CostFilter-AD: Enhancing Anomaly Detection through Matching Cost Filtering

Zhe Zhang, Mingxiu Cai, Hanxiao Wang, Gaochang Wu, Tianyou Chai, Xiatian Zhu
Correspondence to: Gaochang Wu <wugc@mail.neu.edu.cn>, Xiatian Zhu <Eddy.zhuxt@gmail.com>

Background and Motivation

Unsupervised Anomaly Detection (UAD)

🔍 UAD is widely used in industrial inspection, where only normal data is available for training due to the scarcity of anomalies.

🔍 While most methods follow the **"single-model-per-category"** design, **unified multi-class UAD** is emerging as a promising direction.

🔍 Diverse inter-class and subtle intra-class anomalies pose major challenges.

Matching Noise: Ubiquitous Yet Overlooked

❗ UAD relies on image- or feature-level matching, a process inherent to both reconstruction- and embedding-based methods.

❗ Such matching noise impairs the localization of subtle or boundary-adjacent anomalies.

❗ We address this via **cost volume filtering**, inspired by concepts in stereo and flow tasks.

Problem Formulation

The task targets image- and pixel-level anomaly detection using only synthesized anomalies, without access to real defects during training.

We reformulate multi-class UAD as a three-step process:

1. **Feature extraction:** from input and template or reconstructed samples.
2. **Anomaly Cost Volume Construction:** modeling spatial anomaly patterns and channel-wise matching similarity.
3. **Cost Volume Filtering:** with dual-stream attention guidance for noise suppression and anomaly refinement.

Challenges

◆ **Matching Noise vs. Fine Anomalies**
Suppressing matching noise while preserving subtle anomaly cues.

◆ **Subtle and Edge-bound Defects**
Low-contrast or boundary-adjacent anomalies are easily confused with normal regions.

◆ **Identical Shortcut in Reconstruction-based or Embedding-based methods**
The "identical shortcut" effect always replicates anomalies, hindering residual-based detection.

◆ **Category-wise Anomaly Diversity**
Multi-class UAD must handle varying anomaly types across categories, increasing the complexity.

Our contribution

💡 **New Unsupervised Anomaly Detection Formulation**
We reinterpret anomaly detection as a cost filtering process to explicitly address matching noise.

🧩 **CostFilter-AD Method**
A plug-and-play filtering network guided by attention to refine cost volumes and suppress noise.

🧩 **Broad Compatibility**
Our method integrates seamlessly with both reconstruction- and embedding-based models.

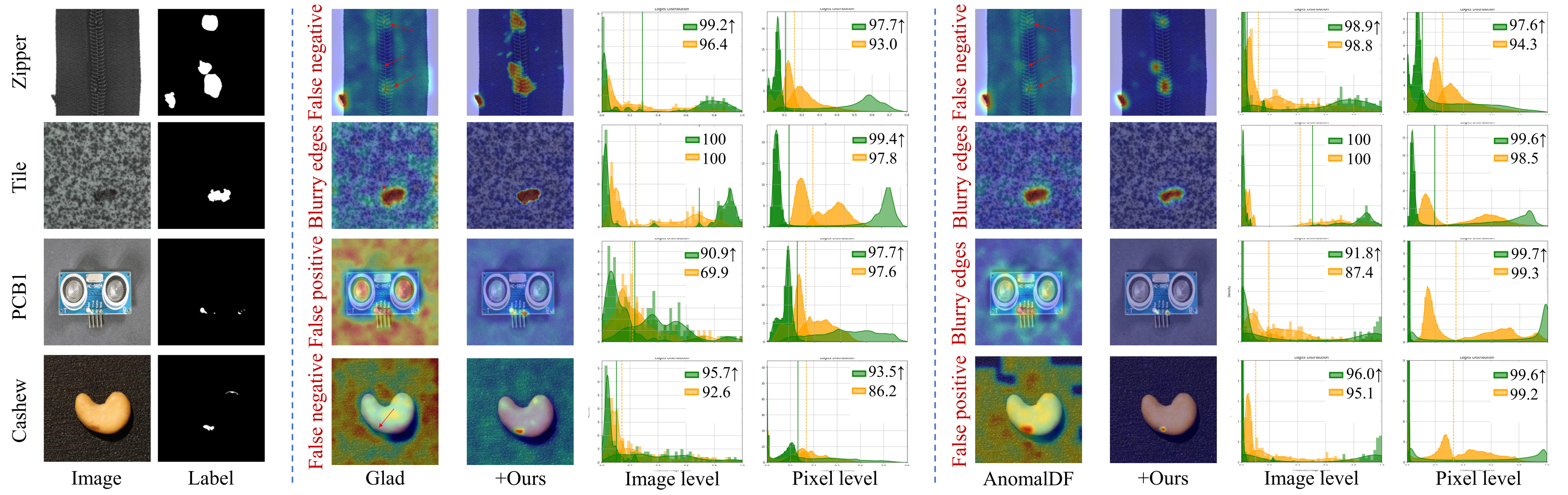
🏆 **Strong Performance Gain**
We enhance 5 baselines across 7 metrics and achieve state-of-the-art results on 4 popular datasets.

Efficiency of Our Plug-and-Play Method

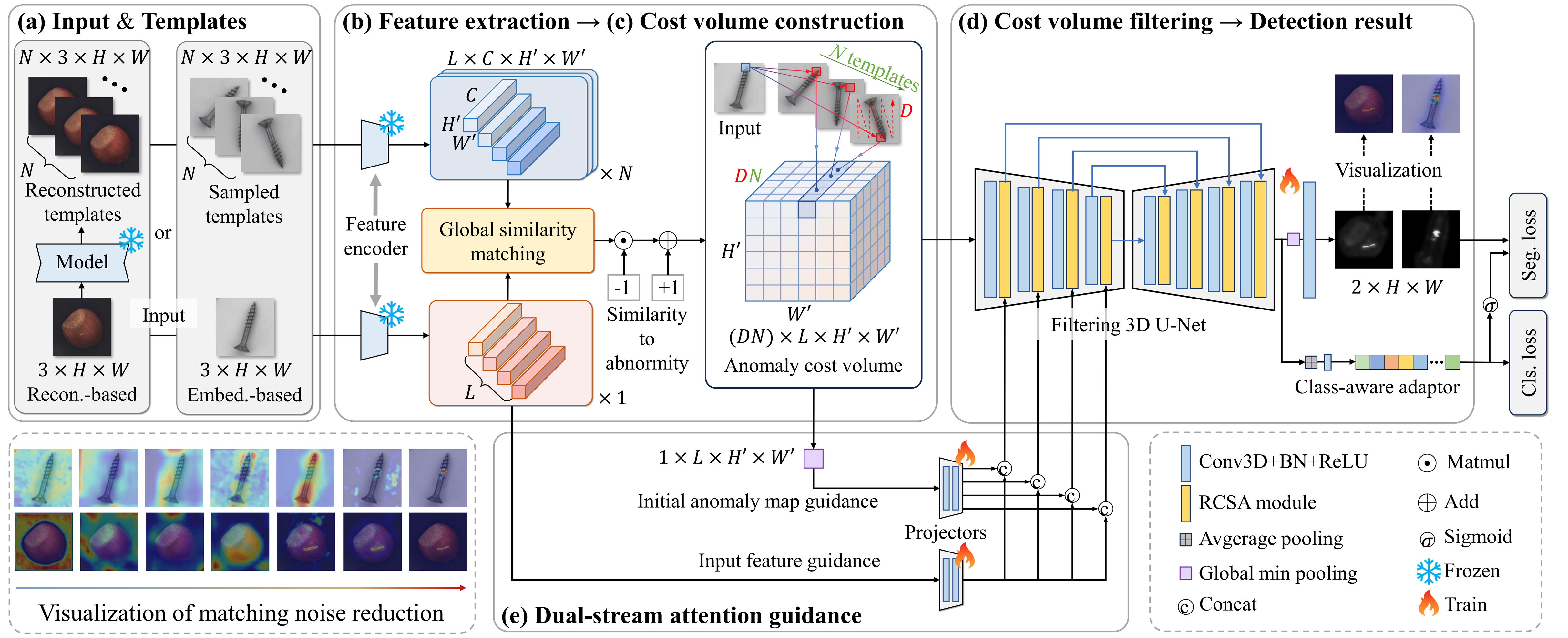
Table 6. Computational efficiency of baselines vs. + Ours.

Method	#Params	FLOPs	Mem. (GB)	Inf. (s/image)
UniAD / +Ours	7.7M / +43.0M	198.0G / 207.8G	4.53 / +0.56	0.01 / +0.04
Glad/+Ours	1.3B / +43.8M	>2.2T / 261.3G	8.79 / +2.07	3.96 / +0.37
HVQ-Trans/+Ours	18.0M / +43.0M	7.4G / 207.8G	4.78 / +0.94	0.05 / +0.07
AnomalDF/+Ours	21.0M / +43.8M	4.9G / 261.3G	3.25 / +0.82	0.31 / +0.32
Dinomaly/+Ours	132.8M / +43.6M	104.7G / 114.6G	4.32 / +1.11	0.11 / +0.05

From Heatmaps to Histograms: Revealing Ubiquitous Matching Noise



From Matching Noise to Filtering: Enhancing UAD via Cost Volume Denoising



Plug-and-Play Boosting of Multi-class UAD on Various Datasets and Metrics

Table 1. Multi-class UAD evaluation on MVTec-AD and MPDD, reporting category-wise mean results for each benchmark.

Benchmark	Method	Image-level AU-ROC	Image-level AP	Image-level F1max	Pixel-level AU-ROC	Pixel-level AP	Pixel-level F1max	Pixel-level AUPRO
MVTec-AD	UniAD (NeurIPS'22)	97.5	99.1	97.0	96.9	44.5	50.5	90.6
	UniAD+Ours	**99.0**	**99.7**	**98.1**	**97.5**	**60.5**	**59.9**	**91.3**
	HVQ-Trans (NeurIPS'23)	97.9	99.3	97.4	97.4	49.4	54.3	91.5
	HVQ-Tran+Ours	**99**	**99.7**	**98.6**	**97.9**	**58.1**	**61.2**	**93.2**
	Glad (ECCV'24)	97.5	98.8	96.8	97.3	58.8	59.7	92.8
	Glad+Ours	**98.7**	**99.6**	**97.8**	**98.2**	**66.8**	**64.4**	**94.1**
	AnomalDF (WACV'25)	96.8	98.6	97.1	98.1	61.3	60.8	93.6
	AnomalDF+Ours	**98.5**	**99.4**	**97.8**	**98.8**	**67.8**	**64.9**	**94.1**
	Dinomaly (CVPR'25)	99.6	99.8	99.0	98.3	68.7	68.7	94.6
	Dinomaly+Ours	**99.7**	99.8	**99.1**	**98.4**	**68.9**	**68.9**	**94.8**
MPDD	HVQ-Trans (NeurIPS'23)	86.5	87.9	85.6	96.9	26.4	30.5	88.0
	HVQ-Tran+Ours	**93.1**	**95.4**	**90.3**	**97.5**	**34.1**	**37.0**	82.9
	Dinomaly (CVPR'25)	97.3	98.5	95.6	99.1	60.0	59.8	96.7
	Dinomaly+Ours	**97.5**	98.5	**95.8**	**99.2**	**60.2**	**59.9**	96.7

Table 2. Multi-class UAD evaluation on VisA and BTAD, reporting category-wise mean results for each benchmark.

Benchmark	Method	Image-level AU-ROC	Image-level AP	Image-level F1max	Pixel-level AU-ROC	Pixel-level AP	Pixel-level F1max	Pixel-level AUPRO
VisA	UniAD (NeurIPS'22)	91.5	93.6	88.5	98.0	32.7	38.4	76.1
	UniAD+Ours	**92.1**	**94.0**	**88.9**	**98.6**	**34.0**	**39.0**	**86.4**
	HVQ-Trans (NeurIPS'23)	91.5	93.4	88.1	98.5	35.5	39.6	86.4
	HVQ-Tran+Ours	**93.4**	**95.2**	**89.3**	**98.6**	**41.4**	**45.0**	**86.8**
	Glad (ECCV'24)	90.1	91.4	86.7	97.4	33.9	39.4	91.5
	Glad+Ours	**93.2**	**94.1**	**89.2**	**98.1**	**40.7**	**43.7**	91.5
	AnomalDF (WACV'25)	90.5	91.4	86.2	97.4	39.6	40.4	86.3
	AnomalDF+Ours	**94.3**	**95.1**	**90.6**	**99.2**	**44.6**	**45.5**	86.3
	Dinomaly (CVPR'25)	98.7	98.9	96.1	98.7	52.5	55.4	94.5
	Dinomaly+Ours	98.7	**99.0**	**96.3**	**98.8**	**53.2**	**55.8**	**94.7**
BTAD	HVQ-Trans (NeurIPS'23)	90.9	97.8	94.8	96.7	43.2	48.7	75.6
	HVQ-Tran+Ours	**93.3**	**98.6**	**96.0**	**97.3**	**47.0**	**50.2**	**76.2**
	Dinomaly (CVPR'25)	95.4	98.5	95.5	97.9	70.1	68.0	76.5
	Dinomaly+Ours	**95.5**	**98.6**	**95.8**	**98.1**	**74.3**	**69.8**	**77.5**

Ablation Studies and Further Analysis

Table 3. Ablation studies of Glad+Ours on MVTec-AD. "$DN\rightarrow$ depth/channel" refers to mapping the matching dimension into the depth/channel dimension of the 3D U-Net. $\mathcal{C}_0$ denotes the volume uisng the final denoising step, $\mathcal{C}_{N-1}$ indicates uisng $N-1$ intermediate steps. SG and MG denote dual-stream attention guidance. $\mathcal{L}_{\mathrm{F}}$ is focal loss, $\mathcal{L}_{\mathrm{CE}}$ corresponds to the class-aware adaptor, and $\mathcal{L}_{\mathrm{S}}$ is the combination of $\mathcal{L}_{\mathrm{SSIM}}$ and $\mathcal{L}_{\mathrm{Soft\text{-}Iou}}$.

$DN\rightarrow$ depth	$DN\rightarrow$ channel $\mathcal{C}_0$	$DN\rightarrow$ channel $\mathcal{C}_{N-1}$	$DN\rightarrow$ channel SG	$DN\rightarrow$ channel MG	$\mathcal{L}_{\mathrm{F}}$	$\mathcal{L}_{\mathrm{CE}}$	$\mathcal{L}_{\mathrm{S}}$	Results
✓	-	-	-	-	✓	-	-	87.8/89.0
-	✓	-	-	-	✓	-	-	96.2/96.8
-	✓	✓	-	-	✓	-	-	96.7/97.3
-	✓	✓	✓	-	✓	-	-	97.8/97.5
-	✓	✓	✓	✓	✓	-	-	98.3/97.8
-	✓	✓	✓	✓	✓	✓	-	98.5/98.0
-	✓	-	✓	✓	✓	✓	✓	98.4/97.6
-	✓	✓	✓	✓	✓	✓	✓	**98.7/98.2**

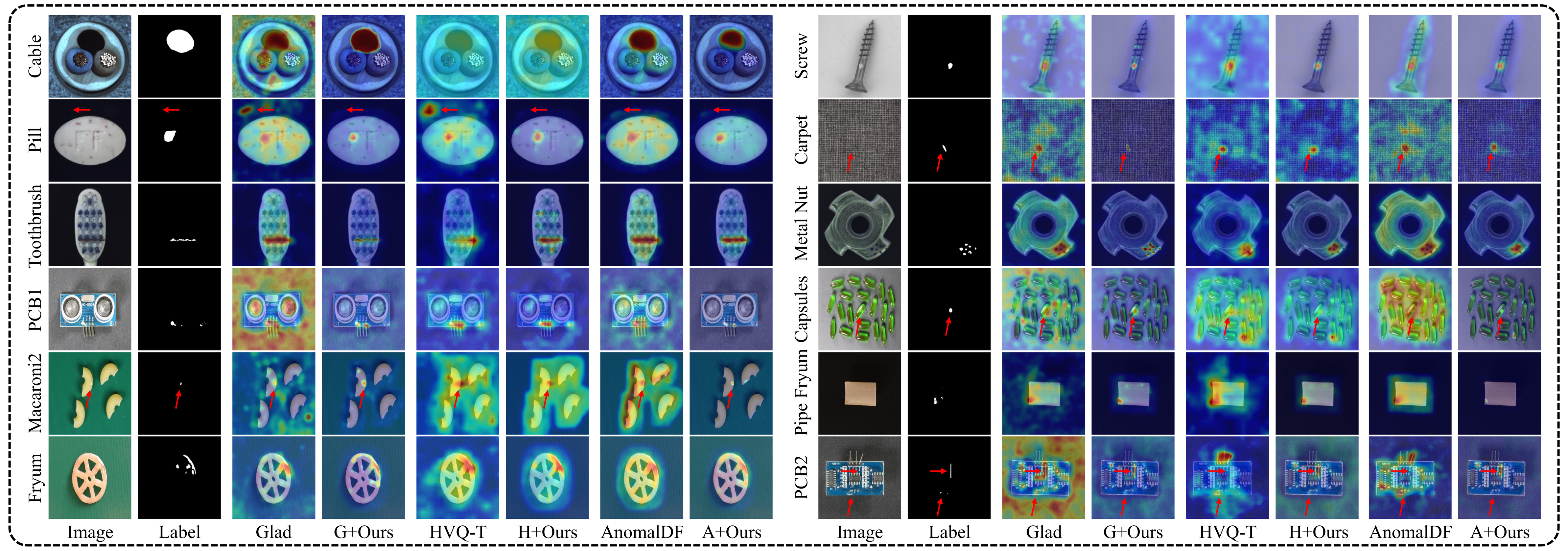
Table 4. Extended studies on single-class UAD with our models.

Benchmark	Method	Image-level AU-ROC	Image-level AP	Image-level F1max	Pixel-level AU-ROC	Pixel-level AP	Pixel-level F1max	Pixel-level AUPRO
MVTec-AD	Glad	99.0	99.7	98.2	98.7	63.8	63.7	95.2
	+Ours	**99.3**	99.7	**98.3**	**98.9**	**66.2**	**65.0**	**96.4**
VisA	Glad	99.3	99.6	97.6	98.3	35.8	42.4	94.1
	+Ours	**99.5**	**99.7**	**98.1**	**98.6**	**37.3**	**45.3**	**94.5**

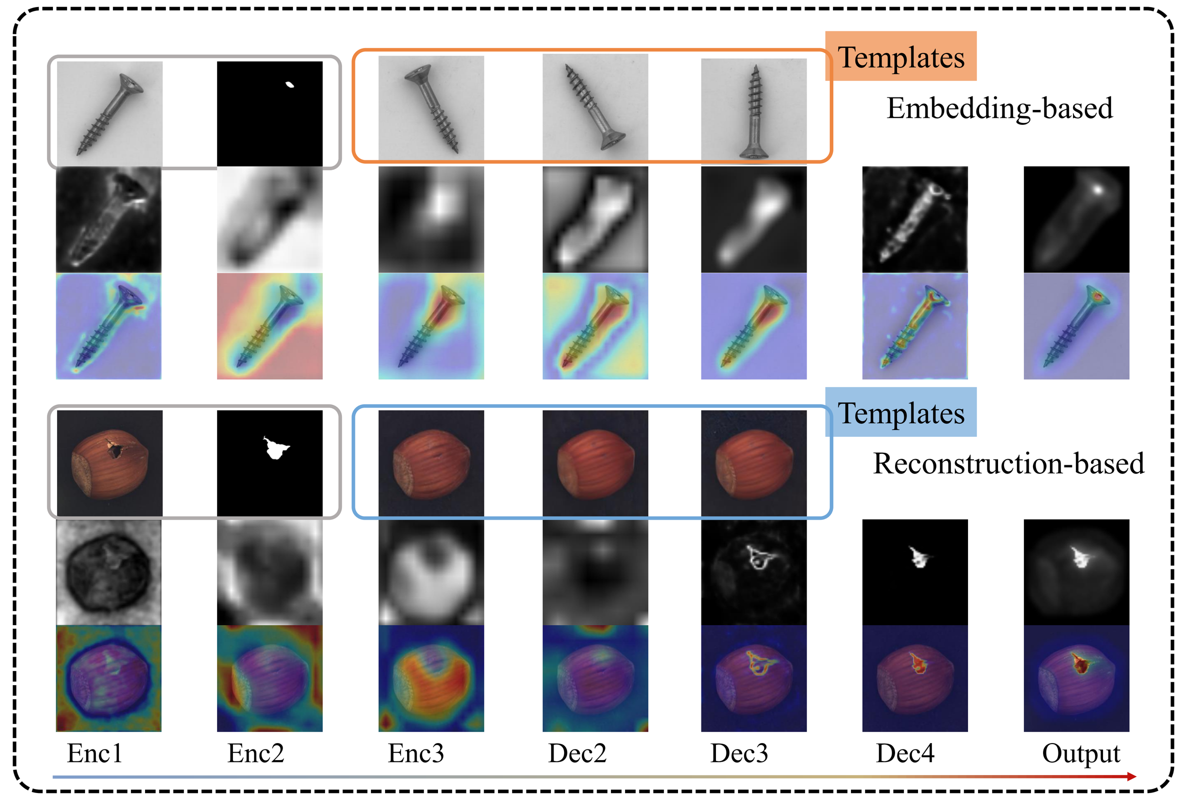
Table 5. Evaluation of our models on various anomaly volumes.

Train \ Test	MVTec-AD Recon.	MVTec-AD Embed.	VisA Recon.	VisA Embed.
Recon.	98.7 / **98.2**	97.5↓ / 97.1↓	**93.2** / 98.1	92.6↓ / 98.0↓
Embed.	94.5↓ / 98.0↓	98.5 / 98.8	85.6↓ / 96.9↓	**94.3** / 99.2
Hybrid	**98.8**↑ / 98.1	**98.6**↑ / **98.9**↑	93.1 / **98.2**↑	92.9 / **99.3**↑

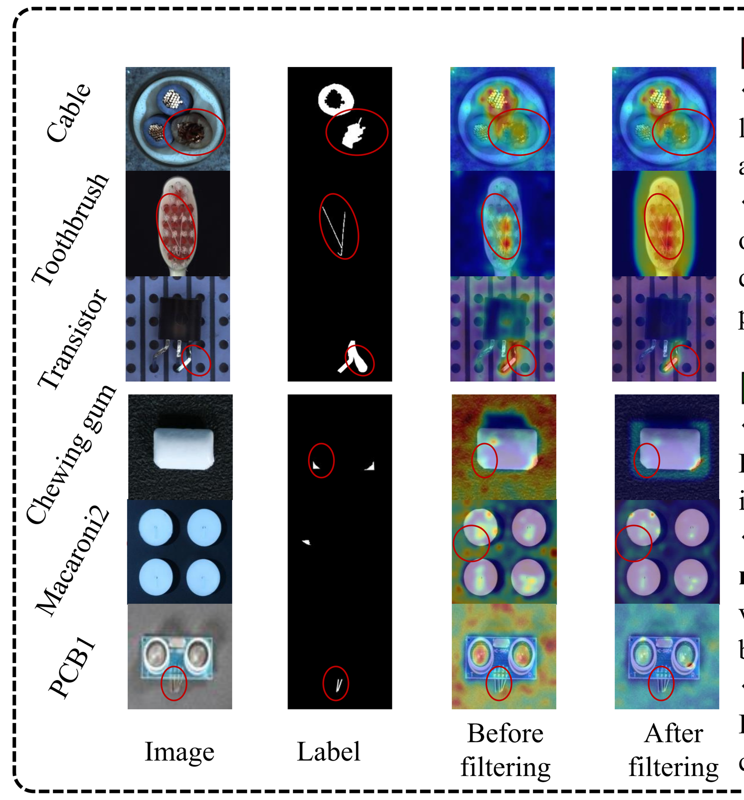
Ours vs. GLAD, HVQ-Trans, and AnomalDF: Localization Visualization



Progressive and Fine-grained Denoising



Failure Cases and Future Direction



Failure Cases

◆ **Subtle Anomalies:** Fails on low-contrast or highly localized anomalies unseen during training.

◆ **Template Sensitivity:** Relies on representative templates; poor quality can degrade detection performance.

Future Directions

◆ **Adaptive Cost Modeling:** Refine matching precision through improved or learned cost functions.

◆ **Spatiotemporal & Multi-modal Extension:** Extend to video or multi-modal inputs for broader applications.

◆ **Hard Negative Mining:** Incorporate challenging normal cases to enhance model robustness.