RAID: Retrieval-Augmented Anomaly Detection

Supplementary Material

In this supplementary material, we present the detailed architecture of the guided MoE filter (Sec. S1), a feasibility analysis of category- and dataset-agnostic retrieval (Sec. S2), additional results **applying RAID to reconstruction-based approaches** (Sec. S3), ablation studies (Sec. S4, S7), Visualization of retrieval-reasoning interaction (Sec. S5), the analysis of expert specialization (Sec. S6), extended quantitative and visualization results (Sec. S8-S10), and an analysis of representative failure cases (Sec. S11).

S1. Detailed Architecture of MoE Filter

The proposed guided MoE filter consists of two stages. The first stage generates a fused guidance map through dual-guidance aggregation, while the second stage filters the anomaly cost volume under this fused guidance. Fig. 4 provides an overview of the detailed architecture.

In the first stage (Fig. 4 (top left)), each expert E_g^i is implemented as a lightweight residual block composed of a 3×3 convolution and a skip connection that concatenates the convolutional output with the expert input $\text{cat}(g_Q, g_s)$. The gated ensemble of all guidance experts produces the fused guidance feature $\tilde{g} \in \mathbb{R}^{H' \times W' \times 3D}$.

In the second stage (Fig. 4 (top right)), the initial anomaly cost volume $\mathcal{C}$ is refined by modeling its semantic affinity with the fused guidance $\tilde{g}$. The concatenated feature $\text{cat}(\tilde{g}, \mathcal{C})$ is passed into a router that generates a dense soft gating distribution over all filtering experts:

$$p = \text{Softmax}(\text{Router}(\text{cat}(\tilde{g}, \mathcal{C}))).$$

Each denoising expert $E_{\mathcal{C}}^i$ (Fig. 4 (bottom)) performs dual-branch filtering, i.e., a semantic-guided cross-attention branch and a confidence-aware convolution branch, to refine the initial anomaly cost volume $\mathcal{C}$ by modeling the semantic affinity with $\tilde{g}$.

Cross-attention branch. The fused guidance $\tilde{g}$ is convolved and projected into query space, while the cost volume $\mathcal{C}$ provides keys and values. The attention map $\mathcal{A}^i$ encodes semantic affinity and enables a semantically weighted refinement:

$$\mathcal{A}^i = \text{Softmax}\left(\frac{(\text{Conv}_{3\times3}(\tilde{g})W_Q^i)(\mathcal{C}W_K^i)^\top}{\sqrt{K}}\right),$$
$$\text{CrossAtt}(\tilde{g}, \mathcal{C}) = \mathcal{A}^i(\mathcal{C}W_V^i),$$

where $\text{Conv}_{3\times3}(\cdot)$ denotes a 3×3 convolutional layer, $W_Q^i, W_K^i, W_V^i \in \mathbb{R}^{K\times K}$ are learnable projections, $\mathcal{A}^i \in \mathbb{R}^{(H'W')\times(H'W')}$ is the attention map, and $\text{CrossAtt}(\tilde{g}, \mathcal{C})$ denotes the output of the cross-attention branch.

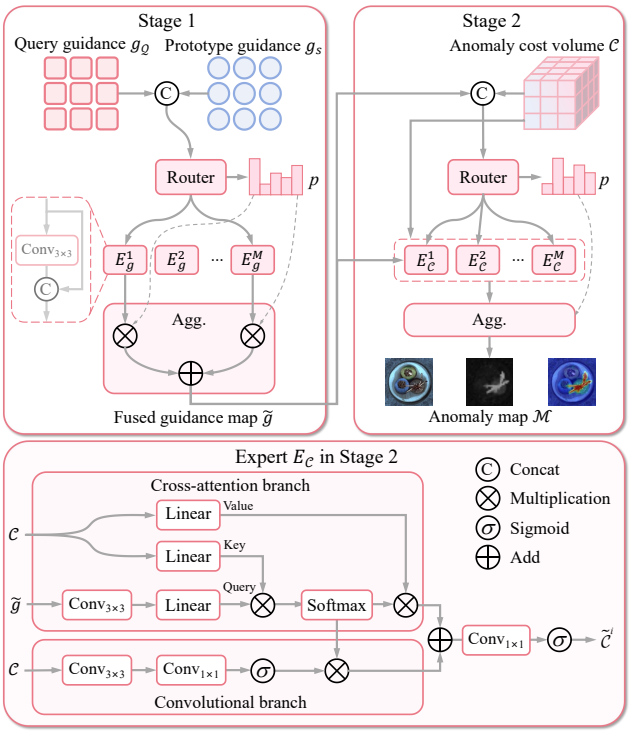

Figure 4. Detailed architecture of the guided MoE filter.

Convolutional branch. This branch estimates and modulates the reliability of the refined matches. A confidence map $R^i \in \mathbb{R}^{H'\times W'\times K}$ is obtained from $\mathcal{C}$ through a pair of convolutional layers:

$$R^i = \text{Sigmoid}(\text{Conv}_{1\times1}(\text{Conv}_{3\times3}(\mathcal{C}))).$$

The confidence map is then used to modulate the attention response from the cross-attention branch via $\mathcal{A}^i R^i$.

Finally, each denoising expert $E_{\mathcal{C}}^i$ outputs a weighted combination of its two branches:

$$\tilde{\mathcal{C}}^i = \text{Sigmoid}(\text{Conv}_{1\times1}(\text{CrossAtt}(\tilde{g}, \mathcal{C}) + \beta \cdot (\mathcal{A}^i R^i))),$$

where $\beta = 0.1$ controls the influence of the reliability modulation. The final anomaly map $\mathcal{M}$ is obtained by aggregating the expert outputs using the gating weights, yielding a spatially coherent, semantically guided anomaly prediction.

S2. Feasibility of Category-Agnostic Retrieval

To assess whether templates from multiple datasets and categories can be jointly organized within our hierarchical vector database using class-level semantic representations, we

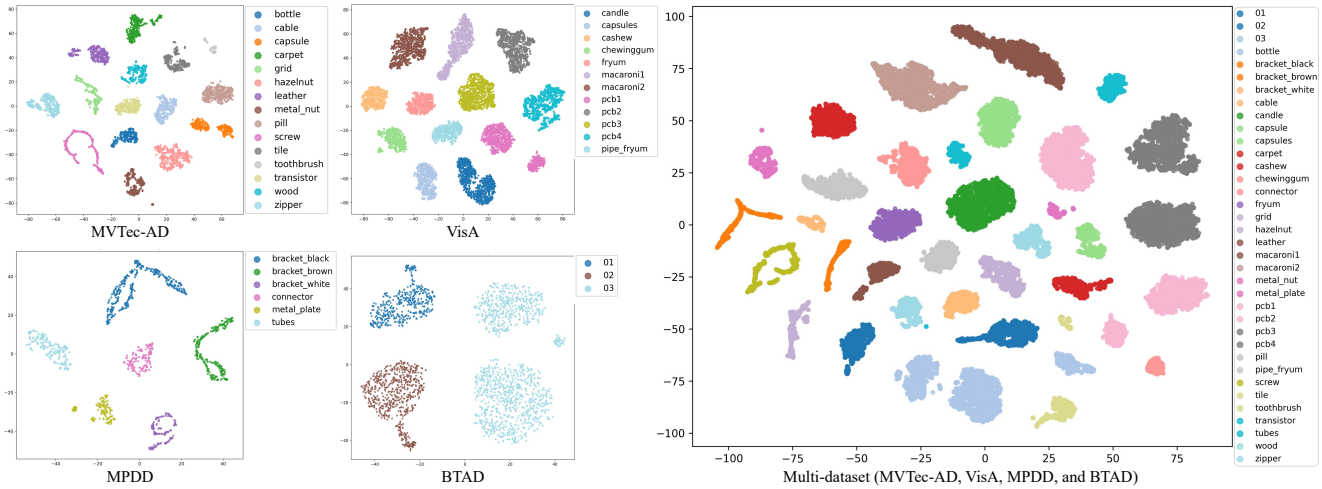

Figure 5. T-SNE visualizations of CLS tokens on MVTec, VisA, MPDD, BTAD, and multi-dataset. Distinct and compact clusters reveal that CLS tokens encode strongly discriminative class-level semantics.

Table 8. Multi-class anomaly detection/localization performance of RAID integrated with reconstruction-based methods on MVTec-AD. All results are obtained by directly applying the full-shot **pretrained RAID model** without any additional fine-tuning. † denotes results using DINOv2-s with a 8×8 patch size.

Method	I-AUROC	I-AP	I-F1max	P-AUROC	P-AP	P-F1max
GLAD† [59]	97.5	98.8	96.8	97.3	58.8	59.7
+ RAID†	**98.2**	**99.2**	**97.0**	**97.5**	**68.2**	**66.0**
DiAD [23]	97.2	**99.0**	**96.5**	96.8	52.6	55.5
+ RAID†	**97.4**	98.8	96.3	**97.7**	**69.8**	**67.2**

visualize the CLS tokens (extracted from the DINOv2-s encoder [46]) using t-distributed Stochastic Neighbor Embedding (t-SNE) on MVTec, VisA, MPDD, BTAD, and multi-dataset , incorporating all samples from each dataset (as shown in Fig. 5).

Across all datasets, the embeddings exhibit compact (intra-class) and well-separated clusters with clear inter-class boundaries, indicating that CLS tokens reliably encode global semantic cues that remain discriminative and stable across diverse categories. This also confirms that the derived class prototype entities are able to serve as robust semantic anchors that faithfully represent each category. These observations provide strong support for adopting CLS tokens as the foundational representation for class-level grouping within our hierarchical database, ensuring that each query can consistently locate its correct semantic cluster during category- and dataset-agnostic retrieval.

S3. Applicability to Reconstruction-based Approaches

In this section, we show that RAID can be seamlessly integrated with reconstruction-based UAD methods such as DiAD [23] and GLAD [59]. During inference, the reconstructed normal image of each test sample is treated as an exemplar for building a dynamic, sample-specific hierarchical vector database. In our experiments, RAID uses features from the final layer of DINO, whereas GLAD [59] and DiAD [23] rely on multi-layer features (four and two layers, respectively). Notably, we directly reuse the full-shot pretrained RAID weights without any fine-tuning.

As reported in Table 8, incorporating GLAD or DiAD into RAID yields substantial performance gains. These results demonstrate that RAID strengthens matching reliability and serves as a flexible, plug-and-play module for enhancing reconstruction-based anomaly detection methods.

S4. Ablation studies of Retrieval

To systematically examine the impact of the retrieval process, we conduct targeted ablation studies on VisA focusing on two aspects: retrieval hierarchy and retrieval hyperparameters.

Retrieval Hierarchy. Removing class proto., semantic proto., and instance tokens as shown in Table 9 causes I-AUROC/P-AP drops of 1.6%/2.0%, 3.0%/5.2%, and 15.4%/15.1%, respectively, with inference times of 0.28s, 0.08s, and 0.07s per image. Specifically, cases are sensitive to different levels: *Capsules*, *Macaroni1*, and *PCB1* drop by 2.9%/1.6% (w/o class), 7.3%/2.9% (w/o semantic), and 41.9%/67.4% (w/o instance), respectively, highlighting the unique contributions of each level.

Retrieval Hyperparameters. We evaluate the sensitivity of retrieval hyperparameters, including clustering seeds, numbers of semantic prototypes ($\#\bar{\mathbf{s}}$), retrieved semantic prototypes (K') and retrieved patch tokens (K). Tables 10 and 11 demonstrate our stable performance (robustness) across the first three settings. Performance drops only with

Table 9. Effectiveness of retrieval hierarchy on VisA using I-AUROC/P-AP.

Class	Semantic	Instance	Result(%)	Inf.(s)
-	✓	✓	93.3 / 43.2	0.28
✓	-	✓	91.9 / 40.0	0.08
✓	✓	-	79.5 / 30.1	0.07
✓	✓	✓	94.9 / 45.2	0.05

Table 10. Ablation of retrieval hyperparameters on VisA using I-AUROC/P-AP.

$\#\bar{\mathbf{s}}$	K'	K	Result(%)
10	1	5	91.3 / 39.6
100	1	5	91.4 / 39.5
50	20	150	94.9 / 45.3
50	5	150	94.9 / 45.2

Table 11. Ablation of clustering seeds on VisA using I-AUROC/P-AP.

Seed	0	2	42
Metrics	95.0 / 45.9	94.9 / 45.2	94.9 / 45.9

insufficient patch retrieval.

Class-level clustering is inherently stable due to large semantic separation between object categories, naturally grouping similar samples and making class prototypes insensitive to initialization. Semantic-level prototypes are derived from many patch tokens, providing sufficient coverage of intra-class variations. Under this regime, the semantic representations remain consistent across different configurations.

By contrast, retrieved patch tokens directly determine the informativeness of the cost volume, which drives MoE routing and guided filtering during generation. Retrieving too few tokens produces an under-informative cost volume, limiting expert routing and denoising, whereas retrieving sufficient tokens enables stable performance and eventual saturation. This highlights that retrieval is tightly coupled with the generation process.

S5. Visualization of the Retrieval-reasoning Interaction

We provide t-SNE visualizations in Fig. 6 comparing token-level representations before & after reasoning (step-level) under hierarchical vs. random retrieval, directly showing that *retrieved evidence actively shapes intermediate representations rather than passively conditioning the model*.

S6. Visualizing Expert Specialization in MoE

We analyze the routing behavior of the second-stage MoE. Specifically, we visualize the Kernel Density Estimation (KDE) distributions of routing weights across: different object categories (Pill and Transistor), and different anomaly types (pill type and scratch). The resulting distributions differ significantly across both dimensions in Fig. 7, indicating that the router consistently assigns distinct importance profiles to experts depending on the semantic context and anomaly characteristics. This shows that, despite dense activation, experts contribute unequally and conditionally, forming a soft but structured specialization pattern.

Table 12. Effectiveness of expert quantity on VisA using I-AUROC/P-AP.

$\#(E_g, E_{\mathcal{C}})$	(2, 3)	(3, 2)	(3, 3)
Metrics	94.0 / 40.5	94.1 / 41.8	94.9 / 45.2

Table 13. Effectiveness of template quantity on VisA using I-AUROC/P-AP.

# template	20	50	100	All
Metrics	91.9 / 44.0	94.0 / 44.7	94.9 / 45.2	94.9 / 47.0

Table 14. Ablation studies of guided MoE filter on VisA using I-AUROC/P-AP. IDs match with Table 6.

ID	1	2	3	4	5	7
Metrics	91.9 / 37.3	93.6 / 41.9	91.4 / 37.4	93.1 / 41.9	92.9 / 39.4	94.9 / 45.2

S7. Ablation Studies on VisA

We conduct ablations on the challenging VisA dataset, evaluating expert number, template quantity, and the guided MoE filter. Results in Tables 12, 13, and 14 consistently validate each component under realistic, non-saturated settings.

S8. Detailed Quantitative Performance under the Full-shot Setting

In this section, we present detailed per-class results of our multi-class model for both image-level anomaly detection and pixel-level anomaly localization on the MVTec-AD and VisA datasets under the full-shot setting. Results for the 15 MVTec-AD categories are reported in Tables 15, 16, 17, and 18, while those for the 12 VisA categories are shown in Tables 19, 20, 21, and 22. Across both benchmarks, our approach consistently achieves strong performance in detection and localization metrics, demonstrating robustness and effectiveness across diverse anomaly categories and establishing state-of-the-art results.

S9. Detailed Quantitative Performance under the Few-shot Setting

In this section, we report the mean and standard deviation of per-category performance under few-shot settings, with all results averaged over five random seeds. During retrieval,

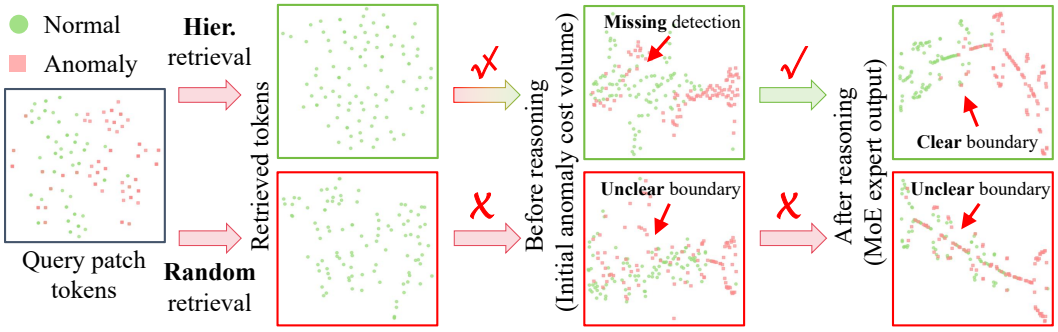

Figure 6. T-SNE visualization of token-level representations before and after step-level reasoning.

Table 15. Multi-class anomaly detection/localization performance on MVTec-AD with I-AUROC/P-AUROC metrics.

	Category	PatchCore [49]	UniAD [60]	SimpleNet	MambaAD [22]	GLAD [59]	DiAD [23]	ViTAD [64]	AnomalyDINO	Costfilter-AD	Ours
Object	Bottle	**100.0** / 99.2	**99.7** / 98.1	**100.0** / 97.2	**100.0** / 98.8	**100.0** / 99.7	98.4 / 98.4	99.5 / 99.0	**100.0** / 99.3	**100.0** / 98.8	**100.0** / 99.3
	Cable	95.3 / 93.6	95.2 / 97.3	97.5 / 96.7	98.8 / 95.8	98.7 / 93.4	94.8 / 96.8	99.7 / **98.6**	99.6 / 98.3	**99.8** / 98.2	98.9 / 97.5
	Capsule	96.8 / 98.0	86.9 / 98.5	90.7 / 98.5	94.4 / 98.4	96.5 / 99.1	89.0 / 97.1	**100.0** / **99.6**	89.7 / 99.1	96.4 / 98.9	96.8 / 99.2
	Hazelnut	99.3 / 97.6	99.8 / 98.1	99.9 / 98.4	**100.0** / 99.0	97 / 98.9	99.5 / 98.3	**100.0** / 96.6	99.9 / 99.6	**100.0** / 99.2	**100.0** / **99.7**
	Metal_nut	99.1 / 96.3	99.2 / 62.7	96.9 / **98.0**	99.9 / 96.7	99.9 / 97.3	99.1 / 97.3	98.7 / 96.4	**100.0** / 96.7	**100.0** / 97.9	**100.0** / 96.5
	Pill	86.4 / 90.8	93.7 / 95.0	88.2 / 96.5	97.0 / 97.4	94.4 / 97.9	95.7 / 95.7	**100.0** / 98.8	97.2 / 98.1	96.9 / 96.5	99.2 / **99.4**
	Screw	94.2 / 98.9	87.5 / 98.3	76.7 / 96.5	94.7 / 99.5	93.4 / **99.6**	90.7 / 97.9	**98.5** / 96.2	74.3 / 97.6	95.3 / 99.0	97.9 / 99.4
	Toothbrush	**100.0** / 98.8	94.2 / 98.4	89.7 / 98.4	98.3 / 99.0	99.7 / 99.2	99.7 / 99.0	95.4 / 98.3	99.7 / 99.2	**100.0** / 98.9	**100.0** / **99.3**
	Transistor	98.9 / 92.3	99.8 / 95.8	99.2 / 97.9	**100.0** / 96.5	99.4 / 90.9	99.8 / 95.1	99.8 / **99.0**	96.5 / 95.8	**100.0** / 97.1	99.8 / 95.9
	Zipper	97.1 / 95.7	95.8 / 96.8	99.0 / 97.9	99.3 / 98.4	96.4 / 93.0	95.1 / 96.2	99.7 / 96.4	98.8 / 94.3	98.9 / **98.3**	**99.7** / 98.1
Texture	Carpet	97.0 / 98.1	99.8 / 98.5	95.7 / 97.4	99.8 / 99.2	97.2 / 98.9	99.4 / 98.6	96.2 / 98.7	99.9 / 99.4	**100.0** / 98.5	**100.0** / **99.6**
	Grid	91.4 / 98.4	98.2 / 63.1	97.6 / 96.8	**100.0** / 99.2	95.1 / 98.2	98.5 / 96.6	91.3 / 99.0	98.7 / 97.8	99.3 / 98.3	**100.0** / **99.4**
	Leather	100 / 99.2	**100.0** / 98.8	**100.0** / 98.7	**100.0** / 99.4	99.5 / **99.7**	99.8 / 98.8	98.9 / 99.1	**100.0** / **99.7**	**100.0** / 99.3	**100.0** / 99.4
	Tile	96.0 / 90.3	99.3 / 91.8	99.3 / 95.7	98.2 / 93.8	100 / 97.8	96.8 / 92.4	98.8 / 93.9	**100.0** / 98.5	**100.0** / 95.0	**100.0** / **99.3**
	Wood	93.8 / 90.8	98.6 / 93.2	98.4 / 91.4	98.8 / 94.4	95.4 / 96.8	**99.7** / 93.3	97.6 / 95.9	97.9 / **97.6**	98.5 / 94.3	98.4 / 97.4
	Mean	96.4 / 95.7	96.5 / 96.8	95.3 / 96.9	98.6 / 97.7	97.5 / 97.3	97.2 / 96.8	98.3 / 97.7	96.8 / 98.1	99.0 / 98.0	**99.4** / **98.6**

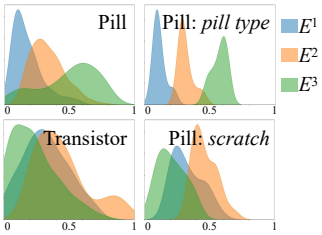

Figure 7. KDE distributions of second-stage MoE routing weights across object categories and anomaly types.

template samples are augmented with rotations to enhance diversity. Tables 23 and 24 summarize the image-level and pixel-level AUROC (I-AUROC and P-AUROC) for each category. The stable improvements observed across all categories and both datasets further highlight the strong generalizability of our method in few-shot anomaly detection.

S10. Extended Visualization Results

To further demonstrate the effectiveness of our approach, we provide multi-class UAD qualitative visualizations of anomaly detection heatmaps on MVTec-AD and VisA in Fig. 8 and Fig. 9. Compared with representative baselines, our method produces cleaner and more structurally coherent anomaly maps, exhibiting reduced noise, fewer spurious activations, and sharper defect boundaries. These results highlight the ability of our RAID model to deliver reliable and fine-grained localization, even for subtle or low-contrast anomalies.

To further demonstrate the effectiveness of RAID, we analyze the distributions of anomaly scores using KDE. Fig. 10 and Fig. 11 show the KDE curves for MVTec-AD and VisA, respectively, where the first two rows correspond to image-level metrics and the last two rows correspond to pixel-level metrics. Pink and blue represent abnormal and normal samples. Compared with a representative reconstruction-based method (GLAD [59]), RAID yields substantially reduced overlap between normal and abnormal score distributions, indicating stronger discriminative power at both the image and pixel levels.

S11. Failure Cases Analysis

As shown in Fig. 12, although the guided MoE filter effectively suppresses matching noise, failures still arise when

Table 16. Multi-class anomaly detection/localization performance on MVTec-AD with I-AP/P-AP metrics.

	Category	UniAD [60]	SimpleNet [39]	MambaAD [22]	GLAD [59]	DiAD [23]	ViTAD [64]	AnomalyDINO	Costfilter-AD	Ours
Object	Bottle	**100.0** / 60.0	**100.0** / 53.8	**100.0** / 79.7	**100.0** / 80.9	96.5 / 52.2	99.9 / 60.5	**100.0** / 87.3	**100.0** / 81.4	**100.0** / **89.6**
	Cable	95.9 / 39.9	98.5 / 42.4	99.2 / 42.2	99.3 / 51.4	98.8 / 50.1	**99.9** / 31.2	99.8 / **69.3**	99.8 / 58.0	99.4 / 62.7
	Capsule	97.8 / 42.7	97.9 / 35.4	98.7 / 43.9	99.2 / 49.1	97.5 / 42.0	**100.0** / 52.1	97.3 / 45.9	99.2 / 49.2	99.0 / **55.3**
	Hazelnut	**100.0** / 55.2	99.9 / 44.6	**100.0** / 63.6	98.2 / 68.0	99.7 / 79.2	**100.0** / 56.4	99.9 / 79.0	**100.0** / 72.4	**100.0** / **87.5**
	Metal_nut	99.9 / 14.6	99.3 / 83.1	**100.0** / 74.5	**100.0** / 81.8	96.0 / 30.0	99.6 / 60.6	**100.0** / 77.2	**100.0** / **79.0**	**100.0** / 74.9
	Pill	98.7 / 44.0	97.7 / 72.4	99.5 / 64.0	99.0 / 73.9	98.5 / 46.0	**100.0** / 79.9	99.5 / 78.6	99.4 / 59.6	99.9 / **84.2**
	Screw	96.5 / 28.7	90.6 / 15.9	97.9 / 49.8	98.0 / 47.8	**99.7** / **60.6**	99.1 / 43.1	88.0 / 12.5	98.3 / 33.5	99.3 / 56.3
	Toothbrush	97.4 / 34.9	95.7 / 46.9	99.3 / 48.5	99.9 / 45.0	99.9 / **78.7**	99.0 / 42.7	99.9 / 46.9	**100.0** / 51.6	99.9 / 71.1
	Transistor	98.0 / 59.5	98.7 / 58.2	**100.0** / 69.4	99.2 / 58.9	99.6 / 15.6	99.9 / 64.6	96.1 / 62.4	**100.0** / **76.6**	99.7 / 64.5
	Zipper	99.5 / 40.1	99.7 / 53.4	99.8 / 60.4	98.9 / 40.9	99.1 / 60.7	99.9 / **75.1**	99.7 / 44.0	99.7 / 55.1	**99.9** / 59.9
Texture	Carpet	99.9 / 49.9	98.7 / 38.7	99.9 / 60.0	99.1 / 72.2	99.9 / 42.2	99.3 / 77.8	**100.0** / 76.2	**100.0** / 64.7	**100.0** / **82.1**
	Grid	99.5 / 10.7	99.2 / 20.5	**100.0** / 47.4	93.6 / 10.2	99.8 / **66.0**	97.0 / 34.0	99.3 / 31.0	99.8 / 34.0	**100.0** / 52.1
	Leather	**100.0** / 32.9	**100.0** / 28.5	**100.0** / 50.3	99.8 / 61.7	99.7 / 56.1	99.6 / 51.3	**100.0** / 60.2	**100.0** / 47.4	**100.0** / **62.2**
	Tile	99.8 / 42.1	99.8 / 60.5	99.3 / 45.1	**100.0** / 70.3	99.9 / 65.7	98.3 / 58.4	**100.0** / 76.4	**100.0** / 56.0	**100.0** / **93.7**
	Wood	99.6 / 37.2	99.5 / 34.8	99.6 / 46.2	98.5 / 70.6	**100.0** / 43.3	99.3 / 42.6	99.3 / 72.7	99.5 / 52.4	99.5 / **80.4**
	Mean	98.8 / 43.4	98.4 / 45.9	99.6 / 56.3	98.8 / 58.8	99.0 / 52.6	99.4 / 55.3	98.6 / 61.3	99.7 / 58.1	**99.8** / **71.7**

Table 17. Multi-class anomaly detection/localization performance on MVTec-AD with I-F1max/P-F1max metrics.

	Category	UniAD [60]	SimpleNet [39]	MambaAD [22]	GLAD [59]	DiAD [23]	ViTAD [64]	AnomalyDINO	Costfilter-AD	Ours
Object	Bottle	**100.0** / 69.2	**100.0** / 62.4	**100.0** / 76.7	**100.0** / 75.5	91.8 / 54.8	99.4 / 64.1	**100.0** / 80.2	**100.0** / 77.6	**100.0** / **83.0**
	Cable	88.0 / 45.2	94.7 / 51.2	95.7 / 48.1	97.3 / 53.4	95.2 / 57.8	**99.1** / 36.7	97.8 / **67.0**	98.4 / 63.0	96.1 / 62.0
	Capsule	94.4 / 46.5	93.5 / 44.3	94.9 / 47.7	96.8 / 51.2	95.5 / 45.3	**100.0** / 55.8	94.3 / 48.9	96.4 / 53.0	98.6 / **57.4**
	Hazelnut	99.3 / 56.8	99.1 / 51.4	**100.0** / 64.4	94.4 / 63.8	97.3 / 80.4	**100.0** / 68.8	99.3 / 75.5	**100.0** / 70.8	99.3 / **82.8**
	Metal_nut	99.5 / 29.2	96.1 / 79.4	99.5 / 79.1	99.5 / **82.4**	91.6 / 38.3	96.7 / 58.3	**100.0** / 79.5	99.5 / 82.1	**100.0** / 74.2
	Pill	95.7 / 95.3	53.9 / 67.7	96.2 / 66.5	94.6 / 69.9	94.5 / 51.4	**100.0** / 75.6	97.1 / 71.1	96.9 / 61.2	98.6 / **78.7**
	Screw	89.0 / 92.6	37.6 / 23.2	94.0 / 50.9	92.2 / 47.6	**97.9** / **59.6**	95.7 / 47.4	87.2 / 19.4	94.5 / 40.5	96.3 / 54.1
	Toothbrush	95.2 / 45.7	92.3 / 52.5	98.4 / 59.2	98.4 / 57.4	99.2 / **72.8**	95.5 / 47.8	98.4 / 57.7	**100.0** / 61.9	98.4 / 71.2
	Transistor	93.8 / 64.6	97.6 / 56.0	**100.0** / 67.1	95.0 / 58.3	97.4 / 31.7	98.6 / 64.0	89.7 / 59.5	**100.0** / **74.1**	97.6 / 59.5
	Zipper	97.1 / 49.9	98.3 / 54.6	97.5 / 61.7	95.6 / 46.2	94.4 / 60.0	98.4 / **77.3**	97.9 / 49.3	98.3 / 59.5	**99.2** / 59.5
Texture	Carpet	99.4 / 51.1	93.2 / 43.2	99.4 / 63.3	96.6 / 67.9	98.3 / 46.4	96.4 / 75.2	99.4 / 67.7	**100.0** / 63.3	99.4 / **75.8**
	Grid	97.3 / 11.9	96.4 / 27.6	**100.0** / 47.7	98.3 / 24.1	97.7 / **64.1**	93.0 / 41.0	96.6 / 37.4	98.2 / 40.6	**100.0** / 54.1
	Leather	**100.0** / 34.4	**100.0** / 32.9	**100.0** / 53.3	98.4 / 60.7	97.6 / **62.3**	96.8 / 61.9	**100.0** / 57.4	**100.0** / 50.0	99.5 / 58.3
	Tile	98.2 / 50.6	98.8 / 59.9	95.4 / 54.8	**100.0** / 71.5	98.4 / 64.1	92.5 / 55.3	**100.0** / 76.6	**100.0** / 63.5	**100.0** / **85.3**
	Wood	96.6 / 41.5	96.7 / 39.7	96.6 / 48.2	95.1 / 65.2	**100.0** / 43.5	97.1 / 50.8	98.4 / 65.4	97.5 / 56.5	97.6 / **72.0**
	Mean	96.4 / 49.5	95.8 / 49.7	97.6 / 59.2	96.8 / 59.7	96.5 / 55.5	97.3 / 58.7	97.1 / 60.8	98.6 / 61.2	**98.7** / **68.5**

the cost volume provides weak or unreliable anomaly cues. Typical examples include reflections in the *Screw* case, irregular speckled patterns in the *Pill* case, and textureless surfaces in the challenging *Candle* case. While RAID can detect subtle anomalies such as the small spot in the *Capsules* case, low-resolution inputs and smooth textures often weaken feature discriminability, leading to ambiguous retrieval. This ambiguity may cause the filter to oversuppress true anomalies or retain residual noise. Addressing such visually ambiguous scenarios remains an important challenge for future work.

Table 18. Multi-class anomaly localization performance on MVTec-AD with AUPRO metrics.

	Category	UniAD [60]	SimpleNet [39]	MambaAD [22]	GLAD [59]	DiAD [23]	ViTAD [64]	AnomalyDINO [11]	Costfilter-AD [68]	Ours
Object	Bottle	93.1	89.0	95.2	96.1	86.6	94.7	97.5	96.1	**97.1**
	Cable	86.1	85.4	90.3	89.6	80.5	**95.8**	94.2	91.7	91.4
	Capsule	92.1	84.5	92.6	96.1	87.2	**97.9**	95.8	92.9	96.4
	Hazelnut	94.1	87.4	95.7	90.8	91.5	87.0	92.5	92.9	**97.6**
	Metal_nut	81.8	85.2	93.7	94.2	90.6	88.0	94.7	93.3	**95.8**
	Pill	95.3	81.9	95.7	94.3	89.0	94.3	96.7	96.1	**98.2**
	Screw	87.9	84.0	**97.1**	96.7	95.0	90.2	89.4	95.4	96.2
	Toothbrush	93.5	87.4	91.7	95.6	95.0	92.0	96.1	89.6	**96.8**
	Transistor	**97.9**	83.2	87.0	86.5	90.0	95.2	84.2	93.7	82.8
	Zipper	92.6	90.7	**94.3**	84.5	91.6	92.4	86.2	93.3	93.5
Texture	Carpet	94.4	90.6	96.7	95.3	90.6	95.3	97.6	95.4	**98.3**
	Grid	92.9	97.0	96.4	92.7	94.0	93.5	90.0	93.4	**97.3**
	Leather	96.8	98.7	93.0	97.0	91.3	90.9	98.5	**98.8**	97.3
	Tile	78.4	80.0	93.0	96.8	90.7	76.8	96.7	85.5	**97.3**
	Wood	86.7	91.2	95.9	86.3	97.5	87.2	93.4	90.0	**96.1**
	Mean	90.7	93.1	93.1	92.8	90.7	91.4	93.6	93.2	**95.5**

Table 19. Multi-class anomaly detection/localization on VisA with I-AUROC/P-AUROC metrics.

	Category	UniAD [60]	SimpleNet [39]	MambaAD [22]	DiAD [23]	GLAD [59]	ViTAD [64]	AnomalyDINO	Costfilter-AD	Ours
Complex Structure	PCB1	92.8 / 93.3	91.6 / 99.2	95.4 / **99.8**	88.1 / 98.7	69.9 / 97.6	95.8 / 99.5	87.4 / 99.3	**96.3** / 99.3	92.0 / 99.4
	PCB2	87.8 / 93.9	92.4 / 96.6	94.2 / **98.9**	91.4 / 95.2	89.9 / 97.1	90.6 / 97.9	81.9 / 94.2	**97.0** / 98.0	92.3 / 98.1
	PCB3	78.6 / 97.3	89.1 / 97.2	93.7 / **99.1**	86.2 / 96.7	93.3 / 96.2	90.9 / 98.2	87.4 / 96.5	89.8 / 97.7	**95.3** / 97.9
	PCB4	98.8 / 94.9	97.0 / 93.9	**99.9** / 98.6	99.6 / 97.0	99.0 / **99.4**	99.1 / 99.1	96.7 / 97.3	98.7 / 97.8	96.1 / 98.1
Multiple instances	Macaroni 1	79.9 / 97.4	85.9 / 98.9	91.6 / 99.5	85.7 / 94.1	93.1 / **99.9**	85.8 / 98.5	88.0 / 98.2	**93.7** / 99.4	92.8 / 99.6
	Macaroni 2	71.6 / 95.2	68.3 / 93.2	81.6 / 99.5	62.5 / 93.6	74.5 / 99.5	79.1 / 98.1	75.9 / 96.9	88.3 / 98.5	**90.6** / **99.6**
	Capsules	55.6 / 88.7	74.1 / 97.1	91.8 / 99.1	58.2 / 97.3	88.8 / **99.3**	79.2 / 98.2	93.6 / 97.0	80.1 / 97.6	**96.6** / 99.1
	Candle	94.1 / 98.5	84.1 / 97.6	96.8 / 99.0	92.8 / 97.3	86.4 / 98.8	90.4 / 96.2	90.3 / 96.1	**97.8** / 99.2	96.0 / **99.6**
Single instances	Cashew	92.8 / 98.6	88.0 / 98.9	94.5 / 94.3	91.5 / 90.9	92.6 / 86.2	87.8 / 98.5	**95.1** / 99.2	94.1 / 99.3	94.8 / **99.4**
	Chewing gum	96.3 / 98.8	96.4 / 97.9	97.7 / 98.1	99.1 / 94.7	98.0 / **99.6**	94.9 / 97.8	98.0 / 99.3	**99.3** / 99.5	98.9 / **99.6**
	Fryum	83.0 / 95.9	88.4 / 93.0	95.2 / 96.9	89.8 / 97.6	**97.2** / 96.8	94.3 / 97.5	93.4 / 96.1	88.9 / 97.8	94.6 / **98.1**
	Pipe fryum	94.7 / 98.9	90.8 / 98.5	98.7 / 99.1	96.2 / 99.4	98.1 / 98.9	97.8 / **99.5**	98.0 / 99.1	96.6 / **99.5**	**98.8** / 99.1
	Mean	85.5 / 95.9	87.2 / 96.8	94.3 / 98.5	86.8 / 96.0	90.1 / 97.4	90.5 / 98.2	90.5 / 97.5	93.4 / 98.6	**94.9** / **99.0**

Table 20. Multi-class anomaly detection/localization on VisA with I-AP/P-AP metrics.

	Category	UniAD [60]	SimpleNet [39]	MambaAD [22]	DiAD [23]	GLAD [59]	ViTAD [64]	AnomalyDINO	Costfilter-AD	Ours
Complex Structure	PCB1	92.7 / 83.6	91.9 / **86.1**	**93.0** / 77.1	88.7 / 49.6	72.5 / 38.0	94.7 / 64.5	84.6 / 81.3	91.6 / 66.9	91.1/78.9
	PCB2	87.7 / 4.2	93.3 / 8.9	**93.7** / 13.3	91.4 / 7.5	88.9 / 6.4	89.9 / 12.6	81.1 / 12.0	92.0 / **21.1**	93.4 / **21.1**
	PCB3	78.6 / 13.8	91.1 / **31.0**	94.1 / 18.3	87.6 / 8.0	94.0 / 25.0	91.2 / 22.4	90.2 / 23.3	82.1 / 28.6	**95.4** / 26.2
	PCB4	98.8 / 14.7	97.0 / 23.9	**99.9** / 47.0	99.5 / 17.6	98.2 / **52.6**	98.9 / 42.9	96.3 / 37.4	97.1 / 39.4	94.6 / 40.3
Multiple instances	Macaroni 1	79.8 / 3.7	82.5 / 3.5	89.8 / 17.5	85.2 / 10.2	**93.1** / 11.0	83.9 / 8.0	88.9 / 10.6	86.7 / **19.7**	91.7 / 14.2
	Macaroni 2	71.6 / 0.9	54.3 / 0.6	78.0 / 9.2	57.4 / 0.9	73.8 / 7.0	74.7 / 3.6	76.2 / 5.5	81.5 / **15.5**	**91.7** / 13.1
	Capsules	55.6 / 3.0	82.8 / 52.9	95.0 / **61.3**	69.0 / 10.0	94.1 / 47.8	87.6 / 30.4	**96.4** / 43.3	79.4 / 51.4	73.3 / 55.1
	Candle	94.0 / 17.6	73.3 / 8.4	96.9 / 23.2	92.0 / 12.8	88.2 / 29.3	91.2 / 16.8	90.2 / 28.1	92.5 / 41.0	**96.4** / **36.5**
Single instances	Cashew	92.8 / 51.7	91.3 / 68.9	97.3 / 46.8	95.7 / 53.1	96.4 / 29.2	94.2 / 63.9	**97.6** / 60.2	91.5 / 66.1	97.4 / **70.3**
	Chewing gum	96.2 / 54.9	98.2 / 26.8	98.9 / 57.5	**99.5** / 11.9	99.1 / 73.9	97.7 / 61.6	96.9 / 64.7	**99.5** / 40.9	**99.5** / **78.5**
	Fryum	83.0 / 34.0	93.0 / 39.1	**97.7** / 47.8	95.0 / **58.6**	98.9 / 36.1	97.4 / 47.1	97.4 / 46.7	86.8 / 54.0	97.6 / 52.4
	Pipe fryum	94.7 / 50.2	95.5 / 65.6	**99.3** / 53.5	98.1 / **72.7**	**99.3** / 50.1	99.0 / 66.0	99.2 / 61.0	93.5 / 71.3	99.3 / 56.2
	Mean	85.5 / 18.4	87.0 / 31.5	94.5 / 37.6	88.3 / 20.3	91.4 / 33.9	91.7 / 36.6	91.4 / 39.6	89.3 / 45.0	**95.5** / **45.2**

Table 21. Multi-class anomaly detection/localization on VisA with I-F1max/P-F1max metrics.

	Category	UniAD [60]	SimpleNet [39]	MambaAD [22]	DiAD [23]	GLAD [59]	ViTAD [64]	AnomalyDINO	Costfilter-AD	Ours
Complex Structure	PCB1	87.8 / 8.3	86.0 / **78.8**	91.6 / 72.4	80.7 / 52.8	70.1 / 44.4	**91.8** / 61.7	82.2 / 68.0	91.6 / 66.9	85.7 / 74.1
	PCB2	83.1 / 9.2	84.5 / 18.6	89.3 / 23.4	84.7 / 16.7	83.3 / 14.4	85.3 / 21.2	76.2 / 23.7	**92.0** / 21.1	85.7 / **35.3**
	PCB3	76.1 / 21.9	82.6 / 36.1	86.7 / 27.4	77.6 / 18.8	87.6 / 27.7	83.9 / 26.4	80.2 / **38.8**	82.1 / 28.6	**90.0** / 34.2
	PCB4	94.3 / 22.9	93.5 / 32.9	**98.5** / 46.9	97.0 / 27.2	98.0 / **52.0**	96.6 / 48.3	91.0 / 30.8	97.1 / 39.4	92.0 / 45.4
Multiple instances	Macaroni 1	72.7 / 9.7	73.1 / 8.4	81.6 / **27.6**	78.8 / 16.7	85.4 / 19.2	76.7 / 19.3	79.5 / 17.3	86.7 / 19.7	**87.6** / 20.8
	Macaroni 2	69.9 / 4.3	59.7 / 3.9	73.8 / 16.1	69.6 / 2.8	71.8 / 19.3	74.9 / 10.4	73.0 / 11.1	81.5 / 15.5	**83.5** / **22.9**
	Capsules	76.9 / 7.4	74.6 / 53.3	88.8 / **59.8**	78.5 / 21.0	85.9 / 53.3	79.8 / 41.4	89.8 / 45.6	79.4 / 51.4	**94.1** / 58.2
	Candle	86.1 / 27.9	76.6 / 16.5	**90.1** / 32.4	87.6 / 22.8	79.8 / 36.6	83.7 / 26.4	82.9 / 30.6	92.5 / 41.0	89.6 / **42.5**
Single instances	Cashew	91.4 / 58.3	84.7 / 66.0	91.1 / 51.4	89.7 / 60.9	90.5 / 38.2	86.1 / 62.7	**92.0** / 60.3	91.5 / 66.1	**92.0** / **66.4**
	Chewing gum	95.2 / 56.1	93.8 / 29.8	94.2 / 59.9	95.9 / 25.8	95.5 / 69.6	91.4 / 58.7	**97.5** / 56.9	96.9 / 64.7	97.0 / **72.3**
	Fryum	85.0 / 40.6	83.3 / 45.4	90.5 / 51.9	87.2 / **60.1**	95.8 / 43.5	90.9 / 50.3	**92.7** / 45.2	86.8 / 54.0	90.8 / 57.2
	Pipe fryum	93.9 / 57.7	88.6 / 63.4	97.0 / 58.5	93.7 / 69.9	97.0 / 55.1	94.7 / 66.5	97.5 / 56.2	93.5 / **71.3**	**98.5** / 60.9
	Mean	84.4 / 27.0	81.8 / 37.8	89.4 / 44.0	85.1 / 33.0	86.7 / 39.4	86.3 / 41.1	86.2 / 40.4	89.3 / 45.0	**90.5** / **49.2**

Table 22. Multi-class anomaly localization on VisA with AUPRO metrics.

	Category	UniAD [60]	SimpleNet [39]	MambaAD [22]	DiAD [23]	GLAD [59]	ViTAD [64]	AnomalyDINO [11]	Costfilter-AD [68]	Ours
Complex Structure	PCB1	64.1	83.6	**92.8**	80.2	88.3	89.6	82.8	88.3	90.2
	PCB2	66.9	85.7	**89.6**	67.0	91.7	82.0	77.7	85.4	78.5
	PCB3	70.6	85.1	89.1	68.9	**94.2**	88.0	79.7	75.6	82.0
	PCB4	72.3	61.1	87.6	85.0	**94.9**	91.8	83.1	84.4	86.5
Multiple Instances	Macaroni 1	84.0	92.0	95.2	68.5	**99.1**	89.2	90.2	96.4	98.0
	Macaroni 2	76.6	77.8	96.2	73.1	97.2	87.2	84.8	94.2	**97.5**
	Capsules	43.7	73.7	91.8	77.9	91.8	75.1	86.1	61.9	**90.1**
	Candle	91.6	87.6	95.5	89.4	92.8	85.2	94.1	95.2	**96.9**
Single Instances	Cashew	87.9	84.1	87.8	61.8	61.1	78.8	91.3	90.5	**96.3**
	Chewing gum	81.3	78.3	79.7	59.5	92.5	71.5	85.7	88.5	**94.5**
	Fryum	76.2	85.1	91.6	81.3	**96.4**	87.8	85.0	86.9	92.3
	Pipe fryum	91.5	83.0	95.1	89.9	**98.0**	94.7	94.7	94.5	97.1
	Mean	75.6	81.4	91.0	75.2	91.5	85.1	86.3	86.8	**91.7**

Table 23. **Few-shot** anomaly detection and localization performance of RAID on MVTec-AD, with per-category I-AUROC and P-AUROC reported as mean (%) ± standard deviation (%).

Object	1-shot		2-shot		4-shot	
	I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
Bottle	99.9±0.1	99.2±0.1	100.0±0.0	99.3±0.1	100.0±0.1	99.3±0.0
Cable	93.5±1.2	92.7±1.4	95.3±0.8	94.0±0.5	95.8±1.1	94.5±0.4
Capsule	79.7±3.9	98.5±0.2	85.8±7.6	98.7±0.2	91.0±2.7	98.9±0.1
Hazelnut	93.5±12.1	99.0±1.0	98.0±2.9	99.5±0.2	99.9±0.2	99.6±0.1
Metal_nut	99.8±0.2	96.3±0.4	99.9 ±0.1	96.4±0.5	100.0±0.0	97.0±0.3
Pill	94.5±1.2	97.5±0.3	96.2±0.9	98.0±0.2	97.2±0.3	98.0±0.3
Screw	84.3±3.0	96.8±0.7	83.3±2.3	95.3±0.3	80.2±3.0	93.2±0.9
Toothbrush	97.8±0.5	99.1±0.3	98.6±0.9	99.3±0.1	99.6±0.5	99.4±0.1
Transisitor	87.3±7.1	86.4±6.5	94.2±2.1	91.7±0.5	92.7±5.0	90.0±3.7
Zipper	97.9±0.5	92.2±2.0	98.3±0.7	92.7±1.0	98.1±0.3	93.2±0.8
Carpet	100.0±0.0	99.4±0.1	100.0±0.0	99.4±0.1	100.0±0.0	99.4±0.1
Grid	99.2±1.1	99.1±0.1	99.9±0.1	99.2±0.1	99.9±0.1	99.2±0.1
Leather	100.0±0.0	99.0±0.1	100.0±0.0	98.9±0.1	100.0±0.1	98.8±0.1
Tile	100.0±0.0	97.0±0.2	100.0±0.0	97.1±0.2	100.0±0.0	97.1±0.1
Wood	97.9±0.7	96.5±0.5	98.8±0.5	96.7±0.4	98.9±0.3	96.4±0.7
Mean	95.1±0.7	96.6±0.5	96.6±0.7	97.1±0.1	96.9±0.3	96.9±0.3

Table 24. **Few-shot** anomaly detection and localization performance of RAID on VisA, with per-category I-AUROC and P-AUROC reported as mean (%) ± standard deviation (%).

Object	1-shot		2-shot		4-shot	
	I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
PCB1	72.1±8.9	97.4±0.2	79.5±0.9	97.9±0.3	78.1±2.6	98.2±0.1
PCB2	81.4±2.0	97.2±0.5	81.5±4.1	97.5±0.2	83.8±1.4	97.4±0.3
PCB3	86.6±1.1	96.6±0.1	89.0±3.0	97.2±0.2	89.0±1.6	98.0±0.1
PCB4	93.4±1.2	96.7±0.7	96.4±0.9	97.3±0.2	95.8±1.2	97.4±0.3
Macaroni 1	85.1±5.0	97.2±1.2	84.3±3.0	96.2±1.9	87.7±2.1	96.4±1.8
Macaroni 2	59.5±4.9	97.5±0.5	70.0±6.3	98.0±0.4	70.6±6.7	97.9±0.5
Capsules	95.0±2.6	98.7±0.1	98.0±1.0	99.0±0.1	95.6±1.7	98.6±0.3
Candle	87.6±1.8	98.5±0.1	88.8±0.4	98.9±0.1	92.2±0.5	99.2±0.0
Cashew	92.0±3.3	98.3±0.4	92.6±2.0	98.0±0.2	93.5±0.9	99.2±0.2
Chewing gum	97.1±0.6	98.9±0.1	97.6±0.9	98.8±0.1	97.2±0.2	98.9±0.1
Fryum	91.9±1.9	96.6±0.5	92.9±1.8	96.9±0.4	93.6±1.4	97.5±0.1
Pipe fryum	87.4±5.8	98.8±0.3	90.9±4.7	99.2±0.0	95.5±0.7	99.4±0.1
Mean	85.8±0.6	97.1±0.1	88.5±0.2	97.9±0.2	89.3±0.5	98.2±0.2

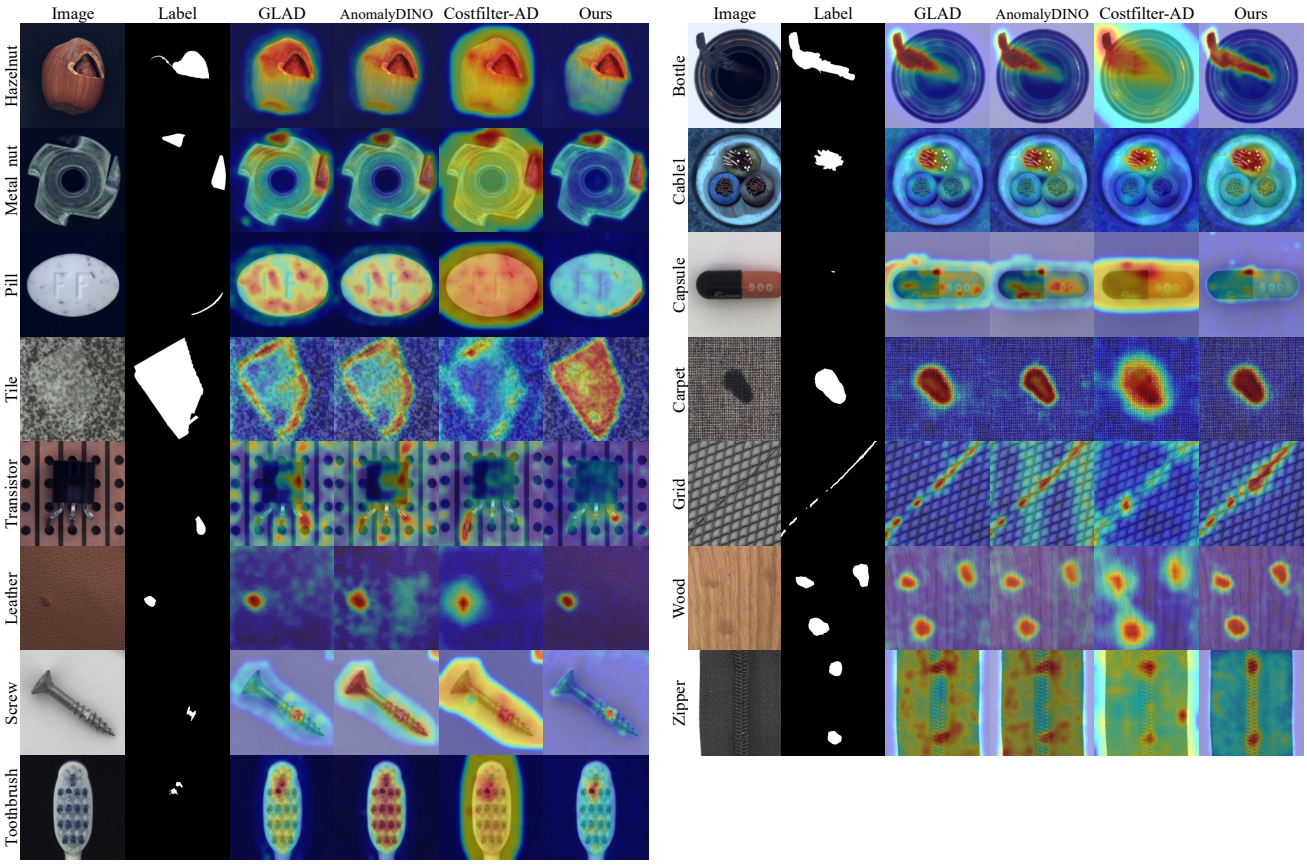

Figure 8. Qualitative comparison of multi-class anomaly localization results on MVTec-AD dataset.

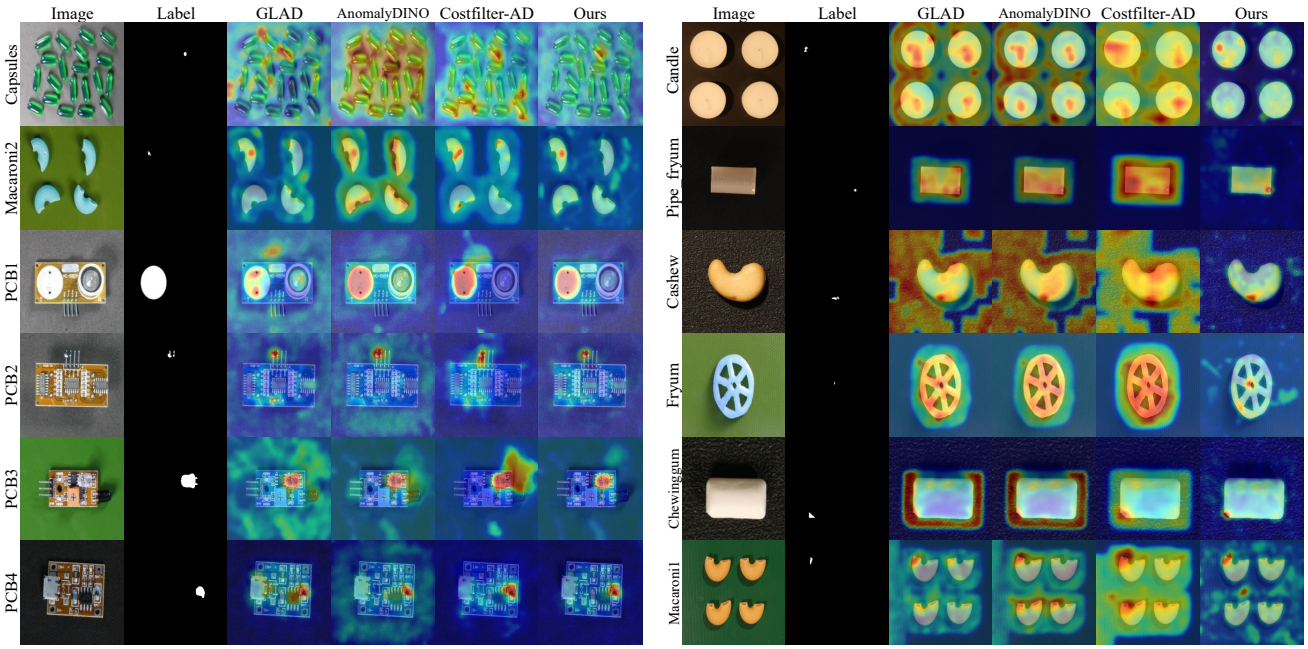

Figure 9. Qualitative comparison of multi-class anomaly localization results on VisA dataset.

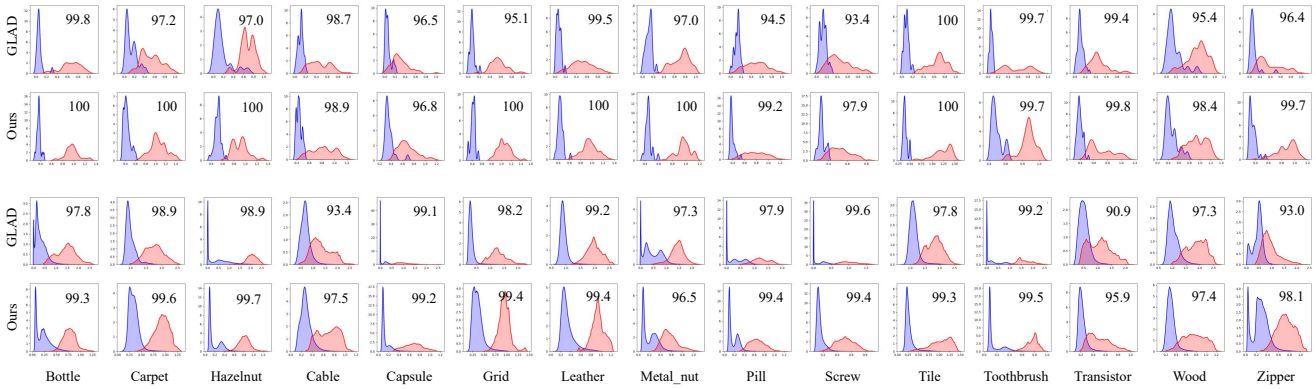

Figure 10. KDE curves of image-level and pixel-level anomaly scores on MVTec-AD. The first two rows correspond to image-level metrics, while the last two rows correspond to pixel-level metrics.

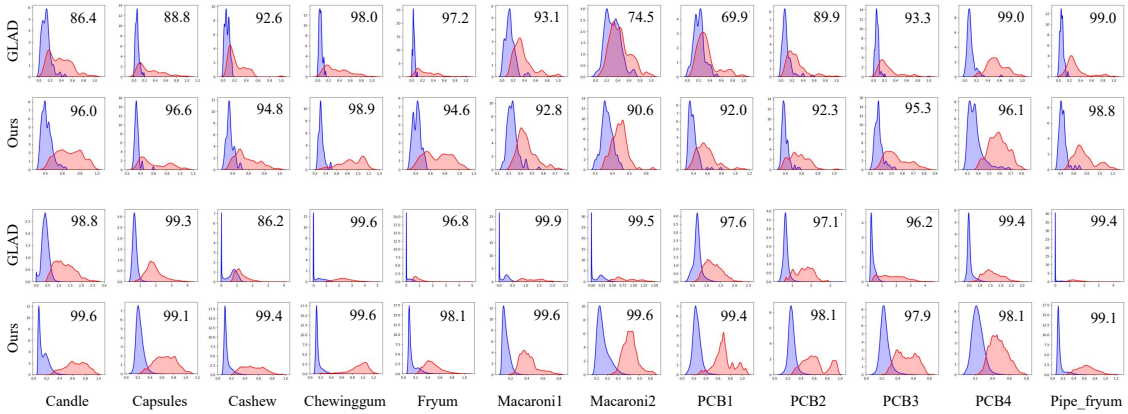

Figure 11. KDE curves of image-level and pixel-level anomaly scores on VisA. The first two rows correspond to image-level metrics, while the last two rows correspond to pixel-level metrics.

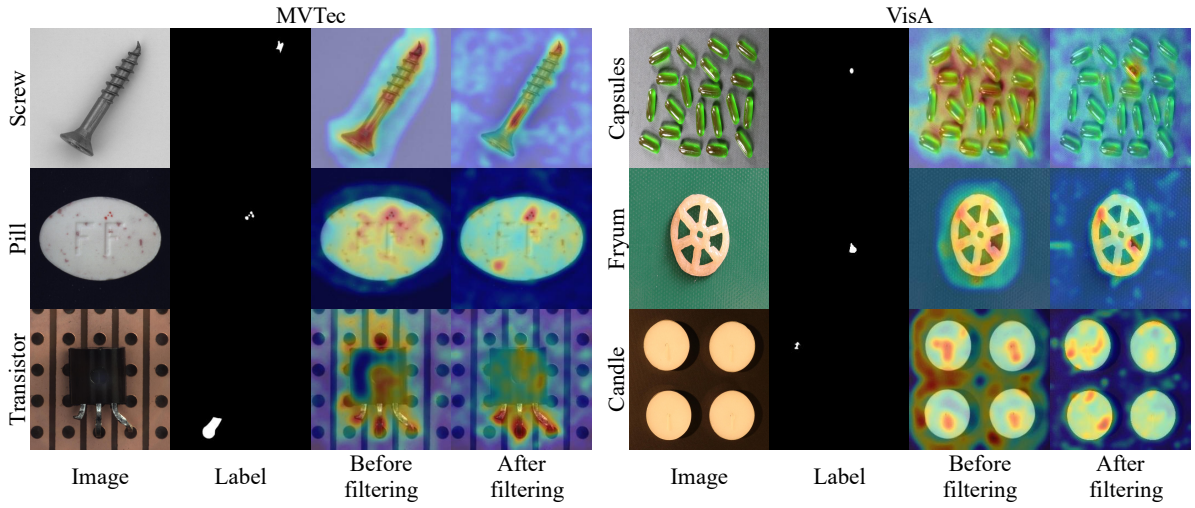

Figure 12. Failure cases on MVTec-AD and VisA. Extremely subtle or low-contrast anomalies may still cause missed or imprecise localization, likely due to insufficient discriminative cues encoded in the cost volume.