-level results on Real3D-AD

Table C.1 reports the category-wise Real3D-AD results underlying the averages in Table 1. MARS attains the best or second-best P-AUROC in 11 of 12 categories, indicating that the overall localization gain is broadly distributed rather than driven by a small subset of objects. These results are consistent with our motivation that controlling pose-induced residuals and aggregating residual evidence across spatial scales enables direct geometric comparison across diverse object shapes.

Table C.1: **Real3D-AD category-wise P/O-AUROC (%) in the 4-shot setting.** Each cell reports point-/object-level AUROC. Methods marked with * require training. **Bold**/underlined values denote **best**/second-best results.

Method	Air.	Car	Candy	Chick.	Diam.	Duck	Fish	Gem.	Sea.	Shell	Star.	Toffee
Non-registration-based												
PatchCore	55.6/84.8	74.0/77.7	74.9/57.0	55.8/85.3	85.4/78.4	65.8/62.8	78.1/83.7	53.9/35.9	80.8/76.7	75.3/66.3	61.3/47.1	54.9/62.6
BTF	73.8/73.0	70.8/64.7	86.4/53.9	69.3/78.9	88.2/70.7	87.5/69.1	70.9/60.2	89.1/68.6	51.2/59.6	57.1/39.6	50.1/53.0	81.5/70.3
IMRNet*	-/76.2	-/71.1	-/75.5	-/78.0	-/90.5	-/51.7	-/88.0	-/67.4	-/60.4	-/66.5	-/67.4	-/77.4
CASL*	84.2/80.8	90.5/79.9	93.2/84.8	71.3/65.7	**98.8**/97.6	89.5/83.6	93.5/93.5	91.6/76.9	81.4/64.3	87.3/79.1	83.9/89.3	93.7/92.4
AARD*	83.5/88.2	80.1/72.5	85.7/86.4	90.6/87.1	88.6/91.7	89.4/**84.6**	91.7/89.1	92.6/76.5	85.2/80.7	84.6/81.1	78.5/75.9	80.6/85.5
Registration-based												
ISMP	75.3/85.8	83.6/73.1	90.7/85.2	79.8/71.4	92.6/94.8	87.6/71.2	88.6/94.5	85.7/46.8	81.3/72.9	83.9/62.3	64.1/66.0	89.5/84.2
PASDF*	77.7/62.8	80.2/95.9	54.6/78.8	76.8/73.9	69.9/89.4	78.2/65.8	83.7/98.9	65.4/63.4	88.7/**100.0**	64.8/85.0	70.3/61.7	83.8/86.6
MVR-DR	-/**89.1**	-/**96.4**	-/79.7	-/69.0	-/88.8	-/70.2	-/98.2	-/62.2	-/95.9	-/84.7	-/72.2	-/77.5
Reg3D-AD	63.1/71.6	71.8/69.7	72.4/68.5	67.6/85.2	83.5/90.0	50.3/58.4	82.6/91.5	54.5/41.7	81.7/76.2	81.1/58.3	61.7/50.6	75.9/82.7
Reg2Inv*	**92.3**/81.8	94.4/75.8	96.9/**100.0**	**91.0/94.4**	97.9/**100.0**	93.7/75.0	84.6/67.2	90.7/73.5	64.5/53.2	90.6/69.2	84.0/84.1	73.7/62.6
PointCore	60.8/66.0	70.6/86.6	76.0/97.6	78.0/84.1	81.0/96.3	71.2/68.4	78.2/99.3	51.5/53.5	84.1/97.3	78.1/88.2	73.6/65.2	74.5/**92.9**
MARS (Ours)	91.9/75.7	**98.2**/89.4	**98.7**/99.9	89.1/88.0	97.9/98.6	**96.2**/79.2	**99.8/100.0**	**94.9/78.2**	**99.3/100.0**	**99.3/95.7**	**99.6/97.1**	**98.5**/92.9

Table C.2: **Anomaly-ShapeNet category-wise P/O-AUROC (%) in the 4-shot setting.** Each cell reports point-/object-level AUROC. Methods marked with * require training. **Bold/**underlined values indicate the best/second-best result for each metric.

Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5	bucket0	bucket1	cap0
Non-registration-based														
PatchCore	59.7/58.7	57.4/57.1	65.4/60.4	68.7/66.7	51.2/57.2	52.4/50.4	53.1/63.9	62.5/61.5	32.7/53.7	72.0/49.4	35.8/55.8	45.9/46.9	57.1/55.1	47.2/58.0
BTF	62.4/42.0	74.6/54.6	64.1/34.4	54.9/54.6	62.2/32.2	71.0/50.9	76.8/66.8	51.8/51.0	59.0/49.0	67.9/60.9	69.9/69.9	40.1/40.1	63.3/63.3	73.0/61.8
IMRNet*	67.1/67.1	66.8/66.0	55.6/55.2	70.2/70.0	64.1/64.0	78.1/68.1	70.5/70.2	68.4/68.5	59.9/59.9	57.6/67.6	71.5/71.0	58.5/58.0	77.4/77.1	71.5/73.7
CASL*	88.7/94.3	96.8/94.8	85.3/95.7	86.2/95.4	91.4/**100.0**	95.8/**100.0**	97.5/93.3	98.2/99.3	99.8/99.6	99.6/94.4	97.4/88.8	79.6/**99.0**	92.7/**91.4**	97.4/**94.8**
AARD*	**97.1**/-	93.1/-	95.5/-	92.7/-	93.4/-	97.3/-	98.2/-	96.6/-	95.8/-	98.1/-	97.7/-	88.2/-	95.1/-	96.0/-
Registration-based														
ISMP	60.3/-	74.7/-	77.0/-	56.8/-	77.5/-	85.1/-	54.6/-	73.6/-	77.3/-	74.0/-	53.4/-	52.4/-	67.2/-	86.5/-
PASDF*	91.9/**100.0**	95.8/99.5	95.1/**100.0**	92.6/**100.0**	94.8/**100.0**	96.3/**100.0**	90.0/94.8	81.6/**100.0**	93.9/**100.0**	86.5/93.3	90.9/91.2	87.5/96.8	82.4/77.5	94.8/85.2
MVR-DR	88.5/**100.0**	93.5/**100.0**	92.6/**100.0**	83.0/99.6	84.9/**100.0**	93.4/**100.0**	84.5/94.4	77.9/99.3	90.9/**100.0**	87.6/95.6	82.6/**100.0**	77.2/85.1	78.5/84.4	84.5/88.5
Reg3D-AD	69.8/59.7	71.5/70.6	88.6/48.6	69.6/69.5	52.5/52.5	77.5/67.1	61.5/52.5	59.3/49.0	65.4/34.8	80.0/66.3	69.1/59.3	61.9/61.0	75.2/75.2	63.2/69.3
Reg2Inv*	78.5/90.0	99.1/**100.0**	99.5/**100.0**	84.9/**100.0**	81.7/**100.0**	98.3/**100.0**	82.8/80.7	82.2/65.6	76.1/58.5	78.8/85.2	82.4/81.8	61.0/81.3	85.5/90.2	86.1/65.9
PointCore	-/**100.0**	-/55.7	-/70.9	-/**100.0**	-/63.5	-/83.0	-/77.4	-/57.0	-/74.4	-/74.1	-/55.3	-/57.5	-/71.1	-/51.1
MARS	96.5/**100.0**	**99.9/100.0**	**99.8/100.0**	**95.6/100.0**	**96.0/100.0**	**99.6/100.0**	**99.5/99.6**	**98.6**/98.2	**99.9/100.0**	**99.8/98.9**	**99.9/100.0**	**95.1**/96.8	**98.0**/89.5	**98.0**/84.1

Method	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1	helmet0	helmet1	helmet2	helmet3	jar0	microphone0
Non-registration-based														
PatchCore	65.3/45.3	59.5/75.7	79.5/79.0	65.5/60.0	59.6/58.6	81.0/65.7	58.3/58.3	46.4/63.7	54.8/54.6	48.9/48.4	45.5/42.5	73.7/40.4	47.8/47.2	48.8/38.8
BTF	65.8/52.2	52.4/52.0	58.6/58.6	79.0/58.6	61.9/61.0	71.9/71.9	62.0/52.0	59.1/49.0	57.5/57.1	74.9/71.9	64.3/54.2	72.4/44.4	42.7/42.4	67.5/67.1
IMRNet*	70.6/77.5	75.3/65.2	74.2/65.2	64.3/64.3	68.8/75.7	54.8/54.8	61.5/72.0	47.6/67.6	59.8/59.7	60.4/60.0	64.4/64.1	66.3/57.3	76.5/78.0	74.2/75.5
CASL*	**97.1**/82.1	94.9/77.2	66.9/55.4	98.0/99.0	79.7/64.8	87.6/99.5	77.0/83.6	90.9/96.7	92.3/78.3	88.2/89.0	92.9/**88.1**	98.0/86.1	97.4/99.5	83.8/98.1
AARD*	96.2/-	**96.8**/-	94.9/-	94.5/-	95.7/-	97.0/-	86.6/-	96.7/-	86.2/-	90.7/-	86.1/-	89.2/-	90.5/-	86.4/-
Registration-based														
ISMP	73.4/-	75.3/-	67.8/-	86.9/-	60.0/-	70.6/-	58.0/-	70.2/-	68.3/-	62.2/-	84.4/-	72.2/-	82.3/-	66.1/-
PASDF*	86.1/64.9	89.4/64.6	92.0/85.3	94.8/97.1	88.4/85.7	94.5/95.2	86.3/**100.0**	89.1/79.5	81.6/81.2	64.6/93.8	80.9/76.5	95.8/84.6	95.9/**100.0**	95.1/**100.0**
MVR-DR	80.9/**93.7**	84.9/**93.3**	86.2/**98.2**	93.5/**100.0**	89.0/**100.0**	91.9/**100.0**	89.2/**100.0**	93.5/**100.0**	73.1/86.4	89.5/**100.0**	82.2/82.3	87.1/93.9	92.9/**100.0**	90.0/**100.0**
Reg3D-AD	71.8/72.5	81.5/64.3	46.7/46.7	68.5/51.0	69.8/53.8	75.5/34.3	58.0/53.7	62.6/61.0	60.0/60.0	62.4/38.1	82.5/61.4	62.0/36.7	59.9/59.2	59.9/41.4
Reg2Inv*	94.5/86.3	86.4/68.1	97.0/90.2	79.8/73.3	88.1/93.3	98.0/**100.0**	94.6/**100.0**	97.0/84.3	92.5/81.7	90.6/98.6	89.1/87.5	95.6/87.6	98.2/**100.0**	**99.2/100.0**
PointCore	-/66.9	-/51.4	-/68.9	-/55.0	-/56.3	-/75.0	-/71.1	-/87.6	-/47.7	-/41.4	-/52.7	-/53.6	-/90.0	-/66.2
MARS	96.0/89.1	96.3/92.6	**98.6**/92.3	**99.8/100.0**	**99.7/100.0**	**99.8/100.0**	**97.8/100.0**	**99.7/100.0**	**95.5/90.1**	**99.6/100.0**	**93.4**/80.9	**99.8/98.2**	**99.9/100.0**	98.7/**100.0**

Method	shelf0	tap0	tap1	vase0	vase1	vase2	vase3	vase4	vase5	vase7	vase8	vase9
Non-registration-based												
PatchCore	61.3/49.4	73.3/75.3	76.8/76.6	65.5/45.5	45.3/42.3	72.1/72.1	43.0/44.9	50.5/50.6	44.7/41.7	69.3/69.3	57.5/66.2	66.3/66.0
BTF	61.9/60.9	56.8/56.0	59.6/54.6	64.2/34.2	61.9/21.9	64.6/54.6	69.9/69.9	71.0/51.0	42.9/40.9	54.0/51.8	66.2/66.8	56.8/26.8
IMRNet*	60.5/60.3	68.1/67.6	69.9/69.6	53.5/53.3	68.5/75.7	61.4/61.4	40.1/70.0	52.4/52.4	68.2/67.6	59.3/63.5	63.5/63.0	69.1/59.4
CASL*	92.8/**83.2**	65.2/76.1	60.2/56.3	89.4/83.8	97.2/89.0	97.1/98.1	93.1/**93.0**	91.3/82.1	73.2/71.4	99.7/**100.0**	**98.9**/93.3	88.6/81.8
AARD*	89.1/-	85.7/-	82.6/-	98.3/-	92.4/-	98.8/-	93.6/-	93.6/-	96.2/-	98.5/-	97.1/-	**98.7**/-
Registration-based												
ISMP	68.7/-	52.2/-	55.2/-	66.1/-	84.3/-	73.3/-	76.2/-	54.5/-	47.2/-	70.1/-	85.1/-	61.5/-
PASDF*	86.5/71.3	88.4/88.2	90.2/79.3	94.4/**100.0**	79.7/92.9	95.6/**100.0**	86.8/80.6	89.9/**91.2**	91.5/**100.0**	95.9/**100.0**	90.9/92.4	86.3/83.6
MVR-DR	72.8/80.0	83.5/91.5	85.3/**92.2**	93.8/**100.0**	81.6/96.7	91.5/**100.0**	76.6/84.9	79.1/90.9	86.8/**100.0**	92.2/**100.0**	80.9/87.6	75.2/82.1
Reg3D-AD	68.8/68.8	58.9/67.6	74.1/64.1	54.8/53.3	60.2/70.2	40.5/60.5	51.1/65.0	75.5/50.0	62.4/52.0	88.1/46.2	81.1/62.0	69.4/59.4
Reg2Inv*	63.2/57.7	91.8/**94.8**	86.9/80.4	98.0/99.6	70.5/60.5	99.7/**100.0**	84.4/84.5	92.7/81.8	87.9/**100.0**	86.3/64.3	93.4/81.8	97.1/**87.3**
PointCore	-/70.1	-/64.5	-/80.0	-/52.5	-/61.2	-/92.0	-/69.8	-/62.4	-/71.0	-/56.7	-/71.7	-/80.0
MARS	**95.1**/80.6	**98.1**/93.0	**96.5**/89.6	**98.8/100.0**	**98.2/100.0**	**99.9/100.0**	**97.8**/90.3	**98.5**/86.7	**99.9/100.0**	**100.0/100.0**	96.9/**94.8**	96.9/80.9

C.2 CATEGORY-LEVEL RESULTS ON ANOMALY-SHAPENET

Table C.2 reports the full category-wise results on the 40 Anomaly-ShapeNet categories corresponding to the summary in Table 1. The consistently high point-level results across diverse complete shapes complement the Real3D-AD results, supporting the same alignment-and-residual-scoring principle across partial observations and complete object geometry.

Table D.1: **Alignment and residual-scoring ablations on Anomaly-ShapeNet.** P/O denote point- and object-level AUROC (%). Latency is normalized by the complete MARS configuration.

(a) Alignment components

Configuration	P/O-AUROC	Rel. latency ↓
No ICP refinement	75.13 / 57.13	0.53×
No roll search	97.62 / 94.54	1.04×
No candidate pruning	**98.25** / 95.45	2.25×
MARS (Ours)	98.19 / **95.66**	**1.00**×

(b) Residual-scoring components

Residual statistics	Aggregation	P/O-AUROC
Pointwise	–	90.36 / 95.44
Multi-scale means	Average	95.91 / 94.64
Means + dispersions	Average	97.62 / 95.26
Means + dispersions	Fisher-style	**98.19 / 95.67**
+ Chamfer refinement	**Fisher-style**	**98.19** / 95.66

D Additional Ablation Studies

D.1 Main-paper ablations on Anomaly-ShapeNet

We repeat the main-paper alignment and residual-scoring ablations on Anomaly-ShapeNet in Table D.1.

Alignment components. Removing ICP refinement causes the largest accuracy degradation. Removing candidate pruning yields nearly identical accuracy but increases latency by more than a factor of two, indicating that staged hypothesis reduction eliminates redundant computation without materially changing detection accuracy.

Residual-scoring components. Multi-scale means improve localization over pointwise residuals, and local dispersions provide an additional gain. Fisher-style aggregation further improves the joint P/O-AUROC result, whereas Chamfer refinement has a negligible effect on this complete-shape benchmark.

D.2 Additional residual analysis

We compare nearest-neighbor residuals from saved registration outputs before anomaly scoring on Real3D-AD and Anomaly-ShapeNet in Tables D.2 and D.3. Each query is evaluated using the same 8,192 scored points and common reference geometry across methods. These residuals measure query-to-reference geometric fit.

Normal-query residuals. MARS achieves the lowest median and narrowest interquartile range on both benchmarks, indicating smaller and more consistent geometric discrepancies on defect-free inputs.

Region-wise residuals. We separately evaluate GT-normal and GT-abnormal regions within each query. MARS achieves the lowest normal-region median on both benchmarks and the largest IQR gap on Real3D-AD, while PASDF has the largest gap on Anomaly-ShapeNet.

D.3 Controlled probes of geometric residuals

We provide experimental details for the probes on the candybar category of Real3D-AD shown in Figure 1. All probes reuse the same four normal references and normal-only calibration without recalibration.

Table D.2: **Registration residual analysis on Real3D-AD.** Each observation is a query-mean nearest-neighbor residual over the relevant region among 8,192 scored points. Panel (a) uses 604 GT-normal queries; panel (b) separates GT-normal and GT-abnormal regions within each query, omitting empty regions. Units are category-dependent voxels. Bold values indicate the lowest normal-region median, the narrowest normal-query IQR, or the largest IQR gap.

(a) Normal-query residuals

Method	Q_1	Median ↓	Q_3	IQR ↓
Reg3D-AD	0.0433	0.0700	0.1179	0.0746
Reg2Inv	0.0424	0.0571	0.0774	0.0350
PointCore	0.0193	0.0244	0.0441	0.0248
PASDF	0.0196	0.0277	0.0505	0.0308
MARS (Ours)	0.0186	**0.0209**	0.0244	**0.0058**

(b) Region-wise residuals

	GT-normal region		GT-abnormal region		
Method	**Median ↓**	$[Q_1, Q_3]$	**Median**	$[Q_1, Q_3]$	**IQR gap ↑**
Reg3D-AD	0.0722	[0.0462, 0.1326]	0.1738	[0.1007, 0.2503]	-0.0319
Reg2Inv	0.0612	[0.0465, 0.0832]	0.1442	[0.0855, 0.2126]	0.0023
PointCore	0.0247	[0.0197, 0.0424]	0.1609	[0.0858, 0.2421]	0.0435
PASDF	0.0283	[0.0199, 0.0485]	0.1376	[0.0827, 0.2218]	0.0342
MARS (Ours)	**0.0214**	[0.0190, 0.0263]	0.1408	[0.0803, 0.2069]	**0.0540**

IQR gap $= Q_1^{\text{abnormal}} - Q_3^{\text{normal}}$; larger values indicate greater separation between the region-wise IQRs.

Table D.3: **Registration residual analysis on Anomaly-ShapeNet.** Each observation is a query-mean nearest-neighbor residual over the relevant region among 8,192 scored points. Panel (a) uses 600 GT-normal queries; panel (b) separates GT-normal and GT-abnormal regions within each query, omitting empty regions. Units are category-dependent voxels. Bold values indicate the lowest normal-region median, the narrowest normal-query IQR, or the largest IQR gap.

(a) Normal-query residuals

Method	Q_1	Median ↓	Q_3	IQR ↓
Reg3D-AD	1.4673	2.3472	3.4704	2.0031
Reg2Inv	0.0676	0.0798	0.1024	0.0348
PointCore	0.0534	0.5099	1.2376	1.1842
PASDF	0.0450	0.0533	0.0611	0.0162
MARS (Ours)	0.0441	**0.0513**	0.0567	**0.0126**

(b) Region-wise residuals

	GT-normal region		GT-abnormal region		
Method	**Median ↓**	$[Q_1, Q_3]$	**Median**	$[Q_1, Q_3]$	**IQR gap ↑**
Reg3D-AD	2.3501	[1.4559, 3.4204]	1.5864	[0.7549, 3.6112]	-2.6656
Reg2Inv	0.0977	[0.0737, 0.1348]	0.2068	[0.1403, 0.3089]	0.0055
PointCore	0.5882	[0.0572, 1.3137]	0.5022	[0.2608, 1.1149]	-1.0529
PASDF	0.0538	[0.0451, 0.0625]	0.2708	[0.1647, 0.4137]	**0.1022**
MARS (Ours)	**0.0530**	[0.0474, 0.0621]	0.2497	[0.1590, 0.3568]	0.0969

IQR gap $= Q_1^{\text{abnormal}} - Q_3^{\text{normal}}$; larger values indicate greater separation between the region-wise IQRs.

Probe setup. Alignment and noise probes each use 50 normal and 50 anomalous queries. We inject a 7° rotation and a translation of $0.03D$, or Gaussian coordinate noise with standard deviation $0.005D$, where D is the unperturbed aligned query's bounding-box diagonal over the 8,192 scored points. The small-defect probe uses the same normal queries and the 25 anomalous queries with

the smallest annotated defect fractions among scored points (0.71–2.22%), selected independently of detection performance.

Scoring comparison. For alignment error, the qualitative example illustrates the residuals caused by inaccurate registration. For scan noise, the example shows residuals before re-registration. Quantitative results for both probes compare scoring after re-registration. The small-defect visualization and evaluation use the saved registration outputs. Within each probe, raw nearest-neighbor distances and complete MARS scoring use identical final registration, 8,192 scored points per query, and original ground-truth labels. MARS improves P-AUROC by 17.96, 27.42, and 16.71 percentage points for alignment error, scan noise, and small defects, respectively.

Visualization and interpretation. The perturbed qualitative examples in Figure 1 show residuals before re-registration, whereas quantitative results compare scoring after re-registration. These category-specific probes demonstrate improved discrimination over raw geometric distance without isolating individual scoring components.

D.4 Reference pooling and normal calibration

We examine reference pooling and empirical channel calibration in Table D.4.

Reference pooling. We isolate the geometry used for residual computation while keeping the registration results, calibration data, and subsequent score processing fixed. Compared with scoring against only the reference selected during registration, pooling all normal references improves P/O AUROC from 96.28/87.84% to 96.95/91.23% on Real3D-AD and from 97.90/93.39% to 98.19/95.66% on Anomaly-ShapeNet. These results show that pooling normal geometry provides a more effective scoring reference than relying on the single instance selected during registration, consistent with reduced sensitivity to reference-specific sampling and surface coverage.

Calibration and evidence aggregation. We compare aggregation after channel-wise leave-one-out calibration with direct aggregation of the raw residual statistics. Calibrated Fisher-style aggregation achieves the highest P-AUROC on Real3D-AD (96.95%) and Anomaly-ShapeNet (98.19%), although it does not maximize every individual metric. On Anomaly-ShapeNet, uncalibrated mean exceeds its O-AUROC by 0.27 percentage points. We retain the calibrated formulation because it maps heterogeneous residual channels to a common empirical scale and provides consistent performance across both benchmarks.

D.5 Score refinement and object-level aggregation

We evaluate score refinement and object-level aggregation in Table D.5.

Chamfer refinement and smoothing. We remove Chamfer refinement and median smoothing individually and jointly. Removing Chamfer refinement leaves Real3D-AD P-AUROC unchanged but reduces O-AUROC by 0.42 percentage points. Its effect on Anomaly-ShapeNet is negligible. This behavior is consistent with the intended role of Chamfer refinement as a targeted correction for residual geometric mismatch, with its benefit appearing more clearly on partial real-world scans.

Object-level aggregation. We compare maximum pooling, global averaging, and arithmetic or harmonic aggregation over the 16 largest point scores. Global averaging performs poorly because defective regions generally occupy only a small fraction of an object. Top-16 harmonic aggregation obtains the highest O-AUROC on both benchmarks.

We further vary the number of point scores used for harmonic aggregation in Table D.5(c). Top-16 obtains the highest O-AUROC on both benchmarks among the evaluated harmonic variants, although performance remains stable across several small top-ranked subsets. In contrast, expanding the subset to the top 1% reduces performance, indicating that including too many point scores can dilute localized anomaly evidence with responses from predominantly normal regions.

Table D.4: **Reference pooling, calibration, and aggregation on Anomaly-ShapeNet and Real3D-AD.** Panels (a) and (b) report P/O-AUROC (%); O uses top-16 harmonic pooling. Panel (c) reports channel-level calibration diagnostics.

(a) Reference pooling

Scoring reference	Real3D-AD P/O	Anomaly-ShapeNet P/O
Matched reference	96.28 / 87.84	97.90 / 93.39
Pooled reference	**96.95 / 91.23**	**98.19 / 95.66**

(b) Calibration and aggregation

Configuration	Real3D-AD P/O	Anomaly-ShapeNet P/O
Calibrated mean	94.73 / 89.93	97.62 / 95.21
Calibrated max	96.24 / 88.43	95.47 / 84.28
Calibrated Stouffer	96.51 / 91.04	98.09 / 95.73
Uncalibrated mean	96.19 / 90.96	98.16 / **95.93**
Uncalibrated max	95.75 / **91.54**	97.87 / 95.58
Calibrated Fisher-style	**96.95** / 91.23	**98.19** / 95.66

(c) Calibration-resolution diagnostics

Diagnostic (%)	Real3D-AD	Anomaly-ShapeNet
Minimum-tail fraction	1.0158	3.8787
Channel evidence pair-tie rate	0.4010	2.2004
Normal point/channel FPR at $\alpha = 0.05$	4.4636	5.7563

E Sensitivity and Robustness Analyses

E.1 Scoring-point sampling and reference count

We examine the scoring-point budget, sampling variation, and number of normal references in Table E.1.

Scoring-point budget. We vary the number of scored query points while reusing the same registration results and normal calibration. Chamfer refinement, smoothing, and object-level aggregation are recomputed for each point set. The selected 8,192-point budget provides the strongest object-level result on Real3D-AD while retaining near-best point-level performance on both benchmarks.

Variation across scoring-point samples. We repeat 8,192-point sampling with two additional deterministic seeds. Across the three sampled subsets, P-AUROC varies by at most 0.02 percentage points on Real3D-AD and 0.09 percentage points on Anomaly-ShapeNet. This limited variation indicates that the reported result is not determined by one particular sampled subset.

Two-reference setting. We additionally evaluate one deterministic two-reference subset formed by selecting the first two references in lexicographic order. The registration bank, pooled reference geometry, leave-one-out calibration, and Chamfer statistics are reconstructed using only these references. This experiment evaluates performance under a reduced reference set but does not estimate variation across different two-reference subsets.

E.2 Input perturbations

Robustness to coordinate noise, point-density changes, and rigid-pose perturbations is summarized in Table E.2.

Table D.5: **Effects of score refinement and object-level aggregation.** Panel (a) reports P/O-AUROC (%), panel (b) compares object-level aggregation rules using O-AUROC (%), and panel (c) examines the number of top-ranked point scores used for harmonic aggregation. All object-level aggregation rules use the same final pointwise scores. Panel (c) reports three decimal places to distinguish closely performing settings.

(a) Chamfer refinement and smoothing

Configuration	Real3D-AD P/O	Anomaly-ShapeNet P/O
No median smoothing	96.88 / 91.02	98.17 / 95.58
No Chamfer refinement	96.95 / 90.81	**98.19 / 95.67**
Neither component	96.88 / 90.73	98.17 / 95.60
Complete MARS	**96.95 / 91.23**	98.19 / 95.66

(b) Object-level aggregation rule

Aggregation	Real3D-AD	Anomaly-ShapeNet
Maximum	91.04	95.63
Global mean	79.78	88.35
Top-16 arithmetic mean	91.18	95.66
Top-16 harmonic mean	**91.23**	**95.66**

(c) Number of scores in harmonic aggregation

Aggregation	n_o	Real3D-AD	Anomaly-ShapeNet
Top-8 harmonic	8	91.172	95.644
Top-16 harmonic	**16**	**91.229**	**95.657**
Top-32 harmonic	32	91.074	95.610
Top-0.01% harmonic	1	91.035	95.632
Top-0.1% harmonic	9	91.205	95.619
Top-1% harmonic	82	90.309	95.330

Gaussian coordinate noise. We add independent Gaussian noise before registration, with standard deviation $\sigma = \alpha d_q$, where d_q is the query bounding-box diagonal. Performance changes only slightly for $\alpha \leq 0.001$ and degrades as the perturbation becomes large relative to the object scale.

Point-density perturbation. We retain a fixed fraction of the original query points without replacement and rerun the complete registration and scoring pipeline. Retaining 10% of the query points changes P/O-AUROC by $-0.06/-0.72$ percentage points on Real3D-AD and $-0.11/-0.37$ percentage points on Anomaly-ShapeNet.

Rigid-pose perturbation. We apply a random rotation in SO(3) and a translation of up to 10% of the query bounding-box diagonal along each axis before registration. Under the tested perturbation, the absolute changes remain at most 0.75 percentage points on the reported metrics. This experiment uses a fixed random seed and should therefore be interpreted as an additional perturbation check rather than an estimate over the full pose distribution.

F Additional Qualitative Results

We provide additional registration and anomaly-localization results on Real3D-AD (Figure F.1) and Anomaly-ShapeNet (Figures F.2 and F.3). In registration overlays, blue and orange denote normal-reference and query points, respectively. Ground-truth defects are shown in red, and anomaly maps use a blue-to-red scale.

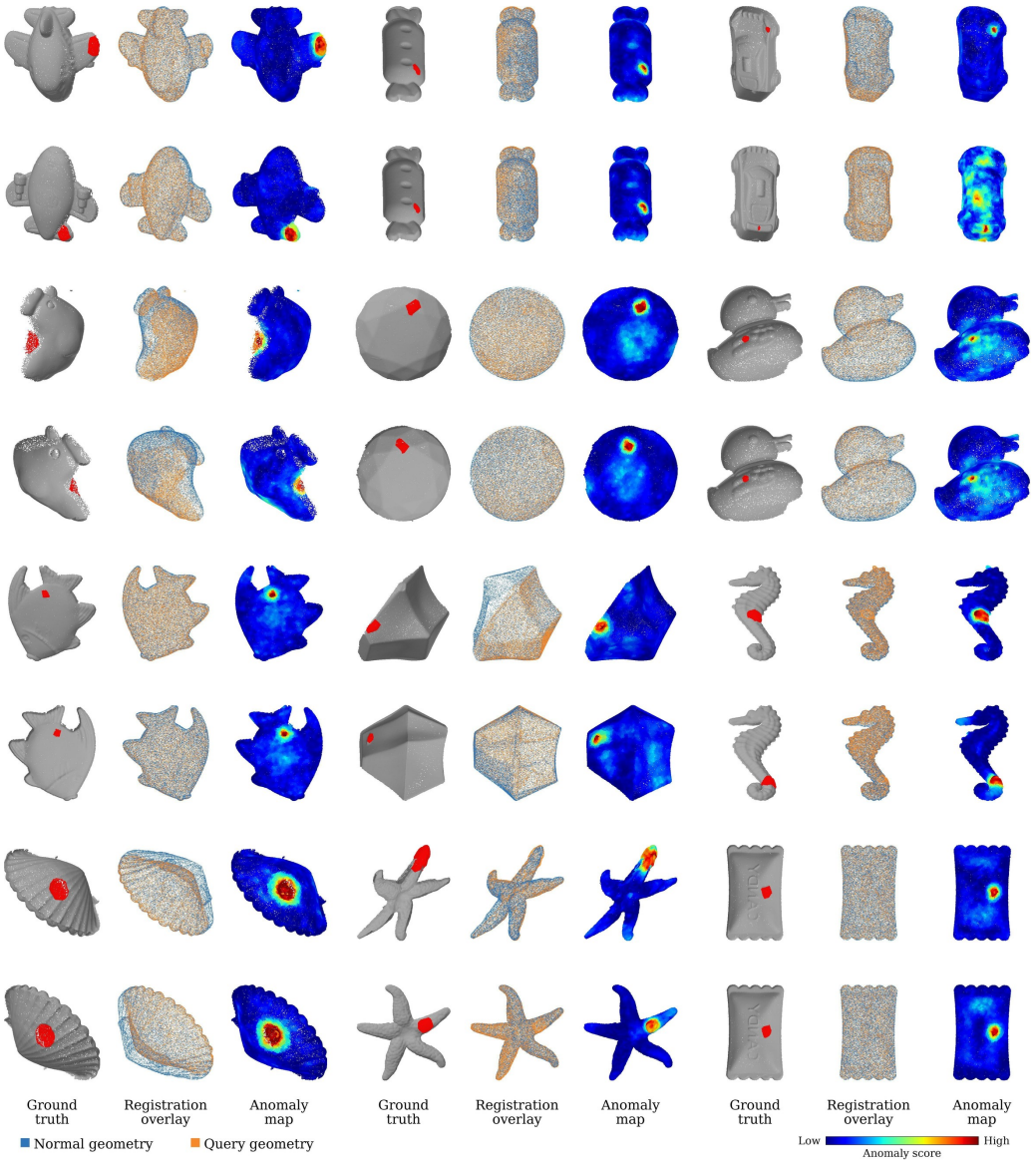

Figure F.1: **Additional qualitative results on Real3D-AD.** The examples cover object categories and defect types not shown in the main paper.

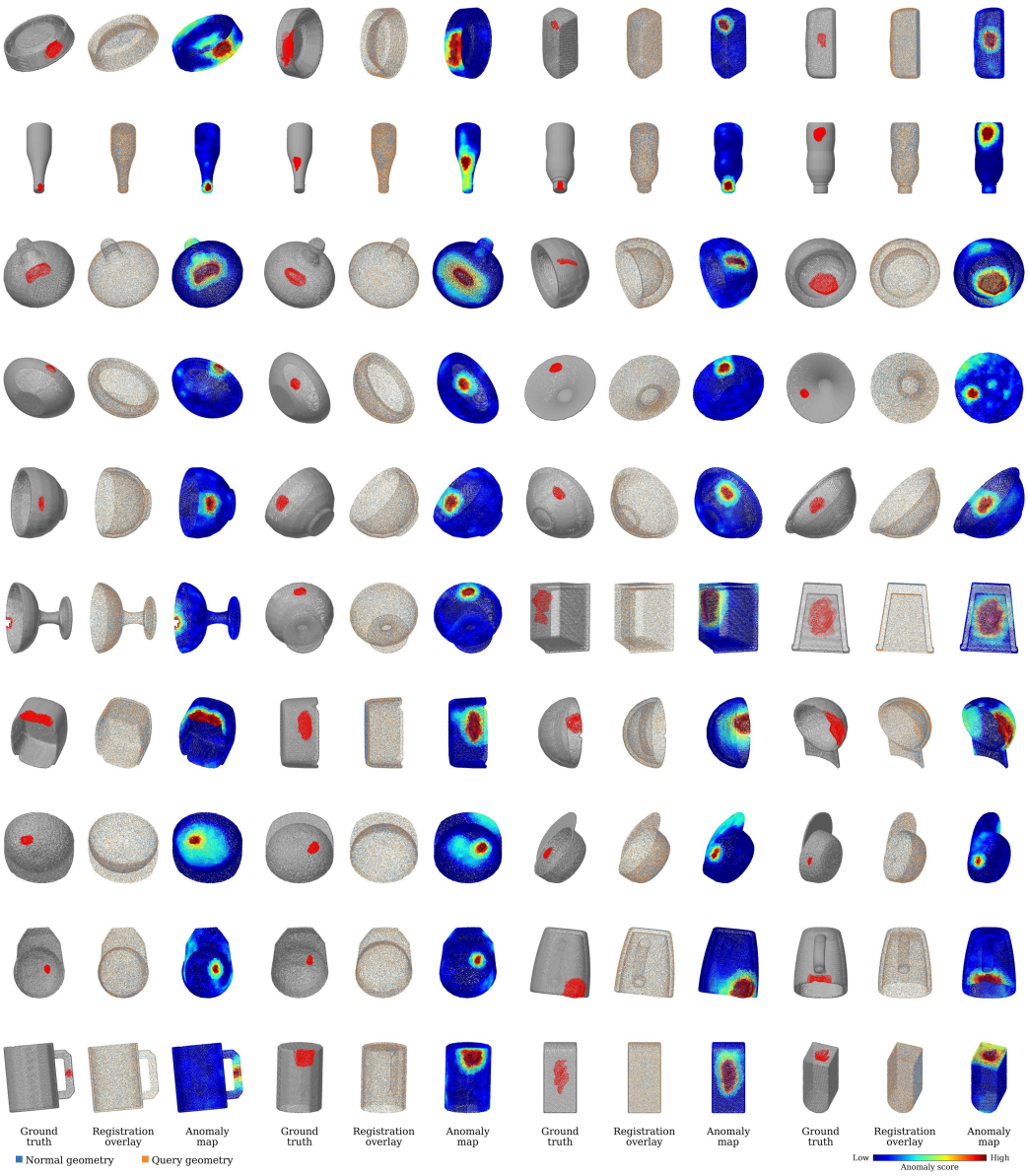

Figure F.2: **Additional qualitative results on Anomaly-ShapeNet (part 1).** The examples cover diverse complete shapes and geometric defect types.

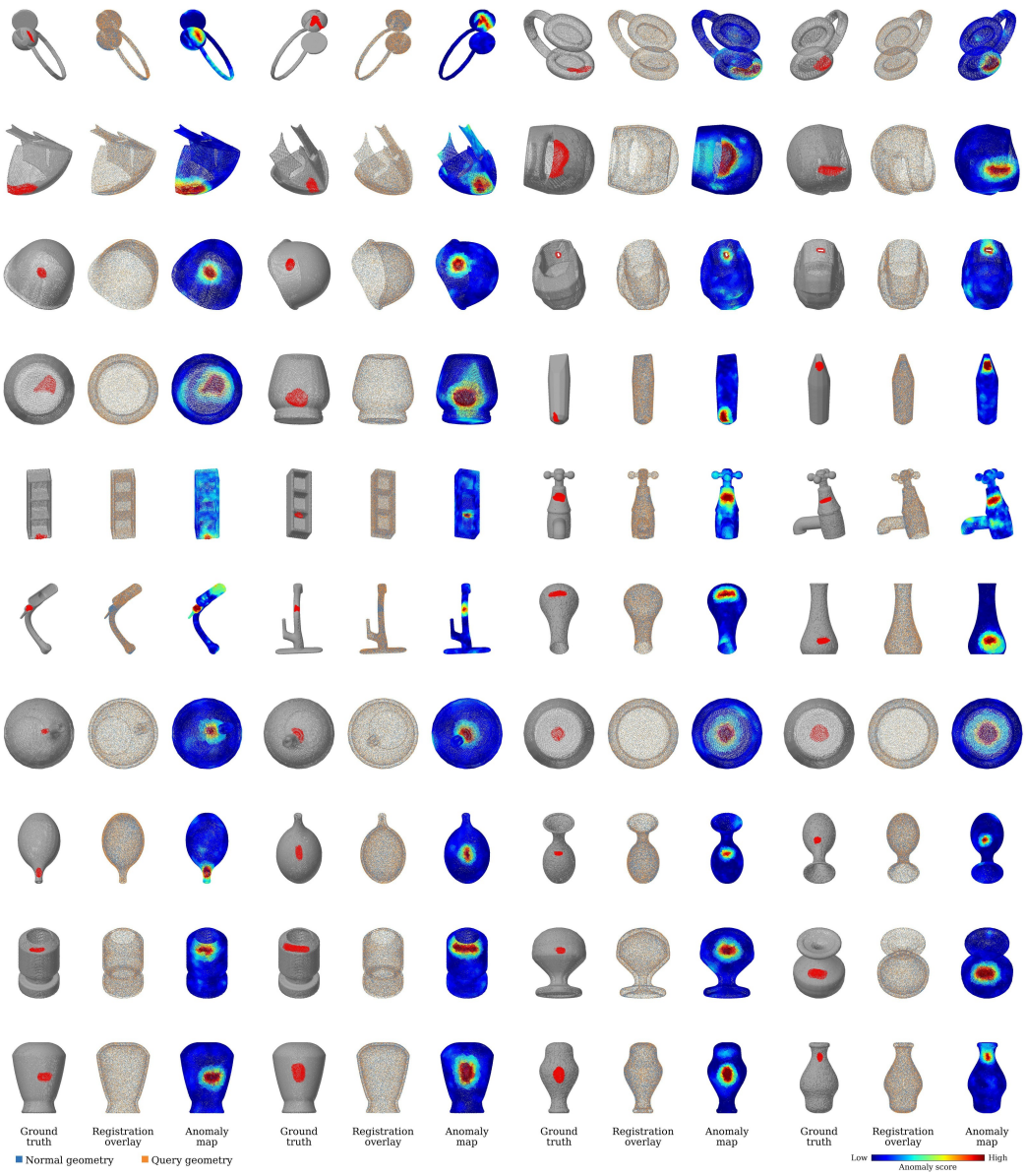

Figure F.3: **Additional qualitative results on Anomaly-ShapeNet (part 2).** The examples further illustrate residual-based localization across categories.

Table E.1: **Sensitivity to scoring-point sampling and reference count.** All entries report P/O-AUROC (%). The reported two-reference result uses one deterministic subset.

(a) Scoring-point budget

Scoring points	Real3D-AD P/O	Anomaly-ShapeNet P/O
1,000	95.18 / 87.22	97.66 / 93.82
4,096	**97.14** / 90.09	**98.21** / 95.14
8,192 (Ours)	96.95 / **91.23**	98.19 / **95.66**

(b) Variation across 8,192-point samples

Sample	Real3D-AD P/O	Anomaly-ShapeNet P/O
Reported sample	96.95 / 91.23	98.19 / 95.66
Alternative seed 1	96.95 / 90.64	98.25 / 95.43
Alternative seed 2	96.93 / 90.80	98.28 / 95.29

(c) Number of normal references

References	Real3D-AD P/O	Anomaly-ShapeNet P/O
2 (deterministic subset)	95.90 / 88.91	98.08 / 95.49
4 (Ours)	**96.95 / 91.23**	**98.19 / 95.66**

Table E.2: **Robustness to input perturbations.** All perturbations are applied before rerunning registration and scoring. Results are reported as P/O-AUROC (%). The rigid-pose result uses one fixed random seed.

(a) Gaussian coordinate noise

α	Real3D-AD P/O	Anomaly-ShapeNet P/O
0	96.95 / 91.23	98.19 / 95.66
0.0005	97.07 / 90.86	98.19 / 95.18
0.001	97.13 / 90.57	98.17 / 95.15
0.0025	95.06 / 88.78	97.45 / 93.89
0.005	86.62 / 84.22	93.08 / 92.22
0.01	58.17 / 57.44	81.09 / 82.72

(b) Point density and rigid pose

Setting	Real3D-AD P/O	Anomaly-ShapeNet P/O
Unperturbed	96.95 / 91.23	98.19 / 95.66
Retain 50%	96.89 / 90.62	98.32 / 95.43
Retain 25%	96.80 / 90.15	98.21 / 95.61
Retain 10%	96.89 / 90.51	98.08 / 95.29
Rigid pose	96.95 / 91.14	98.17 / 94.91