SES-3D: BEYOND SPATIAL CORRESPONDENCE FOR FEW-TEMPLATE 3D ANOMALY LOCALIZATION

Anonymous authors

ABSTRACT

Few-template 3D anomaly localization requires reliable normality estimation from only a handful of reference shapes. Spatial registration establishes local correspondences, but distribution-based scoring remains sensitive to the limited observations available at each location. We propose Spatial Exemplar Scoring (SES-3D), a framework that represents local normality with spatially aligned exemplars and detects anomalies through their nearest valid match. By preserving individual normal configurations, SES avoids local variance estimation and introduces no additional trainable parameters. We integrate SES with a registered patch memory and Gaussian point projection, and systematically separate the effects of spatial retrieval and score estimation. On Real3D-AD, SES achieves 98.54% point AUROC and 63.32% point average precision, improving the matched effect-size baseline by 2.18 and 16.70 percentage points, respectively. On Anomaly-ShapeNet, it achieves 98.64% point AUROC. Experiments with two and three reference templates, alternative estimators, and multiple feature representations demonstrate the effectiveness of exemplar-based normality estimation for data-scarce 3D localization. Code is available at `https://anonymous.4open.science/r/SES_3D-D25E/`.

1 INTRODUCTION

Detecting geometric defects in 3D point clouds is central to automated inspection. Anomaly localization must distinguish small surface deviations from normal variation, often with only a few defect-free references. A successful detector must establish reliable correspondences while retaining the detail needed to identify cracks, bulges, and local deformation.

Recent 3D anomaly detectors address this problem through learned representations, reconstruction, and normal-reference matching (Liu et al., 2023; Zhou et al., 2024; Zheng et al., 2025; Cheng et al., 2026). Registered template methods provide a particularly direct solution: align a query to normal references and compare features at corresponding surface regions (Liu et al., 2025; Tarvo et al., 2026). Spatial correspondence prevents a defective region from being explained by an unrelated normal part elsewhere on the object. However, alignment alone does not determine the anomaly score. Even after accurate registration, each location may contain only two, three, or four normal observations, making local normality estimation a distinct challenge.

This challenge is often obscured by treating retrieval and scoring as one operation. A location-specific effect-size score compares query–reference distances with reference–reference distances and normalizes their difference by an estimated pooled standard deviation. With few references, both the configuration and the normalization influence the anomaly map. Improving features or alignment does not directly resolve this estimation problem.

We introduce *Spatial Exemplar Scoring* for 3D anomaly localization, abbreviated SES-3D. The key idea is to represent normality at each registered location by its observed exemplars, rather than compressing them into a local prototype or moment-normalized statistic. A query patch receives a low anomaly score when at least one spatially corresponding normal exemplar explains it well (Figure 1). This preserves alternative normal configurations and eliminates the need to estimate a local variance. The score can be computed from a single valid reference and integrates directly with a registered patch memory.

To study this design, we separate *WHERE* normal references are retrieved from *HOW* their evidence is scored. The WHERE comparison tests global, shuffled, and corresponding retrieval using the

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.

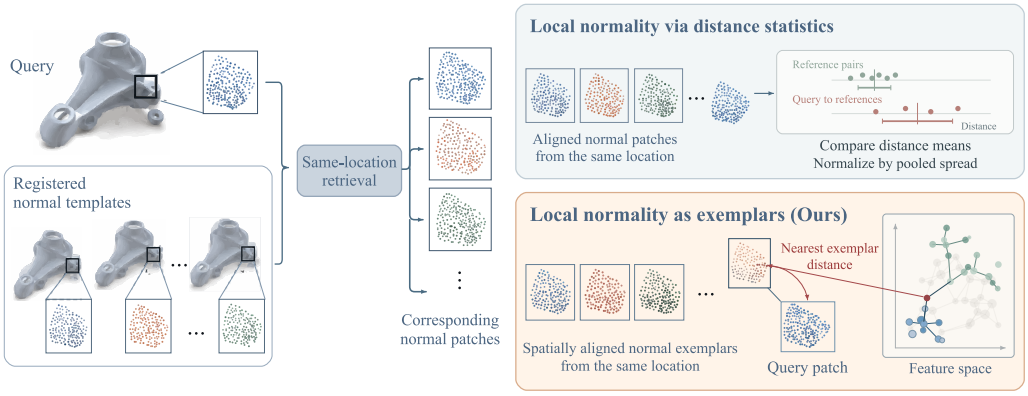

Figure 1: **From distance statistics to spatial exemplars.** Registered retrieval provides normal patches from the same location. Statistical scoring compresses their distances into moments; SES preserves individual exemplars and scores a query by its nearest valid match. The upper-right panel illustrates scalar distance samples and their spreads.

same cosine-distance rule. The HOW comparison fixes the candidate set and evaluates effect-size, variance-shrinkage, centroid, and exemplar scores. This design attributes changes in localization to concrete components of the inference pipeline. It also makes the role of the reference budget explicit, instead of conflating fewer templates with a different backbone or registration procedure.

SES delivers substantial localization gains on both real and synthetic point clouds. On Real3D-AD, it improves point AUROC from 96.36% to 98.54% and point AP from 46.62% to 63.32% under the matched evaluation. On the 40-category Anomaly-ShapeNet benchmark, it reaches 98.64% point AUROC. Three-template experiments improve AP over the raw effect-size score by 16.59 and 18.02 percentage points on ShapeNet and Real3D, respectively. Two-template experiments further demonstrate that explicit exemplar support remains effective when very little normal evidence is available.

Our contributions are threefold:

- We propose SES-3D, a spatial exemplar scoring framework that preserves local normal observations and produces dense anomaly maps without estimating a per-location variance or learning additional scoring parameters.
- We develop a controlled WHERE–HOW evaluation that separates spatial correspondence from normality estimation, revealing the importance of the scoring rule in few-template 3D localization.
- We demonstrate strong localization performance on Real3D-AD and Anomaly-ShapeNet, supported by few-template experiments, estimator comparisons, representation transfer, and qualitative analysis.

2 Related Work

3D anomaly detection and localization. Existing methods exploit geometric descriptors, learned point features, reconstruction, and implicit surfaces. BTF revisits classical 3D features, while CPMF, PointCore, and Group3AD improve local or multimodal representations (Horwitz & Hoshen, 2023; Cao et al., 2024; Zhao et al., 2024; Zhu et al., 2024). Reconstruction-based approaches, including R3D-AD and MC3D-AD, compare an observed shape with an estimated normal counterpart; PO3AD learns point offsets for anomaly detection (Zhou et al., 2024; Cheng et al., 2025; Ye et al., 2025). PASDF uses a continuous geometric representation, Simple3D aggregates high-resolution local features, and CASL introduces curvature-aware representation learning (Zheng et al., 2025; Cheng et al., 2026; Zha et al., 2026). These developments improve the geometric information available to a detector. SES complements them by addressing how a small set of aligned normal features is converted into anomaly evidence.

Normal memories and few-shot detection. Memory-based detection compares query features against stored normal observations. PatchCore uses nearest-neighbor feature distances, whereas PaDiM estimates position-specific feature distributions (Roth et al., 2022; Defard et al., 2020). RegAD and FastRecon address limited normal data through registration and feature reconstruction (Huang et al., 2022; Fang et al., 2023). UniVAD also uses a small set of normal references at inference (Gu et al., 2025); FIND and GPFlow study normal-only multimodal inspection and Gaussian-prototype modeling (Li et al., 2025; 2026). Our setting combines reference scarcity with registered 3D geometry. The spatial restriction preserves correspondence, while the exemplar representation retains the alternatives observed at each local surface region.

Registered 3D template matching. Reg3D-AD and Template3D-AD demonstrate the effectiveness of alignment and template-based geometric comparison (Liu et al., 2023; 2025). COSSAD combines contrastive patch features with spatially indexed memories and effect-size scoring (Tarvo et al., 2026). We build on this registered feature pipeline and introduce an exemplar-based localization rule. Our controlled comparisons share its encoder, alignment, template identities, and point projection. This relationship enables a direct assessment of the scoring component alongside comparisons with other complete 3D detectors.

3 METHOD: SES-3D

Figure 2 presents the framework. A query and its normal references are registered to a common frame, their patch features populate a spatial memory, and nearest-exemplar matching produces local anomaly scores. Gaussian projection then converts these scores into a dense point-level map.

3.1 SPATIAL EXEMPLAR MEMORY

Let $X = \{x_j\}_{j=1}^{N}$ denote a query point cloud and $\{X_t\}_{t=1}^{T}$ a set of normal reference shapes. Registration establishes corresponding local regions, indexed by $i \in \{1, \dots, S\}$. A frozen patch encoder extracts a query feature $q_i \in \mathbb{R}^d$ and reference features $r_{t,i} \in \mathbb{R}^d$. Because partial scans need not observe every region, the valid template set $\mathcal{T}_i$ can vary with location. We construct the memory as

$$\mathcal{R}_i = \{r_{t,i} : t \in \mathcal{T}_i\}, \qquad \mathcal{M} = \{\mathcal{R}_i\}_{i=1}^{S}. \tag{1}$$

A slot denotes a registered local region, rather than a raw point index. This representation separates geometric correspondence from the score used to compare a query with its normal references. In particular, it stores each valid exemplar independently, preserving reference-specific variations that a single local prototype would average together.

We normalize nonzero features to unit length, $\widehat{q} = q/\|q\|_2$ and $\widehat{r} = r/\|r\|_2$, and use cosine distance

$$d(q, r) = 1 - \widehat{q}^{\top}\widehat{r}. \tag{2}$$

The same distance and reference features are used throughout the controlled retrieval and estimator comparisons. Invalid features are omitted from the exemplar set; a location with no valid query–reference comparison contributes no exemplar score.

3.2 SPATIAL EXEMPLAR SCORING

SES defines the local anomaly score as the distance to the nearest valid exemplar at the corresponding location:

$$s_i^{\mathrm{SES}} = \min_{t \in \mathcal{T}_i} d(q_i, r_{t,i}). \tag{3}$$

This rule expresses local normality as support by an observed normal configuration. It preserves the alternatives in $\mathcal{R}_i$ and requires neither a per-slot covariance model nor an estimated variance of pairwise distances. Each comparison remains interpretable: the selected reference identifies the normal observation that most closely explains the query patch.

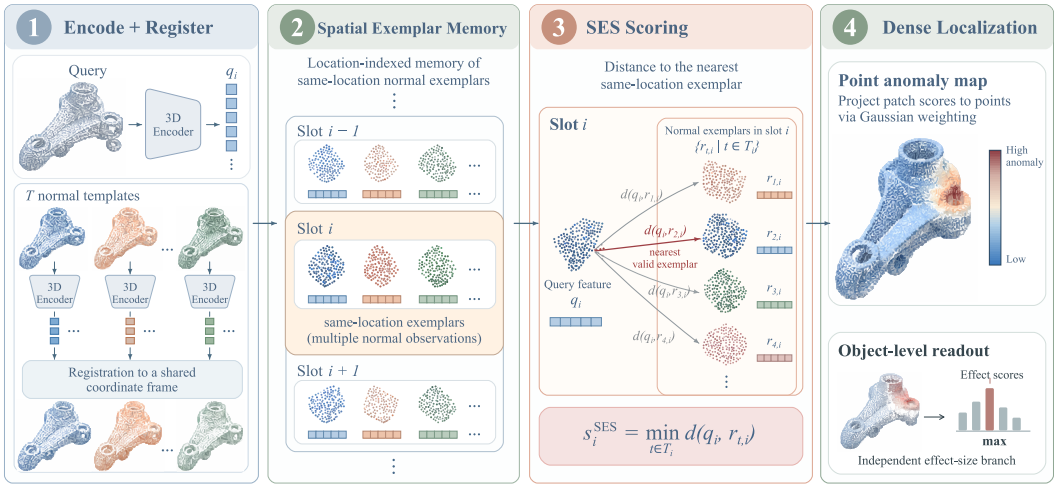

Figure 2: **Overview of SES-3D.** Registered patch features form a spatial exemplar memory. Each query patch is compared with normal exemplars at the same location, and the nearest distance is projected to a dense anomaly map. A separate effect-size branch supplies the inherited object-level readout. Shapes and score illustrations are schematic.

Preserving normal configurations. The difference between exemplars and a prototype can be seen from two unit reference features u and v. If the query reproduces u, SES returns zero. In contrast, the normalized-centroid score equals

$$1 - u^\top \frac{u+v}{\|u+v\|_2} = 1 - \sqrt{\frac{1+u^\top v}{2}}, \qquad u+v \neq 0. \tag{4}$$

Thus a query matching an observed normal exemplar can still receive a positive centroid score when the references differ. SES retains the individual observations and avoids introducing an intermediate reference configuration. This geometric distinction motivates using exemplar support when the available templates capture different normal appearances of the same registered region.

Contrast with moment-based scoring. Let D_i contain query–reference distances and P_i contain distances between pairs of normal references. The effect-size comparator summarizes them through

$$s_i^{\text{eff}} = \frac{|\overline{D}_i - \overline{P}_i|}{\sqrt{v_{\text{pool},i}}}, \tag{5}$$

where $v_{\text{pool},i}$ is the pooled sample variance. With few references, this score depends on both the contrast between distance means and the estimated spread. SES instead measures proximity to normal support directly. We compare these choices with centroid and variance-shrinkage controls in Section 4.4; the exact comparator formulas are given in Appendix A.1.

3.3 Dense Localization and Object Readout

For scored patch centers z_i with valid support V, we project patch evidence onto query point x using normalized Gaussian weights:

$$a(x) = \frac{1}{\sqrt{2\pi}\sigma} \frac{\sum_{i\in V} \exp(-\|x-z_i\|_2^2/(2\sigma^2))s_i}{\sum_{i\in V} \exp(-\|x-z_i\|_2^2/(2\sigma^2))}, \qquad \sigma = 1. \tag{6}$$

The shared projection combines nearby patch responses while retaining spatial detail. All compared point estimators use the same projection parameters. In fixed-budget experiments, we freeze the slot mask before selecting template subsets so that the scoring rules are evaluated on a common set of locations.

The object-level output is obtained from the separate inherited effect-size branch,

$$o(X) = \max_{i\in V_{\text{eff}}} s_i^{\text{eff}}. \tag{7}$$

This branch is fixed across the point-scoring comparisons. Our method and primary evaluation focus on dense localization; complete object-level results and the optional exemplar-evidence readout are reported in Appendix A.1. The primary point branch uses radius-1 features on both benchmarks, while the inherited ShapeNet object branch retains radius 2.

3.4 Properties and implementation

Few-reference operation. SES is defined for any nonempty set of valid references. With one reference, nearest-exemplar, mean-distance, and normalized-centroid scoring coincide. Two references are therefore the smallest budget that distinguishes these estimators. For a fixed query, adding exemplars can only decrease its nearest-reference distance. This makes the score responsive to additional normal configurations, while motivating our evaluation of both reference budget and reference correspondence.

Feature-space stability. For normalized query features u and v and a fixed set of unit reference features, the score satisfies

$$|s^{\mathrm{SES}}(u) - s^{\mathrm{SES}}(v)| \leq \max_{r \in \mathcal{R}} |(u - v)^\top r| \leq \|u - v\|_2. \tag{8}$$

Consequently, small perturbations of the normalized query feature cannot be amplified by an estimated variance denominator. This property concerns the scoring function for fixed correspondences; registration and feature extraction remain essential to the full pipeline.

Computation. The memory stores $O(STd)$ feature values, and local matching requires $O(ST)$ cosine comparisons, or $O(STd)$ arithmetic operations. In practice, normalized reference features are cached, similarities are computed in batches, and the minimum is reduced over valid exemplars. SES adds no trainable parameters and avoids constructing all reference-pair distances at scoring time. Section 4.8 measures its practical cost within the complete system.

4 Experiments

4.1 Experimental setup

Datasets and metrics. We evaluate on Real3D-AD, a real scanned-object benchmark with 12 categories, and Anomaly-ShapeNet, a synthetic geometric-anomaly benchmark (Liu et al., 2023; Li et al., 2024). The main ShapeNet evaluation uses the 40 original categories; the extended 52-category inventory supports few-template analysis. IEC3D-AD (Guo et al., 2026) provides a 15-category transfer benchmark. We report category-macro object and point AUROC (O-AUROC and P-AUROC) and point average precision (P-AP). AP measures the precision of high-scoring anomaly points and complements AUROC for small defects.

Implementation. We use the COSSAD-based RIConv++ patch encoder trained with cross-category anomalous patches (Tarvo et al., 2026). Normal reference objects provide the inference memory. The primary point branch uses the frozen radius-1 feature view and Gaussian bandwidth $\sigma = 1$. Within controlled comparisons, we fix the encoder, registration, reference identities, and projection, and change only the component under study. This eliminates retraining as a source of variation between the scoring methods.

Evaluation protocol. The main Real3D evaluation contains 1,172 successfully registered scans out of 1,206 official test inputs. We use 1,295 scans for the 40-category ShapeNet evaluation. Retrieval and fixed-budget Real3D ablations use the 587-scan processed subset. At budget T, we evaluate all eligible size-T reference subsets, average their metrics within category, and then aggregate categories. Paired category bootstrap provides confidence intervals for estimator differences; full sample counts, support conventions, and intervals appear in Appendices A–E.

4.2 Comparison with state-of-the-art methods

Table 1 compares SES with representative state-of-the-art 3D anomaly detectors. The comparison covers registered matching, learned reconstruction, implicit geometry, and high-resolution feature

Table 1: **Comparison with state-of-the-art methods.** AUROC (%, higher is better). Baselines use published native protocols; SES uses the registered inputs described in the setup. ShapeNet P-AUROC covers all 40 base categories; † its inherited object head covers 39, excluding `shelf0`. Dashes denote unreported results. Sources: Appendix G.

	Real3D-AD		Anomaly-ShapeNet	
Method	O-AUROC	P-AUROC	O-AUROC	P-AUROC
Reg3D-AD (Liu et al., 2023)	70.4	70.5	57.2	66.8
PASDF (Zheng et al., 2025)	80.2	74.5	90.0	89.7
Simple3D (Cheng et al., 2026)	80.4	92.3	86.0	92.9
CASL (Zha et al., 2026)	82.3	88.2	88.7	89.9
Template3D-AD (Liu et al., 2025)	84.4	92.5	–	–
COSSAD (Tarvo et al., 2026)	91.6	95.3	–	–
SES-3D (Ours)	**93.32**	**98.54**	**95.50**†	**98.64**

Table 2: **Effect of the localization score.** P-AUROC and P-AP (%) with identical features, registration, and projection within each pair. SES improves both ranking metrics.

Cohort	Point score	P-AUROC	P-AP
Real3D-12 (1,172 scans)	Raw effect	96.36	46.62
Real3D-12 (1,172 scans)	SES-3D	98.54	63.32
ShapeNet-39 (matched)	Raw effect	97.76	54.20
ShapeNet-39 (matched)	SES-3D	98.66	71.31
ShapeNet-40 (all base)	SES-3D	98.64	71.37

aggregation. SES achieves 98.54% P-AUROC on Real3D and 98.64% on ShapeNet, demonstrating strong localization performance across real scans and synthetic geometric defects. Its performance is particularly notable given that the scoring module introduces no additional trainable parameters.

Localization gains under a matched pipeline. Table 2 isolates the contribution of the scoring rule. On Real3D, SES improves P-AUROC by 2.18 points and P-AP by 16.70 points over the effect-size baseline. The AP gain shows that the improvement extends to the high-scoring points most relevant to identifying defect regions. On the matched 39-category ShapeNet comparison, SES improves P-AUROC from 97.76 to 98.66 and P-AP from 54.20 to 71.31. The all-40-category result is 98.64/71.37. The 39-category comparison excludes the two-template `shelf0` category, for which the raw effect-size implementation is undefined.

Matched comparison with Template3D-AD. We additionally evaluate the available Template3D implementation on the same 587 processed Real3D scans and point labels. Its one-template configuration obtains 92.17 P-AUROC and 49.98 P-AP, while SES with the available reference bank obtains 98.46 and 63.76. This comparison complements the published results by aligning the test inputs. The template counts follow the respective methods; the fixed-budget experiments below examine scoring differences with identical reference subsets.

4.3 Spatial correspondence: WHERE

We first examine the reference candidate set while keeping the nearest-exemplar cosine score fixed. Table 3 compares global retrieval, shuffled slot identities, and correct spatial correspondence. Corresponding retrieval improves P-AUROC by 1.19 points over global retrieval on ShapeNet and 1.25 points on Real3D. Shuffling slot identities causes a larger decrease even though the local candidate counts are preserved. These results show that normal support must be geometrically relevant: a strong feature match from the wrong location can explain an anomalous region incorrectly.

Table 3: **Spatial retrieval ablation.** P-AUROC (%) with the same features and cosine score. Real3D uses the 587-scan processed subset.

Reference search	ShapeNet-40	Real3D-12
Global search	97.45	97.22
Shuffled locations	92.29	94.54
Spatial correspondence	98.64	98.46

Table 4: **Scoring ablation with three templates.** P-AUROC/P-AP (%). All estimators share the same spatial candidates, template subsets, features, and projection. Best values within each dataset are bold.

Dataset	Estimator	P-AUROC	P-AP	Classes
ShapeNetAD	Raw effect	96.85	47.99	50
ShapeNetAD	Shrunk effect	97.84	61.97	50
ShapeNetAD	Centroid	97.67	64.13	50
ShapeNetAD	SES-3D (Ours)	**98.03**	**64.58**	50
Real3D-AD	Raw effect	96.23	44.96	12
Real3D-AD	Shrunk effect	**98.44**	**63.65**	12
Real3D-AD	Centroid	98.19	60.02	12
Real3D-AD	SES-3D (Ours)	98.41	62.98	12

SES therefore combines physical correspondence with flexible exemplar matching: it first restricts the candidate set to the relevant region and then selects the nearest normal observation.

4.4 Local normality estimation: HOW

We next fix the spatial candidates and use exactly three reference templates. Table 4 compares raw effect-size, shrunk effect-size, centroid, and SES scores. SES improves over raw effect-size by 1.18/2.18 P-AUROC points and 16.59/18.02 P-AP points on ShapeNet/Real3D. The paired intervals for these improvements exclude zero. This establishes score estimation as an important source of localization performance after correspondence has already been fixed.

Exemplars versus a centroid. SES improves P-AUROC over centroid scoring by 0.35 points on ShapeNet and 0.22 points on Real3D, with paired 98.75% intervals of [0.13, 0.70] and [0.09, 0.44]. On Real3D, the P-AP improvement is 2.96 points. These results support retaining the observed local configurations instead of replacing them with an averaged prototype. The ShapeNet AP difference is smaller and its interval includes zero, indicating similar precision–recall behavior for the two estimators on that endpoint.

Exemplars versus variance regularization. Shrinkage recovers much of the performance lost by the raw effect-size statistic. SES achieves the highest ShapeNet point estimates, including a 2.61-point AP improvement over shrinkage with interval [0.59, 4.42]. On Real3D, shrinkage and SES are close, with shrinkage slightly higher in the point estimates. Together, these comparisons identify two effective responses to scarce local evidence: regularizing its moments or preserving its exemplars. SES provides a direct solution that avoids selecting a variance prior and remains defined at smaller reference budgets.

4.5 Learning from very few references

Table 5 reports SES at two and three templates. With only two references, SES reaches 97.79/63.47 P-AUROC/AP on the 52-category ShapeNet inventory and 98.27/60.98 on Real3D. Relative to centroid scoring on the same two-template subsets, AUROC improves by 0.29 points on ShapeNet and 0.24 points on Real3D; both nominal 95% intervals exclude zero. Mean-distance scoring gives larger gaps of 0.62 and 0.50 points. This is the smallest reference budget at which these scoring rules differ.

Table 5: **Few-template localization.** SES results and paired AUROC gains over centroid scoring. The 50-category ShapeNet subset enables comparison across budgets; the complete two-template result includes 52 categories.

Dataset	Templates	P-AUROC	P-AP	Δ centroid
ShapeNet-52	2	97.79	63.47	+0.29
ShapeNet-50	2	97.76	63.36	+0.29
ShapeNet-50	3	98.03	64.58	+0.35
Real3D-12	2	98.27	60.98	+0.24
Real3D-12	3	98.41	62.98	+0.22

Increasing from two to three templates improves both metrics on the common ShapeNet-50 and Real3D-12 cohorts. On ShapeNet-50, P-AP rises from 63.36 to 64.58; on Real3D, it rises from 60.98 to 62.98. Additional exemplars therefore improve the coverage of normal variation while the scoring rule remains unchanged. Full four-template results and category-level intervals are reported in Appendix E.1.

4.6 Generalization across representations

We evaluate transfer to IEC3D-AD (Guo et al., 2026) with UCA features (UCA-3DAL Contributors, 2026). At three templates, SES achieves 79.02 P-AUROC across 15 categories, compared with 66.81 for raw effect-size scoring (+12.22 points). On identical effect-defined support, SES reaches 76.62 versus 66.81 (+9.81 points), confirming that its advantage persists when support is shared. Tables 19 and 20 report the complete estimator comparison; Table 21 separately evaluates native UCA and alternative memory designs.

The Point-M2AE probe (Zhang et al., 2022) gives a complementary view. At three templates, replacing effect-size with SES improves Real3D P-AUROC by 6.83 points, while ShapeNet favors effect-size in AUROC and SES in AP. This interaction shows that the normality estimator is an important companion to the feature representation. Our primary registered patch features provide the strongest localization results, while the transfer studies expose how score behavior changes with the geometry of the embedding space.

4.7 Qualitative localization

Figure 3 compares SES with Template3D-AD, PASDF, and the effect-size baseline on matched scans. On the car, SES identifies a small defect that is weak in the effect-size map, reaching 99.24 P-AUROC. On the fish, it concentrates high responses around the bulge and reaches 99.54. On the chicken, it recovers the defect response relative to the effect-size baseline, although Template3D-AD has the higher scan-level AUROC. The starfish example shows a failure mode in which SES activates a broader region while effect-size produces a more selective response. These examples connect the benchmark improvements to local prediction behavior and illustrate the remaining influence of reference support.

4.8 Computational efficiency

SES reduces scoring overhead through direct query–exemplar comparisons. On an RTX 3090, patch scoring takes 5.58 ms on ShapeNet and 2.47 ms on Real3D, compared with 167.20 and 67.69 ms for effect-size scoring: approximately 30× and 27× faster. Registration and shared preprocessing

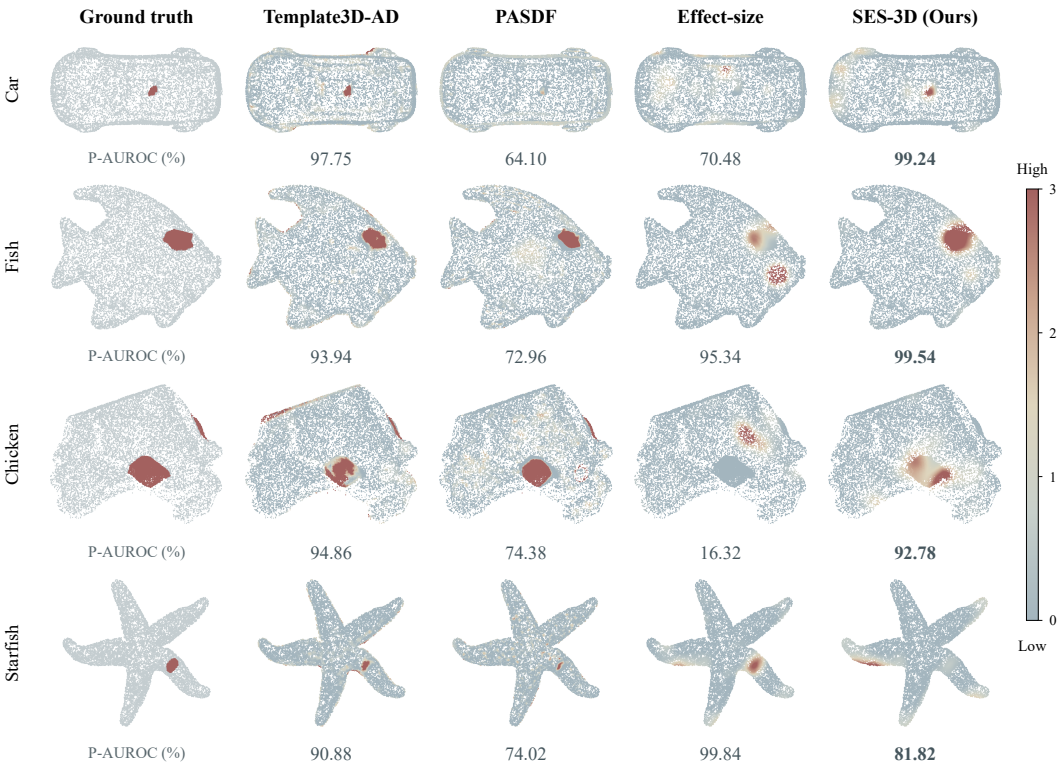

Figure 3: **Qualitative comparison on Real3D-AD.** Methods share the displayed points and view within each row. Numbers denote full-scan P-AUROC (%). Colors use a shared palette and normal-quantile display transform. Appendix B provides rendering details and enlarged views.

dominate total runtime, so end-to-end latency changes by less than 1%. These medians cover two categories and three scans per category; Appendix E provides the timing breakdown.

5 Discussion

Spatial correspondence and normality estimation should be optimized together. Correspondence constrains the normal evidence, while exemplar scoring preserves its alternatives. The large AP gains demonstrate improved concentration of anomaly responses, and competitive shrinkage results further establish the importance of scarce-reference estimation.

SES depends on registration quality and reference coverage. An anomalous patch resembling a normal exemplar can receive a low score, while unobserved normal variation can cause false positives. The starfish example illustrates these limitations. Combining exemplar evidence with reference reliability and improving correspondence under partial observations are promising directions for further work.

6 Conclusion

We presented SES-3D for few-template 3D anomaly localization. SES preserves registered normal exemplars and scores queries through their nearest valid match, avoiding local variance estimation and additional trainable parameters. Experiments on Real3D-AD and Anomaly-ShapeNet demonstrate strong localization accuracy and effective operation with two or three references. Separating spatial retrieval from score estimation reveals a practical route to improved localization beyond correspondence alone.

REPRODUCIBILITY STATEMENT

The appendix provides scoring formulas, evaluation subsets, per-category results, and implementation details. The accompanying source package includes the aggregated experimental evidence and verification procedures. Code is available at `https://anonymous.4open.science/r/SES_3D-D25E/`.

AI USE STATEMENT

Generative AI assisted literature lookup, English editing, figure preparation, and consistency checks during manuscript revision, including comparison of scoring formulas with implementation and verification of reported aggregates.

REFERENCES

Yunkang Cao, Xiaohao Xu, and Weiming Shen. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition*, 156:110761, 2024. doi: 10.1016/j.patcog.2024.110761.

Jiayi Cheng, Can Gao, Jie Zhou, Jiajun Wen, Tao Dai, and Jinbao Wang. MC3D-AD: A unified geometry-aware reconstruction model for multi-category 3D anomaly detection. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 837–845, 2025. doi: 10.24963/ijcai.2025/94.

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. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(5):3327–3334, 2026. doi: 10.1609/aaai.v40i5.37328. URL `https://doi.org/10.1609/aaai.v40i5.37328`.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A patch distribution modeling framework for anomaly detection and localization. *arXiv preprint arXiv:2011.08785*, 2020. URL `https://arxiv.org/abs/2011.08785`.

Zheng Fang, Xiaoyang Wang, Haocheng Li, Jiejie Liu, Qiugui Hu, and Jimin Xiao. FastRecon: Few-shot industrial anomaly detection via fast feature reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17481–17490, 2023. doi: 10.1109/ICCV51070.2023.01603.

Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. UniVAD: A training-free unified model for few-shot visual anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15194–15203, 2025.

Bingyang Guo, Hongjie Li, Ruiyun Yu, Hanzhe Liang, and Jinbao Wang. IEC3D-AD: A 3D dataset of industrial equipment components for unsupervised point cloud anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026. URL `https://arxiv.org/abs/2511.03267`. Accepted. Preprint: arXiv:2511.03267.

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*, 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 *Computer Vision – ECCV 2022*, pp. 303–319, 2022. doi: 10.1007/978-3-031-20053-3_18.

Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, and Yingna Wu. Towards scalable 3D anomaly detection and localization: A benchmark via 3D anomaly synthesis and a self-supervised learning network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22207–22216, 2024. URL `https://arxiv.org/abs/2311.14897`.

Yiting Li, Fayao Liu, Jingyi Liao, Sichao Tian, Chuan-Sheng Foo, and Xulei Yang. Find: Few-shot anomaly inspection with normal-only multi-modal data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 23290–23299, 2025. doi: 10.1109/ICCV51701.2025.02162.

Yiting Li, Xulei Yang, Jingyi Liao, Jing Zhang, and Fayao Liu. Gpflow: Gaussian prototype probability flow for unsupervised multi-modal anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 43103–43112, 2026.

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*, volume 36, pp. 30402–30415, 2023. URL https://arxiv.org/abs/2309.13226.

Yi Liu, Changsheng Zhang, and Yufei Yang. Template3D-AD: Point cloud template matching method based on center points for 3D anomaly detection. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 2025. URL https://www.ijcai.org/proceedings/2025/0182.pdf.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328, 2022. URL https://arxiv.org/abs/2106.08265.

Alexander Tarvo, Colin Acton, Yusen Wan, and Xu Chen. The good, the bad, and the template: Contrastive anomaly detection in 3D. In *Pattern Recognition: ICPR 2026*, pp. 17–32, 2026. doi: 10.1007/978-3-032-31920-3_2. URL https://doi.org/10.1007/978-3-032-31920-3_2.

UCA-3DAL Contributors. UCA-3DAL: A unified contrastive framework with test-time adaptation for robust 3D anomaly localization. Official code repository, snapshot e7616f7, 2026. URL https://github.com/M-3LAB/UCA-3DL.

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*, pp. 1353–1362, 2025. doi: 10.1109/CVPR52734.2025.00134.

Yaohua Zha, Yuerong Xue, Chunlin Fan, Yuansong Wang, Tao Dai, Ke Chen, and Shu-Tao Xia. CASL: Curvature-augmented self-supervised learning for 3D anomaly detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(15):12340–12348, 2026. doi: 10.1609/aaai.v40i15.38226. URL https://doi.org/10.1609/aaai.v40i15.38226.

Renrui Zhang, Ziyu Guo, Rongyao Fang, Bin Zhao, Dong Wang, Yu Qiao, Hongsheng Li, and Peng Gao. Point-M2AE: Multi-scale masked autoencoders for hierarchical point cloud pre-training. *arXiv preprint arXiv:2205.14401*, 2022. URL https://arxiv.org/abs/2205.14401.

Baozhu Zhao, Qiwei Xiong, Xiaohan Zhang, Jingfeng Guo, Qi Liu, Xiaofen Xing, and Xiangmin Xu. Point-Core: Efficient unsupervised point cloud anomaly detector using local-global features. *arXiv preprint arXiv:2403.01804*, 2024. URL https://arxiv.org/abs/2403.01804.

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*, 2025. URL https://arxiv.org/abs/2505.24431.

Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3D-AD: Reconstruction via diffusion for 3D anomaly detection. In *Computer Vision – ECCV 2024*, pp. 91–107, 2024. 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 *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 4680–4689, 2024. doi: 10.1145/3664647.3680919.

A EXACT SCORING AND SUPPORT CONVENTIONS

A.1 COMPARATOR ESTIMATORS AND OBJECT-READOUT COVERAGE

For controlled HOW comparisons, all estimators operate on the same valid same-location reference set $\mathcal{R}_i$. Let $D_i = \{d(q_i, r_{t,i})\}_{t\in\mathcal{T}_i}$ be query–reference distances and $P_i = \{d(r_{t,i}, r_{u,i}) : t < u\}$ be pairwise reference distances. With sample means $\overline{D}_i, \overline{P}_i$ and sample variances $v_{D,i}, v_{P,i}$, the inherited raw effect statistic is

$$s_i^{\text{eff}} = \frac{|\overline{D}_i - \overline{P}_i|}{\sqrt{v_{\text{pool},i}}}, \qquad v_{\text{pool},i} = \frac{(|D_i|-1)v_{D,i} + (|P_i|-1)v_{P,i}}{|D_i| + |P_i| - 2}. \tag{9}$$

The evaluated version requires at least three valid references and positive finite pooled variance. Shared endpoints make the pairwise distances dependent, so this statistic is used descriptively rather than as a classical independent-sample test.

The centroid comparator averages unit reference vectors and renormalizes,

$$m_i = \frac{1}{|\mathcal{T}_i|}\sum_{t\in\mathcal{T}_i} \widehat{r}_{t,i}, \qquad s_i^{\text{ctr}} = 1 - \widehat{q}_i^{\top}\frac{m_i}{\|m_i\|_2}, \tag{10}$$

when $\|m_i\|_2 \geq 10^{-12}$, matching the implementation. The shrunk effect estimator retains the contrast numerator of Equation 9 and regularizes variance with a normal-only prior v_0 of weight $\nu_0 = 2$:

$$\widetilde{v}_i = \frac{\nu_D v_{D,i} + \nu_P v_{P,i} + \nu_0 v_0}{\nu_D + \nu_P + \nu_0}, \qquad \nu_D = |D_i| - 1, \quad \nu_P = |P_i| - 1. \tag{11}$$

The prior (μ_0, v_0) is computed per category and selected reference subset by pooling pairwise cosine distances between valid co-located normal templates over all canonical slots. It uses neither queries nor test labels. The implemented shrunk score is

$$s_i^{\text{shr}} = \frac{|\overline{D}_i - \mu_{P,i}|}{\sqrt{\max(\widetilde{v}_i, 10^{-12}v_0)}}, \qquad \nu_D = \max(|D_i| - 1, 0), \quad \nu_P = \max(|P_i| - 1, 0).$$

Here $\mu_{P,i} = \overline{P}_i$ when a reference pair exists and $\mu_{P,i} = \mu_0$ otherwise. Missing sample variances use v_0 with their nonnegative degrees of freedom. This normal-only prior permits scoring with fewer than three valid references when a valid category prior exists. It must not be conflated with the stricter support mask used in the UCA stratified diagnostic below. The mean-distance comparator is $s_i^{\text{mean}} = |\mathcal{T}_i|^{-1}\sum_t d(q_i, r_{t,i})$.

For the primary fixed-budget experiments, the raw-effect legacy arm retains a zero fallback for undefined patch scores. The separately named common-defined effect/exemplar controls exclude undefined entries from both projection numerator and denominator. The UCA challenge also excludes undefined entries. These are experiment-specific support conventions, not interchangeable treatments of missing evidence.

For object-level coverage analysis, we additionally evaluate an exemplar-evidence-field (EEF) readout that remains defined with two templates. For normal template t at slot i, let $\rho_{t,i} = \min_{u\neq t} d(r_{t,i}, r_{u,i})$ and $d_i^{(-t)} = \min_{u\neq t} d(q_i, r_{u,i})$. The nonnegative slot evidence is $e_i = [\min_t(d_i^{(-t)} - \rho_{t,i})]_+$. EEF aggregates e_i on the fixed slot graph over $k \in \{1, 4, 16, 64\}$ neighboring slots, normalizes by $\sqrt{\min(k, n_{\text{valid}})}$, and takes the maximum over slots and scales. Its settings were fixed before test-label evaluation. EEF is an auxiliary object branch and does not alter the SES-3D point map.

Table 6: **Complete-category object/point coverage.** Both rows per dataset share the frozen SES-3D point scores; only the object readout differs. Anomaly-ShapeNet native object scores use radius 2 and are undefined for `shelf0`; optional EEF uses radius 1 and covers all 40 base categories. Both Real3D-AD readouts use radius 1 on the 1,172 accepted inputs. Cross-row object differences are not treated as point-estimator gains.

Dataset	Object readout	O-AUROC ↑	P-AUROC ↑	O coverage
ShapeNet base 40	Native object	95.50	98.64	39/40
ShapeNet base 40	EEF object	95.79	98.64	40/40
Real3D-AD 12	Native object	93.32	98.54	12/12
Real3D-AD 12	EEF object	94.17	98.54	12/12

On base Anomaly-ShapeNet, EEF provides object scores for `shelf0`, yielding 40/40 object coverage and 95.79 O-AUROC; on Real3D-AD, it gives 94.17 O-AUROC on the same accepted inputs. EEF extends object-level coverage while preserving the SES point map. Its radius-1 features differ from the native ShapeNet object branch, which uses radius 2.

A.2 TEMPLATE-BUDGET STUDY

The template-budget study begins with a fixed shared slot set: a finite query, finite features for the entire available normal bank, and a valid query center. This set is held fixed across every selected template subset. Query and reference features are normalized before cosine distances are formed. Subsets are enumerated from the recorded normal template identifiers; T counts reference objects, not patches or tokens.

For compatibility with the original implementation, the deployed effect branch maps an undefined patch statistic to zero before Equation 6. At $T = 2$, the reference-pair variance is undefined everywhere, so this fallback produces a constant map and is not treated as a meaningful estimate of the effect statistic. With three or four references, nonpositive or nonfinite pooled variance triggers the same fallback. The exemplar-distance score is unaffected by this validity rule.

The Gaussian bandwidth is $\sigma = 1$ in the registered COSSAD coordinate frame. The implementation retains the prefactor $1/(\sqrt{2\pi}\sigma)$ and evaluates normalized weighted sums over query points. This is a projection convention, not a calibrated anomaly probability. Point metrics are computed over concatenated points of the category's evaluated samples. Sample sizes vary between categories, but the final category macro-average gives each category equal weight.

A.3 FIXED-COHORT ESTIMATOR VERIFICATION

The fixed-cohort verification rebuilds score maps for raw effect, shrunk effect, mean distance, centroid distance, and nearest-exemplar distance from the same cached features and template subsets. Unlike the UCA probe below, all five estimators are defined on the reported common-four cohorts. The shrinkage rule follows Equation 11 with settings fixed before evaluation. The original estimator verification uses 20,000 paired category resamples and seed 20260922. The base-39 ShapeNet intervals were regenerated with seed $20260922 + T$; the Real3D-10 intervals retain seed 20260922. The 98.75% intervals correspond to a Bonferroni adjustment for four declared AUROC endpoints under the bootstrap approximation. They do not provide 98.75% simultaneous coverage over all comparisons, datasets, budgets, and metrics.

We additionally regenerated the estimator results in a single numerical environment. The float64 verification agrees with the same-runtime anchor across 604 keys (maximum metric difference 2.92×10^{-7}, tolerance 10^{-5}). The historical E75 CSV comparison still has 119 keys above that tolerance, with maximum AP difference 6.29×10^{-5} in fractional units. A float32 projection recheck has five larger anchor exceptions and maximum AP difference 2.32×10^{-3}; the recorded 95% and 98.75% interval signs remain unchanged. Thus numerical verification supports the reported contrasts without asserting exact historical or cross-precision equality.

A.4 UCA ESTIMATOR CHALLENGE

The UCA challenge reuses frozen tokens, four normal templates, ICP transforms, and token-to-slot mapping from the preceding representation-transfer run. Local interpolation uses three nearest tokens and a radius factor of two. Its projection bandwidth is 1.621 times the median slot nearest-neighbor spacing within a category. This scale rule was inherited from the transfer setup, not tuned among the six estimators.

At a query slot, each distance estimator uses the valid references in the selected subset. The centroid averages unit reference vectors and is renormalized; a mean norm below 10^{-12} makes it undefined. Original effect is left undefined when its conditions fail. Projection excludes each estimator's undefined entries from both numerator and denominator. The common-defined nearest-exemplar control uses exactly the original-effect mask. Shrinkage prior (μ_0, v_0) is computed per category and selected subset by pooling pairwise cosine distances between valid, co-located normal templates over all canonical slots. The final denominator is $\sqrt{\max(\widetilde{v}_i, 10^{-12} v_0)}$, with $\widetilde{v}_i$ from Equation 11.

A.5 POINT-M2AE PROBE

The Point-M2AE probe uses the official pretrained checkpoint and first-level local tokens. It inherits the registered primary samples and the recorded local interpolation rule ($k = 3$, radius factor two), with $\sigma = 1$ and seed 2027. Each nearest-exemplar/effect pair shares its template subset, slot support, and projection. This post-hoc probe does not isolate every preprocessing difference between a contrastive patch encoder and a hierarchical masked autoencoder.

A.6 TEMPLATE3D REPRODUCTION AND MATCHED ADAPTATION

We rerun the official Template3D-AD code and checkpoint on the same raw inventory. The rerun scores 1,206/1,206 official raw files. The 603 processed candidates and all 587 matched samples also score without failures. Aggregated scores match the official report within the stated numerical tolerance. The reported 85.78/92.17 Template3D and 93.27/98.46 SES-3D comparison is restricted to the same 587-sample matched subset and is distinct from the full-scored official inventory.

For the matched comparison, Template3D per-class min–max scaling and product fusion are frozen over the 603 candidate range and then reused unchanged on the 587 subset. SES-3D and raw effect scores remain raw. Category metrics are computed on identical samples and point order, then macro-averaged. The two point endpoints use 20,000 paired category resamples, seed 20260922, with 97.5% Bonferroni intervals.

A.7 QUALITATIVE MAP PROVENANCE

For Figure 3 and the additional maps, we use the shared display transform $\mathrm{clip}((s - \mathrm{med}_n(s))/(q_{0.95,n}(s) - \mathrm{med}_n(s)), 0, 3)$, where n denotes annotated normal points in the same scan. Colors run from muted blue-gray to dusty rose. This normalization is applied identically to each method for visualization and does not alter evaluation scores. All methods share the displayed points, PCA view, and depth order within each row. Figures 4 and 3 use original Real3D-AD point coordinates and masks aligned one-to-one with the measured point scores. The four illustrated scans are a car bulge, a fish bulge, a chicken multi-bulge case, and a starfish sink (the failure case); they are the same four rows reported in Table 7. Template3D point maps are fused with frozen per-category scaling and PASDF scores are recomputed using the original category-specific checkpoints. PASDF registered points retain text-file order, whereas cached arrays retain point-cloud-file order; we match them by source coordinates and verify rigid alignment. The maximum alignment residual is below 10^{-6}. A fixed seed selects up to 12,500 normal and 4,500 anomalous points per file; selection affects illustration only.

B ADDITIONAL SAME-SCAN LOCALIZATION EVIDENCE

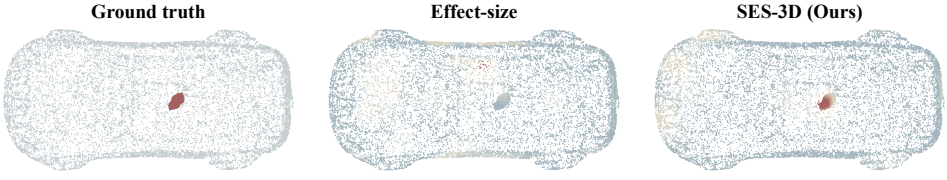

Figure 4: **A small defect on a Real3D-AD car.** Ground truth (red), the matched-setup effect-size point score, and the SES-3D (spatial exemplar) point score. Dusty rose indicates high anomaly scores. All panels use the same measured points and view, with the normal-quantile display transform used in Figure 3.

Table 7 gives full-scan P-AUROC for the four examples in Figures 4 and 3. The scans illustrate successes and a failure; they are not used to estimate a dataset mean. Local PASDF and Template3D models differ in training and reference budgets from SES-3D, so these numbers are descriptive same-scan comparisons.

Table 7: Full-scan point AUROC (%) for the four displayed Real3D-AD examples. All columns use the same point labels after coordinate alignment.

Sample	Template3D-AD	PASDF	Effect-size	SES-3D
Car	97.75	64.10	70.48	99.24
Fish	93.94	72.96	95.34	99.54
Chicken	94.86	74.38	16.32	92.78
Starfish	90.88	74.02	99.84	81.82

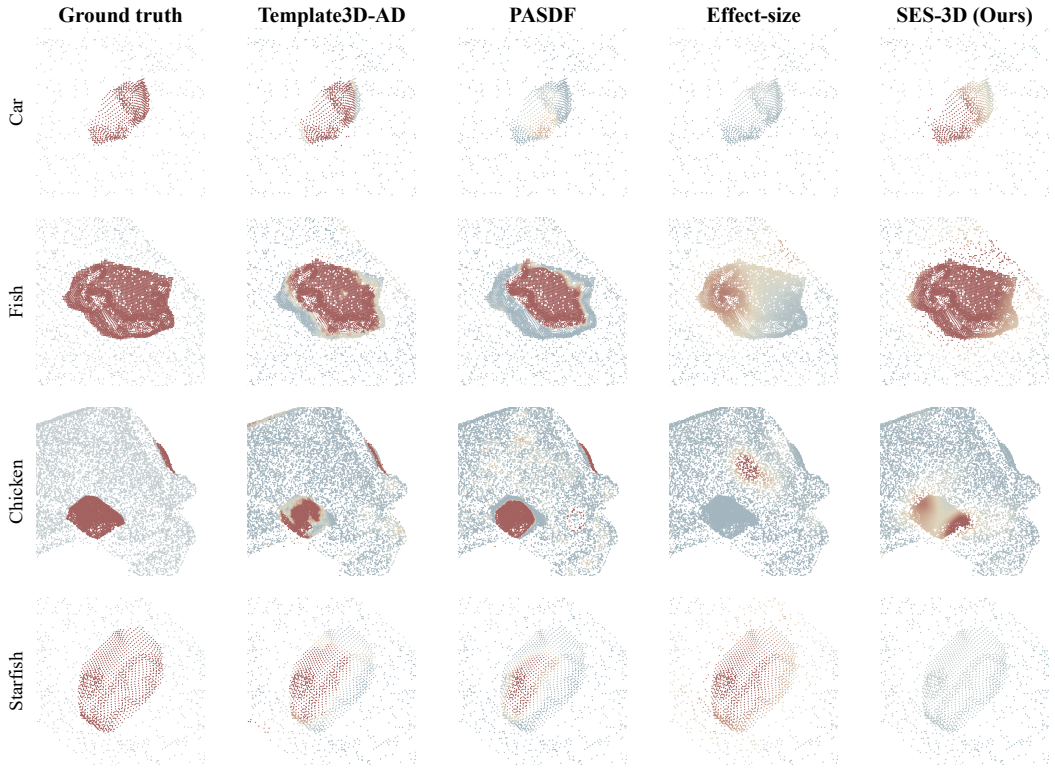

Figure 5: **Defect-region crops for Figure 3.** The rows and score columns match Figure 3; each crop is centered on its annotated defect. Cropping and label-dependent rescaling are for visualization only.

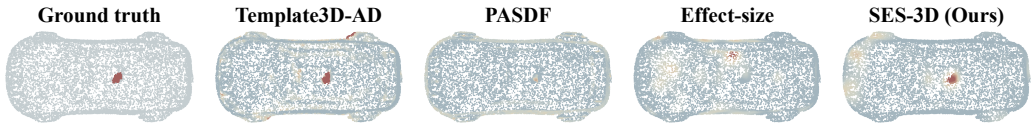

Figure 6: The Real3D-AD car from Figure 4 with local Template3D-AD and PASDF maps added.

C Category scope, coverage, and aggregation

Benchmark categories and fixed-budget cohorts. The local Anomaly-ShapeNet v2 asset pack contains the original 40 base categories and 12 added categories. The main benchmark reports the base 40. The *native* object score is undefined on `shelf0`, so its object AUROC uses 39 classes while point AUROC uses all 40; the optional EEF score is finite on all 40, including `shelf0` (91.15 O-AUROC). Because `shelf0` has only two valid templates, the fixed $T = 3, 4$ point-estimator ablation uses the other 39 base categories. The extended v2 cache has 52 categories; removing `cap1`

and `shelf0` gives its 50-category common-four cohort. In Real3D-AD, `chicken` and `toffees` have three templates, yielding a fixed 10-category cohort for $T = 3, 4$.

Table 8: Nearest-exemplar results on all eligible categories in the template-budget experiment. At $T = 2$ there is no defined original-effect comparator.

Dataset	T	Eligible categories	AUROC / AP (%)
ShapeNet	2	52	97.79 / 63.47
ShapeNet	3	50	98.03 / 64.58
ShapeNet	4	50	98.19 / 65.61
Real3D	2	12	98.27 / 60.98
Real3D	3	12	98.41 / 62.98
Real3D	4	10	99.04 / 66.35

Coverage is conditional on preprocessing. The full official Real3D-AD test inventory has 1,206 files. The method scores 1,172 and explicitly records 34 pairs rejected by the official registration procedure; no rejected pair is assigned a neutral score. The 34 refusals are listed by category and reason in the bundled evidence. This is distinct from the earlier processed inventory of 603 candidates, of which 587 were successfully scored and 16 failed preprocessing; the matched Template3D comparison uses the 587 common samples. ShapeNet has 1,702 cached test samples across its 52 categories.

Scope of the two object readouts. The inherited object branch gives 95.50 O-AUROC on 39 base Anomaly-ShapeNet categories and 93.32 on all 12 Real3D-AD categories; its output is identical across point-score arms. The separately evaluated EEF branch gives 95.79 on all 40 base Anomaly-ShapeNet categories, 94.17 on all 12 Real3D-AD categories, and 94.37 on the extended 52-category Anomaly-ShapeNet inventory. Across datasets, EEF provides broader object-score coverage but does not establish a consistent object-level improvement. On the 50 common extended Anomaly-ShapeNet categories, EEF gives 94.32 O-AUROC versus 94.87 for the deployed native head, but the native and EEF feature supports have different radii; these numbers are not a one-factor object comparison.

Statistical unit and dependence. Normal subsets share reference objects, and points within a scan are spatially dependent. We do not bootstrap points as if each supplied independent evidence. Subset metrics are averaged inside each category before the paired category bootstrap.

D COMPLETE PER-CATEGORY O/P RESULTS

Tables 9 and 10 list every evaluated category with separate native and EEF object columns. The native object score is fixed across point-score arms; EEF is an optional readout. Dashes mark undefined scores. Raw effect is undefined for the two-template categories `shelf0` and `cap1`; their legacy 50.00 fallbacks are excluded. Paired means use only categories with both point scores defined. All-category rows report SES and EEF coverage separately. Native-object and EEF feature radii follow Table 6.

Table 9: Real3D-AD per-category AUROC (%) on the full official inventory: all 12 categories, 1,172 scored inputs, and 34 official registration refusals. Native and optional EEF object scores are distinct readouts on the same radius-1 support. Template counts are not attached to the full-inventory metric export, so T is left blank rather than inferred from a different processed cohort.

Category	n	T	Native O	EEF O	Effect P	SES-3D P
airplane	91	–	90.46	85.22	97.00	97.12
candybar	100	–	99.64	99.28	99.32	99.65
car	100	–	92.32	96.80	97.45	98.96
chicken	105	–	84.22	77.98	74.47	92.90
diamond	100	–	99.48	100.00	98.34	99.19
duck	83	–	93.18	98.78	98.61	99.69
fish	100	–	92.48	99.28	98.27	99.62
gemstone	96	–	96.48	100.00	98.59	99.52
seahorse	100	–	99.52	100.00	98.93	98.86
shell	97	–	83.25	88.22	98.61	98.98
starfish	100	–	94.32	90.00	99.09	98.59
toffees	100	–	94.48	94.48	97.65	99.41
Real3D-AD mean (12 native O, 12 EEF O, 12 P)	1172	–	93.32	94.17	96.36	98.54

Table 10: Anomaly-ShapeNet per-category AUROC (%) for the original 40 and added 12 categories. Native O uses radius 2; EEF O and both P columns use radius 1 ($\sigma = 1$ for P). Paired rows exclude undefined effect scores; all-category rows leave incomplete metrics blank.

Category	n	T	Native O	EEF O	Effect P	SES-3D P
Original 40 categories						
ashtray0	27	4	97.78	100.00	98.62	99.40
bag0	28	4	93.33	91.28	94.49	97.87
bottle0	29	4	100.00	100.00	98.98	99.54
bottle1	32	4	100.00	100.00	95.35	92.82
bottle3	35	4	99.33	99.67	94.43	96.18
bowl0	33	4	100.00	99.63	97.17	98.92
bowl1	33	4	91.85	99.63	99.19	99.72
bowl2	33	4	100.00	99.26	99.09	99.40
bowl3	33	4	100.00	100.00	97.80	99.16
bowl4	33	4	100.00	100.00	99.65	99.77
bowl5	34	4	91.93	100.00	99.00	99.67
bucket0	31	4	100.00	90.00	96.02	98.45
bucket1	36	4	89.21	87.94	95.20	98.25
cap0	33	4	90.74	87.41	97.63	99.00
cap3	34	4	97.89	97.54	98.76	98.88
cap4	34	4	95.09	91.58	98.56	99.44
cap5	34	4	98.95	98.60	99.46	99.75
cup0	29	4	99.52	99.52	99.30	99.07
cup1	29	4	100.00	100.00	98.02	99.54
eraser0	29	4	100.00	72.38	91.26	93.33
headset0	30	4	100.00	100.00	97.40	97.55
headset1	29	4	100.00	100.00	97.80	99.18
helmet0	38	4	93.04	95.07	98.22	97.99
helmet1	29	4	100.00	100.00	98.83	99.48
helmet2	38	4	87.25	94.20	98.85	99.53
helmet3	37	4	92.73	99.09	98.36	98.91
jar0	29	4	99.05	98.57	98.03	99.44
microphone0	29	4	100.00	100.00	99.01	99.26
shelf0	33	2	–	91.15	–	97.88
tap0	37	4	95.76	96.97	98.46	99.22

Continued on next page

Table 10 continued

Category	n	T	Native O	EEF O	Effect P	SES-3D P
tap1	33	4	71.85	94.81	94.86	96.46
vase0	30	4	100.00	100.00	97.94	98.06
vase1	29	4	95.24	100.00	99.10	99.14
vase2	29	4	100.00	80.48	96.77	99.03
vase3	37	4	94.85	94.85	98.33	98.47
vase4	37	4	87.27	88.79	97.07	99.01
vase5	29	4	90.95	98.10	98.90	99.67
vase7	29	4	100.00	100.00	99.55	99.88
vase8	37	4	89.70	95.76	98.82	99.72
vase9	37	4	81.21	89.39	98.30	99.53
Base: paired 39 mean	1262	–	95.50	95.91	97.76	98.66
Base: all 40 mean	1295	–	–	95.79	–	98.64
Added 12 categories						
cabinet0	35	4	91.00	67.67	89.72	92.51
cap1	31	2	–	100.00	–	99.49
cap2	34	4	91.58	87.72	97.89	96.84
chair0	35	4	89.33	95.33	98.57	93.70
cup2	29	4	100.00	98.57	99.22	99.67
desk0	35	4	86.67	82.33	82.66	96.50
knife0	35	4	92.00	93.00	96.24	94.97
knife1	35	4	91.67	95.67	98.07	97.82
microphone1	31	4	91.25	91.25	97.19	98.94
screen0	33	4	93.70	99.26	98.95	99.33
vase10	38	4	93.91	76.81	89.55	93.56
vase6	36	4	98.10	87.94	96.59	98.05
Added: paired 11 mean	376	–	92.66	88.69	94.97	96.54
Added: all 12 mean	407	–	–	89.63	–	96.78
Extended: paired 50 mean	1638	–	94.87	94.32	97.14	98.19
Extended: all 52 mean	1702	–	–	94.37	–	98.21

E ADDITIONAL CONTROLLED RESULTS

E.1 TWO-TEMPLATE SCORING AND LOCAL SUPPORT

Table 11 adds a frozen $T = 2$ head-to-head comparison on identical reference subsets, slots, and Gaussian projection. It uses the same radius-1/SMALL feature view as the primary point experiments. The frozen ShapeNet native object readout separately uses radius 2; no object metric is an endpoint of this test. The raw effect-size statistic is undefined at two templates and is not assigned a zero-filled AUROC. Nearest-exemplar point AUROC exceeds centroid and mean distance on all listed cohorts under nominal 95% paired category intervals (10,000 resamples, seed 2026); the 50-category ShapeNet row is contained in the 52-category row, not an independent replication. For point AP against centroid, the 50-category ShapeNet interval includes zero. On 52 categories, replaying the recorded bootstrap gives a lower endpoint of only +0.0014 pp, which the archived four-decimal fractional export rounds to zero. This near-boundary AP result is not used to claim a robust advantage; the two-template claim concerns point AUROC. The inherited native object branch is fixed and is not an endpoint of this test.

Table 11: **Two-template HOW control.** Absolute nearest-exemplar point AUROC / AP (%); differences versus centroid and mean are point-AUROC percentage points with nominal 95% paired category-bootstrap intervals (10,000 resamples). All arms use the same two normal references, valid slots, and radius-1 features and $\sigma = 1$ projection. ShapeNet-50 is a subset of ShapeNet-52.

Cohort	Classes	SES-3D P-AUROC / P-AP	Δ vs centroid	Δ vs mean
ShapeNet-50	50	97.76 / 63.36	+0.29 [0.14, 0.48]	+0.63 [0.36, 0.97]
ShapeNet-52	52	97.79 / 63.47	+0.29 [0.14, 0.46]	+0.62 [0.36, 0.94]
Real3D-12	12	98.27 / 60.98	+0.24 [0.11, 0.40]	+0.50 [0.24, 0.83]

Table 12 separates low-support availability from accuracy where the compared scores are defined. This UCA / IEC3D-AD diagnostic pools the frozen $T = 3, 4$ subsets, assigns each query point to its nearest registered slot, and applies a strict support-≥ 3 validity rule to shrinkage and raw effect.

Under that rule, nearest-exemplar scoring covers the support-1 and support-2 strata, whereas the two moment-based arms are excluded. On the common support-≥ 3 stratum, its recorded bootstrap contrasts are $+7.36$ [5.34, 9.61] points over shrinkage and $+9.78$ [7.38, 12.27] over raw effect (15-category bootstrap, nominal 95%). The source reports the Monte Carlo bootstrap mean for these differences, which can differ slightly from subtraction of the empirical category means. These are class-level contrasts, not independent-point intervals. The support strata differ in their underlying point populations, so increasing AUROC across strata is not a causal effect of adding references. A nearest-slot assignment labels points for this diagnostic; Gaussian-projected scores may still receive contributions from several slots. The low-support exclusion is specific to this diagnostic: a shrinkage rule with a valid normal-only prior can be defined at two templates, as in the separate E68 slot census. That census finds nearest-exemplar scores defined in all 52 Anomaly-ShapeNet and 12 Real3D-AD categories; raw effect is undefined in the two-template `cap1` and `shelf0` categories.

Table 12: **UCA support-stratified diagnostic.** Category-macro point AUROC (%) on IEC3D-AD-15, with $T = 3, 4$ pooled. Coverage is the percentage of points scored by exemplar / shrinkage under this experiment's strict support-≥ 3 rule. A dash means excluded by that rule, not an imputed score.

Valid slot support	Exemplar	Shrinkage	Raw effect	Coverage (%)
1	71.81	–	–	100 / 0
2	76.01	–	–	100 / 0
≥ 3	79.76	72.40	69.97	100 / 100

Table 13: **Estimator differences across representations** ($T = 3$). Entries are SES-3D minus comparator in point AUROC (percentage points), computed before rounding. Primary rows match Table 4; Point-M2AE and UCA use their respective feature spaces.

Setting	vs. effect	vs. centroid
Primary, ShapeNet-50	+1.18	+0.35
Primary, Real3D-12	+2.18	+0.22
Point-M2AE, ShapeNet-50	-1.18	–
Point-M2AE, Real3D-12	+6.83	–
UCA, IEC3D-AD-15	+12.22	+0.57

E.2 Primary paired intervals and inherited baseline

Table 14 completes the controlled Table 2 on the full official Real3D-AD inventory with the point-AP column. Table 15 reports fixed-cohort nearest-exemplar-minus-centroid contrasts with 98.75% paired category intervals; on the fixed base-39 cohort, the $T = 3$ AUROC difference is +0.22 [+0.10, +0.39] pp and excludes zero, whereas its AP interval includes zero. The all-eligible 50-category AUROC contrast also excludes zero, so the unresolved endpoint is ShapeNet AP, not AUROC. Table 16 gives that fixed-cohort AUROC/AP ablation; its $T = 4$ nearest-exemplar column uses the full available bank. The all-eligible $T = 3$ estimates and paired intervals appear in Tables 4 and 18. Extended 50-category Anomaly-ShapeNet and 10-category Real3D-AD comparisons against the raw effect-size score are in Table 17.

Table 14: Companion to Table 2 on the full official Real3D-AD inventory (1,172 of 1,206 inputs scored; 34 official registration refusals). Category-macro AUROC and AP (%); only the point score differs between the rows.

Point score	O-AUROC ↑	P-AUROC ↑	P-AP ↑
Raw effect	93.32	96.36	46.62
SES-3D (Ours)	93.32	98.54	63.32

Table 15: Nearest-exemplar minus centroid on the fixed base-39 Anomaly-ShapeNet and Real3D-10 cohorts (pp). Brackets are 98.75% paired category-bootstrap intervals over 20,000 resamples.

Cohort	T	Δ point AUROC	Δ point AP
ShapeNet-39	3	+0.22 [+0.10, +0.39]	+0.18 [-0.49, +1.04]
ShapeNet-39	4	+0.25 [+0.10, +0.43]	+0.40 [-0.69, +1.77]
Real3D-10	3	+0.22 [+0.07, +0.48]	+2.79 [+0.95, +5.72]
Real3D-10	4	+0.24 [+0.06, +0.58]	+3.16 [+0.68, +7.54]

Table 16: Fixed-cohort estimator ablation. Each cell is point AUROC / point AP (%); $T = 3$ is the constrained-reference comparison, while $T = 4$ uses the full native bank; its nearest-exemplar result is the full-bank consistency control.

	Anomaly-ShapeNet (39)		Real3D-AD (10)	
Estimator	$T = 3$	$T = 4$	$T = 3$	$T = 4$
Original effect	97.41 / 51.61	97.76 / 54.20	98.33 / 47.29	98.49 / 49.54
Shrunk effect	98.28 / 66.70	98.41 / 66.39	99.11 / 66.81	99.17 / 67.03
Mean distance	98.06 / 69.58	98.12 / 70.08	98.53 / 60.73	98.55 / 61.20
Centroid	98.34 / 70.35	98.41 / 70.91	98.75 / 62.62	98.79 / 63.19
SES-3D (Ours)	98.56 / 70.53	98.66 / 71.31	98.97 / 65.41	99.04 / 66.34

Table 17: Raw effect versus nearest exemplar on fixed common-four cohorts. Point AUROC / AP (%); the final column is nearest-exemplar minus effect AUROC with nominal 95% paired interval (pp).

Cohort	T	Effect	SES-3D	Δ AUROC [95% CI]
ShapeNet-50	3	96.85 / 47.99	98.03 / 64.58	+1.18 [0.61, 1.79]
ShapeNet-50	4	97.14 / 50.41	98.19 / 65.61	+1.05 [0.48, 1.73]
Real3D-10	3	98.33 / 47.29	98.97 / 65.41	+0.64 [0.25, 1.05]
Real3D-10	4	98.49 / 49.54	99.04 / 66.35	+0.55 [0.18, 0.92]

E.3 Independent numerical verification at three templates

An independent float64 verification recomputes the frozen estimator comparisons from the same features and template subsets. The regenerated nearest-exemplar anchor agrees with the primary results within the stated 10^{-5} tolerance. The paired intervals below resample categories 20,000 times with seed 20260923. At $T = 3$, nearest-exemplar minus raw effect is positive on both datasets for both point metrics. Against shrinkage, Anomaly-ShapeNet P-AP favors nearest exemplar, whereas the Real3D-AD AUROC and AP intervals include zero; these differences are therefore treated as unresolved. At $T = 4$, the subset equals the full bank for four-template categories. The recorded comparison against historical native slot-nearest-neighbor metrics has maximum AUROC difference 5.31×10^{-6} and AP difference 4.64×10^{-4} (fractional units); the latter exceeds the separate 10^{-5} same-runtime anchor tolerance. This is approximate numerical consistency, not exact equality at that tolerance. We therefore use $T = 4$ only as an implementation consistency check, not as evidence for the few-template claim.

Table 18: $T = 3$ paired estimator differences (percentage points). Intervals are 98.75% paired category-bootstrap intervals over 20,000 resamples (seed 20260923). A confidence interval containing zero is reported as unresolved, not as equivalence.

Dataset	Contrast	Metric	Δ [98.75% CI]	Classes
ShapeNetAD	SES-3D - raw effect	P-AUROC	+1.18 [+0.47, +2.01]	50
ShapeNetAD	SES-3D - raw effect	P-AP	+16.59 [+11.88, +20.88]	50
ShapeNetAD	SES-3D - shrinkage	P-AUROC	+0.19 [-0.04, +0.46]	50
ShapeNetAD	SES-3D - shrinkage	P-AP	+2.61 [+0.59, +4.42]	50
ShapeNetAD	SES-3D - centroid	P-AUROC	+0.35 [+0.13, +0.70]	50
ShapeNetAD	SES-3D - centroid	P-AP	+0.45 [-0.29, +1.41]	50
Real3D-AD	SES-3D - raw effect	P-AUROC	+2.18 [+0.33, +6.56]	12
Real3D-AD	SES-3D - raw effect	P-AP	+18.02 [+11.69, +24.89]	12
Real3D-AD	SES-3D - shrinkage	P-AUROC	-0.03 [-0.23, +0.26]	12
Real3D-AD	SES-3D - shrinkage	P-AP	-0.66 [-3.14, +1.52]	12
Real3D-AD	SES-3D - centroid	P-AUROC	+0.22 [+0.09, +0.44]	12
Real3D-AD	SES-3D - centroid	P-AP	+2.96 [+1.23, +5.49]	12

E.4 Complete UCA estimator challenge

Table 19 reports all UCA estimators and their defined-slot fractions. The original effect statistic covers only 63.01% and 78.54% of candidate slots at three and four templates.

Table 19: Estimator challenge on UCA / IEC3D-AD-15. Point AUROC / AP (%); defined support is shown for $T = 3/T = 4$.

Estimator	$T = 3$	$T = 4$	Defined support (%)
Original effect	66.81 / 2.57	69.82 / 3.48	63.01 / 78.54
Shrunk effect	73.16 / 4.53	74.07 / 4.66	100.00 / 100.00
Mean distance	78.05 / 6.60	78.75 / 7.26	100.00 / 100.00
Median distance	78.31 / 6.22	78.99 / 6.73	100.00 / 100.00
Centroid distance	78.45 / 6.65	79.18 / 7.33	100.00 / 100.00
Minimum distance	79.02 / 6.30	79.98 / 7.09	100.00 / 100.00

E.5 Common-defined UCA comparison

Table 20: Identical effect-defined support, IEC3D-AD-15. AUROC / AP are percentages; the AUROC difference is in pp.

T	Original effect	Common-defined SES-3D	Δ AUROC
3	66.81 / 2.57	76.62 / 6.26	+9.81
4	69.82 / 3.48	79.49 / 7.51	+9.67

E.6 Spatial transfer to UCA

Table 21: Native UCA and alternative memory designs on IEC3D-AD-15 (%). Each row is a distinct system arm. Native UCA and global memory are not the SES/raw-effect pair in Tables 19–20.

Method	Object AUROC	Point AUROC	Point AP
Native UCA	75.11	87.04	29.41
Global memory	75.63	79.66	11.06
Same-slot minimum	66.20	79.98	7.09

E.7 PER-CATEGORY UCA ESTIMATOR RESULTS

Table 22: Point AUROC (%) by IEC3D-AD category. Each three-template entry averages four subsets.

	$T = 3$			$T = 4$		
Category	Effect	Centroid	SES-3D	Effect	Centroid	SES-3D
ButterflyBolt	60.52	69.53	69.31	63.02	69.84	70.21
ButterflyNut	61.20	70.79	70.44	68.68	71.49	71.33
DoubleEndStud	69.07	75.06	75.89	69.82	75.79	76.43
FenderRing	60.36	68.26	72.83	65.31	68.34	74.76
HexagonalBolt	57.99	75.45	77.02	59.28	76.48	77.74
HexagonalNut	80.72	90.87	90.25	84.97	91.08	90.47
HoleRetainingRing	70.01	81.13	82.15	75.10	81.64	83.16
JointingStud	72.96	91.99	93.38	73.59	92.17	93.60
KNut	62.21	86.82	85.01	64.49	88.53	86.26
PullingNail	53.79	64.89	65.76	55.95	66.36	66.70
RoundNut	69.22	80.53	80.29	70.87	81.50	81.16
SlipknotBolt	68.34	81.47	81.64	70.09	82.00	82.62
SquareWeldedNut	75.50	89.50	88.62	80.59	90.48	89.46
TScrew	61.00	69.96	71.22	64.77	71.33	73.77
Washer	79.20	80.46	81.51	80.79	80.63	82.10
Macro mean	66.81	78.45	79.02	69.82	79.18	79.98

E.8 POINT-M2AE REPRESENTATION PROBE

Table 23: Point-M2AE point AUROC / AP (%). The difference is nearest-exemplar minus effect. This probe changes the representation and is not a score-only primary comparison.

Cohort	T	Original effect	SES-3D	Difference
ShapeNet-50	3	89.62 / 42.72	88.43 / 44.45	-1.18 / +1.72
ShapeNet-50	4	90.80 / 47.64	89.04 / 45.86	-1.76 / -1.78
Real3D-12	3	75.46 / 13.46	82.28 / 14.44	+6.83 / +0.98
Real3D-10	4	76.69 / 19.12	83.25 / 14.33	+6.56 / -4.79

E.9 REPRESENTATIVE EFFICIENCY TIMING

Table 24: **Representative efficiency on an RTX 3090.** Per-dataset medians over two classes and three samples per class, not a full-category throughput claim. Score columns are ms per sample (15 timed repeats); end-to-end columns are seconds per sample and include separately measured ICP (three repeats), shared encoding, scoring, and the same GPU projection. Point-feature extraction used a PyTorch fallback for an unavailable custom point operator. The optional EEF object readout is not timed here.

	Patch scoring (ms) ↓		End to end (s) ↓	
Dataset	Native	SES	Native	SES
ShapeNetAD	167.20	5.58	21.49	21.33
Real3D-AD	67.69	2.47	43.14	43.08

F IMPLEMENTATION AND NUMERICAL REPRODUCIBILITY

Feature support versus object support. The frozen point cache resolver fixes the SMALL arm for all 52 ShapeNet categories and the corresponding radius-1 arm for Real3D. The ten original Real3D cache categories use SMALL; chicken and toffees use a separately extracted radius-1 cache

named BASE. The radius is $r = 1 = 0.5 \times 2$ in the inherited coordinate convention. These naming differences do not imply a different point support radius. E75 reads these E68 feature caches; the E85 and E91 estimator checks reuse them. The full-inventory Real3D coverage export also records support radius 1 and $\sigma = 1$ for every category.

The native object head is read from a separate deployed score cache: ShapeNet BASE at radius 2, Real3D at radius 1. Therefore the reported ShapeNet point result of 98.64 is a radius-1 localization result paired with an independently frozen radius-2 object branch. The optional EEF object branch uses radius 1. Object-branch radius differences must not be assigned to the SES point score.

Connection to published COSSAD. The official COSSAD release reports 91.6/95.3 object/point AUROC on Real3D, whereas the matched local raw-effect configuration gives 93.32/96.36 on the accepted full inventory. Support/projection settings and accepted-sample scope differ. The contributions of those differences to the gap have not been isolated. Only the raw-effect-versus-SES comparison within one frozen configuration is used to attribute a gain to point scoring.

Numerical reproducibility. The accompanying verification recomputes category macro means, coverage totals, and shared object results from the saved per-category outputs. It also reproduces the paired bootstrap with the recorded category order, seed, and resampling count, together with the two-template and stratified aggregates. The complete prediction pipeline additionally uses the category-to-checkpoint map, registration transforms, template identifiers, cached features, and runtime environment.

Interpretation of confidence intervals. The 98.75% E91 intervals use marginal percentile cutoffs 0.00625 and 0.99375. Their use throughout the contrast table does not create a simultaneous guarantee for all twelve displayed $T = 3$ contrasts. No test is interpreted as equivalence merely because its interval contains zero. The few category units, fixed checkpoints, correlated template subsets, and unmeasured acquisition shifts limit broader statistical generalization.

One-reference and candidate-set properties. With one nonzero reference, minimum, mean, and normalized-centroid cosine scores are identical. For a fixed query and nonempty sets $A \subseteq B$, $\min_{r \in A} d(q, r) \geq \min_{r \in B} d(q, r)$: additional exemplars expand the available normal support. The resulting localization accuracy is measured empirically after projection.

G SOURCES OF REPORTED BENCHMARK ROWS

Table 25 collects the published object- and point-level results used in Table 1. Table 26 identifies the source table or official release for each entry. Baselines retain their native training and reference protocols; Table 2 supplies the controlled comparison with shared features and accepted inputs.

Table 25: **Published benchmark results.** Object and point AUROC (%) from the cited papers or official releases, using each method's native protocol. Both metrics appear in the main comparison.

	Real3D-AD		Anomaly-ShapeNet	
Method	O-AUROC ↑	P-AUROC ↑	O-AUROC ↑	P-AUROC ↑
Reg3D-AD (Liu et al., 2023)	70.4	70.5	57.2	66.8
PASDF (Zheng et al., 2025)	80.2	74.5	90.0	89.7
Simple3D (Cheng et al., 2026)	80.4	92.3	86.0	92.9
CASL (Zha et al., 2026)	82.3	88.2	88.7	89.9
Template3D-AD (Liu et al., 2025)	84.4	92.5	–	–
COSSAD (Tarvo et al., 2026)	91.6	95.3	–	–

Table 26: Source locations for the external rows of Table 25. O/P denote object/point AUROC.

Method	Real3D O/P source	ShapeNet O/P source
Reg3D-AD	Simple3D, Table 3	CASL, Table 2
PASDF	PASDF, Table 1	PASDF, Tables 2–3
Simple3D	Simple3D, Table 3	Simple3D, Table 3
CASL	CASL, Table 1	CASL, Table 2
Template3D-AD	Template3D-AD, Table 4	—
COSSAD	Official repository, Real3D mean	—