RIFT: REGISTRATION-INDUCED FEATURE TUNING FOR FEW-SHOT 3D ANOMALY LOCALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot 3D anomaly localization aims to identify subtle geometric defects from only a few normal reference scans. Existing registration-based methods primarily use alignment to reduce spatial ambiguity during feature matching or representation learning, leaving the structure exposed by registered normal references underexplored after the representation has been fixed. We propose RIFT, a lightweight Registration-Induced Feature Tuning framework that turns these aligned normal references from passive matching memories into supervision for post-hoc target-category specialization. RIFT exploits two complementary cues revealed by registration: canonical coordinates provide spatial identity for a coordinate-conditioned residual update, while same-slot patches across aligned templates provide cross-template structural correspondence supervision. The residual formulation preserves the pretrained representation while adapting its local metric without retraining the encoder or requiring anomalous samples and defect annotations. On Real3D-AD, RIFT improves P-AUROC from 94.21% to 96.43% and P-AUPR from 30.29% to 45.95%. On Anomaly-ShapeNet V2, it improves P-AUROC from 94.65% to 95.10% and P-AUPR from 44.40% to 46.19%. Controlled ablations further show that learned spatial conditioning and correct same-slot correspondence are both important, with the latter providing a clear advantage over generic or shuffled correspondence supervision.

1 INTRODUCTION

Three-dimensional anomaly detection is increasingly important for industrial inspection, where subtle geometric defects such as cracks, dents, deformations, and missing structures may compromise product quality and reliability. Compared with image-based inspection, point-cloud anomaly detection must additionally handle irregular sampling, pose variation, and complex local geometry. These challenges become more pronounced in the few-shot setting, where only a small number of defect-free reference scans are available and anomalous regions typically occupy only a small fraction of the observed surface (Bergmann & Sattlegger, 2023; Liu et al., 2023; Li et al., 2024).

Existing 3D anomaly detection methods mainly model normal geometry through feature matching or reconstruction. Feature-memory approaches compare local descriptors of test samples against normal feature banks (Roth et al., 2022; Bergmann & Sattlegger, 2023; Zhao et al., 2024), whereas reconstruction-based methods recover normal structures and detect anomalies from reconstruction discrepancies (Zhou et al., 2024; Ye et al., 2025; Cheng et al., 2025). Registration further reduces spatial ambiguity by aligning test shapes with normal references before local comparison (Liu et al., 2023; Tarvo et al., 2026). More recently, registration has also been incorporated into representation learning. COSSAD (Tarvo et al., 2026) learns contrastive features from registered patch pairs, while Reg2Inv (Yu et al., 2025) couples registration with feature learning to obtain rotation-invariant and locally discriminative representations.

Despite these advances, registration is predominantly used either during feature pretraining or as a geometric preprocessing step for anomaly matching. Its role after the feature encoder has been fixed remains less explored. In a few-shot target category, the available normal references are already aligned to a common canonical frame, and therefore expose additional structure that is not fully exploited by conventional memory-based inference. In particular, registration provides both canonical spatial identity, indicating where a local patch belongs, and cross-template correspondence, indicating

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.

which normal observations describe the same structural role. These cues offer a natural source of target-category supervision without requiring anomalous samples or retraining the backbone. This perspective shifts registration from being only a preprocessing or representation-learning mechanism to a source of category-specific supervision that remains available after the representation has been frozen.

Motivated by this observation, we propose RIFT (Registration-Induced Feature Tuning), a lightweight post-hoc adaptation framework for few-shot 3D anomaly localization. RIFT keeps the pretrained registration-aware encoder fixed and specializes only the local feature metric from the few aligned normal references of the target category. Registered coordinates condition a residual feature update, while same-slot patches across normal templates define which observations should be brought into agreement. The residual formulation preserves the pretrained representation and limits adaptation to a small category-specific transformation. At inference, the adapted representation is used for point-level localization, while the original object-level detector is retained unchanged.

We evaluate RIFT on Real3D-AD and Anomaly-ShapeNet V2. On Real3D-AD, RIFT improves P-AUROC from 94.21% to 96.43% and P-AUPR from 30.29% to 45.95%. On Anomaly-ShapeNet V2, P-AUROC improves from 94.65% to 95.10% and P-AUPR from 44.40% to 46.19%. Controlled ablations show that learned coordinate-conditioned adaptation is more effective than a fixed spatial cue, and that correct same-slot correspondence provides a clear advantage over shuffled or generic feature-neighbor supervision. These results indicate that the benefit of RIFT arises from registration-induced structural supervision rather than from generic adapter fitting alone.

Our main contributions are:

- We formulate post-hoc target-category adaptation for a frozen registration-aware 3D anomaly detector, reusing registered few-shot normal references as supervision rather than only as passive matching memories.
- We instantiate this idea by combining two structures exposed by registration: canonical spatial identity conditions a lightweight residual feature update, while cross-template same-slot correspondence supervises target-category metric specialization without retraining the encoder or using target-category anomaly supervision.
- We validate the proposed supervision mechanism on Real3D-AD and Anomaly-ShapeNet V2 through paired evaluations and controlled ablations, showing that correct same-slot correspondence contributes beyond fixed spatial cues, shuffled correspondence, and generic normal-feature adaptation.

2 Related Work

2.1 3D Anomaly Detection

Existing 3D anomaly detection methods mainly model normality through feature matching or geometry reconstruction. Feature-based approaches represent normal samples with local or global descriptors and identify anomalies according to their distance from a normal feature distribution. PatchCore (Roth et al., 2022) established the effectiveness of nearest-neighbor feature memories for industrial anomaly detection, and subsequent studies extended this paradigm to 3D data. The MVTec 3D-AD benchmark (Bergmann et al., 2022) promoted the study of geometric anomaly detection, with representative methods exploring learned local descriptors (Bergmann & Sattlegger, 2023), rotation-invariant geometric features (Horwitz & Hoshen, 2023), and multimodal RGB–3D representations (Wang et al., 2023; Chu et al., 2023).

More recent methods explicitly exploit spatial structure in point clouds. Real3D-AD (Liu et al., 2023) introduces high-resolution real scans and a registration-based detector for spatially aligned anomaly comparison. PointCore (Zhao et al., 2024) combines local geometric information with global point features, while Template3D-AD (Liu et al., 2025) performs template matching based on center-point geometry. Simple3D (Cheng et al., 2026) targets subtle defects in high-resolution point clouds using scalable local geometric representations, and COSSAD (Tarvo et al., 2026) combines contrastive geometric features with spatially-aware comparison against aligned normal templates. These methods highlight the importance of spatially structured normal references for 3D anomaly detection.

A complementary line of work reconstructs normal geometry and identifies anomalies from reconstruction discrepancies. Anomaly-ShapeNet and IMRNet (Li et al., 2024) study scalable anomaly synthesis and self-supervised reconstruction, R3D-AD (Zhou et al., 2024) employs a diffusion model to recover normal geometry, and PO3AD (Ye et al., 2025) predicts point offsets toward normal structures. MC3D-AD (Cheng et al., 2025) extends geometry-aware reconstruction to the multi-category setting by jointly modeling local and global geometric information, while geometry-aligned reconstruction further incorporates spatial consistency into the recovery process (Wu et al., 2026). In contrast, RIFT does not reconstruct normal geometry or synthesize anomalies; instead, it adapts an existing local feature metric using structural correspondence already exposed by registration.

2.2 Feature Adaptation and Registration-Aware Learning

Generic pretrained representations are not necessarily optimal for anomaly localization, motivating target-specific feature adaptation. LSFA (Tu et al., 2024) introduces lightweight self-supervised adapters for industrial anomaly detection through compactness and cross-modal consistency. COSSAD (Tarvo et al., 2026) learns an RIConv++-based encoder from registered patch tuples with contrastive supervision and then performs spatially restricted matching at inference, demonstrating that registration can provide useful spatial supervision for anomaly representation learning.

Registration has also been integrated more directly into feature learning. Reg2Inv (Yu et al., 2025) jointly exploits registration and representation learning to obtain rotation-invariant and locally discriminative features. RIFT differs from these methods in both stage and supervision: it does not optimize registration or retrain the feature encoder, but instead performs lightweight post-hoc adaptation of an already fixed representation using only the few registered normal templates of the target category. Registered spatial identity conditions the adaptation, while cross-template same-slot correspondence provides the supervisory signal. Accordingly, the "normal-only" setting in this work refers specifically to the RIFT target-category adaptation stage, rather than to pretraining of the underlying COSSAD encoder.

3 Method

RIFT performs post-hoc target-category adaptation on top of a frozen registration-based anomaly detector. Given a few normal references of the target category, the underlying registration procedure aligns them to a common canonical frame and establishes spatially corresponding local regions. RIFT reuses this structure to specialize the local feature metric while leaving the point encoder and registration procedure unchanged.

As illustrated in Fig. 1, RIFT exploits two complementary signals exposed by registration. Registered patch centers provide canonical spatial identity for conditioning a lightweight residual adapter, whereas same-slot observations across normal templates provide correspondence supervision for feature adaptation. The resulting representation remains anchored to the pretrained descriptor and is optimized using normal references only. At inference, the adapted metric is used for point-level localization, while the original object-level scoring branch remains unchanged.

3.1 Canonical Correspondence from Registered Normal Templates

Consider a target category with T registered normal templates $\{X^t\}_{t=1}^{T}$. Following the underlying detector, a fixed set of N patch locations on the reference shape defines canonical slots $i \in \{1, \dots, N\}$. For template t, let P_i^t denote the patch assigned to slot i and $c_i^t \in \mathbb{R}^3$ its registered center. A frozen point encoder f_θ produces the normalized local descriptor

$$e_i^t = \frac{f_\theta(P_i^t)}{\|f_\theta(P_i^t)\|_2}, \qquad \mathcal{M}_i^{\mathrm{F}} = \{e_i^t\}_{t=1}^{T}, \tag{1}$$

where θ remains fixed throughout adaptation and $\mathcal{M}_i^{\mathrm{F}}$ denotes the frozen feature memory associated with canonical slot i.

Registration naturally induces correspondence across normal templates. Specifically, patches occupying the same canonical slot in two different templates form a positive relation,

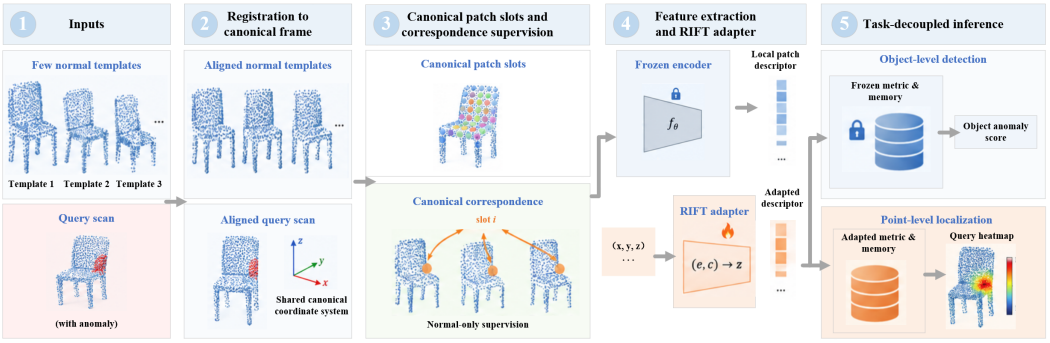


Figure 1: Overview of RIFT. Registered normal templates define a shared canonical frame and cross-template same-slot correspondence. A lightweight residual adapter uses registered coordinates as spatial conditioning and same-slot observations as normal-only supervision. The adapted metric is used for point-level localization, while object-level scoring retains the frozen detector.

$$(t, i) \sim (t', i), \qquad t \neq t', \tag{2}$$

whereas patches from different slots provide nonmatching observations. This construction assumes that the retained normal references are sufficiently well aligned for canonical slot identity to remain meaningful, but requires neither anomalous samples nor defect annotations.

The distinction between geometric similarity and structural identity is important for anomaly localization. Local patches from different object parts may have similar geometry, whereas observations of the same structural part may vary because of sampling density, sensor noise, or small surface variations. Canonical correspondence therefore provides an explicit target-category signal for reducing variation among repeated observations of the same structural role while maintaining separation between different roles.

3.2 Coordinate-Conditioned Residual Feature Tuning

Although the frozen representation captures generic local geometry, it is not explicitly specialized to the normal variations of the target category. RIFT therefore learns a lightweight residual transformation while keeping f_θ fixed.

For a registered center $c \in \mathbb{R}^3$, we first standardize its coordinates using statistics computed from the normal references and then apply Fourier encoding (Tancik et al., 2020):

$$\tilde{c} = \frac{c - \mu_c}{\sigma_c + \epsilon}, \qquad \gamma(c) = \left[\tilde{c},\ \left\{\sin(2^k \pi \tilde{c}), \cos(2^k \pi \tilde{c})\right\}_{k=0}^{F-1}\right]. \tag{3}$$

The frozen descriptor and coordinate encoding are jointly processed by a compact adapter,

$$h = \mathrm{MLP}\left(\mathrm{LN}([e; \gamma(c)])\right), \qquad u = W_u h + b_u, \qquad g = \sigma(W_g h + b_g), \tag{4}$$

where $u \in \mathbb{R}^d$ is a residual feature and $g \in (0, 1)$ is a data-dependent scalar gate. The adapted descriptor is

$$z(e, c) = \frac{e + \alpha g u}{\|e + \alpha g u\|_2}, \tag{5}$$

where α scales the residual update. The residual form keeps the representation anchored to the pretrained feature space instead of replacing it with an unconstrained projection. The output heads are zero-initialized, such that $z(e, c) = e$ at the beginning of adaptation.

Applying the adapter to the registered normal references produces the adapted slot memory

$$\mathcal{M}_i^{\mathrm{R}} = \left\{ z(e_i^t, c_i^t) \right\}_{t=1}^{T}. \tag{6}$$

We optimize this representation using the canonical correspondence defined in Eq. 2. Let z_a be an anchor, z_+ a same-slot observation from another normal template, and $\{z_k^-\}_{k=1}^K$ observations sampled from different canonical slots. Since all descriptors are normalized, their inner products correspond to cosine similarity. The correspondence objective is

$$\mathcal{L}_{\mathrm{align}} = -\log \frac{\exp(\langle z_a, z_+ \rangle / \tau)}{\exp(\langle z_a, z_+ \rangle / \tau) + \sum_{k=1}^{K} \exp(\langle z_a, z_k^- \rangle / \tau)}, \tag{7}$$

where τ is the temperature. We additionally encourage repeated observations of the same canonical role to remain compact, giving the complete objective

$$\mathcal{L}_{\mathrm{RIFT}} = \mathcal{L}_{\mathrm{align}} + \lambda_{\mathrm{c}} \left(1 - \langle z_a, z_+ \rangle\right). \tag{8}$$

Only the residual adapter is optimized. The feature encoder, registration procedure, canonical patch construction, and normal reference geometry remain fixed. RIFT therefore specializes the local metric using structural supervision already available from registered normal references.

3.3 Localization-Only Adaptation at Inference

RIFT is designed to improve fine-grained local discrimination rather than to recalibrate the object-level detector. For a registered query patch P_i^q with center c_i^q, we first obtain its frozen descriptor e_i^q using the same encoder as the normal references and compute its adapted descriptor $z_i^q = z(e_i^q, c_i^q)$.

For point-level localization, the frozen descriptors in the original spatial matcher are replaced by their adapted counterparts. Let $\mathcal{C}(i)$ be the candidate canonical slots considered for query slot i. The local anomaly evidence is

$$d_i^{\mathrm{R}} = \mathcal{D}\left(z_i^q, \{\mathcal{M}_j^{\mathrm{R}} : j \in \mathcal{C}(i)\}\right), \tag{9}$$

where $\mathcal{D}$ is the unchanged matching operator of the base detector. The resulting patch scores are mapped to points using the original patch-to-point interpolation procedure.

Object-level detection remains entirely on the frozen branch, i.e., $S_{\mathrm{obj}}^{\mathrm{RIFT}}(X^q) = S_{\mathrm{obj}}^{\mathrm{F}}(X^q)$. Thus, RIFT changes only the feature metric used for local anomaly comparison, while preserving the original object-level behavior of the pretrained detector.

4 Experiments

4.1 Experimental Setup

Datasets and metrics. We evaluate RIFT on Real3D-AD (Liu et al., 2023) and Anomaly-ShapeNet V2 (Li et al., 2024). Real3D-AD contains 1,254 scans from 12 real-world object categories with four normal references per category, while Anomaly-ShapeNet V2 contains 40 categories, 160 normal training shapes, and 1,312 test shapes. We report object-level AUROC (O-AUROC) for anomaly detection and point-level AUROC (P-AUROC) for localization. Because anomalous points are typically sparse, P-AUPR is also reported for the paired Frozen–RIFT comparison.

Implementation details. RIFT is built on the registration pipeline and frozen RIConv++ representation of COSSAD (Tarvo et al., 2026). Each point cloud is represented by 1,000 canonical patches of 512 points with 64-dimensional descriptors. The residual adapter has a hidden dimension of 128 and uses four Fourier frequencies. We set $\alpha = 0.25$, $K = 16$, $\tau = 0.08$, and $\lambda_{\mathrm{c}} = 0.25$, and optimize the adapter with AdamW at a learning rate of 2×10^{-3}. Only the adapter is updated; the encoder and registration procedure remain frozen.

Table 1: Controlled paired localization results on the two benchmarks. Frozen COSSAD and RIFT share the evaluator and differ only by adaptation. Values are percentages; best values are bold.

Dataset	Frozen COSSAD		RIFT	
	P-AUROC	P-AUPR	P-AUROC	P-AUPR
Real3D-AD	94.21	30.29	**96.43**	**45.95**
Anomaly-ShapeNet V2	94.65	44.40	**95.10**	**46.19**

Table 2: Category-wise Real3D-AD comparison (%). Published values are transcribed as reported; published averages are not recomputed from rounded class-wise entries. Bold denotes the largest numerical value in each column. RIFT keeps the frozen COSSAD object branch; Frozen COSSAD and RIFT form the controlled localization comparison in (b).

(a) O-AUROC ↑

Method	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Avg.
Reg3D-AD (Liu et al., 2023)	71.6	69.7	82.7	85.2	90.0	58.4	91.5	41.7	76.2	58.3	50.6	68.5	70.4
Group3AD (Zhu et al., 2024)	74.4	72.8	84.7	78.6	93.2	67.9	97.6	53.9	84.1	58.5	56.2	79.6	75.1
MC3D-AD (Cheng et al., 2025)	85.0	74.9	83.0	71.5	95.5	83.1	86.5	56.0	71.6	80.3	76.6	73.8	78.2
PASDF (Zheng et al., 2025)	62.8	95.9	78.8	73.9	89.4	65.8	98.9	63.4	**100.0**	85.0	61.7	86.6	80.2
SeDiR (Kim et al., 2026)	86.0	78.3	81.9	72.9	94.8	86.2	93.8	62.7	67.4	77.9	**85.4**	84.5	81.0
COSSAD (Tarvo et al., 2026)	91.2	**96.5**	99.9	**85.4**	**100.0**	**94.8**	**99.4**	**97.1**	80.5	79.1	84.4	88.7	91.6
RIFT (Ours)	**95.83**	94.40	**100.00**	85.08	**100.00**	94.52	98.40	96.63	94.88	**87.23**	84.33	**89.38**	**93.39**

(b) P-AUROC ↑

Method	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Avg.
Reg3D-AD (Liu et al., 2023)	63.1	71.8	72.4	67.6	83.5	50.3	82.6	54.5	81.7	81.1	61.7	75.9	70.5
Group3AD (Zhu et al., 2024)	63.6	74.5	73.8	75.9	86.2	63.1	83.6	56.4	82.7	79.8	62.5	80.3	73.5
MC3D-AD (Cheng et al., 2025)	62.8	81.9	91.0	64.0	94.2	82.2	93.2	45.8	65.9	77.8	69.0	93.4	76.8
PASDF (Zheng et al., 2025)	77.7	80.2	54.6	76.8	69.9	78.4	83.7	65.4	88.7	64.8	70.3	83.8	74.5
SeDiR (Kim et al., 2026)	66.6	90.8	95.1	63.2	97.7	83.4	96.7	54.7	73.9	79.8	70.9	**94.0**	80.6
COSSAD (Tarvo et al., 2026)	96.4	98.2	98.2	87.4	97.5	97.0	96.2	**97.4**	**96.5**	91.2	95.0	92.5	95.3
Frozen COSSAD (paired)	**97.44**	97.61	98.67	80.81	96.34	97.10	95.72	95.95	96.06	**91.33**	93.02	90.46	94.21
RIFT (Ours)	96.33	**99.17**	**99.24**	**94.39**	**97.88**	**98.16**	**99.47**	96.28	96.38	91.10	**95.30**	93.46	**96.43**

For target-category adaptation and model selection, RIFT uses only accepted normal reference templates. One normal reference is held out to select the adaptation budget from candidate checkpoints up to 1,200 updates using normal-reference retrieval statistics; the adapter is then reset and refit on all accepted normal references for the selected number of updates. No test sample, anomaly annotation, anomaly type, or test metric is used during this procedure. The paired Frozen COSSAD baseline uses the same registration, canonical patch construction, matching, and interpolation pipeline.

4.2 Comparison with State-of-the-Art Methods

Table 1 first summarizes the controlled paired comparison that is central to our evaluation. These rows use the same registration and inference pipeline and therefore isolate the effect of RIFT adaptation. The larger relative changes in P-AUPR motivate reporting it alongside P-AUROC rather than treating it only as an auxiliary textual metric.

Results on Real3D-AD. Table 2 reports category-wise O-AUROC and P-AUROC on all 12 Real3D-AD categories. Published baselines are transcribed from the cited papers or their peer-reviewed comparison tables rather than reimplemented in our codebase (Cheng et al., 2025; Zheng et al., 2025; Kim et al., 2026; Tarvo et al., 2026). Because preprocessing, registration, and sampling protocols can differ across methods, these results provide benchmark context rather than a controlled cross-paper ranking. Recent methods Simple3D (Cheng et al., 2026) and AARD (Wu et al., 2026) report macro O/P-AUROC values of 80.4/92.3 and 82.0/86.0, respectively. The Frozen COSSAD and RIFT rows form the controlled comparison under our evaluator.

Results on Anomaly-ShapeNet V2. Under the paired evaluator, RIFT improves P-AUROC from 94.65% to 95.10% and P-AUPR from 44.40% to 46.19%. Positive P-AUROC improvements are

Table 3: Category-wise P-AUROC on Anomaly-ShapeNet V2 (%). Published baselines are transcribed from the cited sources; the last two rows use our paired evaluator. Bold denotes the largest numerical value in each column, while cross-protocol differences are not interpreted as controlled rankings.

Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5	bucket0	bucket1
Reg3D-AD (Liu et al., 2023)	69.8	71.5	88.6	69.6	52.5	77.5	61.5	59.3	65.4	80.0	69.1	61.9	75.2
MC3D-AD (Cheng et al., 2025)	80.1	81.5	89.5	88.3	90.1	82.7	53.4	60.7	78.3	65.4	55.2	79.0	89.5
PO3AD (Ye et al., 2025)	**96.2**	94.9	91.2	84.4	88.0	97.8	91.4	91.8	93.5	96.7	94.1	75.5	89.9
SeDiR (Kim et al., 2026)	76.4	89.8	91.1	89.8	94.5	86.6	70.0	79.5	86.9	80.6	64.6	71.7	88.4
PASDF (Zheng et al., 2025)	91.9	**95.8**	95.1	**92.6**	**94.8**	96.3	90.0	81.6	93.9	86.5	90.9	**87.5**	82.4
Simple3D (Cheng et al., 2026)	92.0	95.4	97.4	72.8	83.8	**98.8**	95.1	93.3	**99.3**	92.9	97.9	72.5	92.1
Frozen COSSAD (paired)	95.98	86.36	98.91	90.16	87.77	95.51	99.32	99.28	99.18	**99.37**	99.38	79.27	95.47
RIFT (Ours)	95.97	85.74	**98.97**	89.28	87.89	93.24	**99.38**	**99.40**	**99.30**	99.34	**99.40**	80.20	**96.30**

Method	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0	headset1	helmet0	helmet1	helmet2	helmet3
Reg3D-AD (Liu et al., 2023)	63.2	71.8	81.5	46.7	68.5	69.8	75.5	58.0	62.6	60.0	62.4	82.5	62.0
MC3D-AD (Cheng et al., 2025)	86.7	92.2	88.0	88.5	82.9	71.5	85.6	66.6	70.9	74.4	58.2	81.7	60.2
PO3AD (Ye et al., 2025)	95.7	94.8	94.0	86.4	90.9	93.2	**97.4**	82.3	90.7	87.8	94.8	93.2	84.6
SeDiR (Kim et al., 2026)	90.0	97.2	94.0	95.0	87.1	84.2	74.8	70.8	68.7	81.0	59.6	88.5	70.1
PASDF (Zheng et al., 2025)	94.8	86.1	89.4	92.0	94.8	88.4	94.5	86.3	89.1	81.6	64.6	80.9	95.8
Simple3D (Cheng et al., 2026)	**98.8**	96.4	97.9	96.4	97.9	93.7	97.0	93.6	95.8	90.2	90.2	94.7	92.8
Frozen COSSAD (paired)	97.83	97.08	99.28	99.57	99.07	96.17	59.84	99.08	98.55	96.96	98.87	**98.30**	96.61
RIFT (Ours)	97.65	**97.50**	**99.35**	**99.61**	**99.32**	**97.62**	74.82	**99.12**	**98.66**	**97.56**	**99.00**	97.54	**98.60**

Method	jar0	microphone0	shelf0	tap0	tap1	vase0	vase1	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Avg.
Reg3D-AD (Liu et al., 2023)	59.9	59.9	68.8	58.9	74.1	54.8	60.2	40.5	51.1	75.5	62.4	88.1	81.1	69.4	66.8
MC3D-AD (Cheng et al., 2025)	84.8	85.4	65.8	52.4	53.7	87.0	72.6	82.2	79.7	78.6	64.7	61.6	88.5	77.6	75.9
PO3AD (Ye et al., 2025)	87.1	81.0	66.3	78.3	69.2	95.5	88.2	97.8	88.4	90.2	93.7	98.2	95.0	95.2	89.8
SeDiR (Kim et al., 2026)	91.1	89.1	67.9	59.7	63.0	84.6	72.2	85.0	83.0	88.4	76.6	77.7	90.0	83.5	81.0
PASDF (Zheng et al., 2025)	95.9	95.1	86.5	88.4	90.2	94.4	79.7	95.6	86.8	89.9	91.5	95.9	90.9	86.3	89.7
Simple3D (Cheng et al., 2026)	**97.8**	96.4	90.1	85.7	81.1	93.4	80.7	**98.6**	91.0	**98.5**	96.6	**99.0**	97.5	91.3	92.9
Frozen COSSAD (paired)	88.36	98.53	96.94	97.95	92.01	98.12	**98.16**	69.43	95.97	95.77	**98.39**	96.80	**98.39**	**98.06**	94.65
RIFT (Ours)	88.74	**98.75**	**96.97**	**98.18**	**92.90**	**98.19**	95.88	71.07	**96.26**	94.96	98.33	97.46	97.69	97.95	**95.10**

observed on 28 of the 40 categories. Table 3 reports the complete category-wise comparison with published baselines.

4.3 Ablation Studies

The main ablation isolates the effects of fixed spatial information, learned coordinate-conditioned adaptation, and correct registration-induced correspondence. Secondary controls are reported in Appendix A, while the representation analysis is presented in Sec. 4.5.

Core contribution ablation. Table 4 isolates the spatial and correspondence cues under the shared Real3D-AD evaluator. Frozen+FixedCoord augments the frozen descriptor with a fixed Fourier encoding of the registered patch center, without learning. Coordinate-only removes descriptor conditioning from the adapter input while retaining the frozen descriptor as the residual anchor, testing how much can be achieved by learning a position-conditioned correction. Coord+Shuffled keeps the coordinate input, adapter architecture, and optimization budget but replaces the true same-slot positive relation with fixed nonidentity slot permutations.

A fixed spatial prior improves P-AUPR from 30.29% to 36.79% while leaving P-AUROC essentially unchanged. Learning the residual correction from registered position is substantially stronger: Coordinate-only reaches 95.92% P-AUROC and 42.41% P-AUPR. Most importantly, under the same coordinate-conditioned adapter, replacing shuffled positives with correct same-slot relations improves P-AUROC from 95.75% to 96.43% and P-AUPR from 37.57% to 45.95%. These results show that correct registration-induced same-slot relations provide complementary supervision beyond a fixed spatial cue or generic position-conditioned adaptation. As an additional matched control, GNA-NN replaces canonical same-slot positives with cross-template frozen-feature nearest neighbors and obtains 93.36% P-AUROC under the same evaluator, compared with 94.21% for Frozen COSSAD and 96.43% for RIFT. Thus, generic normal-feature adaptation does not reproduce the gain from registration-induced correspondence; details are provided in Appendix A.2.

Table 4: Core contribution ablation on Real3D-AD under the shared evaluation protocol. The table focuses on the spatial-conditioning and correspondence components of RIFT; the matched GNA-NN control is reported separately in Appendix A.2. Values are percentages; best values are bold.

Variant	P-AUROC ↑	P-AUPR ↑
Frozen COSSAD	94.21	30.29
Frozen + FixedCoord	94.18	36.79
Coordinate-only	95.92	42.41
Coord + Shuffled	95.75	37.57
Coord + Correct (Full RIFT)	**96.43**	**45.95**

Relative to Frozen COSSAD, Full RIFT improves P-AUROC by 2.22 percentage points and P-AUPR by 15.66 points. Full RIFT exceeds Coordinate-only by 0.51 P-AUROC and 3.54 P-AUPR points, while the more tightly controlled Coord + Shuffled versus Coord + Correct comparison yields +0.68 P-AUROC and +8.38 P-AUPR. The latter comparison is the primary evidence that correct canonical correspondence adds value beyond the shared coordinate-conditioned adaptation framework. As an additional transfer check, Appendix A.3 replaces only the frozen local descriptor backbone with PointMAE while keeping registration, canonical slots, matching, and interpolation unchanged. RIFT remains beneficial under this substantially different representation.

4.4 Efficiency

RIFT introduces only 36,855 trainable parameters and leaves the point encoder and registration pipeline unchanged. Because adaptation operates on cached normal descriptors, it does not require repeated backbone extraction during target-category fitting and takes approximately 7.4 seconds per category on an NVIDIA RTX 3090. At inference, the residual adapter adds about 0.12 ms for 1,000 descriptors. The dominant operations of the base detector—registration, backbone extraction, memory search, and patch-to-point interpolation—remain unchanged. RIFT therefore provides target-category specialization with a small optimization and inference overhead rather than introducing a second category-specific detector.

4.5 Qualitative Results

Anomaly localization. Figure 2 compares representative point-level anomaly maps produced by COSSAD and RIFT on Real3D-AD and Anomaly-ShapeNet V2 using the same test examples and visualization scale. Compared with COSSAD, RIFT generally produces more spatially concentrated responses around annotated anomalous regions while suppressing diffuse activations over normal surfaces. The improvement is particularly visible for small and spatially localized defects, where reducing normal-region responses helps separate anomalous points from the surrounding geometry. This qualitative behavior is consistent with the improvement in P-AUPR, suggesting that RIFT enhances local anomaly discrimination rather than simply increasing response magnitudes globally.

Representation and correspondence analysis. Figure 3 provides an intuitive view of the structural supervision used by RIFT and its effect on the normal feature space. Panel (a) illustrates that registration places normal templates in a shared canonical frame, so that the same slot indexes repeated observations of the same structural role. Panel (b) contrasts this relation with generic feature-nearest matching: local geometric similarity alone can associate patches from different structural roles, whereas same-slot correspondence uses the registered identity of the normal templates. Panel (c) visualizes the resulting representation change on the airplane category of Real3D-AD. After adaptation, observations sharing a canonical slot are more consistently grouped across different normal templates while different slots remain separated. Consistently, held-out same-slot top-1 retrieval increases from 40.2% to 99.7% on Real3D-AD and from 13.0% to 99.9% on Anomaly-ShapeNet V2.

The first two panels are schematic and are intended to explain the source of RIFT supervision rather than provide independent empirical evidence. Likewise, the t-SNE projection is used only as a representation diagnostic. Because the adapter also receives registered coordinates, the visualization

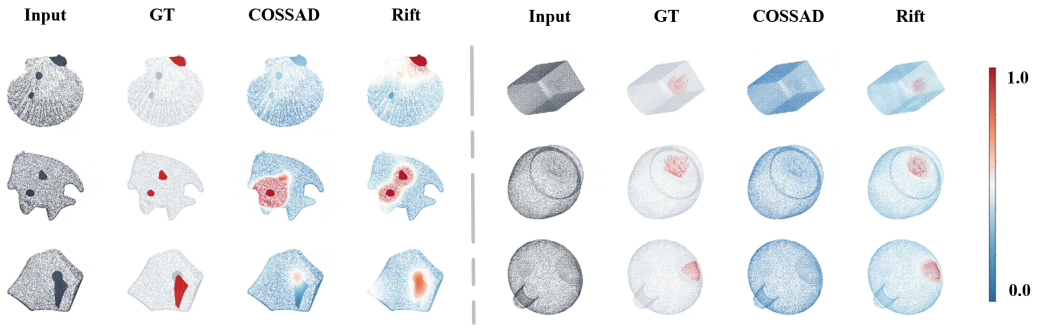


Figure 2: Qualitative point-level anomaly localization on Real3D-AD and Anomaly-ShapeNet V2. Each row shows the input/query, ground truth, COSSAD, and RIFT for the same test example under a common anomaly-score visualization scale.

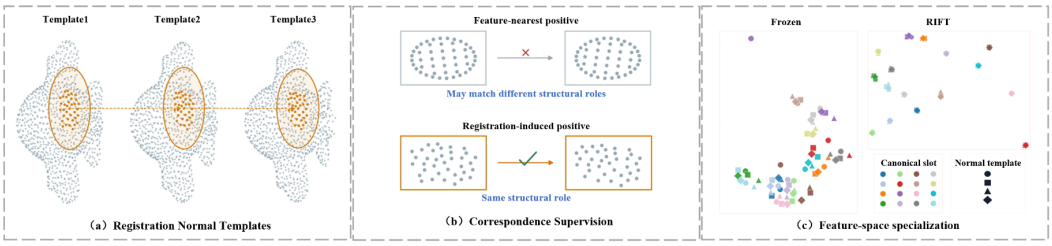


Figure 3: Registration-induced correspondence and feature-space specialization in RIFT. (a) Registered normal templates share a canonical frame, allowing the same slot to identify repeated observations of the same structural role. (b) Schematic comparison between feature-nearest matching, which can associate geometrically similar patches from different structural roles, and registration-induced same-slot correspondence. (c) t-SNE visualization of the same canonical normal patches for the airplane category of Real3D-AD before and after RIFT adaptation. Colors denote canonical slots and marker shapes denote normal templates. Panels (a–b) are conceptual illustrations, while panel (c) is a representation diagnostic rather than an independent performance measure.

alone cannot isolate the effect of correspondence supervision; the controlled Coord+Shuffled versus Coord+Correct comparison in Sec. 4.3 provides the corresponding quantitative evidence.

5 Conclusion

We introduced RIFT, a lightweight post-hoc adaptation framework for few-shot 3D anomaly localization that turns the structure exposed by registered normal references into target-category supervision after the representation has been frozen. Canonical spatial identity conditions a residual feature update, while cross-template same-slot correspondence supervises the specialization of the local metric. The controlled evidence distinguishes this mechanism from generic adaptation: GNA-NN reaches 93.36% P-AUROC under the same evaluator, compared with 94.21% for Frozen COSSAD and 96.43% for RIFT, while replacing shuffled positives with correct same-slot correspondence further improves both P-AUROC and P-AUPR under an otherwise matched adapter. Across Real3D-AD and Anomaly-ShapeNet V2, RIFT improves point-level localization with only a small category-specific adapter and no target-category anomaly supervision. These results support registration-induced structural correspondence as a useful source of post-hoc normal-only supervision for specializing a frozen detector. A limitation is that RIFT inherits its canonical coordinates and same-slot relations from upstream registration; residual misalignment can therefore reduce the reliability of the adaptation signal.

REFERENCES

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

Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The MVTec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), Volume 5: VISAPP*, pp. 202–213, 2022. doi: 10.5220/0010865000003124.

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. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 3327–3334, 2026. doi: 10.1609/aaai.v40i5.37328.

Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6185–6194. PMLR, 2023.

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.

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

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.

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, Datasets and Benchmarks Track*, volume 36, 2023.

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*, pp. 1630–1638, 2025. doi: 10.24963/ijcai.2025/182.

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.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7537–7547, 2020.

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

Yuanpeng Tu, Boshen Zhang, Liang Liu, Yuxi Li, Jiangning Zhang, Yabiao Wang, Chengjie Wang, and Cairong Zhao. Self-supervised feature adaptation for 3d industrial anomaly detection. In *Computer Vision – ECCV 2024*, volume 15060 of *Lecture Notes in Computer Science*, pp. 75–91. Springer, 2024. doi: 10.1007/978-3-031-72627-9_5.

Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8032–8041, 2023.

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*, pp. 14648–14657, 2026.

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, 2025.

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*, volume 38, pp. 5806–5832, 2025. doi: 10.52202/085713-0180.

Baozhu Zhao, Qiwei Xiong, Xiaohan Zhang, Jingfeng Guo, Qi Liu, Xiaofen Xing, and Xiangmin Xu. Pointcore: Efficient unsupervised point cloud anomaly detector using local-global features. *arXiv preprint arXiv:2403.01804*, 2024.

Bozhong Zheng, Jinye Gan, Xiaohao Xu, Xintao Chen, Wenqiao Li, Xiaonan Huang, Na Ni, and Yingna Wu. Bridging 3d anomaly localization and repair via high-quality continuous geometric representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27063–27072, 2025. doi: 10.1109/ICCV51701.2025.02512.

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*, pp. 91–107, 2024.

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 Additional Experimental Analyses

The following material complements the main experiments with implementation details for the fixed-coordinate control and additional analyses of generic normal adaptation, transfer to a different frozen representation, and the auxiliary compactness objective.

A.1 Fixed-coordinate control

For Frozen+FixedCoord, the frozen RIConv++ descriptor is concatenated with an independently normalized Fourier encoding of the registered patch center using a fixed weight of 1.0, followed by joint ℓ_2 normalization. The resulting descriptor is evaluated by the unchanged cosine-based indexed memory matcher and Cohen's effect-size scoring of the base detector. No parameters are trained, and this control uses neither a correspondence objective nor anomaly labels.

A.2 Generic normal adaptation control

GNA-NN is an auxiliary control introduced in this work. It uses the same frozen encoder, coordinate-conditioned residual adapter, optimizer, and normal-only training budget as RIFT, but replaces canonical same-slot positives with cosine-nearest frozen normal descriptors from another normal template. These nearest-neighbor relations are precomputed from frozen descriptors and kept fixed during adaptation.

We re-evaluate GNA-NN with the same category-macro P-AUROC protocol used for the paired Real3D-AD results in the main paper. Under this matched evaluator, Frozen COSSAD, GNA-NN, and RIFT obtain 94.21%, 93.36%, and 96.43%, respectively. Thus, generic nearest-neighbor normal adaptation does not reproduce the improvement of RIFT and slightly degrades P-AUROC relative to the frozen representation.

Table 5: Matched generic-normal-adaptation control on Real3D-AD using the same category-macro P-AUROC evaluator as the main results. Values are percentages; best is bold.

Variant	P-AUROC ↑
Frozen COSSAD	94.21
GNA-NN	93.36
RIFT	**96.43**

Table 6: Cross-backbone transfer control on Real3D-AD. Registration, canonical slots, matcher, and interpolation are unchanged; only the frozen local descriptor backbone is replaced by PointMAE. Values are percentages; best values are bold.

Variant	P-AUROC ↑	P-AUPR ↑
Frozen PointMAE	56.78	2.34
PointMAE + RIFT	**59.70**	**3.34**

Table 7: Compactness-loss ablation on Real3D-AD. Values are percentages; best values are bold.

Variant	P-AUROC ↑	P-AUPR ↑
Frozen COSSAD	94.21	30.29
RIFT w/o compactness	96.38	43.91
Full RIFT	**96.43**	**45.95**

A.3 Cross-backbone transfer to PointMAE

To examine whether the adaptation effect is tied to the RIConv++ representation, we replace only the frozen local descriptor extractor with a PointMAE-based patch encoder while retaining the same COSSAD registration, canonical slot construction, indexed matching, and patch-to-point interpolation.

The PointMAE encoder remains frozen; RIFT is applied post hoc using the same registered-coordinate conditioning and same-slot supervision.

Under this controlled backbone swap, Frozen PointMAE obtains 56.78% P-AUROC and 2.34% P-AUPR, whereas PointMAE + RIFT reaches 59.70% and 3.34%, respectively. The paired category-wise comparison is positive on 10 of 12 Real3D-AD categories. The absolute PointMAE baseline is substantially weaker than the RIConv++ representation in this local-patch pipeline, so we interpret this experiment as supporting evidence that RIFT's adaptation effect is not restricted to the original backbone, rather than as evidence that this PointMAE configuration is itself a competitive detector.

A.4 Compactness ablation

Removing the compactness term ($\lambda_c = 0$) yields 96.38% P-AUROC and 43.91% P-AUPR, whereas the full objective reaches 96.43% and 45.95%, respectively. The similar P-AUROC and higher P-AUPR with the full objective indicate that compactness mainly benefits fine-grained ranking of sparse anomalous points.