ANYNORMAL: LET EVERY POINT CHOOSE ITS NORMAL

Anonymous authors
Paper under double-blind review

ABSTRACT

In few-shot 3D anomaly detection, only a handful of normal examples are available to characterize normality, even though normal objects can differ in local shape and sampling. This raises two questions: how should a geometric difference be judged against the variation among normal objects, and which normal example should be used to assess each test region? We propose AnyNormal. It first compares the normal examples with one another to estimate the range of normal geometric variation. At test time, each query point is matched independently to every aligned normal reference. For each reference, Euclidean, point-to-plane, and normal-angle differences are measured relative to the variation observed among normal examples, at both the point and local-neighborhood levels, and averaged into a score. The minimum score across references is used for each point, allowing different regions of the same object to be assessed against different normal examples. No scoring network is trained: a parameter-free mean of these calibrated geometric features achieves 90.3/95.1%, 94.5/96.2%, and 89.0/87.0% object/point AUROC on Real3D-AD, Anomaly-ShapeNet, and MulSen-AD, respectively. In a transfer experiment, a learned scorer trained on only four normal scans from a single Real3D-AD category reaches essentially the same accuracy across all three datasets without further weight updates. A compressed configuration achieves 85.9/82.4%, 90.1/88.0%, and 79.3/71.4% object/point AUROC on the three datasets, respectively, while reaching 69.74 FPS in continuous-stream evaluation on an RTX 4090. Code for the parameter-free method is available at `https://anonymous.4open.science/r/AnyNormal/`.

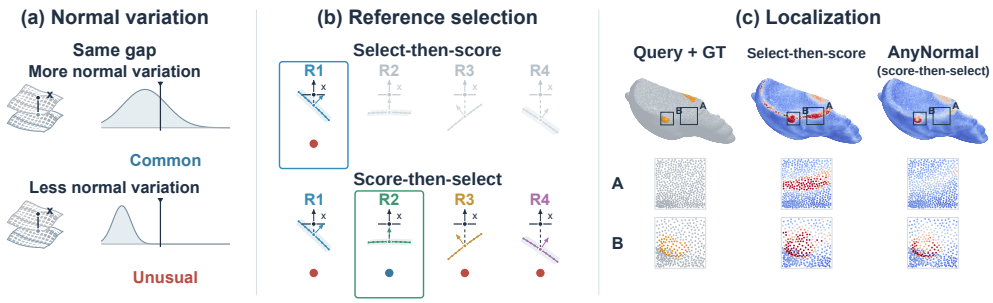


Figure 1: **Normal variation and reference choice.** (a) The same gap can be common or unusual under different normal variation. (b) For query point x, pooled matching merges normal reference points and chooses one match by proximity before scoring; AnyNormal scores each reference before pointwise selection. Faded references are not scored. Panels (a,b) are schematic. (c) Corresponding measured maps share references, calibration, view and linear scale. AnyNormal reduces the response in normal region A while preserving the high response at defect B. Blue/red denote lower/higher scores; orange denotes ground truth (GT). Table 5(b) reports the controlled comparison.

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.

1 Introduction

Few-shot 3D anomaly detection uses a handful of normal point clouds to detect and localize defects. Normal examples can differ in local shape, sampling and surface coverage. A geometric mismatch may therefore reflect normal variation, and no single normal example is guaranteed to explain every region of a test object.

Reference-based scoring must account for normal variation while deciding which reference to use. A natural baseline merges reference points and selects one match by proximity before scoring; we call this *pooled matching*. This early selection can discard another reference that would assign a lower calibrated anomaly score (Figure 1).

We propose AnyNormal. Before testing, normal-to-normal comparisons establish calibration statistics for Euclidean distance, point-to-plane offset and normal-angle discrepancy. At test time, each aligned reference independently matches every query point and averages its calibrated pointwise and neighborhood features into a score. Taking the lowest score lets different regions use different normal references; a point receives a high anomaly score only when all references fit it poorly.

These calibrated per-reference features make a parameter-free mean effective across Real3D-AD and Anomaly-ShapeNet with four references per category, and MulSen-AD with full normal support. A small scorer trained on four normal scans from one Real3D-AD category reaches similar accuracy across all three datasets without further weight updates. Each target category still supplies its own normal references and calibration statistics. Compressing the reference point sets trades accuracy for online speed.

Our contributions are summarized as follows:

- **Normal-Variation Calibration.** Normal-to-normal residuals calibrate geometric differences at point and neighborhood levels. Compared with directly averaging uncalibrated residuals, this calibration improves point AUROC by **10.56/5.55** percentage points on Real3D-AD/Anomaly-ShapeNet.
- **Per-Reference Scoring.** Scoring each reference before pointwise selection preserves alternative normal explanations. With identical references, calibration and total retrieved candidate counts, it improves point AP over pooled matching by **2.50/2.65** percentage points on the same two datasets.
- **Training-Free and Efficient Detection.** Fixed-mean scoring achieves **90.3/95.1%, 94.5/96.2%, and 89.0/87.0%** object/point AUROC on the three benchmarks. A scorer trained on four scans of one source category gives similar cross-dataset accuracy without weight updates. The separate compressed configuration reaches **69.74 FPS** on RTX 4090.

2 Method

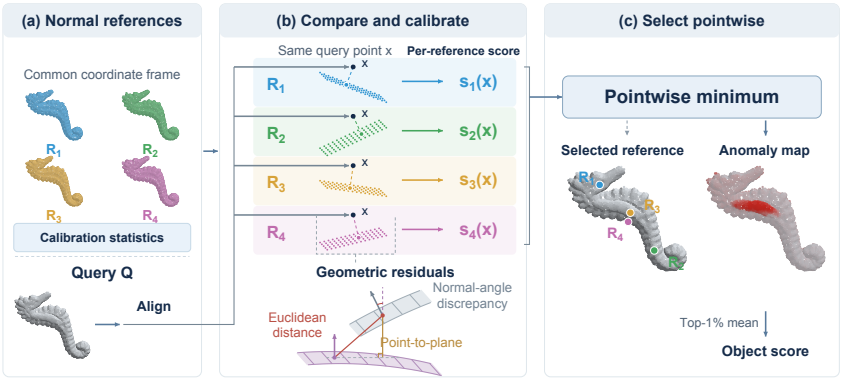


Figure 2: **AnyNormal overview.** (a) Normal references supply geometry and calibration. (b) Calibrated Euclidean, point-to-plane and normal-angle residuals give one fixed-mean score per reference. (c) Pointwise minimum yields the anomaly map; top-1% mean pooling yields the object score. Markers show reference choices. Clouds and scores are measured; local surfaces are schematic.

2.1 Overview

Normal examples serve two roles: their mutual differences quantify normal geometric variation, while each provides a candidate surface against which the query is compared.

Let $\mathcal{R} = \{R_j\}_{j=1}^{m}$ denote the m aligned normal references of a category and $Q = \{x_i\}_{i=1}^{N} \subset \mathbb{R}^3$ a query point cloud with N points. References and queries share a common coordinate frame; Appendix A describes dataset-specific alignment and normalization.

Figure 2 summarizes the procedure. We estimate normal variation offline. At test time, each reference independently matches and scores every query point. The pointwise minimum combines these calibrated scores without learned parameters.

2.2 Per-Reference Geometric Matching

We first establish one geometric correspondence between a query point and each normal reference. Reference R_j stores K_j points $r_{jk} \in \mathbb{R}^3$, their unit normal axes n_{jk} and positive local point spacings h_{jk}, where $k = 1, \ldots, K_j$. For a query point x with estimated unit normal n_x, we consider its 16 nearest points in R_j and retain candidates with compatible surface orientation, $|n_x^\top n_{jk}| \geq 0.5$. Let $\mathcal{C}_j(x)$ contain their indices. The closest remaining candidate is selected:

$$\kappa_j(x) = \arg\min_{k \in \mathcal{C}_j(x)} \|x - r_{jk}\|_2, \qquad \pi_j(x) = r_{j,\kappa_j(x)}. \tag{1}$$

Here $\kappa_j(x)$ identifies the match and $\pi_j(x)$ its position. If no candidate satisfies the orientation constraint, we use the unrestricted nearest point. Equation 1 selects a sampled point, not an orthogonal projection.

Three geometric residuals. With matched normal $n_j(x) = n_{j,\kappa_j(x)}$ and spacing $h_j(x) = h_{j,\kappa_j(x)}$, define three query–reference residuals:

$$\begin{aligned} d_{\mathrm{e},j}(x) &= \|x - \pi_j(x)\|_2 / h_j(x), \\ d_{\perp,j}(x) &= |(x - \pi_j(x))^\top n_j(x)| / h_j(x), \\ d_{\mathrm{n},j}(x) &= 1 - |n_x^\top n_j(x)|. \end{aligned} \tag{2}$$

In Equation 2, Euclidean distance measures overall positional disagreement, point-to-plane offset measures displacement away from the reference surface, and normal-angle discrepancy captures a change in local surface orientation. Dividing the distance terms by local reference spacing reduces sensitivity to point-sampling density; the absolute dot product handles the sign ambiguity of estimated normals. At this stage, these quantities describe geometric differences only; they are not yet treated as anomaly scores.

2.3 Normal-Variation Calibration

The same geometric difference can have different meanings depending on how much normal objects already vary. We estimate this variation by holding out each normal scan and comparing it with another normal reference (Appendix A.1). Before testing, we pool these normal-to-normal residuals and record their 95th and 99.5th percentiles, $q_{95,a}$ and $q_{99.5,a}$, for each family $a \in \{\mathrm{e}, \perp, \mathrm{n}\}$. All references of a category share these statistics.

We retain both the excess above the normal 95th percentile and the residual's magnitude relative to it:

$$\begin{aligned} e_{a,j}(x) &= \left[\frac{(d_{a,j}(x) - q_{95,a})_+}{\max(q_{99.5,a} - q_{95,a}, \epsilon)} \right]_0^c, \\ u_{a,j}(x) &= \left[\log\left(1 + \frac{d_{a,j}(x)}{\max(q_{95,a}, \epsilon_{\mathrm{r}})}\right) \right]_0^c. \end{aligned} \tag{3}$$

Here $(v)_+ = \max(v, 0)$, $[v]_0^c$ clips to $[0, c]$, and log is the natural logarithm. We use $c = 20$, $\epsilon = 10^{-8}$ and $\epsilon_{\mathrm{r}} = 10^{-6}$. In Equation 3, the first coordinate measures how far a residual exceeds the normal 95th-percentile level, normalized by the width of the normal tail. The second measures its magnitude relative to that level, retaining differences below the 95th-percentile level.

To capture local surface context, we average each calibrated coordinate over the query's 16 nearest neighbors whose surface normals satisfy the same orientation threshold of 0.5. If none qualifies, we retain the pointwise value. Calibration precedes neighborhood averaging. Three residual families, two calibration transforms and two aggregation levels (pointwise and neighborhood) therefore yield a 12-dimensional calibrated feature vector for each query–reference pair, $\phi_j(x) \in \mathbb{R}^{12}$.

Table 5(a) evaluates calibration and local context jointly with matching and reference selection fixed (Appendix B.1).

2.4 Parameter-Free Per-Reference Scoring

Each normal reference is scored before we choose among references. For reference R_j, fixed-mean scoring averages its 12 calibrated features:

$$s_j(x) = \frac{1}{12} \sum_{\ell=1}^{12} \phi_{j,\ell}(x). \tag{4}$$

Here $\phi_{j,\ell}(x)$ is coordinate ℓ of $\phi_j(x)$. Equation 4 is the main AnyNormal scorer and contains no learned parameters. Crucially, reference selection occurs only after every reference has produced its own score:

$$s(x) = \min_{j \in \{1,\dots,m\}} s_j(x). \tag{5}$$

Equation 5 gives a low score whenever at least one normal reference explains the point well. Different points of the same query can select different references, whereas a high score arises only when no reference provides a close match.

Unlike pooled matching, this rule scores each reference before selection. In the controlled comparison of Table 5(b), normal data and calibration are held fixed, with 64 nearest-neighbor candidates in total: 64 from the merged set or 16 per reference.

For any point-score threshold τ, the accepted normal points satisfy

$$\{x \in Q : s(x) \le \tau\} = \bigcup_{j=1}^{m} \{x \in Q : s_j(x) \le \tau\}. \tag{6}$$

In Equation 6, the set of accepted query points is the union of the per-reference accepted sets. The references remain geometrically separate; their accepted regions combine through the pointwise minimum.

To obtain an object-level decision, we average the largest 1% of point scores:

$$S(Q) = \text{mean}\left(\text{Top}_{\lceil 0.01N \rceil}\{s(x_i)\}_{i=1}^{N}\right). \tag{7}$$

Here Top_k retains the k largest values, $\lceil \cdot \rceil$ rounds upward and mean is the arithmetic mean. The same scoring, reference-selection and pooling rules are used across categories and datasets.

3 Experiments

We compare AnyNormal with published methods (Table 1), examine geometric cues and scorer transfer, and measure compressed inference. Appendix B.1 provides the controlled design comparisons.

3.1 Experimental Setup

Datasets and normal support. We evaluate AnyNormal on Real3D-AD (Liu et al., 2023) with 12 categories, the standard 40-category subset of Anomaly-ShapeNet (Li et al., 2024), and MulSen-AD (Li et al., 2025) with 15 categories. Our few-shot setting uses four normal training scans per category on Real3D-AD and Anomaly-ShapeNet. On MulSen-AD, the main accuracy evaluation uses the full normal training set of 1,391 objects. Extended Anomaly-ShapeNet results are reported separately in the appendix.

Table 1: **Detection and localization across published point-cloud methods.** Object (O) and point (P) AUROC (%, higher is better). (a) Four-normal support on Real3D-AD and standard Anomaly-ShapeNet. (b) Full normal support on MulSen-AD. Rows report venue and training setting; no group uses real anomaly training examples. AnyNormal uses fixed-mean scoring without scorer training. Bold/underline mark best/second-best target-specific results across both training groups in (a), and across all rows in (b).

(a) Four normal scans per category

Method	Venue	Real3D-AD		Anomaly-ShapeNet	
		O	P	O	P
Target-specific: no synthetic-anomaly training					
Reg3D-AD (Liu et al., 2023)	NeurIPS'23	70.4	70.5†	57.2†	66.8†
IMRNet (Li et al., 2024)	CVPR'24	72.5	–	66.1	65.0
ISMP (Liang et al., 2025a)	AAAI'25	76.7	83.6	–	69.1
Reg2Inv (Yu et al., 2025)	NeurIPS'25	78.0	87.8	86.1	88.2
PASDF (Zheng et al., 2025)	ICCV'25	80.2	74.5	90.0	89.7
3DKeyAD (Wang et al., 2026)	IEICE'26	80.1	86.1	–	–
Simple3D (Cheng et al., 2026)	AAAI'26	80.4	92.3	86.0	92.9
RIF (Liang et al., 2026)	PR'26	–	78.3	–	84.5
Dual-residual (Yang et al., 2026)	Neurocomp.'26	82.0	–	**95.0**	85.6
PointCore (Zhao et al., 2026)	Neural Netw.'26	82.9	73.1	67.7	–
Template3D-AD$^{\triangle}$ (Liu et al., 2025)	IJCAI'25	84.4	92.5	86.5	–
MVR (Sun et al., 2025)	SMC'25	<u>89.6</u>	**95.7**	–	–
AnyNormal	–	**90.3**	<u>95.1</u>	<u>94.5</u>	**96.2**
Target-specific: synthetic-anomaly training					
Group3AD (Zhu et al., 2024)	ACM MM'24	75.1	73.5	–	–
R3D-AD (Zhou et al., 2024b)	ECCV'24	73.4	59.2	74.9	–
PO3AD (Ye et al., 2025)	CVPR'25	76.5	65.0	83.9	89.8
DUS-Net (Liang et al., 2025b)	ACM MM'25	79.5	84.7	79.7	71.2
MAD-3D (Lv et al., 2026)	IVC'26	77.3	83.6	90.3	<u>95.0</u>
AARD (Wu et al., 2026)	CVPR'26	82.0	86.0	–	93.6
CASL (Zha et al., 2026)	AAAI'26	82.3	88.2	88.7	89.9
Point-Patch (Kang et al., 2026)	CVPR'26	84.2	87.4	87.6	91.2
Shared multi-category models					
MC3D-AD§ (Cheng et al., 2025a)	IJCAI'25	78.2	76.8	84.2	74.8
SeDiR§ (Kim et al., 2026)	CVPR'26	81.0	80.6	93.3	81.0

–: no protocol-matched result reported. $\triangle$: one normal scan serves as the template. §: one model across categories. †: later Reg3D-AD results from Simple3D (Real3D-AD) and Reg2Inv (Anomaly-ShapeNet). Baselines retain published preprocessing.

(b) Full normal support: MulSen-AD

Method	Venue	O-AUROC	P-AUROC
M3DM (PointMAE) (Wang et al., 2023)	CVPR'23	62.8	58.7
M3DM (Point-BERT) (Wang et al., 2023)	CVPR'23	70.5	61.1
BTF (FPFH) (Horwitz & Hoshen, 2023)	CVPRW'23	72.1	72.9
Reg3D-AD (Liu et al., 2023)	NeurIPS'23	74.9	64.1
GLFM* (Cheng et al., 2025b)	TASE'25	78.5	66.5
IMRNet (Li et al., 2024)	CVPR'24	60.1	46.7
PatchCore (FPFH) (Roth et al., 2022)	CVPR'22	86.0	64.0
PatchCore (FPFH+raw) (Roth et al., 2022)	CVPR'22	83.3	62.0
PatchCore (PointMAE) (Roth et al., 2022)	CVPR'22	84.0	60.5
PointCore (Zhao et al., 2026)	Neural Netw.'26	83.7	–
Simple3D (Cheng et al., 2026)	AAAI'26	<u>88.2</u>	<u>80.3</u>
MAD-3D* (Lv et al., 2026)	IVC'26	<u>88.2</u>	80.1
AnyNormal	–	**89.0**	**87.0**

Point clouds only. PointMAE and Point-BERT denote the backbones of Pang et al. (2022); Yu et al. (2022). Baselines follow published MulSen-AD evaluations (Li et al., 2025; Cheng et al., 2026); $*$: synthetic perturbations. AnyNormal uses all 1,391 normal training objects.

Input and preprocessing. Main accuracy experiments use precomputed 16,384-point farthest-point-sampled clouds (Qi et al., 2017). Surface normals are estimated by local principal component analysis (PCA) over 16 neighbors, excluding the query point. Canonical alignment and normalization are established from normal training geometry; test clouds are aligned to this frame without using their annotations. Test labels are used exclusively for evaluation. Appendix A gives dataset-specific preprocessing and registration details.

Scoring configuration. Unless stated otherwise, AnyNormal uses all three residuals with normal-variation calibration. Fixed-mean scoring averages their pointwise and neighborhood features; the pointwise minimum selects across reference scores.

To test whether learned scoring improves on the fixed mean, we compare positive-weight linear combinations and monotone multilayer perceptrons (MLPs). By construction, these scorers' outputs are monotonically non-decreasing in every calibrated feature (Table 5(c); Appendix B.3). Section 3.4 evaluates MLP transfer across categories and datasets.

Metrics. We report category-averaged area under the receiver operating characteristic curve (AUROC) and average precision (AP) for object-level detection (O) and point-level localization (P). Accuracy values are percentages and differences are percentage points. Object scores average the largest 1% of point scores (Equation 7). AUROC and AP evaluate rankings without selecting a decision threshold.

3.2 Comparison with Existing Methods

Few-shot benchmarks. Table 1(a) compares published point-cloud methods on Real3D-AD and standard Anomaly-ShapeNet, with four normal objects per target category. We group target-specific methods by whether they train with synthetic anomalies and list shared multi-category models separately. Baselines retain their published preprocessing and training protocols, so the table compares end-to-end benchmark performance rather than isolated components.

With no learned scoring network, AnyNormal reaches 90.3/95.1% object/point AUROC on Real3D-AD and 94.5/96.2% on Anomaly-ShapeNet. Among the listed target-specific methods, it achieves the highest Real3D-AD object AUROC and Anomaly-ShapeNet point AUROC. MVR reports the highest Real3D-AD point AUROC (95.7%), while Dual-residual reports the highest Anomaly-ShapeNet object AUROC (95.0%).

Full-support benchmark. With all 1,391 normal training objects, AnyNormal reaches 89.0% object AUROC and 87.0% point AUROC on MulSen-AD (Table 1(b)). Appendix C provides per-category results, and Table 6 reports AP.

3.3 Analysis of Geometric Cues

Table 2 evaluates Euclidean distance, point-to-plane offset and normal-angle discrepancy individually, keeping the calibration and scoring procedure fixed. No single cue dominates across datasets. Point-to-plane gives the highest object AUROC on Anomaly-ShapeNet (95.86%) and MulSen-AD (91.17%), whereas normal-angle discrepancy gives the highest MulSen-AD point AUROC (89.18%). Combining all three gives the strongest object/point results on Real3D-AD (90.30/95.06%), but does not outperform every individual cue on the other benchmarks.

Table 2: **Complementary geometric cues.** Object/point AUROC (%). Normal-variation calibration, neighborhood features and fixed-mean scoring are retained. Anomaly-ShapeNet uses 40 categories; MulSen-AD uses full support. Shading marks the default.

Cue	Real3D-AD	Anomaly-ShapeNet	MulSen-AD
Point-to-plane	86.42 / 89.43	95.86 / 96.02	91.17 / 84.54
Normal-angle discrepancy	89.42 / 94.77	83.86 / 91.45	87.99 / 89.18
Euclidean	86.57 / 86.29	94.22 / 92.86	89.88 / 81.45
All three	90.30 / 95.06	94.50 / 96.24	88.96 / 87.04

These differing strengths support the use of complementary geometric residuals (Section 2.2). Table 7 reports the extended evaluation.

Design-choice analysis. Figure 3 evaluates the two main design choices—calibration with local context, and per-reference scoring. The former improves point AUROC in all 12 Real3D-AD and 40 Anomaly-ShapeNet categories; the latter improves point AP in 12/12 and 36/40 categories, respectively.

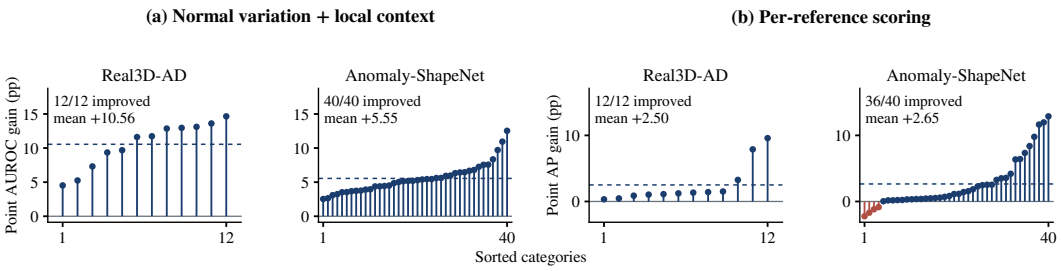


Figure 3: **Category-level gains.** (a) Normal variation and local context: point-AUROC gain. (b) Per-reference scoring: point-AP gain. Protocols follow Table 5(a,b). Categories are independently sorted; dashed lines mark means and solid lines mark zero. Units are percentage points (pp), with separate scales for the two metrics.

3.4 Generalization Across Source Categories and Datasets

Having analyzed the individual residuals, we ask whether learned scoring depends on the categories used to train it. We train a small monotone MLP on one, four or all categories of Real3D-AD or MulSen-AD, using four normal scans per source category, and hold its weights fixed when evaluating all three target datasets. Each target category retains its own normal references and calibration statistics; fixed-mean AnyNormal requires no scorer training.

Table 3: **How much source data does a learned scorer need?** Object/point AUROC (%). Fixed mean requires no scorer training; learned rows average three runs and use four scans per source category. K is the number of source categories; scans count scorer-training examples, not target normal support. All rows retain the same target references and calibration: four per category for Real3D-AD and standard Anomaly-ShapeNet, full support for MulSen-AD. Blue marks fixed mean; gray marks single-category training.

	Scorer training		Target: object / point AUROC		
Scorer / source	K	Scans	Real3D-AD	Anomaly-ShapeNet	MulSen-AD
Fixed mean (no training)	0	0	90.30 / 95.06	94.50 / 96.24	88.96 / 87.04
Real3D-AD	1	4	90.32 / 95.05	94.48 / 96.25	88.99 / 87.03
	4	16	90.22 / 95.05	94.54 / 96.25	89.01 / 87.02
	12 (all)	48	90.14 / 95.04	94.54 / 96.24	88.99 / 87.01
MulSen-AD	1	4	90.31 / 95.06	94.47 / 96.25	88.99 / 87.03
	4	16	90.31 / 95.06	94.46 / 96.25	88.98 / 87.03
	15 (all)	60	90.32 / 95.05	94.48 / 96.25	89.02 / 87.02

Compared with the fixed-mean baseline in Table 3, learned scoring shows no consistent gain from increasing the number of source categories. A single scorer trained on the four normal *toffees* scans from Real3D-AD reaches 90.31/95.05%, 94.46/96.25% and 88.99/87.03% object/point AUROC on Real3D-AD, Anomaly-ShapeNet and MulSen-AD, respectively, without further weight updates. These results are close to the fixed mean; the table separately reports averages over three runs.

The results support a division of roles: target normal examples provide category-specific shape information, while the rule for combining calibrated geometric features can be reused across datasets. Within-dataset averages include source training categories; cross-dataset tests hold scorer weights fixed but still use target normal support. Appendix D provides the results for all source-category counts and training schedules.

Localization and reference responses. Figure 4 shows per-reference responses (a,b) alongside localization examples across datasets (c): normal regions A/B favor different references, while defect C responds strongly to all four. Appendix E provides six objects per dataset and detailed reference maps.

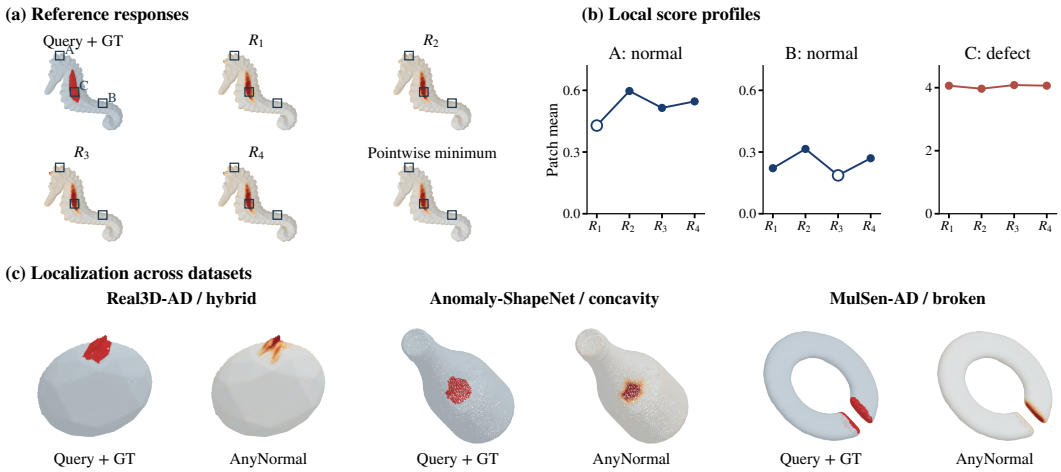


Figure 4: **Reference choice and anomaly localization.** (a) One query, four reference responses and their pointwise minimum share a view and score scale. (b) Normal regions A/B favor R_1/R_3; defect C responds strongly to all references. Patch means average scores within marked regions; C is annotated with a larger patch. (c) Query–prediction pairs labeled by dataset and defect type. Red on queries marks ground truth (GT); warmer map colors indicate higher scores.

3.5 Efficient Inference

The preceding accuracy results use full-resolution references. To reduce latency, we compress each of four references to 4,096 points and retain only point-to-plane scoring, which avoids query-normal estimation. Six sparse ICP steps align 2,048 query points; scoring and pointwise selection use all 16,384. References, calibration and GPU state are prepared offline.

Table 4: **Accuracy and efficiency of compressed inference.** RTX 4090; standard Anomaly-ShapeNet. (a) AUROC and host-memory-to-decision latency using stored sampling indices; p50/p95 denote percentiles. (b) Batch-one FPS (objects/s; mean±standard deviation, five trials per class) and peak allocated/reserved CUDA memory.

(a) Accuracy and latency

Dataset	AUROC (%) ↑ Object / Point	Latency (ms) ↓ p50 / p95
Real3D-AD	85.9 / 82.4	20.57 / 32.57
Anomaly-ShapeNet	90.1 / 88.0	20.41 / 32.14
MulSen-AD	79.3 / 71.4	20.37 / 32.04
All objects	–	**20.46 / 32.29**

(b) Continuous stream

Category	Objects per trial	FPS ↑ mean ± std.
bottle3	360	73.22 ± 0.34
screw	410	69.16 ± 1.56
piggy	400	67.48 ± 0.15
Overall	5,850	**69.74**

(a) Categories/objects: Real3D-AD 12/1,206; Anomaly-ShapeNet 40/1,312; MulSen-AD 15/644; total 67/3,162. (b) Fast/middle/slow workloads: Anomaly-ShapeNet bottle3, MulSen-AD screw/piggy. All use **18.0/1,782.0 MiB** peak allocated/reserved memory. Overall FPS = 5,850 objects / total time over 15 trials; see Appendix B.4 for selection.

Accuracy and resource cost. Table 4 reports the accuracy–speed trade-off, including the reduction to four MulSen-AD references. Continuous-stream throughput is 69.74 FPS (Table 4(b)); 5,850 objects take 83.879 s across 15 trials. Peak allocated memory is 18.0 MiB, with 1.74 GiB reserved for tensors, caching and graph pools.

Single-object latency. Timing starts with a cloud in host memory, already in the category coordinate frame, and includes sampling by stored indices, padding, transfers, sparse ICP, scoring, pooling and thresholding through Boolean output. Coarse alignment and normalization are preprocessing; the threshold is loaded offline. Across 3,162 objects, p50/p95 latency is 20.46/32.29 ms. This measures latency, not threshold quality; Table 9 gives stage costs.

4 Related Work

Reference-based 3D anomaly detection. Reference-based methods compare a test sample with stored normal examples. PatchCore retrieves nearest normal patch features (Roth et al., 2022), while BTF uses hand-crafted fast point feature histograms (FPFH) for 3D geometry (Horwitz & Hoshen, 2023; Rusu et al., 2009). Reg3D-AD combines registration with coordinate and PointMAE memories (Liu et al., 2023; Pang et al., 2022). PointCore stores coordinates and semantic features jointly (Zhao et al., 2026); Simple3D combines multi-scale neighborhood descriptors with local feature aggregation (Cheng et al., 2026). Reg2Inv and RIF develop rotation-robust representations (Yu et al., 2025; Liang et al., 2026). Template3D-AD uses center-point template matching and curvature-based local features (Liu et al., 2025). Closely related to our setting, 3DKeyAD aligns multiple normal prototypes, pools their corresponding keypoint-centered clusters and scores test points against the pooled clusters (Wang et al., 2026). AnyNormal instead retains a separate match and calibrated score for every reference before pointwise selection.

Geometric modeling and anomaly scoring. 3D anomaly scores can arise from learned representations or direct geometric comparison. 3D-ST detects disagreement between a student and a pretrained geometric-descriptor teacher (Bergmann & Sattlegger, 2023). Reconstruction methods use masked prediction, multi-view reconstruction, continuous surfaces or diffusion (Li et al., 2024; Sun et al., 2025; Zheng et al., 2025; Zhou et al., 2024b; Wu et al., 2026). PO3AD learns point offsets, MAD-3D learns direct pointwise discrimination, and DUS-Net preserves group centers during down-up reconstruction (Ye et al., 2025; Lv et al., 2026; Liang et al., 2025b). Point-to-plane distance and robust registration are established geometric tools (Chen & Medioni, 1992; Chetverikov et al., 2002). Yang et al. (2026) combine geometric and feature residuals after coarse-to-fine template registration without learned features. AnyNormal also uses direct geometry, but calibrates its Euclidean, point-to-plane and normal-angle residuals against variation between normal objects before combining them with a fixed mean.

Few-shot and shared anomaly detection. Few-shot anomaly detection characterizes normality from limited target data. Real3D-AD and Anomaly-ShapeNet provide four normal scans per category (Liu et al., 2023; Li et al., 2024). RegAD studies transferable few-shot reference comparison for images (Huang et al., 2022), while MC3D-AD and SeDiR learn shared 3D models across multiple categories (Cheng et al., 2025a; Kim et al., 2026). PointAD transfers vision–language knowledge to zero-shot 3D anomaly detection (Zhou et al., 2024a). AnyNormal separates category-specific geometry from the scoring rule: target normal examples provide the shape information and calibration, while the same scoring rule is used across categories and datasets. The default rule requires no training. As a control, a learned scorer trained on four normal scans from one source category is reused across datasets without weight updates. This constitutes scorer transfer with target normal support, not detection without any target data.

5 Conclusion

AnyNormal shows that effective 3D anomaly detection does not require a learned scoring network once normal geometry is used appropriately. Normal-to-normal comparisons provide the scale for interpreting geometric differences, and per-reference scoring allows different regions of a query to rely on different normal references. Together, these two choices support a parameter-free scoring rule that performs strongly across three benchmarks, while a scorer trained from only one source category transfers with essentially unchanged accuracy. A compressed configuration further reaches 69.74 FPS on an RTX 4090. These results suggest that target-specific shape and variation can remain in the normal references, leaving the final scoring rule simple, shared, and efficient.

Limitations. Geometric misalignment or normal variation absent from the references may be mistaken for defects. Because AnyNormal uses geometry alone, it does not address appearance-only defects.

Reproducibility Statement

Section 2 defines matching, residual calibration and scoring; Section 3.1 states datasets, support sizes and metrics. Appendix A documents preprocessing, neighborhoods, calibration and single-category scorer transfer. Additional quantitative, per-category and visual results appear in Appendices B–E. The anonymous repository linked in the abstract provides parameter-free inference, compressed CUDA inference, raw-data preprocessing, evaluation and tests.

AI Use Statement

AI tools assisted with grammar checking, local language refinement, organizing the presentation of experimental results, and structuring the code repository. The authors reviewed and verified the AI-assisted material and take responsibility for the final manuscript, reported results, and released code.

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 (WACV)*, pp. 2613–2623, 2023.

Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992. doi: 10.1016/0262-8856(92)90066-C.

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 *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 837–845, 2025a. doi: 10.24963/ijcai.2025/94.

Yuqi Cheng, Yunkang Cao, Dongfang Wang, Weiming Shen, and Wenlong Li. Boosting global-local feature matching via anomaly synthesis for multi-class point cloud anomaly detection. *IEEE Transactions on Automation Science and Engineering*, 22:12560–12571, 2025b. doi: 10.1109/TASE.2025.3544462.

Yuqi Cheng, Yihan Sun, Hui Zhang, Weiming Shen, and Yunkang Cao. Towards high-resolution 3D anomaly detection: A scalable dataset and real-time framework for subtle industrial defects. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 3327–3334, 2026. doi: 10.1609/aaai.v40i5.37328.

Dmitry Chetverikov, Dmitry Svirko, Dmitry Stepanov, and Pavel Krsek. The trimmed iterative closest point algorithm. In *International Conference on Pattern Recognition (ICPR)*, volume 3, pp. 545–548, 2002. doi: 10.1109/ICPR.2002.1047997.

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 Workshops (CVPRW)*, pp. 2968–2977, 2023. doi: 10.1109/CVPRW59228.2023.00298.

Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *European Conference on Computer Vision (ECCV)*, pp. 303–319, 2022. doi: 10.1007/978-3-031-20053-3_18.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24258–24267, 2026.

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 (CVPR)*, pp. 33036–33045, 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 (CVPR)*, pp. 22207–22216, 2024.

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

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 (AAAI)*, volume 39, pp. 5146–5154, 2025a. doi: 10.1609/aaai.v39i5.32546.

Hanzhe Liang, Jie Zhang, Tao Dai, Linlin Shen, Jinbao Wang, and Can Gao. Taming anomalies with down-up sampling networks: Group center preserving reconstruction for 3D anomaly detection. In *ACM International Conference on Multimedia*, pp. 7133–7141, 2025b. doi: 10.1145/3746027.3754492.

Hanzhe Liang, Jie Zhou, Can Gao, Bingyang Guo, Jinbao Wang, and Linlin Shen. A lightweight 3D anomaly detection method with rotationally invariant features. *Pattern Recognition*, 174:112924, 2026. doi: 10.1016/j.patcog.2025.112924.

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 *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, volume 36, pp. 30402–30415, 2023.

Yi Liu, Changsheng Zhang, and Yufei Yang. Template3D-AD: Point cloud template matching method based on center points for 3D anomaly detection. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1630–1638, 2025. doi: 10.24963/ijcai.2025/182.

Xiaoling Lv, Min Cao, Qiong Xie, and Silong Peng. From comparison to discrimination: A one-step framework for 3D anomaly detection and localization. *Image and Vision Computing*, 172:106032, 2026. doi: 10.1016/j.imavis.2026.106032.

Yatian Pang, Wenxiao Wang, Francis E. H. Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *European Conference on Computer Vision (ECCV)*, pp. 604–621, 2022. doi: 10.1007/978-3-031-20086-1_35.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pp. 5099–5108, 2017.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14318–14328, 2022. doi: 10.1109/CVPR52688.2022.01392.

Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (FPFH) for 3D registration. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3212–3217, 2009. doi: 10.1109/ROBOT.2009.5152473.

Yihan Sun, Yuqi Cheng, Yunkang Cao, Yuxin Zhang, and Weiming Shen. Multi-view reconstruction with global context for 3D anomaly detection. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 6397–6402, 2025. doi: 10.1109/SMC58881.2025.11343209.

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 (CVPR)*, pp. 8032–8041, 2023.

Zi Wang, Katsuya Hotta, Koichiro Kamide, Yawen Zou, Chao Zhang, and Jun Yu. 3DKeyAD: High-resolution 3D point cloud anomaly detection via keypoint-guided point clustering. *IEICE Transactions on Information and Systems*, E109.D(8):1261–1269, 2026. doi: 10.1587/transinf.2025EDP7152.

Linchun Wu, Qin Zou, Yuanhao Yue, and Zhongyuan Wang. Geometry-aligned and anomaly-aware reconstruction for 3D anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14648–14657, 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*, 684:133558, 2026. doi: 10.1016/j.neucom.2026.133558.

Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, and Kaizhu Huang. PO3AD: Predicting point offsets toward better 3D point cloud anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1353–1362, 2025.

Xumin Yu, Lulu Tang, Yongming Rao, Tiejun Huang, Jie Zhou, and Jiwen Lu. Point-BERT: Pre-training 3D point cloud transformers with masked point modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19313–19322, 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 (NeurIPS)*, volume 38, pp. 5806–5832, 2025. doi: 10.52202/085713-0180.

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 *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 12340–12348, 2026. doi: 10.1609/aaai.v40i15.38226.

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*, 197:108446, 2026. doi: 10.1016/j.neunet.2025.108446.

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 (ICCV)*, pp. 27063–27072, 2025. doi: 10.1109/ICCV51701.2025.02512.

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 *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 84866–84896, 2024a.

Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3D-AD: Reconstruction via diffusion for 3D anomaly detection. In *European Conference on Computer Vision (ECCV)*, pp. 91–107, 2024b. doi: 10.1007/978-3-031-72764-1_6.

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 International Conference on Multimedia*, pp. 4680–4689, 2024. doi: 10.1145/3664647.3680919.

A Protocol Details

A.1 Geometry and normal-variation calibration

Aligned inputs. Accuracy evaluation uses precomputed aligned and sampled point clouds. Anomaly-ShapeNet and MulSen-AD align test objects to a category template built by combining normal training geometry, then apply the stored category normalization. These alignment templates are distinct from the separate scoring references. Anomaly-ShapeNet uses robust bounding-box centering of each input and a shared center and scale fitted to aligned normal training scans. MulSen-AD refits normalization from all aligned normal training scans and disables per-object scale fitting. Its preprocessing uses fast point feature histograms (FPFH) for coarse alignment, followed by point-to-point and point-to-plane ICP. The compressed configuration instead includes six sparse ICP steps in online timing.

Local spacing. Spacing starts from the nearest non-self distance within each reference. Nonfinite values or distances at most 10^{-12} are replaced by the median of finite distances greater than 10^{-12}, denoted h_0 (fallback: 1). Values are clipped below at the maximum of their 5th percentile, $0.25h_0$ and 10^{-8}, and above at the larger of their 95th percentile and that lower bound. These values define h_{jk} in Section 2.2.

Normal-variation calibration. For each held-out normal scan, we select one remaining normal reference using correspondences with compatible surface-normal orientations (Section 2.2). The point cost is the spacing-normalized point-to-plane residual plus 0.25 times normal-angle discrepancy, with a penalty of 2 when no candidate satisfies the orientation constraint. Each reference's cost averages the lowest $\lceil 0.8N_{\text{hold}} \rceil$ point costs, where N_{hold} is the number of held-out points. We choose the lowest-cost reference; ties within 10^{-12} use sorted reference identifiers. Finite residuals pooled across the selected normal pairs determine the calibration quantiles. The matched experiment in Table 5 instead uses pooled remaining references for calibration in both variants.

Pointwise and neighborhood features. Neighborhood averaging retrieves 16 query neighbors and retains those with compatible surface-normal orientations. If none qualifies, it retains the pointwise value. The four feature blocks are pointwise excess, neighborhood excess, pointwise relative magnitude and neighborhood relative magnitude. Each block orders the residual families as point-to-plane, normal-angle discrepancy and Euclidean distance, giving the 12 features averaged in Equation 4.

A.2 Single-category training and transfer

Section 3.4 uses a monotone MLP with widths $12 \to 32 \to 16 \to 1$, trained on the four normal scans of Real3D-AD's *toffees* category, without anomaly examples or test-based selection. Its weights are then held fixed across all 12 Real3D-AD and 15 MulSen-AD categories and the standard 40 Anomaly-ShapeNet categories. We report the extended Anomaly-ShapeNet evaluation separately. The Real3D-AD average includes the training category; Anomaly-ShapeNet and MulSen-AD are cross-dataset evaluations.

Target normal geometry supplies preprocessing, references and calibration: four scans per category on Real3D-AD and Anomaly-ShapeNet, and all 1,391 normal training objects on MulSen-AD. Scores use top-1% mean pooling. These results describe one trained scorer, not a training-seed average. Main-text transfer and published-method comparisons both use the standard 40-category Anomaly-ShapeNet subset.

A.3 Framework display

Figure 2 uses measured normal and query objects, point scores and selected reference identities. Its local correspondence and residual sketches illustrate the geometry rather than a particular measured neighborhood. The anomaly map uses a logarithmic display clipped at the final map's 98.5th percentile. Four colored markers denote actual pointwise reference choices; they do not partition the object into four fixed regions.

B Additional Quantitative Results

Tables 6–9 supplement the main results with AP, geometric-cue and scorer controls, and stage timings.

B.1 Controlled design comparisons

Table 5 expands the controls in Section 2. In (b), pooled matching retrieves 64 nearest neighbors from the merged references, selects one match and scores it. Independent matching retrieves 16 neighbors per reference, selects and scores one match in each, then takes the minimum. Both retrieve 64 candidates using the same normal data and calibration; they score one and four selected matches, respectively. The control equates retrieved candidate counts, not scoring counts or total computation.

Gains precede rounding. The 95% intervals use the 2.5th/97.5th percentiles of 20,000 paired category-bootstrap means, not training-seed uncertainty. Figure 3 shows category-level gains.

Table 5: **What makes simple geometric scoring work?** (a) Normal variation and local context versus direct residuals, with matching and selection fixed; object/point AUROC. (b) Pooled versus independent matching with 64 retrieved candidates per query point. (c) Scorers share inputs and support (definitions in Appendix B.3); learned rows use four scans from each of 15 MulSen-AD categories and average three runs. Values are percentages; gains and 95% category-bootstrap intervals are percentage points. Anomaly-ShapeNet uses 40 categories; shading marks the proposed settings.

(a) How should geometric differences be interpreted?

	Real3D-AD		Anomaly-ShapeNet	
Variant	Object	Point	Object	Point
Direct geometric differences	86.39	84.50	94.83	90.69
Using normal variation + local context	**90.30**	**95.06**	94.50	**96.24**
Gain	+3.91	+10.56	−0.33	+5.55
95% CI	[1.36, 7.15]	[8.66, 12.29]	[−0.90, 0.15]	[4.93, 6.23]

(b) How should normal examples be used?

Real3D-AD (12 categories)

Metric	Pooled	Indep.	Gain	95% CI
O-AUROC	87.51	**89.73**	+2.21	[0.77, 4.23]
O-AP	86.11	**88.06**	+1.95	[0.56, 3.83]
P-AUROC	94.30	**94.75**	+0.45	[0.29, 0.63]
P-AP	43.06	**45.57**	+2.50	[1.09, 4.34]

Anomaly-ShapeNet (40 categories)

Metric	Pooled	Indep.	Gain	95% CI
O-AUROC	94.26	**94.69**	+0.43	[−0.03, 0.91]
O-AP	96.66	**97.01**	+0.35	[0.03, 0.74]
P-AUROC	95.88	**96.21**	+0.33	[0.20, 0.49]
P-AP	63.37	**66.02**	+2.65	[1.56, 3.88]

Pooled: merge before matching; Indep.: score separately, then choose. O/P denote object/point metrics.

(c) Does the final scorer need learning?

Scoring rule	Real3D-AD	Anomaly-ShapeNet	MulSen-AD
	Object / point AUROC (%)		
Fixed mean	90.30 / 95.06	94.50 / 96.24	88.96 / 87.04
Positive linear	90.32 / 95.06	94.50 / 96.26	88.99 / 87.02
Trained MLP (constant init.)	90.32 / 95.05	94.48 / 96.25	89.02 / 87.02
Trained MLP (random init.)	90.31 / 95.05	94.49 / 96.26	89.03 / 87.03

B.2 Average precision

Table 6: **Average precision of fixed-mean AnyNormal.** Category-averaged object and point AP (%, higher is better). The standard 40-category subset and extended 52-category evaluation of Anomaly-ShapeNet are reported separately; normal support is listed for each dataset.

Evaluation	Categories	Normal support / category	Object AP	Point AP
Real3D-AD	12	4	88.48	46.62
Anomaly-ShapeNet (standard)	40	4	96.90	65.20
Anomaly-ShapeNet (extended)	52	4	95.81	63.55
MulSen-AD	15	full	89.42	21.52

B.3 Residual families and scoring

Table 7: **Additional residual-family results.** Anomaly-ShapeNet uses the extended 52-category evaluation here. All rows use fixed-mean scoring and retain both calibration transforms and pointwise and neighborhood features for the included families. O/P: object/point AUROC (%, higher is better). Shading marks the default; bold marks the best per column. MulSen-AD uses full normal support.

	Real3D-AD		Anomaly-ShapeNet		MulSen-AD	
Residual family	O	P	O	P	O	P
All three families (default)	**90.30**	**95.06**	92.82	**95.85**	88.96	87.04
Point-to-plane only	86.42	89.43	**94.17**	95.47	**91.17**	84.54
Normal-angle discrepancy only	89.42	94.77	82.97	91.92	87.99	**89.18**
Euclidean only	86.57	86.29	92.44	92.44	89.88	81.45

Table 8: **Additional scorer controls.** Learned scorers are trained separately on each evaluation dataset: four scans from every Real3D-AD or Anomaly-ShapeNet category, or all normal MulSen-AD training scans. Anomaly-ShapeNet uses 52 categories. In contrast, Table 5(c) transfers one MulSen-AD-trained scorer across datasets. All rows use the same 12 features and pointwise minimum. O/P: object/point AUROC (%). Shading marks fixed-mean scoring.

	Real3D-AD		Anomaly-ShapeNet		MulSen-AD	
Scorer	O	P	O	P	O	P
Fixed mean (default)	90.30	95.06	92.82	95.85	88.96	87.04
Untrained MLP (constant init.)	90.31	95.06	92.82	95.85	88.99	87.04
Untrained MLP (random init.)	90.31	95.06	92.81	95.86	89.00	87.05
Trained MLP (constant init.)	90.14	95.04	92.92	95.86	88.96	87.06
Positive linear	90.38	95.15	92.64	95.91	89.09	86.99
Trained MLP (random init.)	90.14	95.04	92.92	95.86	88.96	87.06

The displayed scores span at most 0.28 object-level and 0.11 point-level AUROC percentage points within each dataset.

Scorer definitions. The positive linear scorer learns a positive-weight sum and a bias. All MLPs have widths $12 \to 32 \to 16 \to 1$, positive weights and softplus hidden activations, making their scores nondecreasing in each input feature. We parameterize weights by applying softplus to unconstrained values. Constant initialization sets these values to -2 and layer biases to zero. Random initialization draws them independently from a normal distribution with mean -2: standard deviation 0.5 with zero layer biases for untrained controls, or 0.1 with bias standard deviation 0.01 for trained controls. The final output offset starts at -2. Trained rows average three runs; untrained random rows average five initializations.

What the MLP controls establish. Among 36 trained scorers with constant initialization, hidden units remained identical: each network reduces to a strictly increasing function of one positive linear projection of the 12 features. We tested these functions on 16,501 synthetic nonnegative 12-dimensional feature vectors, not point clouds. Their Spearman correlation with fixed-mean scoring is at least 0.9981 (mean 0.9996). This symmetry limits those networks' expressiveness; Table 8's random-initialization controls test whether similar detection performance persists without it. Point-score rank preservation does not imply identical object rankings after nonlinear transformation and top-1% mean pooling.

B.4 Online latency

Table 9: **Where online time is spent.** Stage wall times in milliseconds for the 3,162-object compressed benchmark in Table 4(a). Timing starts with sampling by stored indices and ends with Boolean output. Stage percentiles are computed separately and do not sum to the end-to-end percentile.

Online stage	p50	p95	Mean
Sample / pad 16,384 points	0.495	1.285	0.674
Pin host memory	0.252	0.321	0.255
Host-to-device transfer	0.061	0.081	0.063
Registration and scoring (GPU)	19.019	30.818	21.168
Device-to-host transfer	0.169	0.242	0.178
Threshold and output	0.002	0.004	0.003
Other wall-time overhead	0.216	0.283	0.220
Host memory to decision	**20.458**	**32.291**	**22.562**

For stream trials, we rank median single-object latency across 12 Real3D-AD, 52 Anomaly-ShapeNet and 15 MulSen-AD categories, and select workloads at the 10th, 50th and 90th percentiles. All three selected categories also belong to the standard evaluation. FPS is measured separately, not computed from median latency.

C Per-Category Main Results

Tables 10 and 11 report fixed-mean AnyNormal on all 79 evaluated categories. Macro rows are unweighted means of category-level metrics. The standard 40-category subset and extended 52-category evaluation of Anomaly-ShapeNet are listed separately; these tables do not report the single-category-trained MLP.

Table 10: **Fixed-mean results by category: Real3D-AD and MulSen-AD.** O/P denote object/point AUROC (%, ↑). Real3D-AD uses four normal references per category; MulSen-AD uses full normal training support. Macro rows give unweighted means of category-level AUROC.

Real3D-AD (12 categories)			MulSen-AD (15 categories)		
Category	O↑	P↑	Category	O↑	P↑
airplane	88.1	94.4	button cell	82.3	92.4
candybar	97.6	95.5	capsule	100.0	93.8
car	81.6	94.7	cotton	37.2	82.1
chicken	86.5	91.6	cube	98.7	83.3
diamond	99.4	96.9	flat pad	98.5	78.5
duck	79.4	94.9	light	100.0	91.4
fish	100.0	97.1	nut	92.9	95.1
gemstone	77.0	91.5	piggy	100.0	93.8
seahorse	99.6	95.8	plastic cylinder	94.1	92.9
shell	94.6	96.9	screen	100.0	77.3
starfish	94.4	96.5	screw	99.8	86.7
toffees	85.4	94.9	solar panel	52.4	84.4
			spring pad	100.0	90.1
			toothbrush	89.8	88.8
			zipper	88.7	74.9
Macro	**90.3**	**95.1**	**Macro**	**89.0**	**87.0**

Table 11: **Fixed-mean results by category: Anomaly-ShapeNet.** O/P denote object/point AUROC (%, ↑), using four normal references per category. All 52 evaluated categories are listed; the final rows separately average the standard 40-category subset and the extended 52-category set.

Category	O↑	P↑	Category	O↑	P↑	Category	O↑	P↑
ashtray0	100.0	95.8	cap4	87.0	97.4	microphone1	77.5	98.9
bag0	96.7	98.9	cap5	86.0	98.4	screen0	94.1	95.7
bottle0	100.0	99.6	chair0	86.0	85.9	shelf0	82.9	92.9
bottle1	100.0	95.1	cup0	100.0	99.6	tap0	83.9	96.3
bottle3	100.0	97.6	cup1	100.0	96.3	tap1	77.8	96.3
bowl0	100.0	99.3	cup2	100.0	99.6	vase0	99.2	98.2
bowl1	95.2	93.6	desk0	83.0	95.7	vase1	100.0	85.6
bowl2	95.9	95.2	eraser0	100.0	98.8	vase10	91.6	91.4
bowl3	100.0	98.8	headset0	100.0	95.9	vase2	100.0	97.7
bowl4	93.7	96.7	headset1	100.0	97.4	vase3	90.3	96.8
bowl5	91.6	96.9	helmet0	85.2	93.0	vase4	87.6	96.0
bucket0	97.1	89.4	helmet1	100.0	96.7	vase5	100.0	95.6
bucket1	84.8	94.9	helmet2	92.5	95.8	vase6	89.1	93.0
cabinet0	84.0	91.0	helmet3	97.3	98.9	vase7	100.0	99.5
cap0	88.9	97.6	jar0	100.0	93.2	vase8	87.9	97.5
cap1	94.7	98.6	knife0	78.0	93.9	vase9	87.3	93.6
cap2	77.2	94.5	knife1	91.7	96.4			
cap3	91.2	94.2	microphone0	100.0	98.6			
Macro (standard 40)							**94.5**	**96.2**
Macro (extended 52)							92.8	95.9

D Scorer Transfer Across Categories and Datasets

Table 12 reports transfer as the number of scorer-training categories increases: Anomaly-ShapeNet uses 52 categories for both scorer training and evaluation. This differs from the standard 40-category evaluation in Table 3. Cells report object/point AUROC as mean±sample standard deviation (SD) over three runs. Within-dataset averages include categories used for scorer training; cross-dataset evaluation retains target normal references but performs no weight updates.

For Real3D-AD and Anomaly-ShapeNet, we use three predetermined category orders, adding categories as K increases; the full-category setting uses three training seeds. These runs use 30 epochs, so optimizer updates grow with K. The four-scan MulSen-AD runs use 120 updates at every K and select four scorer-training scans per source category; their normal-variation calibration uses the full normal training set. These counts describe scans used for scorer training; target references and calibration are separate.

Table 12: **Scorer transfer with varying source-category counts.** Object/point AUROC (%, mean±SD over three runs). Every row uses four scorer-training scans per source category. $K = 1$ is one source category; the last row of each block uses all source categories. Shading marks $K = 1$.

K	Scans	Real3D-AD	Anomaly-ShapeNet	MulSen-AD
Source: Real3D-AD				
1	4	90.32±0.01 / 95.05±0.00	92.86±0.03 / 95.86±0.00	88.99±0.01 / 87.03±0.01
2	8	90.29±0.00 / 95.06±0.02	92.92±0.05 / 95.86±0.01	89.01±0.03 / 87.02±0.02
4	16	90.22±0.00 / 95.05±0.01	92.93±0.03 / 95.86±0.00	89.01±0.03 / 87.02±0.01
8	32	90.15±0.01 / 95.03±0.00	92.94±0.02 / 95.85±0.00	89.01±0.03 / 87.01±0.01
12	48	90.14±0.01 / 95.04±0.00	92.96±0.02 / 95.85±0.00	88.99±0.02 / 87.01±0.00
Source: Anomaly-ShapeNet				
1	4	90.31±0.01 / 95.05±0.01	92.86±0.06 / 95.86±0.01	88.98±0.01 / 87.03±0.02
2	8	90.29±0.01 / 95.05±0.01	92.89±0.03 / 95.85±0.01	89.02±0.02 / 87.03±0.01
4	16	90.22±0.02 / 95.05±0.01	92.90±0.03 / 95.85±0.00	89.01±0.02 / 87.03±0.01
8	32	90.16±0.03 / 95.04±0.00	92.96±0.05 / 95.86±0.01	89.04±0.01 / 87.01±0.01
16	64	90.13±0.02 / 95.04±0.01	92.98±0.06 / 95.85±0.01	89.02±0.03 / 87.01±0.01
32	128	90.01±0.02 / 95.01±0.04	92.92±0.12 / 95.86±0.01	89.01±0.10 / 87.00±0.06
52	208	89.92±0.01 / 95.00±0.02	92.92±0.10 / 95.86±0.00	89.00±0.06 / 86.98±0.03
Source: MulSen-AD				
1	4	90.31±0.00 / 95.06±0.01	92.85±0.04 / 95.86±0.00	88.99±0.00 / 87.03±0.02
2	8	90.31±0.01 / 95.06±0.01	92.83±0.02 / 95.86±0.00	88.99±0.02 / 87.04±0.01
4	16	90.31±0.00 / 95.06±0.00	92.84±0.01 / 95.86±0.00	88.98±0.01 / 87.03±0.00
8	32	90.31±0.00 / 95.05±0.00	92.85±0.00 / 95.85±0.00	88.99±0.00 / 87.03±0.00
15	60	90.32±0.00 / 95.05±0.00	92.88±0.00 / 95.86±0.00	89.02±0.01 / 87.02±0.00

Table 13 reports complementary MulSen-AD experiments trained with every available normal scan in each selected source category. Evaluation references and pooling are unchanged. These results are separate from the four-scans-per-category setting above.

Table 13: **Scorer training with all available MulSen-AD normal scans.** Object/point AUROC (%, mean±SD over three category orders). Each selected category contributes all its available normal training scans; all 15 categories provide 1,391 scans in total. Target evaluation covers every category.

K	Real3D-AD	Anomaly-ShapeNet	MulSen-AD
1	90.10±0.01 / 95.03±0.02	93.04±0.03 / 95.86±0.02	89.04±0.03 / 86.99±0.01
2	90.08±0.05 / 95.00±0.03	93.05±0.02 / 95.85±0.02	89.07±0.11 / 86.98±0.03
4	90.06±0.01 / 95.00±0.02	92.95±0.07 / 95.83±0.02	89.04±0.08 / 87.00±0.03
8	90.05±0.01 / 94.99±0.01	92.84±0.04 / 95.80±0.01	88.99±0.03 / 87.04±0.02
15	90.05±0.00 / 94.99±0.01	92.77±0.02 / 95.79±0.01	88.96±0.01 / 87.06±0.01

D.1 Effect of source-category count

Figure 5 plots the extended 52-category Anomaly-ShapeNet evaluation and all source-category counts in Table 12. Normalizing the horizontal axis by the total number of source categories allows the three source datasets to be shown together. The first marker is one source category, not zero training data. Each selected category contributes four scorer-training scans.

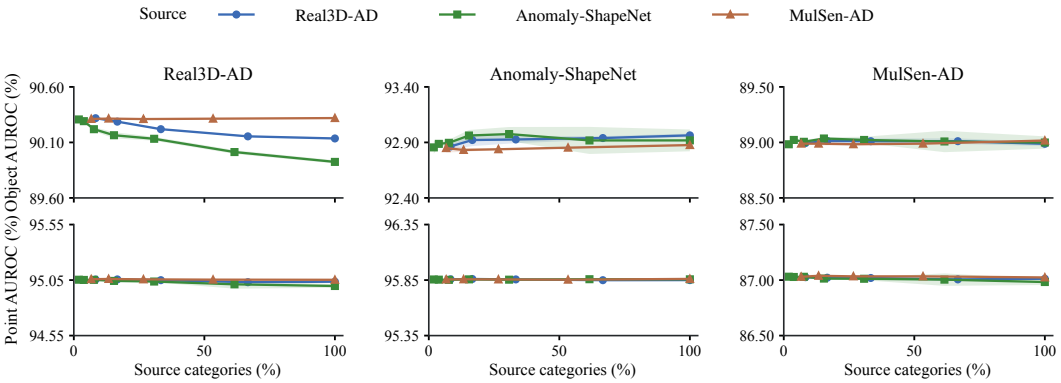


Figure 5: **More source categories do not consistently improve the scorer.** Columns identify the complete evaluation dataset; colors and marker shapes identify the source training dataset. Top/bottom rows report object/point AUROC. Lines show three-run means and bands show sample standard deviations, without smoothing. The horizontal axis is the percentage of available source categories, not the target reference count. All vertical axes span one AUROC percentage point, with different centers. Exact category counts and training schedules appear in Appendix D.

Interpretation. The curves support scorer reuse under fixed target geometry, not a claim that training on fewer categories is universally better. In particular, source-category coverage and the number of optimizer updates vary together in the Real3D-AD and Anomaly-ShapeNet runs. The main fixed-mean method avoids fitting this scorer altogether.

E Additional Visual Evidence

We show AnyNormal localization maps and inspect individual reference responses.

E.1 Anomaly localization

Figure 6(a–c) shows 18 objects selected for category/defect diversity and high AnyNormal point AP. We use four normal references for Real3D-AD and Anomaly-ShapeNet, and full normal support for MulSen-AD, following the main evaluation protocol.

Each column shows the annotated query, the AnyNormal anomaly map and a local enlargement. Boxes use annotated defect locations. The full map and its enlargement share the same view and fixed linear score scale.

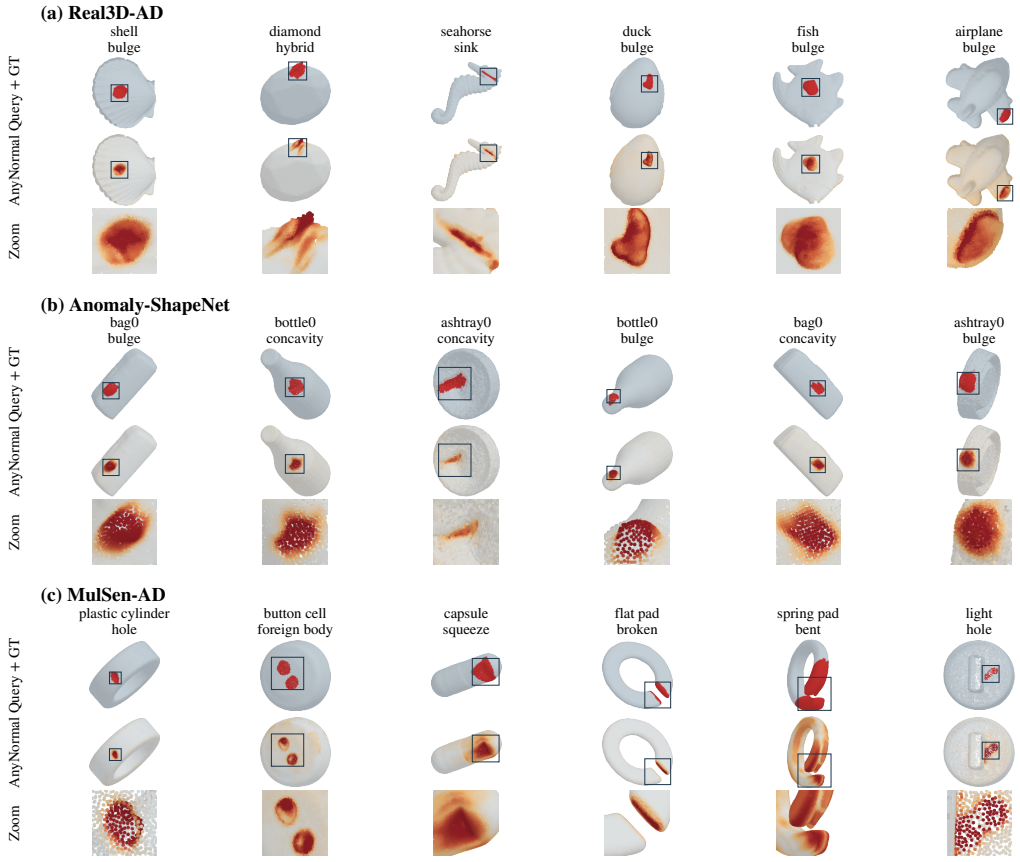


Figure 6: **AnyNormal localization across three benchmarks.** Six objects each from (a) Real3D-AD, (b) Anomaly-ShapeNet and (c) MulSen-AD. Rows show annotated queries, anomaly maps and enlarged boxed regions. Red on queries marks ground-truth defects; warmer map colors indicate higher anomaly scores. Views and linear score scales are fixed within each object.

E.2 Reference choices and responses

Figure 7(a) shows the selected references across complete queries; its colors encode reference choices rather than anomaly labels. Panel (b) expands the query in Figure 4 with the four normal references and local maps. Normal regions A/B have different lowest-scoring references, while defect C has high average responses against every reference. Selection remains pointwise.

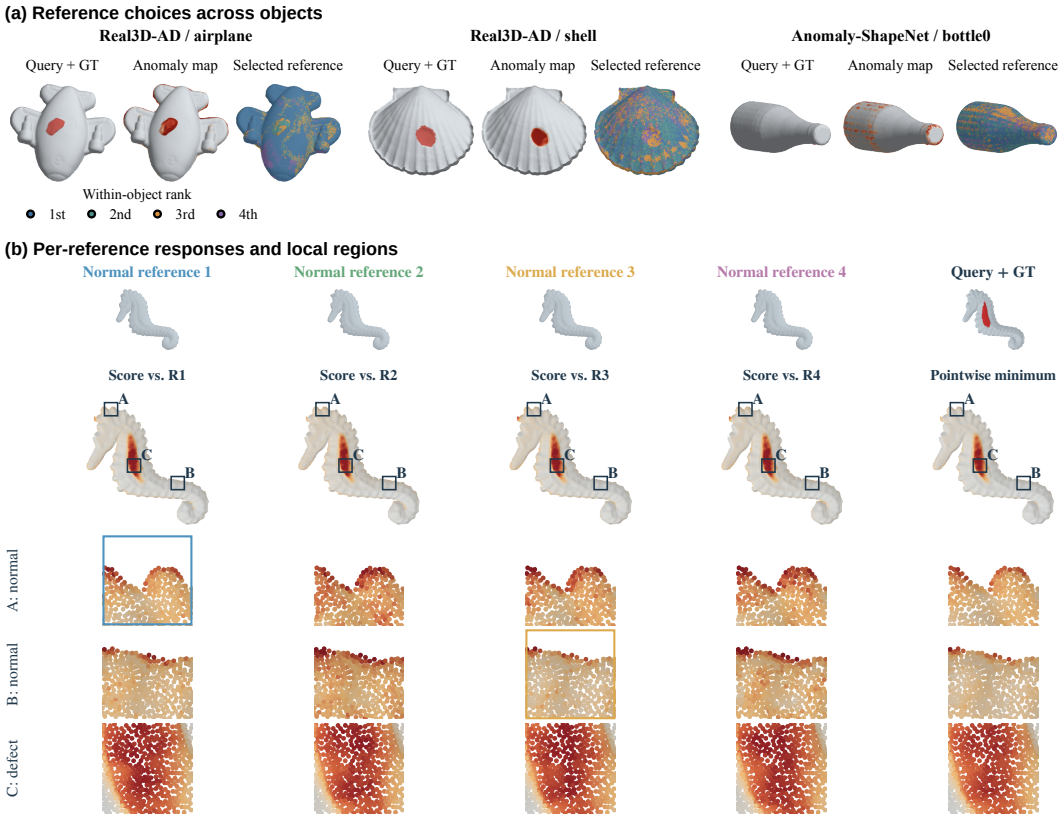


Figure 7: **Pointwise choices and their geometric evidence.** (a) Annotated queries, anomaly scores and selected references for three queries, each with four references. Reference colors denote within-object selection-frequency rank, not shared identities across objects; score colors are normalized per object for display. (b) Four normal references and the query, followed by reference-specific responses and their pointwise minimum. Normal regions A/B favor R_1/R_3; C contains a defect. Each row shares a linear scale, with separate ranges for the local rows. Highlighted maps have the lowest patch mean; selection remains pointwise.