Review

A Survey of Recent Advances in Industrial Anomaly Detection: From Normal-Only Training to Foundation-Model Priors

Xi Jiang *, Bingzhang Hu , Feng Zheng *

Posted Date: 9 June 2026

doi: 10.20944/preprints202606.0719.v1

Keywords: industrial anomaly detection; foundation models; vision-language models; zero-shot anomaly detection; anomaly synthesis; benchmark evaluation

 doi:10.20944/preprints202606.0719.v1

Review

A Survey of Recent Advances in Industrial Anomaly Detection: From Normal-Only Training to Foundation-Model Priors

Xi Jiang [1], Bingzhang Hu [2] and Feng Zheng [1,*]

[1] Department of Computer Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen, China
[2] Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China
* Correspondence: jiangx2020@mail.sustech.edu.cn

Abstract

Industrial anomaly detection, the computer-vision core of automated quality inspection, had by 2023 settled into a stable method taxonomy organized around a single assumption: a separate detector trained per product on defect-free images alone, evaluated on closed-set benchmarks. The 2024–2026 wave breaks this assumption. Web-pretrained foundation models (vision encoders, multimodal large language models, and diffusion models) generalize across product categories without per-product retraining, recasting anomaly detection from per-category density estimation into a *foundation-model-prior* problem: detection from frozen pretrained features, defect reasoning in natural language, mixture-of-experts models shared across modalities, and language-controlled synthesis of the missing abnormal data. Because the reconstruction, embedding, distillation, and memory taxonomy of earlier surveys no longer absorbs this literature, we reorganize it around one question (*what replaces category-specific normal-only training?*) into five emerging generalization mechanisms: visual, reasoning, geometric and multimodal, universal, and synthesis priors. We then audit the field's benchmark and evaluation frontier, catalog eight cross-cutting bottlenecks that no current method resolves at once, and propose a five-problem research agenda. We conclude that, despite real progress on all five mechanisms, current methods still generalize within curated benchmarks more than under real-world deployment; closing that gap is the agenda's central aim.

Keywords: industrial anomaly detection; foundation models; vision-language models; zero-shot anomaly detection; anomaly synthesis; benchmark evaluation

1. Introduction

Industrial anomaly detection (IAD) accumulated a clean method taxonomy through 2023—reconstruction, embedding, distillation, memory-bank, one-class classification—anchored to MVTec-AD [1] and unified by the *category-specific normal-only* training assumption: a per-product detector trained on defect-free samples (history covered in [2–4]). The 2024–2026 wave—hereafter *the wave*—breaks that assumption via foundation-model priors: frozen CLIP/DINOv2/SAM encoders [5–7] for zero/few-shot detection; MLLMs post-trained with RL for defect reasoning; mixture-of-experts across modalities; language-controlled synthesis; and a new benchmark cohort [8–11] targeting specific limitations of the MVTec-AD-era protocol. The shift first took shape in 2024 and research turned to it decisively across 2025–2026, the cohort this survey centers on (Figure 1).

 doi:10.20944/preprints202606.0719.v1

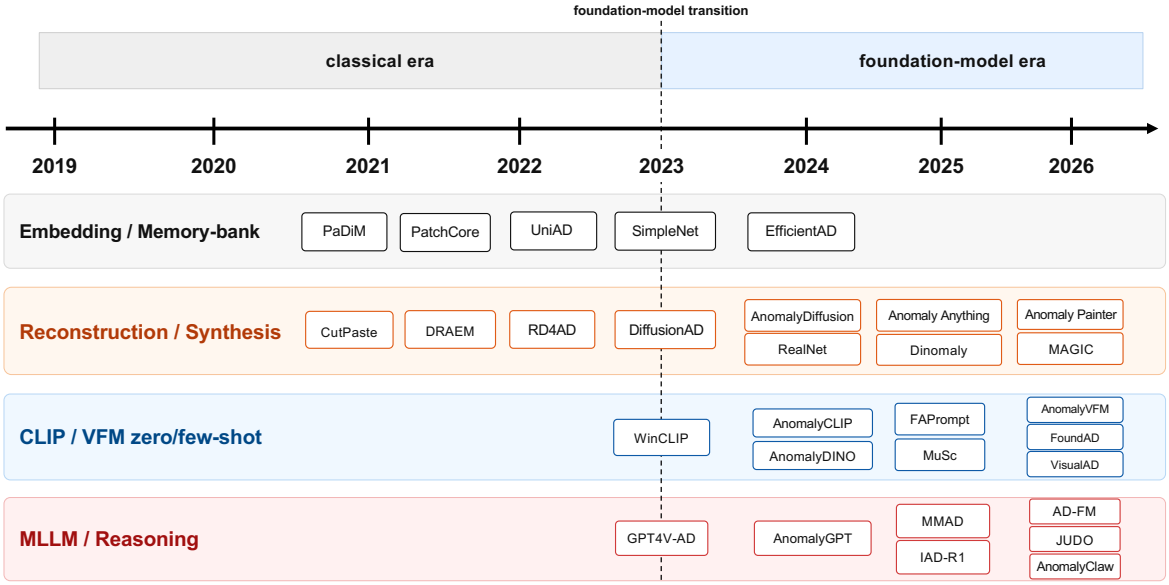

Figure 1. Timeline of IAD method families, 2019–2026. The 2023 line marks the foundation-model transition; recent activity concentrates in the CLIP/VFM, MLLM, and synthesis lanes, while the embedding/memory-bank line slows.

It is natural to read this as a single paradigm shift; we argue more cautiously. The pivot is not one shift but *five concurrent generalization mechanisms*, each replacing a different limitation of the closed-set, normal-only baseline and each short of deployment validation on the harder benchmarks the wave itself produced. A single question motivates this survey: *what replaces category-specific normal-only training?* We give five answers, one per mechanism, developed in the chapters that follow.

1.1. What Is Industrial Anomaly Detection?

Industrial anomaly detection is the computer-vision task at the heart of automated industrial quality inspection: on a production line, it must decide whether a manufactured item is defective and localize the defect, be it a surface scratch, dent, contamination, missing or misplaced part, or structural or logical irregularity. Because defects are rare, diverse, and costly to collect at scale, the dominant formulation learns the appearance of normal (defect-free) products and flags deviations from it. Formally, given normal training images $\mathcal{X}_{\text{train}} = \{x_1, \ldots, x_N\}$ (and, in supervised variants, a small labeled anomalous subset $\mathcal{X}^{-}_{\text{train}}$), an IAD method produces for each test image x_t an image-level anomaly score $s(x_t) \in \mathbb{R}$ and a pixel-level anomaly map $M(x_t) \in \mathbb{R}^{H \times W}$, thresholded against deployment operating points. The task admits three *supervision settings*—unsupervised / normal-only (PatchCore [12]), one-class density estimation, and supervised with a small labeled abnormal subset [13]—and four *shot regimes*: full-shot ($\gtrsim$100 normals/category, the MVTec-AD default), few-shot (1–20 refs), one-shot, and zero-shot (no per-category training, WinCLIP [14]). The wave shifts toward few- and zero-shot.

Methods are evaluated by four core metrics: *I-AUROC* (image-level binary detection); *P-AUROC* and *PRO* [15] (pixel-level localisation); and *bound-FPR metrics* (AP at FPR$\leq$1%, recall at fixed precision) for deployment-faithful operating points where false positives drive production stoppages [16]. Three cross-cutting axes shape methods and benchmarks: single- vs. multi-class, closed-set vs. open-world, and single-modality vs. multimodal (RGB / point cloud / depth / multi-view / text). The wave populates every position on each axis.

1.2. The Five Generalization Mechanisms

We organize the wave around *what source of generalization information replaces normal-only per-category training*, yielding five mechanisms: **M1** pretrained visual priors (Section 3), **M2** language and reasoning priors (Section 4), **M3** geometric and multimodal priors (Section 5), **M4** universal task priors

 doi:10.20944/preprints202606.0719.v1

(Section 6), and **M5** generated abnormal priors (Section 7). Each replaces a different limitation of the closed-set, normal-only baseline (anchors and predecessors mapped in Table 2).

The mechanisms are not orthogonal and do not live at the same conceptual level: M1–M3 operate primarily at the *representation level* (visual, linguistic, geometric features pretrained outside IAD), M4 at the *task / formulation level* (one detector spanning categories, modalities or datasets, learned within IAD), and M5 at the *data / prior-construction level* (synthesized abnormal samples). The partition is best read as an *analytic lens*, not a clean taxonomy—cross-cutting hybrids are common throughout (e.g., MoECLIP combines M1+M4, BTP/GS-CLIP combine M3+M1, AnomalyPainter/Anomagic combine M5+M2), and each chapter flags such cross-references inline.

1.3. Why This Article

The prior surveys and benchmark studies [2–4,17–19] remain the right reference for the era they organise, and we do not aim to replace them. The case for a new review is structural: the normal-only per-category assumption that gave the closed-set taxonomy its coherence has been broken across the entire wave (Section 2), so the new methods no longer fit as a "next chapter" of the same taxonomy. Organising them requires a different question—*what source of generalisation replaces category-specific normal-only training?*—and a different mechanism partition (Section 1.2); that reorganisation is the contribution of this article. We review the foundation-model-driven slice; medical-imaging AD [20,21] appears only as cross-domain comparison (Section 9.8), and a comprehensive medical-AD review belongs to a different article.

1.4. Comparison With Prior IAD Surveys

Table 1 maps coverage of this review against seven prior IAD surveys spanning 2023–2025 along two complementary sets of axes: (a) the topical axes around which the wave has organized itself (columns 1–7), and (b) two reciprocal axes where prior surveys have genuine advantages over this review (columns 8–9). The three-symbol coverage key is defined in the table caption.

Most prior surveys are deep on their declared topic and shallow on the rest—expected for specialist surveys—with one exception: Lin et al. [17] is broad across detection paradigms (2D, 3D, multimodal, FM, MLLM) and the closest in axis-coverage to this article. The difference is positioning: that survey organizes by *modality* (RGB / 3D / multimodal as top-level chapters), whereas the present article organizes by *generalization mechanism* and reserves open-world benchmarking (column 6) and evaluation audit (column 7) as first-class axes—columns where every prior survey, *including* Lin et al. [17], is weak or absent. The reciprocal-axis columns (earlier-era method depth, cross-domain/medical) document where prior work remains stronger; readers needing those should continue with [2,17].

Table 1. Coverage of this review against seven prior IAD surveys/benchmark studies, along the topical axes of the wave (cols 1–7) and two reciprocal axes where prior work retains advantages (cols 8–9). Markings are from a section-level audit (table of contents plus chapter-level reading). ● = deep coverage (≥1 dedicated chapter or major subsection, ≥5% of body, comparing ≥3 methods/datasets); ○ = brief coverage (a paragraph or table entry naming representative methods, no chapter-level treatment);—= not covered (absent or a passing citation). † IM-IAD is a benchmark/protocol study, included for completeness.

		Topical axes of the wave							Reciprocal axes	
Prior work	Year	Closed-set 2D	FM era (CLIP/VFM)	MLLM-AD	3D / multimodal	Synthesis	Open-world bench	Eval audit	Earlier-era method depth	Cross-domain (med./aerial)
Liu et al. [2]	2024	●	○	—	○	○	—	●	●	—
Xie et al. (IM-IAD)† [3]	2024	●	○	○	—	○	—	●	●	—
Cao et al. [4]	2024	●	●	○	●	○	—	○	●	—
Liang et al. [18]	2025	—	○	○	●	○	—	—	○	—
Lin et al. [17]	2025	●	●	●	●	●	—	●	●	—
Wang et al. (IAS) [19]	2025	—	—	○	—	●	—	—	—	—
Cheng et al. (RWIDD) [22]	2026	●	○	○	●	○	○	—	●	—
This review	**2026**	○	●	●	●	●	●	●	○	—

 doi:10.20944/preprints202606.0719.v1

1.5. Scope And Corpus

We survey the 2024–2026 IAD literature across three cohorts: (i) top-venue IAD method papers from the 2025–2026 cycle (CVPR/ICLR/AAAI 2026, NeurIPS/ICCV/ICML/KDD/ACM MM 2025, CVPR 2025 retained as predecessor), industrial as primary domain and advancing detection methodology rather than reporting a dataset alone; (ii) benchmark/dataset releases at top venues/journals 2024–2026, enumerated in Table 5; (iii) a complementary selection of 2026 preprints. We date works by publication (venue) year, so a 2024 method published in 2025 falls within the 2024–2026 wave. **Out of scope:** video AD as primary, network-traffic AD, tabular AD, and medical-imaging AD (touched only as cross-domain comparison in Section 9.8). The full paper list and updates are maintained in the companion repository[1]; conflict-of-interest disclosures are itemized in the Disclosures at the end of the paper.

1.6. Roadmap

Table 2 summarizes the article's organization: each chapter from Section 2 to Section 10 is paired with the generalization mechanism or audit theme it covers, its anchor methods, the closed-set, normal-only baseline it replaces, and the central tension on which the chapter's honest-assessment subsection turns. Read column-wise, the table is a quick-reference map; read row-wise, it tracks one chapter's thesis from problem to open question. The remainder of the article elaborates each row in turn.

Table 2. Section overview of this review. "Anchor methods" lists representative entries from the wave; "replaces what" identifies the earlier-era baseline the chapter is positioned against; "central tension" is the open question the chapter's honest-assessment subsection foregrounds.

Section	Mechanism / topic	Anchor methods	Replaces what	Central tension
Section 2	Established baseline	PatchCore, EfficientAD, DRAEM	—	MVTec-AD saturated
Section 3	M1 Visual priors	CLIP, DINOv2, AnomalyVFM, MoECLIP	WinCLIP CLIP-adapter saturation	Saturation band crossed only by architectural change
Section 4	M2 Reasoning priors	IAD-R1, JUDO, AD-FM, EAGLE	AnomalyGPT SFT-only template	MMAD gains regress binary AUC at deployment FPR
Section 5	M3 Geometric / multimodal	UniMMAD, GS-CLIP, SiM3D	M3DM/CPMF memory-bank monoculture	MVTec-3D saturated; CAD-to-production gap unresolved
Section 6	M4 Universal / MoE	AnomalyMoE, AdaptCLIP, UniSpector	UniAD per-dataset	Cross-domain transfer wall (UniSpector AP_{50} 69.1%→14.1%)
Section 7	M5 Synthesis priors	ARAS, QARAD, MAGIC, FAST	AnomalyDiffusion per-category	No tripartite synthesis benchmark adopted
Section 8	Evaluation frontier	Kaputt, MMR-AD, ASBench	MVTec/VisA single-tier evaluation	Tier-1 benchmarks wide open
Section 9	Cross-cutting bottlenecks	—(8 gaps catalogued)	—	No paper addresses >1 bottleneck
Section 10	Agenda + reforms	—	—	Evidence needed for strong-version paradigm shift

We treat the five-mechanism partition as a summary of the current literature, not a deep-structural claim: the wave contains real method-level advances along all five axes but not yet the deployment-level evidence required for a strong-version paradigm shift; whether the next iteration uses softer or stronger framing is empirical, answered by whether the five-problem agenda (10.1) closes.

[1] https://github.com/M-3LAB/awesome-industrial-anomaly-detection

 doi:10.20944/preprints202606.0719.v1

2. The Established Baseline Being Replaced

Each generalization mechanism in Section 3–7 is a response to a specific limitation of an established method line. This section gives the minimum baseline; readers are referred to prior surveys [2,4] for the complete pre-wave method taxonomy, to the IM-IAD benchmark study [3] for deployment-realistic protocols, to two recent living surveys covering the same window from different angles [19,22], and to two specialty surveys for 3D and multimodal [17,18]. Their conclusions on pre-wave work are largely consistent with ours, which lets this article concentrate on the subsequent transition. The differentiator here is the *five-mechanism* organization plus the eight-bottleneck cross-cutting matrix (Section 9); we frame this article as a complement to, not a replacement for, those references.

2.1. Established Method Lines and Their Saturation

The closed-set, normal-only era and its saturation.

The closed-set, normal-only paradigm trains one detector per category from normal samples: memory-bank [12,23], distillation [24,25], reconstruction [26–28], with SuperSimpleNet [13] later unifying unsupervised and supervised training under the same backbone. PatchCore reached 99.1% I-AUROC on MVTec-AD [1] in 2022; Section 8.1 shows this number has not been beaten in any apples-to-apples comparison through 2026—the MVTec-AD ceiling sits inside the test-set noise floor, and every subsequent generalization mechanism tries to *escape* this saturated axis rather than climb it.[2]

The first multi-class wave: UniAD and one-model-per-dataset.

UniAD [29] and descendants [30–32] formalized multi-class unsupervised AD: one model across all categories within one dataset. Section 6 extends this line toward genuine cross-dataset transfer.

The first foundation-model bridge: WinCLIP descendants and their saturation.

WinCLIP [14] projected image patches into a frozen CLIP [5] space, scored against handcrafted defect-text prompts, and reached 91.8% MVTec-AD zero-shot I-AUROC. The first foundation-model bridges [33–38] added learnable prompts, visual adapters, and SAM-driven masks—a 0.2–0.5% gain over 18 months: prompt-learning saturation. Section 3 covers the response.

The first MLLM bridge: AnomalyGPT and MMAD.

AnomalyGPT [39] introduced a LLaVA-style 7B VLM fine-tuned on simulated MVTec-style VQA, producing detection plus natural-language description in one forward pass. Successors [40–43] established the SFT-on-defect-prompts pattern. The most consequential contribution of this wave was MMAD [44]: a seven-subtask VQA benchmark of 39,672 questions over 8,366 images from four public IAD datasets (38 product classes) and now the de-facto MLLM-AD leaderboard. Section 4 papers either improve on MMAD [45–47] or position themselves as successor benchmarks [9,48,49].

Multimodal and 3D foundations.

M3DM [50] fused Point-MAE [51] with DINO [52] via a memory bank; CPMF [53] refined this with per-modality scoring; Real3D-AD [54] released the first full-point-cloud benchmark beyond MVTec-3D-AD [55]. PointAD [56] combined ULIP [57]/PointCLIP-v2 with WinCLIP-style zero-shot scoring. Section 5 covers harder 3D benchmarks (MiniShift [58], SiM3D [59]), language-grounded industrial 3D defects (IMDD-1M [10]), and generalist 3D AD with mixture-of-experts (UniMMAD [60]).

[2] **IAD metric vocabulary.** *I-AUROC* (image-level AUROC): whole-image normal/abnormal classification. *P-AUROC* (pixel-level AUROC): same, per pixel; measures localization. *PRO* (Per-Region Overlap, 15): mean predicted/ground-truth region overlap, more deployment-faithful than P-AUROC on small defects. *AP* (Average Precision): area under the precision–recall curve. *Bound-FPR metrics* (e.g., AP at FPR$\leq$1%, recall at fixed precision): performance at manufacturing operating points, where false positives stop production.

 doi:10.20944/preprints202606.0719.v1

Anomaly synthesis foundations.

CutPaste [61], DRAEM [26], AnomalyDiffusion [62], RealNet [63], and AnomalyXFusion [64] defined the pre-wave synthesis pattern: train a category-specific LDM, hand-craft a noise mask, sample ∼1000 DDIM steps, evaluate by FID/KID and downstream detection lift. Section 7 covers the subsequent break-up into language grounding, training-free / few-shot generation, 100× speedup [65], joint image+mask outputs [65,66], and architecture diversification beyond LDM UNet.

2.2. The Established Benchmark Ecosystem

The established benchmark set the reader needs to interpret subsequent claims:

- **MVTec-AD** [1]—15 categories, saturated above 99% I-AUROC since 2022.
- **VisA** [67]—12 categories, harder per-pixel metrics, still active.
- **MPDD** [68], **BTAD** [69]—smaller niche benchmarks.
- **Real-IAD** [70]—30 categories, ∼150K images, five viewpoints per part.
- **MVTec-3D-AD** [55], **Real3D-AD** [54], **Anomaly-ShapeNet** [71]—3D and RGB-D benchmarks.
- **MMAD** [44]—the MLLM-AD VQA leaderboard.
- **IM-IAD** [16]—the deployment-realistic protocol (FPR-bound recall).

These benchmarks share the assumptions the wave sets out to break—curated category sets, predominantly full-shot normal-only training, lab-controlled imaging, and a single dominant modality—and are being superseded for two reasons. First, the headline detection sets (MVTec-AD, VisA, MVTec-3D-AD) are *saturated*: top scores now sit inside the test-set noise floor (Section 8.1) and no longer separate competing methods. Second, even where unsaturated, they probe only a narrow slice of deployment reality—none exposes open-world product diversity, pose and identity variation, distribution drift, cross-domain transfer, or false-positive-bounded operating points, the failure modes that actually break deployed inspection (Section 9). The wave therefore forks the ecosystem along precisely these axes; Section 8.2 catalogues the 20+ new benchmarks, each built to expose one such failure mode.

3. Generalization Mechanism 1—Pretrained Visual Priors

The largest cluster of papers in the 2024–2026 wave draws generalization power from pretrained vision encoders—CLIP, DINOv2, DINOv3, RADIO, point-language models—on the premise that pretraining already encodes enough appearance knowledge to separate normal from anomalous patches given the right adaptation. Figure 2 summarises how the field reacts to CLIP-adapter saturation: the 91–93% MVTec-AD I-AUROC band the 2023–2025 adapter generations sit in is crossed only in 2026, and only by methods that change the architecture rather than the adapter (MoECLIP MoE routing, MRAD retrieval memory, CoPS learned subspaces). These responses divide into architectural branches—pure-VFM, retrieval-augmented, and training-free (Section 3.2)—and two cross-modal bridges, to reasoning and to 3D (Section 3.3).

 doi:10.20944/preprints202606.0719.v1

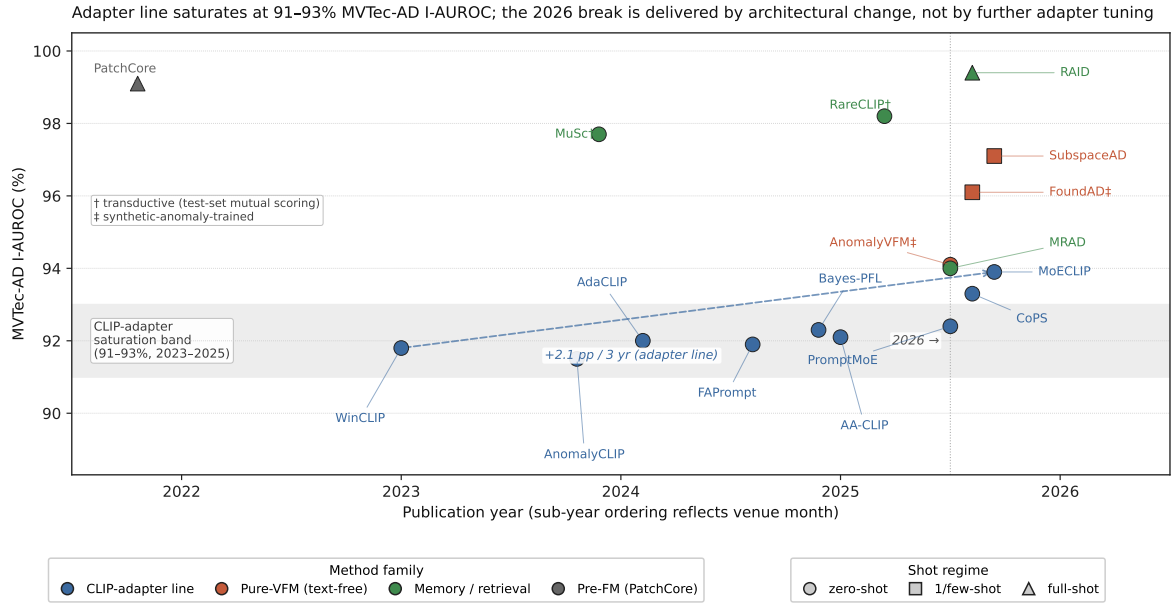

Figure 2. Visual-prior responses to CLIP-adapter saturation on MVTec-AD I-AUROC. Color = mechanism family; marker = shot regime. The 2023–2025 adapter generations sit in the shaded 91–93% band; only the 2026 architectural cluster (MoECLIP, MRAD, CoPS) crosses it. Scores above the band are off-regime: † transductive (MuSc, RareCLIP); ‡ synthetic-anomaly-trained (AnomalyVFM, FoundAD); SubspaceAD/FoundAD use 1–4 references. See Section 3.1.

3.1. The CLIP Prompt-Engineering Line and Its Saturation

WinCLIP [14] established the blueprint—frozen CLIP + handcrafted normal/defect text pairs + sliding-window patch-text similarity at 91.8% MVTec-AD zero-shot I-AUROC. Text-side refinements: AnomalyCLIP [34] learnable object-agnostic prompts (91.5%), PromptAD [33] multi-shot template distillation. Visual-side: AdaCLIP [35] static-plus-dynamic visual context tokens (89.2%), VCP-CLIP [36] nearest-neighbour visual prompts. AdaptCLIP [72] extends the branch to twelve-dataset adapter training (cross-domain behaviour in Section 6).

The wave's papers continue both axes without escaping the band. FAPrompt [73] uses K=10 orthogonally-constrained abnormality prompt sets plus a data-dependent prior (91.9/90.6/83.3% I-AUROC/P-AUROC/PRO). Bayes-PFL [74] swaps deterministic prompt fitting for a Bayesian flow with calibrated uncertainty—useful when bound-FPR metrics depend on threshold sharpness. AA-CLIP [75] adds an anomaly-aware re-projection without extra text supervision. **PromptMoE** [76] routes among eight prompt experts per layer (92.4% mean over seven industrial datasets); **MoECLIP** [77] adds patch-level routing with Frozen Orthogonal Feature Separation and an ETF loss (93.9% MVTec-AD). CoPS [78] samples per-image class/prototype tokens via a VAE (+1.4/+1.9 pp over FAPrompt); FB-CLIP [79] disentangles foreground/background tokens.

The aggregate trajectory—WinCLIP 91.8 → AdaCLIP 89.2 → FAPrompt 91.9 → MoECLIP 93.9—spans only ∼2 pp MVTec-AD I-AUROC over three years, non-monotone and with no generation clearing 94% (Figure 2). The 2026 break comes from architectural change—MoE routing, retrieval memory, learned subspaces—motivating the three branches below.

Regime caveat. The 91–93% band describes *strict zero-shot single-image* inference. Several Section 3.2 numbers above the band are measured under different regimes: MuSc [80] and RareCLIP [81] are transductive; AnomalyVFM [82] and FoundAD [83] train LoRA / a Manifold Projector on synthetic anomaly pairs before their "zero-shot" test phase; SubspaceAD [84] and FoundAD require 1–4 normal references at test. Matched-regime evidence: AnomalyVFM 94.1, MRAD 94.0 vs. MoECLIP 93.9—the visible 4–6 pp gap collapses to ∼0.1–0.2 pp.

 doi:10.20944/preprints202606.0719.v1

3.2. Architectural Responses

The pure-VFM wave (text-free).

Three papers in the wave test whether the visual encoder, not vision-language alignment, is load-bearing. **SubspaceAD** [84]: PCA over DINOv2-G layer-22–28 patches from k=1–4 normals with 30 rotation augmentations, no text branch, no trained params—97.1/93.4% MVTec-AD/VisA 1-shot I-AUROC (+6.0 pp over AnomalyDINO [85]). **FoundAD** [83]: an 11.8M Manifold Projector trained via CutPaste synthesis maps patches back to the normal manifold, one projector for all categories—96.1% MVTec-AD I-AUROC, 99.7% VisA P-AUROC at 97.8M total params vs. 1.3B for prior leaders. **AnomalyVFM** [82]: rank-64 LoRA on RADIOv2.5 ViT-L/16 trained with 10K FLUX/RePaint synthetic pairs—94.1% mean industrial zero-shot (+3.3% over Bayes-PFL) at 20.5 ms/image. **VisualAD** [86] (cross-ref Section 6) drops the text branch with only −0.33% mean cost vs. CLIP methods at 99% parameter reduction. Together they reject text-as-load-bearing for ZSAD. Caveats: SubspaceAD's PCA is optimal only for Gaussian features, and AnomalyVFM's "zero-shot" conflates gains with RADIO strength + synthesis supervision.

Retrieval-augmented and memory-driven.

A second response extends PatchCore with cross-dataset labeled memory and learned cost-volume denoising. **MRAD** [87] stores labeled patch prototypes from auxiliary VisA and scores via soft-max cross-attention; the train-free variant beats WinCLIP, and MRAD-CLIP reaches 94.0% MVTec-AD I-AUROC and 92.7% P-AUROC/PRO across sixteen datasets. **RareCLIP** [81] accumulates cross-image statistics online via Sequential Coreset + Patch-image Similarity banks with k-NN rarity proxies (98.2/94.4% MVTec-AD/VisA I-AUROC at 59.4 ms streaming). **RAID** [88] frames UAD as RAG over a three-level DINOv2-S vector database with MoE cost-volume denoising (99.4/98.6% MVTec-AD, 71.7% pixel-AP vs. AnomalyDINO 61.3%). **FastRef** [89] is a plug-in test-time adaptation converging in two iterations (+2.5/+7.0 pp WinCLIP P-AUROC on MVTec-AD/MPDD 4-shot). The cluster narrows the zero-shot/full-shot gap; what remains is fine-grained localisation, not image-level classification.

Training-free, non-conformity, and geometric variants.

DictAS [90] reformulates few-shot AS as dictionary lookup with self-supervised K/V generators + Sparse Probability Module (98.4% P-AUROC, 92.2% PRO across five industrial benchmarks 4-shot; 98.4% on MVTec-3D without 3D adaptation). Three CVPR 2026 papers instantiate non-conformity scoring in non-Euclidean spaces (graph Laplacian, semantic graviton fields, hyperbolic feature synthesis). **Odd-One-Out** [91] replaces stored-normality matching with nearest-neighbour comparison, sidestepping the per-category training assumption that fails on Kaputt-style open-world deployment. **DPDL** [92] handles the supervised open-set case via distribution-prototype diffusion with calibrated scores. Shared trait: generalization decoupled from trainable parameters.

3.3. Cross-Modal Bridges

LMM-assisted zero-shot segmentation (bridge to Section 4).

AG-VAS [93] is the first 2026 7B-MLLM zero-shot *segmentation* system, combining LLaVA-OneVision-7B, SAM-ViT-H, three learnable anchor tokens ([SEG]/[NOR]/[ANO]), Semantic-Pixel Alignment, and an Anchor-Guided Mask Decoder, to reach MVTec-AD AP 51.0%, IoUano 44.8%, IoUnor 87.7% (vs. Bayes-PFL 29.9%) plus 92.7% AP on ISIC without medical training. Absolute pixel scores trail CLIP adapters, but normal-region reasoning and instruction-following segmentation are qualitatively distinct.

3D-meets-VFM (bridge to Section 5).

Two CVPR 2026 papers extend zero-shot VFMs to 3D. **BTP** [94] uses a ULIP2 Point-Language Model + Multi-Granularity Feature Embedding (PLM + learned geometric + CLS features), reaching 84.5% Real3D-AD point-AUROC (+11.0 pp over PointAD) but only 61.4% O-AUROC—localization gains do not transfer to global detection. **GS-CLIP** [95] keeps 2D projection but adds geometry-aware

 doi:10.20944/preprints202606.0719.v1

prompts (PointNet++ + Geometric Defect Distillation) and a LoRA-adapted depth branch, +1–2 pp O-AUROC across four datasets. Neither closes the 3D–2D zero-shot gap.

3.4. Honest Assessment

(i) Genuine. Across few-/zero-shot regimes text supervision is not load-bearing: SubspaceAD 97.1% MVTec-AD 1-shot I-AUROC with PCA over DINOv2-G (no text, no trained params), AnomalyVFM ties the adapter cluster zero-shot (94.1% vs MoECLIP 93.9%), and AG-VAS shows reasoning-grade segmentation from a 7B model with $\leq$3 anchor tokens.
(ii) Oversold or unresolved. The headline "pure-VFM beats CLIP-adapters by 4–6 pp" is largely a regime-mismatch artifact: once MuSc/RareCLIP are flagged transductive and AnomalyVFM/FoundAD synthetic-trained, the matched-regime gap is $\sim$0.1–0.2 pp. None is evaluated on Kaputt or any Tier-1 open-world benchmark at bound-FPR operating points (Section 8.3); UniSpector's 69.1%$\rightarrow$14.1% cross-domain drop (Section 6.4.0.2) quantifies the kind of degradation to expect.
(iii) Minimum new evidence. A Section 3 method reporting on Kaputt under the $\leq$3-reference protocol *and* a bound-FPR metric (AP at FPR$\leq$1% or F1 at fixed precision) alongside AUROC.

4. Generalization Mechanism 2—Language and Reasoning Priors

The AnomalyGPT–MMAD line (Section 2.1.0.4) established a 2024 template: fine-tune a 7B VLM on supervised industrial VQA, produce anomaly maps plus short descriptions, evaluate on the MMAD seven-subtask benchmark.[3] The 2024–2026 wave identifies three weaknesses in that template—SFT stalls at reasoning-answer inconsistency, models lack inference-time domain knowledge, and standard architectures encode comparison images *independently*—and builds partially overlapping replacement mechanisms (Figure 3).

4.1. Capability Axes: Detection versus Reasoning

Detection means producing a binary label or pixel anomaly score evaluated by AUROC, the axis on which PatchCore [12], EfficientAD [25], and IM-IAD [16] operate. **Reasoning** means producing a structured response covering one or more of the seven MMAD subtasks [44]—discrimination, classification, localization, description, analysis, object classification, object analysis—evaluated by VQA accuracy. The two axes are related but not equivalent, and the wave's papers occupy four positions in this space (Table 3). Provenance disclosures for MMAD and several M-3LAB-adjacent works mentioned below are consolidated in Disclosures.

[3] **MLLM-AD training vocabulary.** *MLLM / VLM*: a large language model with a vision encoder, jointly processing image and text. *VQA*: visual question-answering protocol. *SFT* (supervised fine-tuning): training on (input, gold-answer) pairs with cross-entropy. *GRPO* (group relative policy optimization, from DeepSeek-R1): an RL-style post-training step rewarding correct, well-structured answers over the model's own samples—applied *after* SFT. *RAG* (retrieval-augmented generation): retrieves knowledge into context at inference. *DPO*, *RLVR*: other post-training objectives. *CoT*: chain-of-thought.

 doi:10.20944/preprints202606.0719.v1

Table 3. Recent MLLM-AD methods: base model, training paradigm, MMAD accuracy (1-shot), and binary detection performance. All numbers are as reported in the respective papers; the entries themselves note when a metric is non-standard (relative gain, non-MMAD subset, or qualitative).

Method	Venue	Base model	Paradigm	MMAD (1-shot)	Binary detection
AnomalyGPT	AAAI 2024	LLaVA-7B	SFT	36.52%	∼60% acc.
AnomalyR1	preprint 2025	Qwen2-VL-3B	SFT+GRPO	76.96%	∼70% acc.
IAD-R1	AAAI 2026	LLaVA-OV-7B	PA-SFT+SC-GRPO	—	**86.1% acc.**
JUDO	ICLR 2026	Qwen2.5-VL-7B	SFT×2+GRPO	**81.20%**	65.04%
AD-FM	AAAI 2026	Qwen2.5-VL-7B	GRPO (GIoU)	**83.56%**	73.15%
ADSeeker	CVPR 2026	Qwen2.5-VL-7B	RAG (no RL)	69.90%	∼74%
Reason-IAD	preprint	Qwen3-VL-8B	Training-free	79.43%	—
EAGLE	preprint 2026	PatchCore+MLLM	Tuning-free	not reported	expert-aug.
Triad	ICCV 2025	LLaVA-OV-7B	SFT	—	94.1%
SAGE	ACM MM 2025	InternVL2	SFT+DPO	not reported	MANTA/MPDD
AD-Copilot	preprint	Qwen2.5-VL-7B	4-stage+RLVR	**82.3%**	3.35× BBox
MAU-GPT	AAAI 2026	∼4B LLaVA-style	SFT (AMoE-LoRA)	61.41%	—
ReADL	CVPR 2026	unspecified	GRPO+CGRO	—	competitive

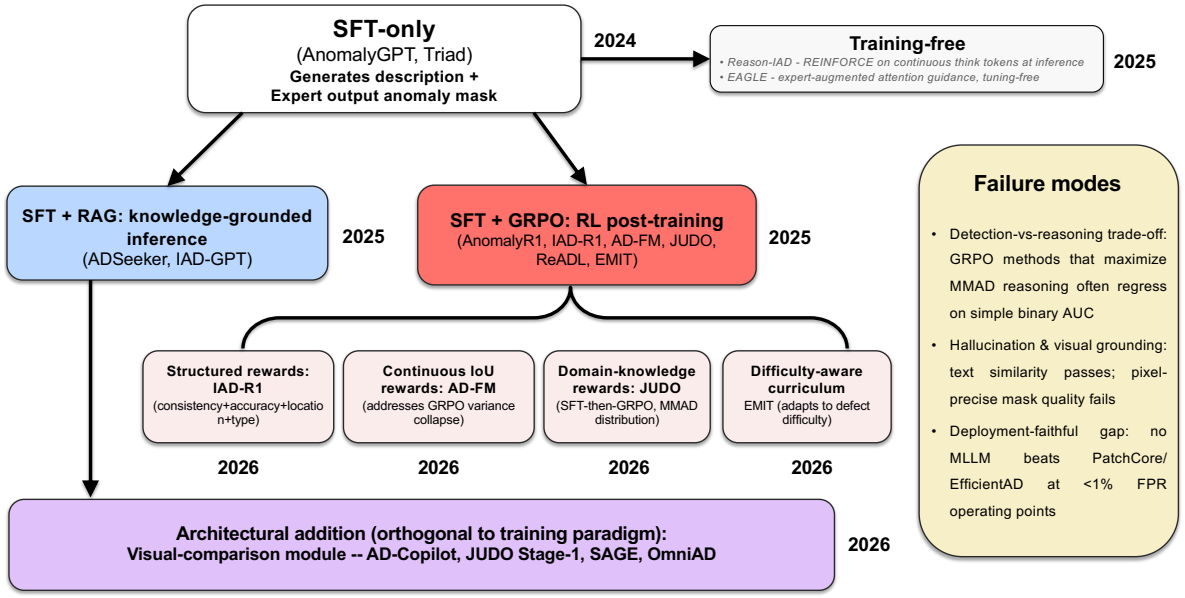

Figure 3. MLLM-AD training paradigm tree. The SFT-only baseline (AnomalyGPT, Triad) branches three ways: SFT+RAG (knowledge-grounded; ADSeeker, IAD-GPT), SFT+GRPO (RL post-training, the dominant branch; AnomalyR1, IAD-R1, AD-FM, JUDO, ReADL, EMIT), and tuning-free outliers (Reason-IAD, EAGLE). The GRPO branch splits into four reward designs (IAD-R1 structured, AD-FM continuous-IoU, JUDO domain-knowledge, EMIT difficulty-aware). Orthogonal to training, the 2026 architectural addition is the visual-comparison module (VCM, purple), layered on any paradigm above (AD-Copilot, JUDO, SAGE, OmniAD). Right panel: three failure modes (Section 4.4).

4.2. Three Replacement Mechanisms

Training-paradigm shift: SFT to GRPO/RL.

AnomalyGPT-era SFT [39] trained on supervised VQA pairs with no signal about reasoning-path validity. Four groups in the wave independently converge on GRPO post-training: **IAD-R1** [45] pairs PA-SFT with Structured-Control GRPO using four simultaneous rewards (86.1% mean binary across six datasets); **AD-FM** [47] uses a continuous GIoU reward that fixes GRPO reward-variance collapse (83.56% MMAD, LoRA only); **JUDO** [46] runs a three-stage curriculum—juxtaposed segmentation, GPT-4o domain QA, GRPO—reaching 81.20% MMAD, on par with Gemini-2.5-Pro; **ReADL** [96]'s CGRO reward recovers pixel localisation from image-level supervision. These are balanced binary numbers, not AUROC at low FPR (see §4.5). SFT-only baselines **Triad** [97] and **MAU-GPT** [98] trail on MMAD breadth, and **AnomalyR1** [99] is the preprint that first established the GRPO recipe.

 doi:10.20944/preprints202606.0719.v1

Knowledge-grounded inference.

A second cluster retrieves domain knowledge at inference rather than during training. **ADSeeker** [100] is a plug-and-play RAG with Q2K retrieval over SEEK-M&V (69.90% MMAD; 94.3/91.5% MVTec/VisA zero-shot AUROC). **Reason-IAD** [101] is training-free latent reasoning over continuous think tokens with RAKI category retrieval (79.43% MMAD on Qwen3-VL-8B). **EAGLE** [102] is tuning-free: a frozen PatchCore expert wired into a frozen MLLM via Distribution-Based Thresholding and Confidence-Aware Attention Steering. **IAD-GPT** [103] is the train-time analogue. The track is useful when training is infeasible but trails GRPO: JUDO's ablation gives SFT+RAG 76.29% vs. SFT+GRPO 77.29%; AD-Seeker has 17% retrieval failure on test-split MVTec/VisA categories; Reason-IAD's entropy reward is self-referential.

Visual comparison: the new architectural frontier.

Standard MLLM architectures encode query and reference images independently, leaving comparison to language-space attention—insufficient for subtle pixel-level differences. Four groups address this: **AD-Copilot** [49] adds a DETR-style Comparison Encoder under a four-stage curriculum (82.3% MMAD, 3.35× BBox gain); **JUDO** [46] Stage-1 juxtaposed segmentation alone contributes +11.84 pp MMAD localisation—the single largest per-stage ablation gain; **Triad** [97] pipes MuSc/AnomalyCLIP through an Expert-Guided RoI tokeniser (+1.4–2.5% trained, −14–23% zero-shot); **SAGE** [104] and **OmniAD** [105] add Self-Guided Fact Enhancement and unified detection-plus-explanation respectively. Directional agreement across four institutions: language-space comparison is the binding constraint, and the next iteration likely needs end-to-end differentiable comparison rather than frozen-expert injection.

4.3. Benchmarks: From MMAD to the Second Generation

MMAD [44]—39,672 multiple-choice questions over 8,366 images across seven subtasks—made the wave's MLLM-AD research coherent by replacing binary AUROC with a richer multi-subtask protocol.

Second-generation benchmarks. MMR-AD [9] is the largest training-and-evaluation set: 127K samples, 188 categories, 395 anomaly types, 112,875 manual bounding boxes with Qwen2.5-VL-72B-generated CoT rationales. **MAU-Set** [98] emphasises breadth (35 product types, 224K QA pairs over 6 industrial domains); MAU-GPT scores 82.12% discriminative on MAU-Set but only 61.41% on MMAD, suggesting distribution-locality. **Chat-AD** [49] (620K+ samples, 327 categories) is the largest training corpus but remains partly proprietary. **M3-AD** targets the multimodality gap (RGB-D, thermal, point cloud). Dataset scale moved from MMAD's 8K to 620K within 18 months.

4.4. Failure Modes

Detection-reasoning trade-off. GRPO-trained models scoring high on MMAD reasoning often show degraded binary discrimination: JUDO reports 65.04% post-training vs. 71.39% for base Qwen2.5-VL-7B (−6.35 pp on the simplest IAD task). No paper resolves the tension; JUDO Stage 1 and AD-FM's `<rethink>` narrow but do not close the gap, which matters because binary detection is the primary deployment function.

Hallucination and visual grounding. Linguistically plausible but visually ungrounded responses pass MMAD subtask scoring—text similarity for description, a coarse grid for localisation—while failing pixel-precise inspection. ADSeeker's 17% retrieval failure has no recovery path; Reason-IAD's entropy reward incentivises confident wrongness; the grid-vs-pixel gap is evaluated only in MMAD-BBox, reported so far only by AD-Copilot.

Benchmark circularity. MMAD's creators currently hold the top MMAD score; JUDO's Stage 2 derives training QA from MMAD's knowledge snippets; ADSeeker's SEEK-M&V knowledge base covers exactly MMAD's MVTec/VisA categories. These structural overlaps do not automatically invalidate the numbers but make it hard to disentangle reasoning capability from MMAD-distribution adaptation. Cross-group benchmark governance is needed (Section 10.2).

 doi:10.20944/preprints202606.0719.v1

Deployment-faithful evaluation gap. No Section 4 method evaluates precision-recall at industrial false-positive rates (e.g., <1% FPR, the IM-IAD [16] protocol); none is compared to PatchCore [12] or EfficientAD [25] at deployment-realistic operating points. This is the most critical open gap in the entire MLLM-AD subfield.

4.5. Honest Assessment

(i) Genuine. SFT→GRPO is real: four independent groups (IAD-R1, AD-FM, JUDO, ReADL) converge across different backbones and reward designs, and ADSeeker 69.90% → JUDO 81.20% on the same Qwen2.5-VL-7B is larger than any SFT-era MMAD improvement; convergent visual-comparison gains (+11.84 pp JUDO Stage-1 localization, 3.35× AD-Copilot BBox) confirm paired-image architectures outperform single-image ones.
(ii) Oversold or unresolved. MMAD benchmark circularity (creators top the leaderboard) plus the JUDO −6.35 pp binary detection regression after GRPO post-training; no MLLM in this batch beats PatchCore or EfficientAD at deployment-faithful operating points, and SALAD [106] reaches 96.1% AUROC on MVTec LOCO with no language component at all—MLLM-AD has advanced *reasoning about* anomalies, not detection at industrial operating points.
(iii) Minimum new evidence. Cross-group evaluation on MMR-AD (by groups not on MMAD authorship) plus an MLLM-AD method matching PatchCore-level binary detection on VisA at FPR≤1% while exceeding 80% MMAD-overall accuracy.

5. Generalization Mechanism 3—Geometric and Multimodal Priors

The third mechanism treats sensing modality as inductive bias, building representations from point clouds, depth, multi-view, and image-text pairs.[4] The pre-wave scaffolding (M3DM [50], CPMF [53], Real3D-AD [54], PointAD [56]) established frozen per-modality encoders with memory-bank scoring on MVTec-3D-AD; the 2024–2026 wave responds to benchmark saturation, mismatched geometric representations, and missing language grounding. *Scope:* SCoNE [107] (tabular) and Wave-MambaAD [108] (2D) are excluded.

[4] **3D/multimodal vocabulary.** *Point cloud*: unordered 3D points; *RGB-D*: paired RGB + depth; *MV*: multi-view RGB. *O-AUROC / P-AUROC*: object- and point/pixel-level AUROC. *Synth2real*: train on synthetic (e.g., CAD) data, test on real scans. *Point-MAE / Point-BERT / PointGPT*: 3D self-supervised pretraining. *ULIP / PointCLIP*: 3D–language alignment. *SDF*: signed-distance function. *FPFH*: hand-crafted 3D feature (fast point-feature histogram). *MoE*: mixture-of-experts routing.

 doi:10.20944/preprints202606.0719.v1

Table 4. Representative 3D and multimodal AD methods, 2023–2026 (foundation baselines in the first block; the curated 2025–2026 cohort below). AS = Anomaly-ShapeNet; PC = point cloud.

Method	Venue	Modality	Benchmark(s)	Key metric	Notes
M3DM [50]	CVPR 2023	RGB-D	MVTec-3D-AD	94.0% O-AUROC	Foundation baseline
CPMF [53]	PR 2024	RGB-D	MVTec-3D-AD	>95% O-AUROC	Per-modality scoring
PointAD [56]	NeurIPS 2024	Point cloud	Real3D-AD, AS	Zero-shot 3D AD	PLM zero-shot scoring
G^{2}SF [109]	ICCV 2025	RGB-D	MVTec-3D-AD, Eyecandies	SOTA AUPRO@1%	Anisotropic memory metric
PASDF [110]	ICCV 2025	Point cloud	Real3D-AD, AS	80.2% O-AUROC	First detect-plus-repair via continuous SDF
SiM3D [59]	ICCV 2025	RGB+PC, multi-view	SiM3D	Baselines fail synth2real	Single-instance; CAD-to-real benchmark
Reg2Inv [111]	NeurIPS 2025	Point cloud	Real3D-AD, AS	SOTA on both	Registration-as-learning
CASL [112]	AAAI 2026	Point cloud	Real3D-AD, AS	+5.6% O-AUROC vs. SSL	Multi-scale curvature beats PointMAE
MiniShift [58]	AAAI 2026	Point cloud	MiniShift, Real3D-AD, AS	80.4/92.3% O-AUROC	500k-pt, <1% anomaly cov.; deep nets near-random
U-MV [113]	AAAI 2026	Multi-view RGB	Real-IAD, MANTA	SOTA on both	Homography alignment; pose-free; 2D only
HPF-APC [114]	CVPR 2026	Point cloud	Real3D-AD, AS	84.2% O-AUROC	Patch codebook; angular/planar deformation
SeDiR [115]	CVPR 2026	Point cloud	Real3D-AD, AS	+2.8/+9.1% vs. MC3D-AD	Multi-class; category-entanglement fix
UniMMAD [60]	CVPR 2026	12 modalities	9 datasets, 66 classes	SOTA on all 9	Cross-MoE decoder; anchor paper
IB-IUMAD [116]	CVPR 2026	RGB-D	MVTec-3D-AD, Eyecandies	91.0% O-AUROC	Info-bottleneck fusion; incremental
PIRN [117]	ICLR 2026	RGB-D	MVTec-3D-AD, Eyecandies	Best 10-shot Eyecandies	Prototype codebook; few-shot multimodal
IMDD-1M [10]	CVPR 2026	RGB+text	IMDD-1M	<5% data ≈ expert	1.24M image-text pairs; 2D only
BTP [94]	CVPR 2026	PC+CLIP 2D proj.	Real3D-AD, AS	Zero-shot; >PointAD pt-level	Cross-tag Section 3.3
GS-CLIP [95]	CVPR 2026	PC+CLIP 2D proj.	Real3D-AD, AS	Zero-shot SOTA	Geometry-aware prompts; cf. Section 3.3

5.1. The MVTec-3D-AD Saturation Problem

MVTec-3D-AD [55] was pushed above 97% by CFM [118], leaving little headroom; G^{2}SF [109] and IB-IUMAD [116] extract only marginal gains. **MiniShift [58]** introduces 2,577 point clouds at 500k points with sub-1% anomaly coverage; deep backbones approach random on the hard tier while handcrafted multi-scale descriptors outperform transformers at >20 FPS. **SiM3D [59]** trains on a single nominal instance and adds a CAD-to-real synth2real protocol where all five adapted baselines (BTF, M3DM, CFM, AST, EfficientAD) degrade substantially, using voxel-based Anomaly Volumes instead of 2D maps. Both expose that resolution has been too coarse and training scenarios too generous (Figure 4).

 doi:10.20944/preprints202606.0719.v1

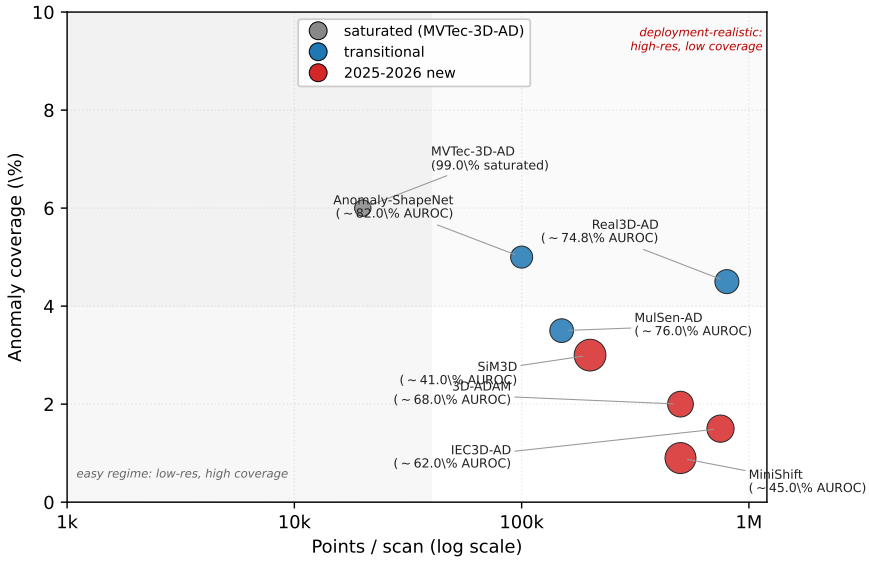

Figure 4. 3D-AD benchmark space: scan resolution (log scale) vs. anomaly coverage. Point size encodes headroom (100 − best AUROC). MVTec-3D-AD (top-left, gray, saturated ∼99%) sits in the easy regime; the newer benchmarks MiniShift, 3D-ADAM, IEC3D-AD, SiM3D move toward the deployment-realistic corner (bottom-right, high-res, low anomaly coverage).

5.2. Stronger Geometric and Multimodal Representations

Rotation-invariance and geometric-feature rethinking.

Five papers argue the PointMAE + FPFH + Euclidean-memory pipeline misidentifies the AD-relevant signal. **CASL** [112] diagnoses a coordinate-to-coordinate SSL shortcut and substitutes multi-scale curvature (+5.6% O-AUROC on Real3D-AD). **PO3AD** [119] predicts per-point offsets toward a learnt normal field. **Reg2Inv** [111] jointly trains registration and AD via a Geometric Transformer with RIConv++, beating Reg3D-AD [120] and Group3AD [121]. **PASDF** [110] replaces discrete points with a continuous SDF doubling as a repair template (80.2/90.0% O-AUROC on Real3D-AD/AS), the first 3D detection+repair. **HPF-APC** [114] targets planar/angular deformations via MinkUNet34C + multi-scale patch codebook with RoPE (84.2/87.6%). None offers zero-shot or language grounding (Table 4).

RGB-D fusion quality.

G^{2}SF [109] replaces isotropic Euclidean distance with an anisotropic Local Scale Prediction metric (SOTA AUPRO@1% on MVTec-3D / Eyecandies). **PIRN** [117] attacks few-shot multimodal where CFM/M3DM degrade, using a prototype codebook with balanced OT and adaptive GRU, dominating baselines at 1–10 shots on Eyecandies. **CIF** [122] uses hypergraph-enhanced memory for query/few-shot co-occurrence. **U-MV** [113] introduces progressive homography-guided alignment so multi-view 3D AD no longer needs known camera pose.

5.3. Generalist, Language-Grounded, and Zero-Shot 3D

Multi-class unified 3D and multimodal AD.

The UniAD-style wave reaches 3D/multimodal. **SeDiR** [115] fixes Inter-Category Entanglement with Coarse-to-Fine Tokenisation, Category-Conditioned Contrastive, and a Geometry-Guided Decoder, beating MC3D-AD [123] by 2.8/9.1%. **UniMMAD** [60] trains one model on 9 datasets / 12 modalities / 66 categories via a Cross-MoE decoder—the first confirmed transfer of NLP MoE scaling into 3D/multimodal IAD. **IB-IUMAD** [116] pairs Information Bottleneck Fusion with a Mamba decoder in the incremental setting: 91.0% O-AUROC on 10-class MVTec-3D vs. IUF [124] 88.7%, 41× faster than M3DM. All three need category labels at training.

 doi:10.20944/preprints202606.0719.v1

Language-grounded industrial defects.

IMDD-1M [10] supplies 1.24M image-text pairs across 63 manufacturing domains and 421 defect types—roughly 18× Real-IAD [70]. An 860M text-conditioned diffusion U-Net with implicit captioner and Mask2Former head handles segmentation, detection, retrieval, and captioning, reportedly matching specialised models with <5% task data. 2D-only, but supplies the text substrate future native-3D grounding will need. Cross-tag: Section 3.2 (pure-VFM few-shot) and Section 4.3 (MLLM-AD training data), addressing the knowledge-base gap of ADSeeker [100] and Reason-IAD [101].

3D-meets-VFM revisited.

BTP [94] (cross-tag Section 3.3) projects point clouds to 2D for CLIP plus geometric point features—zero-shot but discards 3D information like PointAD. **GS-CLIP [95]** (cross-tag Section 3.3) augments CLIP with curvature/normal prompts, beating PointAD zero-shot on Real3D-AD and AS but still projection-dependent.

5.4. Open Problems

Zero-shot 3D AD with native language grounding. No method in the wave grounds language in 3D coordinates without 2D projection; IMDD-1M supplies data but no jointly-trained 3D-language space exists. **CAD-to-production generalization.** SiM3D's synth2real protocol breaks every adapted baseline; the CAD-first deployment workflow—train on the CAD model before any real scan exists, the standard industrial line-changeover scenario—cannot yet be supported. This is the most practically consequential, leaderboard-invisible open problem in 3D AD. **Detection and repair as a unified task.** PASDF [110] is the first attempt via SDF but qualitative and limited to clean registration. **Benchmark governance.** MVTec-3D-AD remains common ground despite saturation; Real3D-AD and AS are inconsistently adopted, and MiniShift/SiM3D have not yet become secondary standards.

5.5. Honest Assessment

(i) Genuine. Three durable advances: curvature/registration objectives (CASL +5.6% O-AUROC on Real3D-AD, PASDF detection-plus-repair 80.2/90.0% O-AUROC); MoE multi-modality (UniMMAD 12 modalities × 66 classes in one checkpoint); language grounding at scale (IMDD-1M's 1.24M industrial-defect image-text pairs).
(ii) Oversold or unresolved. MVTec-3D-AD is saturated (CFM 97%+) yet papers continue mining marginal AUPRO; "zero-shot 3D" bridges (GS-CLIP, BTP) evaluate only on per-instance-aligned-scan benchmarks; no Section 5 method reports on SiM3D's CAD-to-production synth2real protocol (which degrades adapted baselines by ∼14–30 pp, method-dependent) or bound-FPR metrics.
(iii) Minimum new evidence. SiM3D synth2real result from a new method in this section, plus IMDD-1M-grounded zero-shot 3D evaluation that does not collapse to a 2D-projection baseline.

6. Generalization Mechanism 4—Universal Task Priors

Unlike the mechanisms in Section 3–5 that borrow generalization from external sources—visual backbones, language models, 3D priors—the methods here attempt to build a single internal representation spanning all anomaly types, categories, and modalities, a *universal task prior* learned from variation within IAD itself. The 2024–2026 wave's answer is partial: MoE routing, CLIP adapters, and visual-only meta-learning each advance the frontier, but none transfers to a truly unseen industrial domain without target-domain data.

Cross-tag note. Three MoE-based papers primarily housed elsewhere have their routing mechanics unpacked here: PromptMoE [76] and MoECLIP [77] from Section 3.1, and UniMMAD [60] from Section 5.3.

Figure 5 sketches both the design-pattern shift (per-dataset detector stack → single backbone + convergent MoE routing) and the class-count axis the section operates along (C=1 specialists to the CCL scaling wall at C=81).

 doi:10.20944/preprints202606.0719.v1

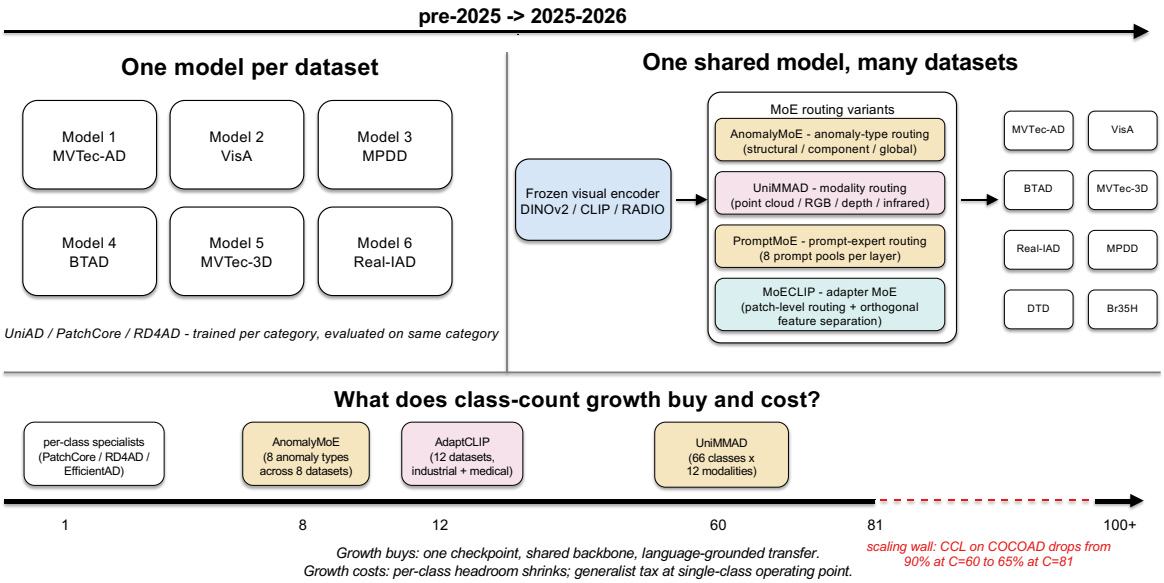

Figure 5. From per-class detectors to generalist + MoE, with the class-count scaling axis. **Top:** one model per dataset (UniAD-era) → a single backbone plus MoE routing across datasets, via four convergent variants—AnomalyMoE (anomaly-type), UniMMAD (modality), PromptMoE (prompt-expert), MoECLIP (adapter MoE). **Bottom:** the implicit class-count axis, from per-class specialists (C=1) through AnomalyMoE (C=8), AdaptCLIP (C=12), UniMMAD (C=66) to the CCL scaling wall at C=81 (Section 6.5).

6.1. The Arc From "One Model Per Class" To "One Model Per Domain"

The closed-set era trained per-category detectors [12,24,25]; UniAD [29] broke this by joint training across MVTec-AD via neighbour-masked attention, refined by HVQ-Trans [30], OneNIP [31], and MetaUAS [32] but without cross-dataset transfer claims. The subsequent generalist phase—driven by WinCLIP-style CLIP transfer [14] and cross-dataset meta-learning [125,126]—attempts the per-class → per-dataset → across-domains step, though "cross-domain" in 2026 still means transfer between curated suites sharing an ImageNet-like prior (Section 6.5).

6.2. Routing and Adapter Generalists

MoE as the convergent design pattern.

The wave's most structurally significant finding is the independent convergence of four papers across three IAD sub-fields on mixture-of-experts routing for handling heterogeneous inputs in a single model. **AnomalyMoE** [127] routes through three expert groups—structural, component, global-logical—on a frozen DINOv2 backbone with Top-K sparse gating, mutual-information repulsion, and selection balancing. Routing specialization is empirically real (removing the global expert costs −2.8 pp on MVTec LOCO; the patch expert alone loses 9.3); a single unified model reaches 99.5 / 98.1 I-AUROC on MVTec-AD / VisA across MVTec LOCO, BrainMRI, LiverCT, RESC, MVTec-3D, and UCSD Ped2. **UniMMAD** [60] extends MoE to modality (RGB / depth / point-cloud / infrared); **PromptMoE** [76] routes prompt experts; **MoECLIP** [77] replaces CLIP adapters with MoE layers. Common signal: when inputs are genuinely heterogeneous, sparse routing beats a shared representation.

CLIP-adapter universal detectors.

This sub-branch treats CLIP [5] as a frozen universal encoder and asks how little fine-tuning suffices for cross-domain IAD. **AdaptCLIP** [72] is the broadest cross-domain CLIP-adapter to date, evaluating on 12 datasets across industrial and medical domains in zero/1/2/4-shot settings; three adapters—visual MLP, text prompts, prompt-query residual—attached at CLIP's input/output, trained *alternately* (joint training degrades), 0.4M added params, zero-shot MVTec-AD 93.5 / Br35H 94.8 I-AUROC. **AnoPLe** [128] uses class names alone with bidirectional text-visual prompt coupling (−6.3% on held-out MVTec vs. −26.2% for INP-Former [129]); **UniVAD** [130] is the extreme training-free point.

 doi:10.20944/preprints202606.0719.v1

MultiADS [131] extends binary to K+1 states via a manually-curated KBA for zero-shot multi-type defect segmentation, with severe per-type localisation gaps (VisA MTAS F1 22.3%). Shared weakness: the domain boundary is CLIP's, not one set by the IAD method.

6.3. Reference- and Prototype-Based Generalists

Language-free visual generalists.

VisualAD [86] challenges the necessity of language: removing AnomalyCLIP's [34] text encoder causes negligible degradation while cutting trainable parameters by over 99%. Two learnable global tokens—anomaly + normal—prepend to a frozen CLIP ViT-L/14 with Spatial-Aware Cross-Attention and a cosine-margin contrastive loss. Implication: learned text prompts can be replicated by two visual prototypes; cross-modal alignment is not the mechanism but an indirect route. Limitations: backbone-dependent (CLIP ViT), and BrainAD 80.8 shows visual generalization is not uniform across medical settings.

Meta-learning with richer reference sets.

NAGL [132] extends InCTRL/ResAD meta-learning with one abnormal reference alongside K normal references (K1+A1). Naive score fusion *degrades* (90.7 vs. 93.2 on MVTec); gain requires a learned Residual Mining module plus cross-attention. On a frozen ViT-S (3M trained params): 17.1 FPS vs. 1.2 for InCTRL; cross-domain N1+A1 reaches MVTec 95.8 / VisA 88.5 but BraTS only 78.1. NAGL's advantage applies only when the abnormal reference matches the test defect type; otherwise it collapses to the normal-only variant.

6.4. Multi-Class Unified and Open-Set Tracks

Multi-class unified track.

Two papers remain in the UniAD lineage without universality claims but clarify why naive joint training does not scale. **MaskAD** [133] uses MAE "identical shortcut" as a measurement device—Sinkhorn distance between reconstruction-error distributions under parallel masking—reaching MVTec-AD 98.6 I-AUROC and +1.2 AUPRO over MambaAD [134] (+3.6 over UniAD/DiAD); coreset tied to training distribution. **CCL** [135] diagnoses inter-class confusion and fixes it via K-Means class-aware contrastive learning, with the harder COCOAD revealing a scaling wall: 65.2% at C=81 vs. above 90% at C=15. **Dinomaly** [136] pushes a "less is more" recipe (DINOv2 ViT-B + minimal linear head + joint training, 99.6/98.7/89.3% I-AUROC on MVTec-AD/VisA/Real-IAD); **Dinomaly2** [137] extends it across single-/multi-/few-shot under one checkpoint—the cleanest evidence that the multi-class bottleneck is a backbone problem, not an architecture problem. INP-Former [129], UniNet [138], CostFilter-AD [139], DecAD [140], and DeCo-Diff [141] add complementary angles; none demonstrates transfer outside the curated suite.

Open-set defect recognition.

UniSpector [142] addresses open-set defect *recognition* via visual prompting from annotated exemplars, using a DETR-style architecture with spatial-spectral and angular-margin contrastive prompt encoders. On the new InsA benchmark (six datasets), in-domain Real-IAD scores 69.1 AP_{50b}, but cross-domain 3CAD collapses to 14.1—a near-fivefold drop and the cleanest quantitative evidence for Section 6.5's main finding. IMDD-1M [10] and MAU-GPT [98] are the open-vocabulary siblings of UniSpector (Section 8.3).

6.5. Honest Assessment

(i) Genuine. Three durable advances: AnomalyMoE's heterogeneous MoE routing yields task-adaptive specialization confirmed by gate ablation; VisualAD closes the WinCLIP debate (negligible degradation without the text branch); NAGL shows abnormal references help only through learned residual cross-attention, not score fusion. ADPretrain [143] sharpens the underlying claim by training

 doi:10.20944/preprints202606.0719.v1

AD-specific representations on Real-IAD with angle/norm-oriented contrastive losses, demonstrating that ImageNet pretraining is a sub-optimal prior even when frozen.

(ii) Oversold or unresolved. "Generalist" in 2026 means multi-class within a curated benchmark suite, not zero-data deployment: AnomalyMoE is trained jointly on all eight of its evaluation datasets; UniSpector drops 69.1%→14.1% AP50b cross-domain; CCL's COCOAD reveals a sharp scaling wall (90.6% I-AUROC at C=60 to 65.2% at C=81); not one paper demonstrates convincing zero-data transfer to a domain outside the CLIP/ImageNet prior.

(iii) Minimum new evidence. >90% AUROC on a held-out industrial domain (wafer / X-ray / audio anomaly) without target-domain fine-tuning, on a class-count axis above C=81.

7. Generalization Mechanism 5—Generated Abnormal Priors

The fifth mechanism is a data-supply argument: if category-specific normal-only training is the bottleneck, generate the missing abnormal examples. The 2024–2026 wave breaks decisively with the AnomalyDiffusion [62] monoculture—per-category LDM fine-tuning with hand-crafted masks and 1,000-step DDIM—along five simultaneous dimensions, producing a sub-field barely recognizable as its descendant.[5]

7.1. Break from the AnomalyDiffusion Era: Five Simultaneous Dimensions

The wave's papers diverge along five axes simultaneously (Figure 6): *(i) language grounding* via text/VLM conditioning—CLIP in AnoStyler [144] and Anomagic [145], GPT-4V in AnomalyPainter [146], Qwen2.5-VL-32B in ARAS [147], Qwen2-VL in SynSur [148]; *(ii) training-free or few-shot* regimes that eliminate or shrink fine-tuning (O2MAG [149], TF-IDG [150] freeze the diffusion prior; MAGIC [151] and SeaS [66] need 5–20 reference images); *(iii) speed*: FAST [65] compresses 1,000 sampling steps to 10 via its AIAS trajectory, ARAS reports a 5× generation speedup over diffusion; *(iv) joint image-and-mask outputs* from a single forward pass in SeaS and FAST; *(v) architecture diversification* beyond LDM UNet—SDXL + ControlNet (AnomalyPainter), Anydoor (TF-IDG), Infinity-8B autoregressive tokens (ARAS), Flux.1-dev DiT (SynSur), and a CLIP-guided U-Net with no diffusion at all (AnoStyler).

[5] **Diffusion / synthesis vocabulary.** *LDM* (latent diffusion model): diffusion in a learned latent space, e.g., Stable Diffusion. *DDIM*: deterministic sampler, ∼1000 (or fewer) steps. *SDXL*: a larger LDM with two text encoders. *ControlNet*: conditions LDM output on edges, depth or masks. *DiT* (diffusion transformer): transformer diffusion backbone, e.g., Flux. *LoRA*: low-rank parameter-efficient fine-tuning. *DreamBooth*: subject-specific LDM fine-tuning from a few references. *CLIPScore / DreamSim*: image–text / image–image similarity metrics. *FID / KID / LPIPS / IS*: generation-quality metrics (lower better for FID/KID/LPIPS, higher for IS).

 doi:10.20944/preprints202606.0719.v1

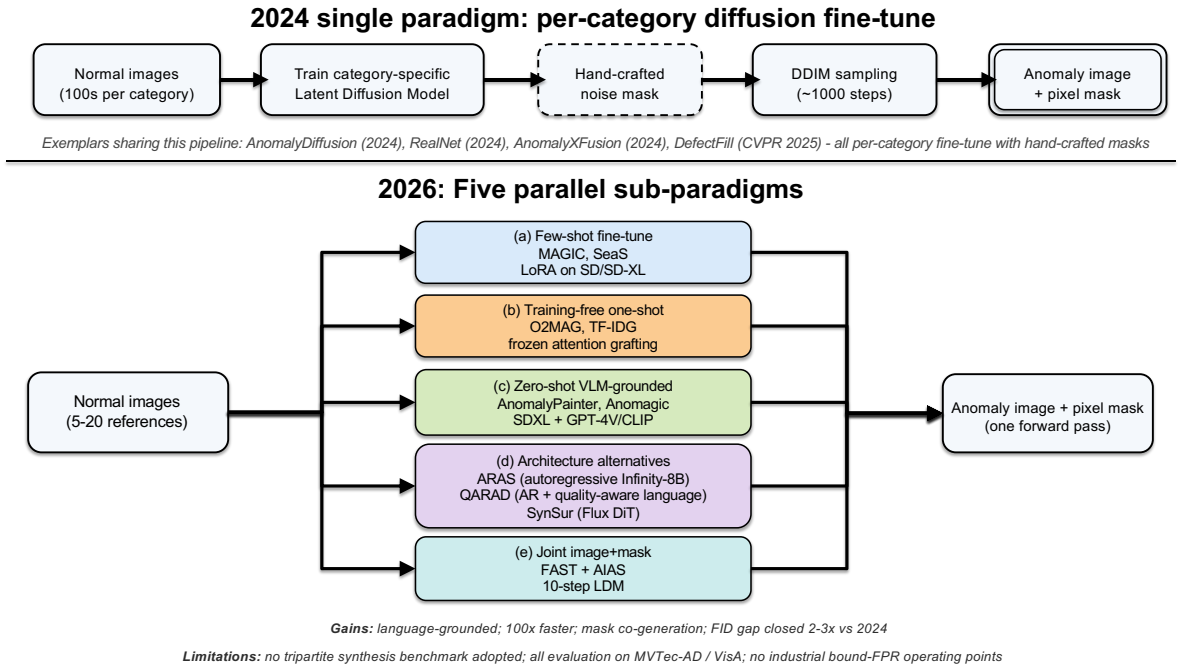

Figure 6. Synthesis pipeline diversification: from the 2024 single-paradigm era (top) to five parallel 2026 sub-paradigms (bottom). Top: per-category diffusion fine-tune → hand-crafted mask → ∼1000-step DDIM (AnomalyDiffusion, RealNet, AnomalyXFusion, DefectFill). Bottom: reference images branch into (a) few-shot LoRA fine-tune, (b) training-free attention grafting, (c) zero-shot VLM-grounded generation, (d) architecture alternatives (autoregressive ARAS/QARAD; DiT SynSur), (e) joint image+mask in one pass. Footer: the field's gains (language grounding, 100× speedup, mask co-generation, FID gap closed 2–3×) against its limits (no tripartite benchmark; evaluation confined to MVTec-AD/VisA; no bound-FPR operating points).

7.2. Method Families

Few-shot fine-tuning: MAGIC and SeaS.

MAGIC [151] fine-tunes SD2-inpainting via DreamBooth, Gaussian Prompt Perturbation, Spatially Adaptive Guidance, and Context-Aware Mask Alignment, reaching KID 40.27 on MVTec-AD (vs. 96+ in the AnomalyDiffusion era) and I-AUROC 99.36/94.28% on MVTec-AD/VisA; **DefectFill** [152] is the sibling inpainting baseline. **SeaS** [66] fine-tunes SD v1.4 under a separation-and-sharing principle—Unique-Anomaly tokens for defects, shared tokens for texture, plus Direct-Anomaly / Normal-Alignment / Reverse-Mask losses—posting the cohort-best generation quality (KID 0.04, IS 1.88 on MVTec-AD) and +8.66 pp pixel AP; **DualAnoDiff** [153] is the dual-diffusion alternative. Both still pay per-category fine-tuning: manageable at 15 categories, problematic for Real-IAD's 30 or open-world deployment.

Training-free one-shot: O2MAG and TF-IDG.

O2MAG [149] freezes SD v1.5 and grafts reference self-attention keys/values into target-texture and background branches (Triple Attention Grafting) with Attention-Guided Optimization (KID 45.55, P-AUROC 99.2% on MVTec-AD). **TF-IDG** [150] adapts Anydoor with Sinkhorn-aligned DINOv2 features + dynamic Anomaly Area Masking + AdaIN (Local IS 3.32, I-/P-AUC/PRO 99.6/99.1/95.8 MVTec-AD). Convergence: two independent teams land within 0.1–0.5 pp of fine-tuned MAGIC/SeaS downstream, suggesting one reference + frozen-diffusion attention nearly suffices. Fine-tuning still buys image quality (KID 45.55→0.04), but their inference-time loops (500 AGO steps; iterative Sinkhorn) are an unreported cost.

Zero-shot VLM-grounded: AnomalyPainter, Anomagic, AnoStyler.

AnomalyPainter [146] pairs Tex-9K texture retrieval, GPT-4V defect captions, and SDXL + ControlNet-Canny with Texture-Aware Layout Injection, lifting RealNet on VisA from 90.2/92.8 to

 doi:10.20944/preprints202606.0719.v1

93.9/96.2 I-/P-AUC (GPT-4V latency unanalysed). **Anomaly Anything** [154] is the prompt-engineering lower bound (no training, no retrieval). **Anomagic** [145] pre-trains a Crossmodal Prompt Encoder on AnomVerse (12,987 triplets across 13 datasets) driving frozen SD v1.5 + LoRA, reaching VisA + INP-Former++ PRO 95.92% / F1 54.00% (+1.39 pp over AnoGen). **AnoStyler** [144] counters it: a 0.61M CLIP-guided U-Net style-transfer with SAM, single-pass, IS 2.04, I-/P-AUC/PRO 98.0/94.4/88.3 on MVTec-AD—text guidance without billion-parameter diffusion.

Architecture alternatives: ARAS/QARAD and SynSur.

ARAS/QARAD [147]—the autoregressive synthesis editor (ARAS) and quality-aware detector (QARAD) of a single work—replaces diffusion with the Infinity-8B autoregressive token model under local token masking, conditioned by Qwen2.5-VL-32B (I-/P-AUROC 99.7/99.8 on MVTec-AD at 1.49 s/image vs. 7.51 s for the diffusion baseline). The same work adds quality-aware language conditioning with local generation control—low-CLIPScore samples are down-weighted rather than discarded—but reports no FID/KID and sits at the AUROC saturation ceiling.

SynSur [148] fine-tunes the 12B Flux.1-dev DiT with LoRA, using Qwen2-VL captions, DreamSim/CLIPScore filtering, and SAM-3 masks. Its real contribution lies in evaluation domain and its negative result—discussed next.

7.3. Evaluating Synthesis

Synthesis-as-benchmark: ASBench.

ASBench [11] evaluates synthesis rather than detection methods, and reports a methodologically central finding: generation-quality metrics (FID/SSIM/LPIPS) do not significantly correlate with downstream detection lift across the methods examined. It proposes a tripartite protocol—(a) FID/KID generation realism on a held-out real anomaly reference set, (b) mask-IoU against ground-truth pixel annotations, and (c) downstream detection delta on a *fixed, named* detector backbone—as the minimum credible evaluation. We adopt this framing in Section 9.6. ASBench originates within the broader M-3LAB ecosystem (Disclosures); independent re-evaluation by outside groups would strengthen the protocol's standing.

Reality check: what happens off MVTec-AD.

Nine of ten synthesis papers in this cohort report positive downstream lift, eight exceed 1 pp on a named benchmark, and the gains span independent implementations and detector architectures—not noise. But every paper bar one evaluates primarily on MVTec-AD or VisA.

SynSur is that exception. On BSData (ball-screw pitting) and MSD (surface scratches)—production-adjacent datasets with realistic surface, lighting, and defect-size variation—synthesis augmentation lifts AP from 0.652 to 0.667 (+0.015 absolute over 150 epochs), while a synthetic-only regime collapses to 0.393 (-0.259). The +0.015 AP gain is positive but an order of magnitude smaller than the 3–12 pp lifts on curated benchmarks. Two interpretations remain plausible: MVTec-AD/VisA may be in-distribution for pretrained diffusion priors, or synthesis benefit may be real but heavily benchmark-dependent; no controlled cross-benchmark ablation exists. What SynSur establishes is a concrete upper bound—$\approx$+1.5% relative AP—in one realistic deployment domain. We return to it in Section 9.6 and Section 10.2.

Note on InvAD: the manifest places InvAD [155] in the synthesis bucket, but PF-ODE inversion is an anomaly scoring mechanism, not generation; it is therefore outside this chapter's scope.

7.4. Honest Assessment

(i) Genuine. The AnomalyDiffusion monoculture is broken along five simultaneous axes (language grounding, training-free regimes, 100$\times$ speedup, joint image+mask in one forward pass, architecture diversification beyond LDM UNet); SeaS posts KID 0.04 / IS 1.88 and FAST compresses 1000 DDIM steps to 10 with comparable downstream lift.

 doi:10.20944/preprints202606.0719.v1

(ii) Oversold or unresolved. FID/SSIM/IS/LPIPS do not predict downstream I-AUROC (ASBench's 19,680-configuration finding); the tripartite protocol has not been adopted even by papers post-ASBench; evaluation is confined to MVTec-AD/VisA, and SynSur's BSData test—the one production-adjacent point—shows the lift collapses to +0.015 AP (an order of magnitude smaller than curated-benchmark gains).
(iii) Minimum new evidence. Any synthesis paper post-ASBench reporting the tripartite (FID/KID + mask-IoU + downstream-detection-delta on a fixed-named detector) on a non-MVTec/VisA benchmark with bound-FPR operating points.

8. Evaluation Frontier—What Counts as a Real Advance?

Each Section 3–7 chapter notes that its mechanism delivers real method-level advances but routinely overstates them in headline numbers. This section asks: against what evaluation is a 2024–2026 IAD method genuinely advancing the state of the art?

8.1. Background: Established Benchmarks Are Saturated Or Constrained

Well-established, and not our contribution: MVTec-AD [1], VisA [67], and MVTec-3D-AD [55] now have headline I-AUROC differences inside the test-set noise floor. PatchCore [12] reported 99.1% on MVTec-AD in 2022; the best apples-to-apples numbers from the wave (Dinomaly [136] 99.6%, AnomalyMoE [127] 99.5%, CCL [135] 99.1%) sit within 0.5 pp of that figure on a benchmark with ∼115 test images per category. IM-IAD [16] already argued the same point in 2024. The corollary we adopt as a methodological premise: a 2026 paper reporting +0.3% I-AUROC over a 2022 baseline on MVTec-AD is communicating less than its numbers suggest. PRO [15] and bound-FPR metrics are not saturated—best 2026 PRO sits at 94–96%—and correlate better with deployment behavior.

8.2. The New Benchmark Catalog

The 2024–2026 benchmark corpus produces, by our count, more than twenty new IAD benchmarks—the largest three-year output in the field's history. Table 5 groups them by the deployment failure mode each was built to expose; methods claiming to address a failure mode should report on benchmarks from the corresponding row.

 doi:10.20944/preprints202606.0719.v1

Table 5. Candidate post-MVTec benchmarks from the 2024–2026 IAD corpus, grouped by the deployment failure mode each was built to expose. *Pri. task* columns: 2D detection, 3D point-cloud, RGB-D, MV (multi-view), Text (image–text), VQA (multimodal QA). Methods claiming to address a given failure mode should report on at least one benchmark in the corresponding row.

Failure mode exposed	Recommended benchmarks	Pri. task	What is hard
Open-world product diversity	Kaputt [8]; Real-IAD Variety [156]; Real-IAD D^3 [157]; 3CAD [158]; PKU-GoodsAD [159]; MANTA [160]	2D, RGB-D, MV, Text	48k items, arbitrary poses, $\leq$3 refs; viewpoint+material variation, tiny objects, real 3C parts
Multi-domain breadth, "one model many domains"	ADNet (380 categories); MMR-AD [9]; Omni-AD [161]; M3-AD [48]; MMAD [44]	2D, VQA	380+ categories across 8+ domains; reflection-aware multi-dimensional
High-resolution / 3D realism	MiniShift [58]; 3D-ADAM [162]; IEC3D-AD [163]; SiM3D [59]; MulSen-AD [164]; Real-IAD D^3 [157]	3D / RGB-D	500k pts, $<$ 1% anomaly cov.; synth-to-real (CAD$\rightarrow$scan); single-instance multi-view
View-illumination interplay / harder 2D realism	M^2AD [165]; MVTec-AD 2 [166]; VISION Datasets [167]; Texture-AD [168]; HSS-IAD [169]	2D	illumination drift, harder logical anomalies, heterogeneous same-sort
Domain-specific specialized inspection	CPS2D-AD [170] (IC substrates); WFDD [171] (fabrics); InsPLAD [172] (power lines); CableInspect-AD [173]; CrashCar101; PeanutAD; CID; CXR-AD; RAD	2D	micrometer-scale defects; specialized appearance and supervision
MLLM / VQA / reasoning evaluation	MMAD [44]; MMR-AD [9]; M3-AD [48]; MAU-Set [98]; Chat-AD [49]; Anomaly-OV [42]	VQA	multiple-choice and free-form defect QA; reasoning consistency
Open-vocabulary defect understanding	IMDD-1M [10] (1.24M pairs, 63 domains); UniSpector [142]; MAU-Set [98]	2D, Text	language-grounded defect classes; cross-domain open-set
Synthesis evaluation (decoupled from detection)	ASBench [11]; Defect Spectrum [174]	2D	synthesis quality vs. downstream detection delta
Long-tail / online / deployment dynamics	LTOAD [175] (eight streaming configs)	2D	long-tailed online updates; class-agnostic concepts

The volume of new benchmarks is the strongest *quantitative* indicator that the field is in transition: a mature field consolidates around two or three benchmarks; a field in transition forks them to expose new failure modes.

Key numerical anchors. Five numbers from this catalog recur throughout the article and anchor the structural-shift claim; each is reported by the cited source paper (not recomputed by us):

- *MVTec-AD saturation* (Section 8.1): PatchCore's 99.1% I-AUROC (2022) is within 0.5 pp of the best 2026 numbers, inside the test-set noise floor [12,127,135,136].
- *Open-world ceiling*: Kaputt best unsupervised AUROC 56.96% under the prescribed $\leq$3-reference protocol over 48,376 unique items—no method clears 60% [8].
- *Cross-domain transfer wall*: UniSpector InsA drops from 69.1% to 14.1% AP_{50} between Real-IAD and 3CAD—a five-fold drop in the same model across imaging conditions [142].
- *Class-count scaling wall*: CCL drops from 90.6% I-AUROC in the all-in-all C=60 setting (MVTec+VisA+BTAD) to 65.2% on the COCO-derived COCOAD (C=81)—a sharp bend rather than gradual decay [135].
- *MLLM detection-vs-reasoning trade-off*: JUDO 65.04% Anomaly-Discrimination (base Qwen2.5-VL-7B 71.39%, $-$6.35 pp) after MMAD-targeted post-training; SALAD's 96.1% MVTec LOCO I-AUROC (no language component) tops any MLLM in our corpus at deployment FPR [46,106].

8.3. A Three-Tier Evaluation Recommendation

We recommend three evaluation tiers structured by what each diagnoses; a method topping one tier without reporting the others has not been shown to advance along the missing axes.

 doi:10.20944/preprints202606.0719.v1

Tier 1—Deployment-realistic open-world. Required for any open-world claim. Candidates: Kaputt [8], Real-IAD-Variety [156], Real-IAD D^3 [157], 3CAD [158], MANTA [160], M^2AD [165]—each releases a different MVTec-AD-era assumption (identity, viewpoint, tiny-object, illumination, real-3C distribution). A method beating one does not automatically beat the others; reporting a single Tier-1 number is uninformative. Methods must report AUROC and a bound-FPR metric. Quantitative anchor: no method clears 60% on Kaputt under the $\leq$3-reference protocol (Section 8.2), making progress measurable. Section 10.1 problem 1 is defined against this tier.

Tier 2—General-purpose multi-class/multi-domain. Required for any generality claim. Anchors: Real-IAD D^3 [157], Omni-AD [161], MVTec-AD 2 [166], Real-IAD [70]. Tier-2 floor sits in the 80–90% AUROC band on Real-IAD D^3 (PRO several pp lower)—multi-class progress is still measurable. MVTec-AD remains reportable but marked saturated. PRO and bound-FPR metrics required; methods training jointly on all evaluation domains (e.g., AnomalyMoE [127]) must disclose the joint training, not imply zero-data deployment. Section 10.2 problem 5 targets this tier.

Tier 3—Domain-specific specialized inspection. Required for any specialized capability claim. Heterogeneous candidates exposing distinct deployment regimes: CPS2D-AD [170] (IC-substrate); BraTS / BrainMRI (medical cross-domain); 3D-ADAM [162] / IEC3D-AD [163] / SiM3D [59] (high-res 3D, SiM3D's synth-to-real breaking every adapted baseline); LTOAD [175] (long-tail online); ASBench [11] (synthesis-vs-detection lift); CableInspect-AD / InsPLAD / WFDD / HSS-IAD (niche industries). *Open-vocabulary defect understanding*: IMDD-1M [10], UniSpector [142], MAU-Set [98]—evaluation frontier cutting across Section 3–6, with UniSpector's 69.1%→14.1% cross-domain mAP drop as the cleanest quantitative anchor. Headline metrics should match the target industry (e.g., micrometer-scale recall for IC, regulator-bound FPR for medical), not generic AUROC. The largest absolute gains live in Tier 3 (Section 9.8).

The honest summary of a method paper is a multi-row report card across Tier-1/2/3 benchmarks from Table 5—not a single number on a single benchmark.

Protocols and threshold rationale

The thresholds above are directional, not statistical; we make their derivation explicit so downstream papers can adjust them as evidence accrues.

Metric formulas. (a) *Bound-FPR AP*: average precision computed over the portion of the precision–recall curve with FPR ≤ 0.01 (false-positive rate at the operating threshold $\leq$ 1% of the normal set, the IM-IAD [16] recommendation). (b) *F1 at fixed precision*: F1-score at the operating threshold that yields precision ≥ 0.95 on a held-out validation split, mimicking the false-stop tolerance of an in-line production system. (c) *PRO*: mean of the per-region intersection-over-union restricted to the ground-truth defective regions, with the standard [15] pre-PRO area-under-curve normalization. Reporting AUROC alone is deployment-incompatible: it averages over operating points the line cannot tolerate.

Reference-budget rationale (Tier 1). Kaputt caps reference images at $\leq$3 per identity: with 48,376 unique items a larger budget is economically infeasible (a fulfillment-center conveyor cannot acquire 100 references per SKU). This is the published Kaputt value, not one we set.

Threshold rationale. All three numbers are *recommended challenge thresholds meant to make progress observable*, not acceptance criteria or statistical claims. The 60% Tier-1 value sits just above Kaputt's current best (56.96% AUROC), so clearing it marks a publication-worthy advance on the strongest open-world benchmark. The Section 10.1 deployment target (>80% AUROC plus >80% bound-FPR AP) follows from in-line stop tolerance: at 80% bound-FPR AP and a 1% defect rate, $\sim$1 false alarm per 99 normal items, comparable to human inspection. The 80–90% Tier-2 band on Real-IAD D^3 reflects the current state of the art (MaskAD, AnomalyMoE) and the range where multi-class progress stays measurable below MVTec saturation. All will need updating as the field reports a richer bound-FPR evidence base.

Benchmarks spanning tiers. A benchmark's tier follows the *protocol used*, not its identity: Real-IAD D^3 is Tier-1 under the $\leq$3-reference protocol but Tier-2 under the full-shot 30-domain protocol,

 doi:10.20944/preprints202606.0719.v1

and SiM3D is Tier-1 only under its synth-to-real protocol, else Tier-3. Beating Real-IAD D^3 full-shot does not address the open-world failure mode Tier-1 diagnoses.

8.4. Summary: Evaluation Is The Bottleneck

The five mechanisms (Section 3–7) are advancing, but the benchmarks they have been evaluated on are no longer the right tools to measure the advance. The three-tier recommendation in Section 8.3 (anchored in Table 5) is our most actionable suggestion. Section 9 catalogs the cross-cutting failure modes that no single mechanism's evaluation tier yet resolves.

9. Cross-Cutting Bottlenecks and Failure Modes

Each Section 3–7 chapter closes with an honest-assessment subsection; this section aggregates them. Eight bottlenecks recur across mechanisms, no paper in our corpus resolves more than one at a time, and they are structurally coupled in ways single-mechanism advances cannot address: real method-level progress (Section 3–7) coexists with evaluation infrastructure that has not kept up (Section 8) and eight problems still unsolved at deployment-relevant operating points (Figure 7).

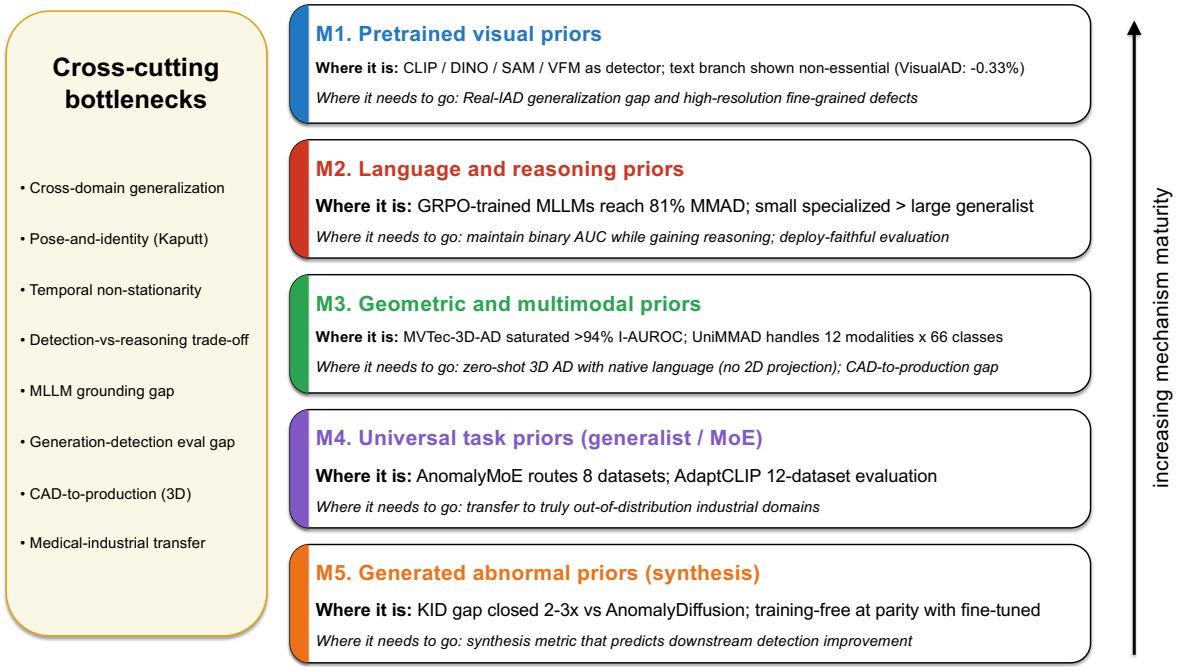

Figure 7. Cross-cutting bottlenecks aggregated across the five mechanisms. Left: the eight recurring bottlenecks (§9.1–§9.8). Right: for each mechanism M1–M5, *where it is* (strongest result to date) and *what it still needs* (closest unresolved bottleneck), along an increasing-maturity axis. The pattern is consistent: each mechanism has one empirical anchor at the frontier and at least one bottleneck no paper in its family addresses.

9.1. The Cross-Domain Generalization Gap

Every chapter contains the same finding at a different scale: the 2024–2026 methods generalize within curated benchmark suites but break under genuine domain shift.

- Section 3—VisualAD [86] and CLIP-adapters (AdaptCLIP [72], FAPrompt [73]) degrade moving from industrial to medical benchmarks; AnomalyVFM [82]'s 94.1% zero-shot drops to 88.0% on Real-IAD—still industrial.
- Section 4—GRPO-trained MLLM-AD (IAD-R1 [45], JUDO [46], AD-FM [47]) all train and evaluate on MMAD or derivatives; no held-out industrial domain reported.
- Section 5—IB-IUMAD and SiM3D [59] explicitly diagnose cross-modal and synth2real gaps.
- Section 6—UniSpector [142]'s in-domain mAP 69.1% → 14.1% cross-domain is the cleanest single demonstration; AnomalyMoE [127] requires joint training on all 8 datasets.

 doi:10.20944/preprints202606.0719.v1

- Section 7—SynSur [148] is the only synthesis paper on non-MVTec/VisA; +1.4 pp gain on MVTec-AD collapses to +0.015 AP on BSData.

The cross-domain gap is the article's most consistent finding and the structural reason "paradigm shift" is the wrong framing.

9.2. The Pose-And-Identity-Variation Problem

Kaputt [8] pairs 48,376 unique product items with arbitrary poses and $\leq$3 reference images each; PatchCore drops from 99.1% on MVTec-AD to 54.69% on Kaputt, and no method clears 60%. The bottleneck couples *pose variation* (breaking per-pose normality in memory banks) with *identity variation* (violating the assumption that learned normality covers the test-time category). Few-reference learning at this scale needs either pose-invariant features beyond CLIP/DINO/VFM or a non-image supervision modality (e.g., product-catalog descriptions)—the structural promise of Section 4 MLLM-AD.

9.3. Temporal Non-Stationarity

Real production lines have non-stationary inputs—new SKUs without labels, distribution drift, supplier substitution—that the static benchmark monoculture does not test. LTOAD [175] attacks this directly with +4.63% offline I-AUROC but leaves online deployment largely open; ReplayCAD [176] addresses continual class growth without scaling to fulfilment-style daily SKU churn; MeDS [177] attacks the related noise-robust setting with memory-distilled selection, reaching 99.16% MVTec-AD I-AUROC at 40% training-data contamination and SOTA on VisA / Real-IAD noisy variants. No single corpus solution suffices; Section 7 synthesis has plausible leverage—realistic defects for newly-introduced categories from a single normal image could keep an online detector calibrated as the catalogue grows.

9.4. The Detection-Vs-Reasoning Trade-Off In MLLM-AD

JUDO [46] reports binary anomaly-discrimination dropping from 71.39% (base Qwen2.5-VL) to 65.04% after GRPO post-training on MMAD reasoning ($-$6.35 pp on the easiest subtask); IAD-R1 [45] reports similar degradation. The pattern is consistent: GRPO-trained reasoners over-verbalise because the gradient favours plausible reasoning chains over the binary answer at the deployment operating point. No corpus paper has a clean solution; Section 10 Problem 2 targets it.

9.5. The MLLM Grounding Gap

A second Section 4 failure mode: MLLMs produce plausible language outputs not visually grounded. AD-Copilot's [49] Comparison Encoder ablation lifts MMAD bounding-box IoU 3.35$\times$, yet its MMAD-overall accuracy ($\sim$82.3%) is only modestly above strong general-purpose visual MLLMs (GPT-4o, Qwen2.5-VL-72B; $\sim$75%)—the absolute gain from anomaly-specific training stays small. Triad [97] and SAGE [104] attack it with ROI tokens and entropy-aware fact-enhancement, but gains stay incremental.

9.6. The Generation–Detection Evaluation Gap

ASBench [11]: across 19,680 configurations spanning twelve synthesis methods, FID/SSIM/IS/LPIPS do not predict downstream detection improvement, and the lowest-FID method does not yield the largest I-AUROC lift. Combined with (a) downstream lift not generalising beyond MVTec-AD/VisA (SynSur on BSData, +0.015 AP) and (b) synthesis methods rarely reporting on harder benchmarks, *synthesis usefulness for deployment* has not been measured by anyone in our corpus. Section 10 Problem 4 targets it.

9.7. The CAD-to-production Gap (3D)

SiM3D [59] introduced the first benchmark requiring training on CAD models and testing on real scans. Existing 3D AD methods [50,118,178] drop to near-baseline under synth2real shift, and the closing strategies known for RGB synth2real (domain randomisation, gradient reversal) do not

 doi:10.20944/preprints202606.0719.v1

transfer cleanly to 3D point clouds. PASDF [110]'s continuous SDF representation is the closest 2026 candidate but is not evaluated on the SiM3D synth2real protocol—the most under-served research gap in the 3D AD literature.

9.8. The Medical-Industrial Transfer Gap

PDD [20] achieves +11.8% AUROC on HeadCT, +5.1% on BrainMRI, and +8.5% on ZhangLab; FDP [21] achieves +17.63% DICE on BraTS20. These are among the largest single-paper gains in the entire IAD literature reviewed here—yet they happen on medical benchmarks where MVTec-AD-saturated methods do not compete, and use *none* of the foundation-model mechanisms reviewed here. Foundation models are one tool, not the toolkit.

9.9. Failure-Mode Summary Matrix

Table 6. Cross-cutting bottleneck × mechanism matrix. × = applies directly; (×) = derivative/indirect form;—= does not apply. M1–M5 are the five mechanisms (Section 3–7). Three bottlenecks (cross-domain generalization, temporal non-stationarity, medical-industrial transfer) affect every column directly; five more affect multiple mechanisms. The detection-vs-reasoning trade-off (Section 9.4) is most explicit in M2 GRPO-trained MLLMs, but M4 generalist VQA+detection heads inherit the same gradient imbalance, and M5 detectors on synthetic "explainable" artifacts can learn generation cues rather than real defect cues.

Bottleneck	**M1** Visual	**M2** Reasoning	**M3** Geometric	**M4** Universal	**M5** Synthesis
Cross-domain generalization (Section 9.1)	×	×	×	×	×
Pose-and-identity (Section 9.2)	×	×	—	×	—
Temporal non-stationarity (Section 9.3)	×	×	×	×	×
Detection-vs-reasoning (Section 9.4)	—	×	—	(×)	(×)
MLLM grounding (Section 9.5)	—	×	—	—	—
Generation–detection evaluation (Section 9.6)	—	—	—	—	×
CAD-to-production (Section 9.7)	—	—	×	—	×
Medical-industrial transfer (Section 9.8)	×	×	×	×	×

No paper resolves more than one column simultaneously. The cross-domain and temporal-non-stationarity rows in particular are universal failures of the foundation-model generalization mechanisms—which directly contradicts the strong-version "paradigm shift to open-world IAD" narrative. Section 10 translates this matrix into a concrete five-problem agenda.

10. Outlook and Research Agenda

We close with a five-problem agenda—each linking a Section 9 bottleneck to a measurable target—plus structural changes that make progress trackable, built on Figure 7.

10.1. Five Concrete Problems

Problem 1: Open-world performance on Tier-1 benchmarks

Gap (Section 9.1+9.2): no method clears 60% unsupervised AUROC on any Tier-1 benchmark. *Target:* >80% AUROC and AP@bound-FPR on ≥2 Tier-1 benchmarks with ≤3 reference images per identity. *Leverage:* Section 3 pure-VFM detectors, Section 4 MLLM-AD with product-catalog reasoning, Section 6 meta-learning with abnormal references.

Problem 2: Maintain detection AUC while gaining MLLM reasoning accuracy

Gap (Section 9.4+9.5): GRPO post-training regresses MLLM binary detection below the base model, far short of PatchCore (99.1% MVTec-AD, 91.2% VisA I-AUROC). *Target:* an MLLM matching PatchCore binary detection on VisA while exceeding 80% MMAD. *Leverage:* Section 4 head-decoupling (frozen detector backbone + MLLM reasoning head); AD-Copilot's Comparison Encoder is the closest precedent.

 doi:10.20944/preprints202606.0719.v1

Problem 3: Zero-shot 3D AD with language grounding (no 2D projection)

Gap (Section 9.7): no zero-shot 3D detector grounds language without 2D projection; direct-3D BTP reaches 84.5% point-AUROC but only 61.4% O-AUROC. *Target:* >85% O-AUROC and >90% point-AUROC via direct 3D + language. *Leverage:* Section 5+3 PLM fine-tuned on IMDD-1M.

Problem 4: Predictive synthesis-quality metric

Gap (Section 9.6): generation-quality metrics (FID/SSIM/IS/LPIPS) do not predict downstream detection lift. *Target:* a metric $M(\cdot)$ over (synthetic, real, detector) with rank correlation > 0.7 across $\geq$3 detectors. *Leverage:* measure boundary realism in the detector's own representation—no new generator needed, only careful study of ASBench's matrix.

Problem 5: Generalist with >90% AUROC on a held-out industrial domain

Gap (Section 9.1+9.8): "generalist" methods are trained jointly on their evaluation domains and collapse under genuine domain shift. *Target:* >90% I-AUROC on an industrial domain not seen in training, without target-domain fine-tuning. *Leverage:* Section 3+6+7 combined; plausibly an industrial-defect-grounded foundation model that does not yet exist. Solving this retires the "paradigm shift is overclaimed" critique.

10.2. Three Structural Recommendations

R1: Adopt three-tier evaluation (Section 8.3)

Report against the Section 8.3 tiers—Tier-1 for open-world claims, Tier-2 for generality, Tier-3 for specialized capability—with MVTec-AD marked saturated. Reviewer-facing: treat methods reporting only on MVTec-AD or VisA as not validated for the open-world setting they claim—benchmark adoption follows reviewer expectation.

R2: Mandate tripartite synthesis evaluation

For any anomaly-synthesis paper, require the Section 9.6 tripartite report—generation quality (FID/KID on held-out real anomalies), mask-fidelity (IoU), and downstream-detection delta on a fixed detector. It costs nothing to adopt, aligns with ASBench, and forecloses the "pretty pictures, no downstream lift" failure mode within one review cycle.

R3: Cross-group governance for MLLM-AD benchmarks

Concrete proposals: blinded leaderboards, community-curated held-out test sets, mandatory pre-registration of training data and prompts. MMAD, MMR-AD, M3-AD should be governed by a multi-group consortium, not single-group benchmark/method co-development. First move: a held-out MMAD test set curated by groups *not* publishing on it—made in light of the disclosed MMAD authorship (Disclosures), on which Section 4's long-term credibility depends.

10.3. Adjacent Emerging Directions IAD Will Likely Absorb Next

Several near-term directions align with Section 9's bottlenecks—one IAD-native, the rest borrowed from the broader AI landscape. Persistent caveat: production inspection runs at <100 ms/image, so methods one or two orders of magnitude slower default to offline or root-cause analysis, not in-line detection (EfficientAD [25] remains the baseline).

Open-vocabulary defect understanding (IAD-native). The clearest near-term chapter: IMDD-1M [10] supplies the million-scale image-text substrate, UniSpector [142] the cross-domain failure diagnostic, and MAU-Set [98] multi-domain VQA evaluation; a 2026–2027 paper combining all three with a Tier-1 benchmark like Kaputt would be the clearest single demonstration of foundation-model open-world IAD yet, likely within 12–18 months.

Embodied / VLA active inspection. OpenVLA / RT-2 / Pi-0 / Helix map pixels+language to manipulator actions. Kaputt's pose-and-identity problem (Section 9.2) is structurally an active-perception task, which no current IAD system performs—detectors and agents alike stay perceptually

 doi:10.20944/preprints202606.0719.v1

passive over fixed views, with no active viewpoint control. Likely 2026–2027 contribution: an industrial-VLA trained on inspection trajectories. Latency of 5–50 Hz fits cell-based or scan-on-demand inspection, not 30k-units/hr conveyor; sim2real is the blocker.

Unified generative–perception models. Show-o / Transfusion / Janus-Pro / Emu3 / Chameleon collapse synthesis and understanding into one backbone, structurally merging Section 3+4+7. ARAS [147] (autoregressive Infinity-8B) and AD-Copilot's [49] Comparison Encoder are early signals; the detection head factors out as a 50–200 ms forward pass while reasoning and synthesis heads stay lazy or offline. First likely instance: an industrial fine-tune of Show-o or Janus-Pro on IMDD-1M + MMR-AD.

Reasoning distilled into fast detectors. o1 / DeepSeek-R1 / K1.5 / Gemini-Thinking established inference-time CoT scaling, incompatible with deployment (30 s/image breaks throughput). The viable IAD adaptation is R1-Distill: generate traces offline, distill into a small fast model. IAD imported GRPO (Section 4.2) but not distillation; offline root-cause analysis at 30 s/image is the complementary niche.

World models for industrial scenes. Genie-2 / Cosmos / Sora 1.5 simulate physically plausible visual sequences, attacking Section 9.3 temporal non-stationarity and synthesis-for-new-SKUs structurally; throughput is not binding because world models run offline. No IAD paper yet trains an industrial world model; SiM3D's [59] CAD-to-real synth2real is the closest precedent (Section 9.7).

Multi-agent orchestrated inspection. A fast detector, a synthesis module, and a reasoning MLLM collaborate via tool-use; AnomalyClaw [179], which invokes a catalog of vision tools and frozen expert probes through multi-turn refutation, is the IAD-adjacent representative. Throughput is clean—only the fast detector runs in-line, slow modules invoke on-demand. Interpretability/modularity vs. single-pass efficiency is the trade-off against unified models; both lines should coexist for 12–18 months.

Industrial video AD. The still-image scope here excludes a distinct line where the anomaly is a temporal or physical-dynamics event no single frame resolves: IPAD [180], Phys-AD [181] (with a VLM-oriented PAEval metric), and Lab-VAD [182]. These belong with the Section 9.3 temporal axis and an as-yet-unformalised physics-prior; we flag them as a future-iteration extension, not a co-equal mechanism.

Confidence. Most likely 2027 mainstream: unified generative-perception and reasoning-distilled-fast-models. *Likely deployment niches:* VLA (robotic cells, off-line teardown) and world models (offline training-data synthesis). *Product-architecture-not-research:* multi-agent orchestration. The five-mechanism partition will not survive 2027 unchanged: reorganisation around whole-system axes (embodied, unified, agentic) is more likely than a sixth column.

10.4. On The "Paradigm Shift" Framing

Our reading of the evidence:

- The shift is *real* at the method level for at least three of the five mechanisms: SFT→RL in Section 4; pure-VFM-replaces-CLIP-text in Section 3; the MoE convergent design pattern in Section 6; the training-free/VLM-grounded synthesis explosion in Section 7; the language-grounded industrial-defect substrate (IMDD-1M) in Section 5.
- The shift is *not yet real* at the deployment level: none of the five mechanisms beats classical methods (PatchCore, EfficientAD, DRAEM) at deployment-faithful operating points on hard benchmarks.
- The shift is *partially circular* in evaluation infrastructure (Section 4.4 MMAD/AD-Copilot overlap; Section 8.1 MVTec saturation; Section 9.6 metric mismatch).

We therefore reserve "paradigm shift" for the post-Kaputt, post-cross-domain era and name what the wave has done instead as *five emerging generalization mechanisms*. The stronger framing becomes defensible only when independent groups (not a benchmark's own authors) clear at least three of the five Section 10.1 Problems—most plausibly Problems 2 and 4, both within reach in 12–18 months—plus an MMAD result reproduced and exceeded by a non-author group (cf. R3). The verdict is unchanged:

 doi:10.20944/preprints202606.0719.v1

progress is real at the method level, not yet at deployment, and the Section 10.2 structural changes would close the gap faster than method-only advances.

10.5. Deployment Economics And The Certification Frontier

The five-mechanism axis reasons about methods independently of deployment cost. Four cost axes the literature underspecifies constrain which mechanism survives at which scale.

Per-line inference cost. EfficientAD's sub-100 ms is the baseline; 7B MLLM-AD is an order of magnitude higher per image. Inspecting 10^4–10^7 images/line/day favours small distilled models or cascaded triage; any "deployment-realistic" claim without cost-per-image is incomplete.

Edge vs. cloud, data sovereignty. Defect data is proprietary, so methods depending on GPT-4V/Gemini APIs fail this constraint. On-prem MLLM needs data-centre hardware most manufacturers lack, so the trend to small distilled MLLMs is partly sovereignty-driven.

Certification and audit trails. Automotive (AEC-Q), medical (FDA/CE-MDR), aerospace (DO-178C), and food (HACCP) require traceable provenance plus frozen-model deployment. MLLM-AD chain-of-thought helps audit trails, but its sampling non-determinism creates a certification problem no IAD paper addresses; deterministic distilled detectors certify far more easily than generative reasoners.

Long-tail SKU onboarding. High-mix manufacturing (consumer electronics, contract assembly) is governed by cost-per-new-SKU. Few-shot foundation-model methods (Section 3+6) attack this axis structurally; evaluation should report onboarding time per new SKU alongside AUROC. Together these axes form a structural gap likeliest closed by industry-academia collaboration, not method papers alone.

11. Conclusions

The 2024–2026 IAD wave is best read as *five emerging generalization mechanisms* layered onto the established baseline, not as a single paradigm shift. Each mechanism—visual, reasoning, geometric/multimodal, universal, and synthesis priors—produced real method-level advances against a specific limitation of the closed-set, normal-only era, and the cohort also produced a larger benchmark fork than the field has seen in any prior three-year window. Yet the deployment-level evidence is not yet there: no method clears 60% unsupervised AUROC on Kaputt, no MLLM-AD method matches PatchCore at industrial false-positive rates, no generalist transfers cleanly to a held-out industrial domain, and Section 9 catalogues eight cross-cutting bottlenecks no current paper resolves more than one at a time.

We therefore frame the five mechanisms as the field's current best response to the harder question *what replaces category-specific normal-only training?*, and the Section 10 five-problem agenda as the concrete evidence threshold under which a stronger paradigm-shift framing becomes defensible. Progress will be faster if the next 12–18 months target Tier-1 deployment benchmarks, cross-group MLLM-AD evaluation infrastructure, and tripartite synthesis evaluation rather than further MVTec-AD saturation.

Disclosures.

Several cited works originate in the authors' research group and should be read with that provenance in mind: the MMAD benchmark [44], the AD-Copilot assistant [49], the Anomaly-Claw agent [179], the earlier IAD survey [2], the IM-IAD benchmark and deployment-realistic protocol [3,16], and the Real3D-AD dataset [54]; the group also maintains the curated IAD list at https://github.com/M-3LAB/awesome-industrial-anomaly-detection used as the corpus index in Section 1.5. Other resources come from outside the group, used on methodological merit—e.g., Real-IAD [70] (a collaborators' dataset) and ASBench [11] (a related group's)—where independent outside re-evaluation would further strengthen the conclusions.

 doi:10.20944/preprints202606.0719.v1

References

1. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. MVTec AD – A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9592–9600. https://doi.org/10.1109/CVPR.2019.00982.
2. Liu, J.; Xie, G.; Wang, J.; Li, S.; Wang, C.; Zheng, F.; Jin, Y. Deep Industrial Image Anomaly Detection: A Survey. *Machine Intelligence Research* **2024**, *21*, 104–135. https://doi.org/10.1007/s11633-023-1459-z.
3. Xie, G.; Wang, J.; Liu, J.; Lyu, J.; Liu, Y.; Wang, C.; Zheng, F.; Jin, Y. IM-IAD: Industrial Image Anomaly Detection Benchmark in Manufacturing. *IEEE Transactions on Cybernetics* **2024**, *54*, 2720–2733. Benchmark+protocol paper, not a survey, https://doi.org/10.1109/TCYB.2024.3357213.
4. Cao, Y.; Xu, X.; Zhang, J.; Cheng, Y.; Huang, X.; Pang, G.; Shen, W. A Survey on Visual Anomaly Detection: Challenge, Approach, and Prospect, 2024, [arXiv:cs.CV/2401.16402]. Cite key retained as 'xieanomaly-survey2024' for backward compatibility; actual first author is Cao, https://doi.org/10.48550/arXiv.2401.16402.
5. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021, Vol. 139, *Proceedings of Machine Learning Research*, pp. 8748–8763.
6. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.V.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research* **2024**.
7. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2023, pp. 4015–4026.
8. Höfer, S.; Henning, D.F.; Amiranashvili, A.; Morrison, D.; Tzes, M.; Posner, I.; Matvienko, M.; Rennola, A.; Milan, A. Kaputt: A Large-Scale Dataset for Visual Defect Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 24224–24233.
9. Yao, X.; Qian, Z.; Shi, C.; Song, J.; Zhang, C. MMR-AD: A Large-Scale Multimodal Dataset for Benchmarking General Anomaly Detection with Multimodal Large Language Models. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 43072–43082.
10. Ni, T.C.; Chen, C.C.; Yang, Y.F. Towards Open-Vocabulary Industrial Defect Understanding with a Large-Scale Multimodal Dataset. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 13059–13068.
11. Zhang, Q.; Zhang, S.; Liu, J.; Wang, J.; Lei, X.; Xie, G.; Jiang, G.; Lu, Z. ASBench: Image Anomalies Synthesis Benchmark for Anomaly Detection. *IEEE Transactions on Artificial Intelligence* **2026**. Accepted for publication.
12. Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; Gehler, P. Towards Total Recall in Industrial Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 14318–14328. https://doi.org/10.1109/CVPR52688.2022.01392.
13. Rolih, B.; Fučka, M.; Skočaj, D. SuperSimpleNet: Unifying Unsupervised and Supervised Learning for Fast and Reliable Surface Defect Detection. In Proceedings of the Pattern Recognition: 27th International Conference, ICPR 2024, Proceedings, Part X. Springer, 2025, Vol. 15310, *Lecture Notes in Computer Science*, pp. 47–65. https://doi.org/10.1007/978-3-031-78192-6_4.
14. Jeong, J.; Zou, Y.; Kim, T.; Zhang, D.; Ravichandran, A.; Dabeer, O. WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 19606–19616. https://doi.org/10.1109/CVPR52729.2023.01878.
15. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. Uninformed Students: Student-Teacher Anomaly Detection With Discriminative Latent Embeddings. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020, pp. 4183–4192. https://doi.org/10.1109/CVPR42600.2020.00424.
16. Xie, G.; Wang, J.; Liu, J.; Lyu, J.; Liu, Y.; Wang, C.; Zheng, F.; Jin, Y. IM-IAD: Industrial Image Anomaly Detection Benchmark in Manufacturing. *IEEE Transactions on Cybernetics* **2024**, *54*, 2720–2733. https://doi.org/10.1109/TCYB.2024.3357213.

 doi:10.20944/preprints202606.0719.v1

17. Lin, Y.; Chang, Y.; Tong, X.; Yu, J.; Liotta, A.; Huang, G.; Song, W.; Zeng, D.; Wu, Z.; Wang, Y.; et al. A survey on RGB, 3D, and multimodal approaches for unsupervised industrial image anomaly detection. *Information Fusion* **2025**, *121*, 103139. https://doi.org/10.1016/j.inffus.2025.103139.
18. Liang, H.; Guo, B.; Huang, Y.; Lyu, J.; Gao, C.; Cao, Y.; Wang, J.; Yu, R.; Shen, L.; Li, P. 3D Anomaly Detection: A Survey. ResearchGate preprint, 2025. Living survey accompanying M-3LAB awesome-3d-anomaly-detection repo, https://doi.org/10.13140/RG.2.2.21218.39361.
19. Wang, Y.; Xu, X.; Liu, J.; Lei, X.; Xie, G.; Jiang, G.; Lu, Z. A Survey on Industrial Anomalies Synthesis, 2025, [arXiv:cs.CV/2502.16412]. https://doi.org/10.48550/arXiv.2502.16412.
20. Lu, X.; Liu, H.; Shang, F.; Hui, Y.; Wan, L. PDD: Manifold-Prior Diverse Distillation for Medical Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 28534–28544.
21. Li, H.; Zhuang, Z.; Lin, J.; Liu, Y.; Chen, Y.; Peng, Q.; Yu, L.; Wang, L. FDP: A Frequency-Decomposition Preprocessing Pipeline for Unsupervised Anomaly Detection in Brain MRI. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 6118–6126. https://doi.org/10.1609/aaai.v40i8.37536.
22. Cheng, Y.; Cao, Y.; Yao, H.; Luo, W.; Jiang, C.; Zhang, H.; Shen, W. A comprehensive survey for real-world industrial surface defect detection: Challenges, approaches, and prospects. *Journal of Manufacturing Systems* **2026**, *84*, 152–172. https://doi.org/10.1016/j.jmsy.2025.11.022.
23. Bachem, O.; Lucic, M.; Krause, A. Practical Coreset Constructions for Machine Learning, 2017, [arXiv:stat.ML/1703.06476].
24. Deng, H.; Li, X. Anomaly Detection via Reverse Distillation From One-Class Embedding. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 9737–9746.
25. Batzner, K.; Heckler, L.; König, R. EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 2024, pp. 127–137. https://doi.org/10.1109/WACV57701.2024.00020.
26. Zavrtanik, V.; Kristan, M.; Skočaj, D. DRAEM - A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, pp. 8330–8339. https://doi.org/10.1109/ICCV48922.2021.00822.
27. Zavrtanik, V.; Kristan, M.; Skočaj, D. DSR - A Dual Subspace Re-Projection Network for Surface Anomaly Detection. In Proceedings of the Computer Vision – ECCV 2022. Springer, 2022, Vol. 13691, *Lecture Notes in Computer Science*, pp. 539–554. https://doi.org/10.1007/978-3-031-19821-2_31.
28. Liu, Z.; Zhou, Y.; Xu, Y.; Wang, Z. SimpleNet: A Simple Network for Image Anomaly Detection and Localization. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2023, pp. 20402–20411.
29. You, Z.; Cui, L.; Shen, Y.; Yang, K.; Lu, X.; Zheng, Y.; Le, X. A Unified Model for Multi-class Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2022, Vol. 35, pp. 4571–4584.
30. Lu, R.; Wu, Y.; Tian, L.; Wang, D.; Chen, B.; Liu, X.; Hu, R. Hierarchical Vector Quantized Transformer for Multi-class Unsupervised Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2023, Vol. 36.
31. Gao, B.B. Learning to Detect Multi-class Anomalies with Just One Normal Image Prompt. In Proceedings of the Computer Vision – ECCV 2024. Springer, 2024, pp. 454–470. https://doi.org/10.1007/978-3-031-72855-6_26.
32. Gao, B.B. MetaUAS: Universal Anomaly Segmentation with One-Prompt Meta-Learning. In Proceedings of the Advances in Neural Information Processing Systems, 2024, Vol. 37.
33. Li, X.; Zhang, Z.; Tan, X.; Chen, C.; Qu, Y.; Xie, Y.; Ma, L. PromptAD: Learning Prompts with only Normal Samples for Few-Shot Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 16838–16848. https://doi.org/10.1109/CVPR52733.2024.01594.
34. Zhou, Q.; Pang, G.; Tian, Y.; He, S.; Chen, J. AnomalyCLIP: Object-Agnostic Prompt Learning for Zero-Shot Anomaly Detection. In Proceedings of the International Conference on Learning Representations (ICLR), 2024.
35. Cao, Y.; Zhang, J.; Frittoli, L.; Cheng, Y.; Shen, W.; Boracchi, G. AdaCLIP: Adapting CLIP with Hybrid Learnable Prompts for Zero-Shot Anomaly Detection. In Proceedings of the Computer Vision – ECCV

 doi:10.20944/preprints202606.0719.v1

2024. Springer Nature Switzerland, 2024, Vol. 15093, *Lecture Notes in Computer Science*, pp. 55–72. https://doi.org/10.1007/978-3-031-72761-0_4.

36. Qu, Z.; Tao, X.; Prasad, M.; Shen, F.; Zhang, Z.; Gong, X.; Ding, G. VCP-CLIP: A Visual Context Prompting Model for Zero-Shot Anomaly Segmentation. In Proceedings of the Computer Vision – ECCV 2024, 2024, Vol. 15127, *Lecture Notes in Computer Science*, pp. 301–317. https://doi.org/10.1007/978-3-031-72890-7_18.
37. Cao, Y.; Xu, X.; Cheng, Y.; Sun, C.; Du, Z.; Gao, L.; Shen, W. Personalizing Vision-Language Models With Hybrid Prompts for Zero-Shot Anomaly Detection. *IEEE Transactions on Cybernetics* **2025**, *55*, 1917–1929. https://doi.org/10.1109/TCYB.2025.3536165.
38. Li, S.; Cao, J.; Ye, P.; Ding, Y.; Tu, C.; Chen, T. ClipSAM: CLIP and SAM collaboration for zero-shot anomaly segmentation. *Neurocomputing* **2025**, *618*, 129122. https://doi.org/10.1016/j.neucom.2024.129122.
39. Gu, Z.; Zhu, B.; Zhu, G.; Chen, Y.; Tang, M.; Wang, J. AnomalyGPT: Detecting Industrial Anomalies Using Large Vision-Language Models. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2024, Vol. 38, pp. 1932–1940. https://doi.org/10.1609/aaai.v38i3.27963.
40. Cao, Y.; Xu, X.; Sun, C.; Huang, X.; Shen, W. Towards Generic Anomaly Detection and Understanding: Large-scale Visual-linguistic Model (GPT-4V) Takes the Lead, 2023, [arXiv:cs.CV/2311.02782].
41. Zhang, Y.; Cao, Y.; Xu, X.; Shen, W. LogiCode: An LLM-Driven Framework for Logical Anomaly Detection. *IEEE Transactions on Automation Science and Engineering* **2025**, *22*, 7712–7723. https://doi.org/10.1109/TASE.2024.3468464.
42. Xu, J.; Lo, S.Y.; Safaei, B.; Patel, V.M.; Dwivedi, I. Towards Zero-Shot Anomaly Detection and Reasoning with Multimodal Large Language Models. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 20370–20382.
43. Deng, H.; Luo, H.; Zhai, W.; Guo, Y.; Cao, Y.; Kang, Y. VMAD: Visual-Enhanced Multimodal Large Language Model for Zero-Shot Anomaly Detection. *IEEE Transactions on Automation Science and Engineering* **2026**, *23*, 3607–3618. https://doi.org/10.1109/TASE.2025.3591656.
44. Jiang, X.; Li, J.; Deng, H.; Liu, Y.; Gao, B.B.; Zhou, Y.; Li, J.; Wang, C.; Zheng, F. MMAD: A Comprehensive Benchmark for Multimodal Large Language Models in Industrial Anomaly Detection. In Proceedings of the The Thirteenth International Conference on Learning Representations (ICLR), 2025.
45. Li, Y.; Cao, Y.; Liu, C.; Xiong, Y.; Dong, X.; Huang, C. IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence* **2026**, *40*, 6583–6591. https://doi.org/10.1609/aaai.v40i8.37588.
46. Kang, H.; Lee, W.; Kim, J.; Park, H. JUDO: A Juxtaposed Domain-Oriented Multimodal Reasoner for Industrial Anomaly QA. In Proceedings of the International Conference on Learning Representations (ICLR), 2026.
47. Liao, J.; Su, Y.; Tu, R.C.; Jin, Z.; Sun, W.; Li, Y.; Xu, X.; Tao, D.; Yang, X. AD-FM: Multimodal LLMs for Anomaly Detection via Multi-Stage Reasoning and Fine-Grained Reward Optimization. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 15234–15242. https://doi.org/10.1609/aaai.v40i18.38548.
48. Huang, C.; Li, Y.; Cao, Y.; Wang, W.; Huang, H.; Wen, J.; Ren, W.; Cao, X. M3-AD: Reflection-aware Multi-modal, Multi-category, and Multi-dimensional Benchmark and Framework for Industrial Anomaly Detection, 2026, [arXiv:cs.LG/2603.00055].
49. Jiang, X.; Guo, Y.; Li, J.; Liu, Y.; Gao, B.B.; Deng, H.; Liu, J.; Zhao, H.; Wang, C.; Zheng, F. AD-Copilot: A Vision-Language Assistant for Industrial Anomaly Detection via Visual In-context Comparison, 2026, [arXiv:cs.CV/2603.13779].
50. Wang, Y.; Peng, J.; Zhang, J.; Yi, R.; Wang, Y.; Wang, C. Multimodal Industrial Anomaly Detection via Hybrid Fusion. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 8032–8041.
51. Pang, Y.; Wang, W.; Tay, F.E.H.; Liu, W.; Tian, Y.; Yuan, L. Masked Autoencoders for Point Cloud Self-Supervised Learning. In Proceedings of the Computer Vision – ECCV 2022. Springer, 2022, Vol. 13662, *Lecture Notes in Computer Science*, pp. 604–621. https://doi.org/10.1007/978-3-031-20086-1_35.
52. Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging Properties in Self-Supervised Vision Transformers. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, pp. 9630–9640. https://doi.org/10.1109/ICCV48922.2021.00951.
53. Cao, Y.; Xu, X.; Shen, W. Complementary Pseudo Multimodal Feature for Point Cloud Anomaly Detection. *Pattern Recognition* **2024**, *156*, 110761. https://doi.org/10.1016/j.patcog.2024.110761.

 doi:10.20944/preprints202606.0719.v1

54. Liu, J.; Xie, G.; Chen, R.; Li, X.; Wang, J.; Liu, Y.; Wang, C.; Zheng, F. Real3D-AD: A Dataset of Point Cloud Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2023, Vol. 36. Datasets and Benchmarks Track.
55. Bergmann, P.; Jin, X.; Sattlegger, D.; Steger, C. The MVTec 3D-AD Dataset for Unsupervised 3D Anomaly Detection and Localization. In Proceedings of the Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 5: VISAPP. SciTePress, 2022, pp. 202–213. https://doi.org/10.5220/0010865000003124.
56. Zhou, Q.; Yan, J.; He, S.; Meng, W.; Chen, J. PointAD: Comprehending 3D Anomalies from Points and Pixels for Zero-shot 3D Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2024, Vol. 37, pp. 84866–84896.
57. Xue, L.; Gao, M.; Xing, C.; Martín-Martín, R.; Wu, J.; Xiong, C.; Xu, R.; Niebles, J.C.; Savarese, S. ULIP: Learning a Unified Representation of Language, Images, and Point Clouds for 3D Understanding. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2023, pp. 1179–1189.
58. Cheng, Y.; Sun, Y.; Zhang, H.; Shen, W.; Cao, Y. Towards High-Resolution 3D Anomaly Detection: A Scalable Dataset and Real-Time Framework for Subtle Industrial Defects. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 3327–3334. https://doi.org/10.1609/aaai.v40i5.37328.
59. Costanzino, A.; Ramirez, P.Z.; Lella, L.; Ragaglia, M.; Oliva, A.; Lisanti, G.; Di Stefano, L. SiM3D: Single-instance Multiview Multimodal and Multisetup 3D Anomaly Detection Benchmark. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 20944–20953.
60. Zhao, Y.; Pang, Y.; Zhang, L.; Liu, H.; Zuo, J.; Lu, H.; Zhao, X. UniMMAD: Unified Multi-Modal and Multi-Class Anomaly Detection via MoE-Driven Feature Decompression. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 28502–28511.
61. Li, C.L.; Sohn, K.; Yoon, J.; Pfister, T. CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, pp. 9664–9674. https://doi.org/10.1109/CVPR46437.2021.00954.
62. Hu, T.; Zhang, J.; Yi, R.; Du, Y.; Chen, X.; Liu, L.; Wang, Y.; Wang, C. AnomalyDiffusion: Few-Shot Anomaly Image Generation with Diffusion Model. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2024, Vol. 38, pp. 8526–8534. https://doi.org/10.1609/aaai.v38i8.28696.
63. Zhang, X.; Xu, M.; Zhou, X. RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 16699–16708.
64. Hu, J.; Huang, Y.; Lu, Y.; Xie, G.; Jiang, G.; Zheng, Y.; Lu, Z. AnomalyXFusion: Multi-modal Anomaly Synthesis with Diffusion, 2024, [arXiv:cs.CV/2404.19444].
65. Xu, X.; Wang, Y.; Wang, J.; Lei, X.; Xie, G.; Jiang, G.; Lu, Z. FAST: Foreground-aware Diffusion with Accelerated Sampling Trajectory for Segmentation-oriented Anomaly Synthesis. In Proceedings of the Advances in Neural Information Processing Systems, 2025.
66. Dai, Z.; Zeng, S.; Liu, H.; Li, X.; Xue, F.; Zhou, Y. SeaS: Few-shot Industrial Anomaly Image Generation with Separation and Sharing Fine-tuning. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 23135–23144.
67. Zou, Y.; Jeong, J.; Pemula, L.; Zhang, D.; Dabeer, O. SPot-the-Difference Self-supervised Pre-training for Anomaly Detection and Segmentation. In Proceedings of the Computer Vision – ECCV 2022, 2022, Vol. 13690, *Lecture Notes in Computer Science*, pp. 392–408. https://doi.org/10.1007/978-3-031-20056-4_23.
68. Jezek, S.; Jonak, M.; Burget, R.; Dvorak, P.; Skotak, M. Deep Learning-Based Defect Detection of Metal Parts: Evaluating Current Methods in Complex Conditions. In Proceedings of the 2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2021, pp. 66–71. https://doi.org/10.1109/ICUMT54235.2021.9631567.
69. Mishra, P.; Verk, R.; Fornasier, D.; Piciarelli, C.; Foresti, G.L. VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization. In Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), June 2021, pp. 1–6. https://doi.org/10.1109/ISIE45552.2021.9576231.
70. Wang, C.; Zhu, W.; Gao, B.B.; Gan, Z.; Zhang, J.; Gu, Z.; Qian, S.; Chen, M.; Ma, L. Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection. In Proceedings of the

 doi:10.20944/preprints202606.0719.v1

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 22883–22892.

71. Li, W.; Xu, X.; Gu, Y.; Zheng, B.; Gao, S.; Wu, Y. Towards Scalable 3D Anomaly Detection and Localization: A Benchmark via 3D Anomaly Synthesis and A Self-Supervised Learning Network. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 22207–22216. https://doi.org/10.1109/CVPR52733.2024.02096.
72. Gao, B.B.; Zhou, Y.; Yan, J.; Cai, Y.; Zhang, W.; Wang, M.; Liu, J.; Liu, Y.; Wang, L.; Wang, C. AdaptCLIP: Adapting CLIP for Universal Visual Anomaly Detection. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 4095–4103. https://doi.org/10.1609/aaai.v40i6.42404.
73. Zhu, J.; Ong, Y.S.; Shen, C.; Pang, G. Fine-grained Abnormality Prompt Learning for Zero-shot Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 22241–22251. https://doi.org/10.1109/ICCV51701.2025.02065.
74. Qu, Z.; Tao, X.; Gong, X.; Qu, S.; Chen, Q.; Zhang, Z.; Wang, X.; Ding, G. Bayesian Prompt Flow Learning for Zero-Shot Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 30398–30408. https://doi.org/10.1109/CVPR52734.2025.02830.
75. Ma, W.; Zhang, X.; Yao, Q.; Tang, F.; Wu, C.; Li, Y.; Yan, R.; Jiang, Z.; Zhou, S.K. AA-CLIP: Enhancing Zero-Shot Anomaly Detection via Anomaly-Aware CLIP. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 4744–4754.
76. Shao, Y.; Wang, L.; Li, C.; Chen, P.; Liu, Q. PromptMoE: Generalizable Zero-Shot Anomaly Detection via Visually-Guided Prompt Mixtures. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 8878–8886. https://doi.org/10.1609/aaai.v40i11.37842.
77. Park, J.Y.; Seo, J.; Kang, M.; Park, Y.R. MoECLIP: Patch-Specialized Experts for Zero-shot Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 35534–35544.
78. Chen, Q.; Qu, Z.; Luo, W.; Yao, H.; Cao, Y.; Jiang, Y.; Duan, Y.; Luo, H.; Lv, C.; Zhang, Z. CoPS: Conditional Prompt Synthesis for Zero-Shot Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Findings, June 2026.
79. Hu, M.; Huo, Y.; Dou, M.; Yin, J.; Zhao, P.; Wang, Y.; Hu, C.; Hu, B.; Wang, Q. FB-CLIP: Fine-Grained Zero-Shot Anomaly Detection with Foreground-Background Disentanglement, 2026, [arXiv:cs.CV/2603.19608].
80. Li, X.; Xue, F.; Zhou, Y. MuSc-V2: Zero-Shot Multimodal Industrial Anomaly Classification and Segmentation with Mutual Scoring of Unlabeled Samples. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2026**. Early Access, https://doi.org/10.1109/TPAMI.2026.3688174.
81. He, J.; Cao, M.; Peng, S.; Xie, Q. RareCLIP: Rarity-aware Online Zero-shot Industrial Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 24478–24487.
82. Fučka, M.; Zavrtanik, V.; Skočaj, D. AnomalyVFM – Transforming Vision Foundation Models into Zero-Shot Anomaly Detectors. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 35555–35566.
83. Zhai, G.; Zhou, Y.; Deng, X.; Heckler-Kram, L.; Navab, N.; Busam, B. Foundation Visual Encoders Are Secretly Few-Shot Anomaly Detectors. In Proceedings of the International Conference on Learning Representations (ICLR), 2026.
84. Lendering, C.; Akdag, E.; Bondarau, E. SubspaceAD: Training-Free Few-Shot Anomaly Detection via Subspace Modeling. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 28557–28566.
85. Damm, S.; Laszkiewicz, M.; Lederer, J.; Fischer, A. AnomalyDINO: Boosting Patch-Based Few-Shot Anomaly Detection with DINOv2. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), February 2025, pp. 1319–1329. https://doi.org/10.1109/WACV61041.2025.00136.
86. Hou, Y.; Li, P.; Liu, Z.; Wang, Y.; Ruan, Y.; Qiu, J.; Xu, K. VisualAD: Language-Free Zero-Shot Anomaly Detection via Vision Transformer. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026.

 doi:10.20944/preprints202606.0719.v1

87. Xu, C.; Lv, C.; Chen, Q.; Zhang, F.; Zhang, Z. MRAD: Zero-Shot Anomaly Detection with Memory-Driven Retrieval. In Proceedings of the The Fourteenth International Conference on Learning Representations (ICLR), 2026.
88. Cai, M.; Zhang, Z.; Wu, G.; Chai, T.; Zhu, X. RAID: Retrieval-Augmented Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 21367–21378.
89. Tian, L.; Li, Y.; Dai, Y.; Chen, W.; Liu, X.; Chen, B. FastRef: Fast Prototype Refinement for Few-Shot Industrial Anomaly Detection, 2025, [arXiv:cs.CV/2506.21398].
90. Qu, Z.; Tao, X.; Gong, X.; Qu, S.; Zhang, X.; Wang, X.; Shen, F.; Zhang, Z.; Prasad, M.; Ding, G. DictAS: A Framework for Class-Generalizable Few-Shot Anomaly Segmentation via Dictionary Lookup. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 20519–20528.
91. Bhunia, A.; Li, C.; Bilen, H. Odd-One-Out: Anomaly Detection by Comparing with Neighbors. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 20395–20404.
92. Wang, F.; Zhang, T.; Wang, Y.; Qiu, Y.; Liu, X.; Guo, X.; Cui, Z. Distribution Prototype Diffusion Learning for Open-set Supervised Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 20416–20426. https://doi.org/10.1109/CVPR52734.2025.01901.
93. Qu, Z.; Tao, X.; Bao, X.; Wang, D.; Qu, S.; Zhang, Z.; Wang, X. AG-VAS: Anchor-Guided Zero-Shot Visual Anomaly Segmentation with Large Multimodal Models. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026, pp. 14126–14136.
94. Li, K.; Li, G.; Zhou, M.; Li, M.; Han, D.; Wan, J. Back to Point: Exploring Point-Language Models for Zero-Shot 3D Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 14167–14177.
95. Deng, Z.; Liu, A.; Wang, Y. GS-CLIP: Zero-shot 3D Anomaly Detection by Geometry-Aware Prompt and Synergistic View Representation Learning. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 35587–35596.
96. Jin, Y.; Feng, Y.; Zhang, J.; Wang, P.; Liu, Q.; Wang, Y. Reasoning-Driven Anomaly Detection and Localization with Image-Level Supervision. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026.
97. Li, Y.; Yuan, S.; Wang, H.; Li, Q.; Liu, M.; Xu, C.; Shi, G.; Zuo, W. Triad: Empowering LMM-based Anomaly Detection with Expert-guided Region-of-Interest Tokenizer and Manufacturing Process. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 21917–21926.
98. Wang, Z.; Fan, Z.; Tan, S.; Zhong, Y.; Yuan, Y.; Li, H.; Jiang, H.; Zhang, W.; Shao, F.; Wang, H.; et al. MAU-GPT: Enhancing Multi-type Industrial Anomaly Understanding via Anomaly-aware and Generalist Experts Adaptation. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 26787–26795. https://doi.org/10.1609/aaai.v40i31.39889.
99. Chao, Y.; Liu, J.; Tang, J.; Wu, G. AnomalyR1: A GRPO-based End-to-end MLLM for Industrial Anomaly Detection, 2025, [arXiv:cs.CV/2504.11914].
100. Zhang, K.; Zhang, Z.; Sun, X.; Wang, A.; Nie, J.; Chen, Q.; Hao, H.; Guo, J.; Zhang, J. ADSeeker: A Knowledge-Grounded Reasoning Framework for Industry Anomaly Detection and Reasoning. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026, pp. 21379–21388.
101. Chen, P.; Huang, C.; Cao, Y.; Liu, C.; Wang, W.; Wang, W.; Yang, M.; Shen, L.; Ren, W.; Cao, X. Towards Explainable Industrial Anomaly Detection via Knowledge-Guided Latent Reasoning, 2026, [arXiv:cs.CV/2602.09850].
102. Peng, X.; Huang, X.; Choi, S.H. EAGLE: Expert-Augmented Attention Guidance for Tuning-Free Industrial Anomaly Detection in Multimodal Large Language Models, 2026, [arXiv:cs.CV/2602.17419].
103. Li, Z.; Yu, Z.; Ye, Q.; Xie, W.; Zhuo, W.; Shen, L. IAD-GPT: Advancing Visual Knowledge in Multimodal Large Language Model for Industrial Anomaly Detection. *IEEE Transactions on Instrumentation and Measurement* **2025**, *74*, 1–12. https://doi.org/10.1109/TIM.2025.3635334.
104. Zang, G.; Li, X.; Di, D.; Nie, L.; Zhan, D.; Song, Y.; Fan, L. SAGE: A Visual Language Model for Anomaly Detection via Fact Enhancement and Entropy-aware Alignment. In Proceedings of the Proceedings of the

 doi:10.20944/preprints202606.0719.v1

33rd ACM International Conference on Multimedia, 2025, pp. 5030–5039. https://doi.org/10.1145/3746027.3755725.

105. Zhao, S.; Lin, Y.; Han, L.; Zhao, Y.; Wei, Y. OmniAD: Detect and Understand Industrial Anomaly via Multimodal Reasoning, 2025, [arXiv:cs.CV/2505.22039].
106. Fučka, M.; Zavrtanik, V.; Skočaj, D. SALAD – Semantics-Aware Logical Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 21843–21852.
107. Xu, Y.; Zhang, H.; Ma, Y.; Zhu, Y.; Ting, K.M. SCoNE: Spherical Consistent Neighborhoods Ensemble for Effective and Efficient Multi-View Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence* **2026**, *40*, 16083–16090. https://doi.org/10.1609/aaai.v40i19.38643.
108. Zhang, Q.; Shao, M.; Chen, X.; Lv, X.; Xu, K. Wave-MambaAD: Wavelet-driven State Space Model for Multi-class Unsupervised Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025, pp. 20868–20877.
109. Tao, C.; Cao, X.; Du, J. G2SF: Geometry-Guided Score Fusion for Multimodal Industrial Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 20551–20560.
110. Zheng, B.; Gan, J.; Xu, X.; Chen, X.; Li, W.; Huang, X.; Ni, N.; Wu, Y. Bridging 3D Anomaly Localization and Repair via High-Quality Continuous Geometric Representation. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025.
111. Yu, Y.; Chen, Z.; Xu, X.; Zhang, L.; Yang, H.; Nie, Y.; He, S. Registration is a Powerful Rotation-Invariance Learner for 3D Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2025.
112. Zha, Y.; Xue, Y.; Fan, C.; Wang, Y.; Dai, T.; Chen, K.; Xia, S.T. CASL: Curvature-Augmented Self-supervised Learning for 3D Anomaly Detection. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 12340–12348. https://doi.org/10.1609/aaai.v40i15.38226.
113. Chen, X.; Xu, X.; Zheng, B.; Liu, Y.; Wu, Y. Unsupervised Multi-View Visual Anomaly Detection via Progressive Homography-Guided Alignment. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 3065–3073. https://doi.org/10.1609/aaai.v40i4.37299.
114. Kang, X.; Li, Z.; Lan, T.; Gong, D.; Khoshelham, K.; Nan, L. Hierarchical Point-Patch Fusion with Adaptive Patch Codebook for 3D Shape Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026.
115. Kim, S.; Lee, W.; Cho, M. A Semantically Disentangled Unified Model for Multi-category 3D Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 33036–33045.
116. Long, K.; Ma, L.; Liu, J.; Liu, L.; Xie, G. Towards an Incremental Unified Multimodal Anomaly Detection: Augmenting Multimodal Denoising From an Information Bottleneck Perspective. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 14116–14125.
117. Li, Y.; Yang, X.; Zhang, J.; Tian, S.; Liao, J.; Liu, F. PIRN: Prototypical-based Intra-modal Reconstruction with Normality Communication for Multi-modal Anomaly Detection. In Proceedings of the The Fourteenth International Conference on Learning Representations (ICLR), 2026.
118. Costanzino, A.; Ramirez, P.Z.; Lisanti, G.; Di Stefano, L. Multimodal Industrial Anomaly Detection by Crossmodal Feature Mapping. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 17234–17243. https://doi.org/10.1109/CVPR52733.2024.01631.
119. Ye, J.; Zhao, W.; Yang, X.; Cheng, G.; Huang, K. PO3AD: Predicting Point Offsets toward Better 3D Point Cloud Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 1353–1362.
120. Huang, C.; Guan, H.; Jiang, A.; Zhang, Y.; Spratling, M.W.; Wang, Y.F. Registration Based Few-Shot Anomaly Detection. In Proceedings of the Computer Vision – ECCV 2022. Springer, 2022, Vol. 13684, *Lecture Notes in Computer Science*, pp. 303–319. https://doi.org/10.1007/978-3-031-20053-3_18.
121. Zhu, H.; Xie, G.; Hou, C.; Dai, T.; Gao, C.; Wang, J.; Shen, L. Towards High-resolution 3D Anomaly Detection via Group-Level Feature Contrastive Learning. In Proceedings of the Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 4680–4689.

 doi:10.20944/preprints202606.0719.v1

122. Lin, Y.; Yan, H.; Tong, X.; Chang, Y.; Wang, H.; Zhou, Z.; Gao, S.; Wang, Y.; Zhang, W. Commonality in Few: Few-Shot Multimodal Anomaly Detection via Hypergraph-Enhanced Memory. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 7015–7023. https://doi.org/10.1609/aaai.v40i9.37636.
123. Cheng, J.; Gao, C.; Zhou, J.; Wen, J.; Dai, T.; Wang, J. MC3D-AD: A Unified Geometry-aware Reconstruction Model for Multi-category 3D Anomaly Detection. In Proceedings of the Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI), 2025, pp. 837–845. https://doi.org/10.24963/ijcai.2025/94.
124. Tang, J.; Lu, H.; Xu, X.; Wu, R.; Hu, S.; Zhang, T.; Cheng, T.W.; Ge, M.; Chen, Y.C.; Tsung, F. An Incremental Unified Framework for Small Defect Inspection. In Proceedings of the Computer Vision – ECCV 2024, 2024, Lecture Notes in Computer Science, pp. 307–324. https://doi.org/10.1007/978-3-031-72751-1_18.
125. Zhu, J.; Pang, G. Toward Generalist Anomaly Detection via In-context Residual Learning with Few-shot Sample Prompts. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 17826–17836.
126. Yao, X.; Chen, Z.; Gao, C.; Zhai, G.; Zhang, C. ResAD: A Simple Framework for Class Generalizable Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2024, Vol. 37, pp. 125287–125311.
127. Gu, Z.; Zhu, B.; Zhu, G.; Chen, Y.; Ge, W.; Tang, M.; Wang, J. AnomalyMoE: Towards a Language-free Generalist Model for Unified Visual Anomaly Detection. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 4348–4356. https://doi.org/10.1609/aaai.v40i6.42432.
128. Lee, Y.; Kim, S.; Moon, D.; Jang, S.; Yoon, H. Bidirectional Multimodal Prompt Learning with Scale-Aware Training for Few-Shot Multi-Class Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 35577–35586.
129. Luo, W.; Cao, Y.; Yao, H.; Zhang, X.; Lou, J.; Cheng, Y.; Shen, W.; Yu, W. Exploring Intrinsic Normal Prototypes within a Single Image for Universal Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 9974–9983.
130. Gu, Z.; Zhu, B.; Zhu, G.; Chen, Y.; Tang, M.; Wang, J. UniVAD: A Training-free Unified Model for Few-shot Visual Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 15194–15203.
131. Sadikaj, Y.; Zhou, H.; Halilaj, L.; Schmid, S.; Staab, S.; Plant, C. MultiADS: Defect-aware Supervision for Multi-type Anomaly Detection and Segmentation in Zero-Shot Learning. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 22978–22988.
132. Wang, Y.; Wang, X.; Gong, Y.; Xiao, J. Normal-Abnormal Guided Generalist Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2025.
133. Lu, R.; Liu, G.; Li, K.; Tian, L.; Zhang, J. MaskAD: Parallel Masked Autoencoder for Multi-class Unsupervised Anomaly Detection. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 15457–15465. https://doi.org/10.1609/aaai.v40i18.38573.
134. He, H.; Bai, Y.; Zhang, J.; He, Q.; Chen, H.; Gan, Z.; Wang, C.; Li, X.; Tian, G.; Xie, L. MambaAD: Exploring State Space Models for Multi-class Unsupervised Anomaly Detection. In Proceedings of the Advances in Neural Information Processing Systems, 2024, Vol. 37.
135. Fan, L.; Huang, J.; Di, D.; Su, A.; Song, T.; Pagnucco, M.; Song, Y. Salvaging the Overlooked: Leveraging Class-Aware Contrastive Learning for Multi-Class Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 21419–21428. https://doi.org/10.1109/ICCV51701.2025.01989.
136. Guo, J.; Lu, S.; Zhang, W.; Chen, F.; Li, H.; Liao, H. Dinomaly: The Less Is More Philosophy in Multi-Class Unsupervised Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 20405–20415. https://doi.org/10.1109/CVPR52734.2025.01900.
137. Guo, J.; Lu, S.; Fan, L.; Li, Z.; Di, D.; Song, Y.; Zhang, W.; Zhu, W.; Yan, H.; Chen, F.; et al. One Dinomaly2 Detect Them All: A Unified Framework for Full-Spectrum Unsupervised Anomaly Detection, 2025, [arXiv:cs.CV/2510.17611].
138. Wei, S.; Jiang, J.; Xu, X. UniNet: A Contrastive Learning-guided Unified Framework with Feature Selection for Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 9994–10003.

 doi:10.20944/preprints202606.0719.v1

139. Zhang, Z.; Cai, M.; Wang, H.; Wu, G.; Chai, T.; Zhu, X. CostFilter-AD: Enhancing Anomaly Detection through Matching Cost Filtering. In Proceedings of the Proceedings of the 42nd International Conference on Machine Learning. PMLR, 2025, Vol. 267, *Proceedings of Machine Learning Research*, pp. 74540–74564.
140. Wang, X.; Wang, X.; Bai, H.; Lim, E.G.; Xiao, J. DecAD: Decoupling Anomalies in Latent Space for Multi-Class Unsupervised Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 21568–21577.
141. Beizaee, F.; Lodygensky, G.A.; Desrosiers, C.; Dolz, J. Correcting Deviations from Normality: A Reformulated Diffusion Model for Multi-Class Unsupervised Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 19088–19097. https://doi.org/10.1109/CVPR52734.2025.01778.
142. Kim, G.; Kim, M.; Lee, K.; Kim, M.; Jeon, H.; Han, J.; Lim, H.; Yim, J. UniSpector: Towards Universal Open-set Defect Recognition via Spectral-Contrastive Visual Prompting. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 6261–6270.
143. Yao, X.; Luo, Y.; Qian, Z.; Zhang, C. ADPretrain: Advancing Industrial Anomaly Detection via Anomaly Representation Pretraining. In Proceedings of the Advances in Neural Information Processing Systems, 2025, Vol. 38.
144. So, Y.; Kang, S. AnoStyler: Text-Driven Localized Anomaly Generation via Lightweight Style Transfer. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 15734–15742. https://doi.org/10.1609/aaai.v40i18.38604.
145. Jiang, Y.; Luo, W.; Zhang, H.; Chen, Q.; Yao, H.; Shen, W.; Cao, Y. Anomagic: Crossmodal Prompt-driven Zero-shot Anomaly Generation. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 5485–5493. https://doi.org/10.1609/aaai.v40i7.37466.
146. Lai, Z.; Lu, Y.; Li, X.; Lin, J.; Qu, Y.; Li, M.; Cao, L. AnomalyPainter: Vision-Language-Diffusion Synergy for Realistic and Diverse Unseen Industrial Anomaly Synthesis. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 5800–5808. https://doi.org/10.1609/aaai.v40i7.37501.
147. Qian, L.; Zhu, B.; Chen, Y.; Tang, M.; Wang, J. Quality-Aware Language-Conditioned Local Auto-Regressive Anomaly Synthesis and Detection. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2026, Vol. 40, pp. 15626–15634. https://doi.org/10.1609/aaai.v40i18.38592.
148. Kühn, P.J.; Pommeranz, M.; Kuijper, A.; Sinha, S.N. SynSur: An end-to-end generative pipeline for synthetic industrial surface defect generation and detection, 2026, [arXiv:cs.CV/2604.26633].
149. Rao, H.; Wang, Z.; Si, C.; Lyu, Y.; Duan, Y.; Zhao, F.; Shan, C. One-to-More: High-Fidelity Training-Free Anomaly Generation with Attention Control. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026.
150. Xu, R.; Chiu, Y.T.; Chen, T.I.; Chew, O.; Chuang, Y.Y.; Cheng, W.H. Training-Free Industrial Defect Generation with Diffusion Models. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 24214–24223.
151. Choi, J.; Kim, M.; Hong, J.H. MAGIC: Few-Shot Mask-Guided Anomaly Inpainting with Prompt Perturbation, Spatially Adaptive Guidance, and Context Awareness. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Findings, 2026.
152. Song, J.; Park, D.; Baek, K.; Lee, S.; Choi, J.; Kim, E.; Yoon, S. DefectFill: Realistic Defect Generation with Inpainting Diffusion Model for Visual Inspection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 18718–18727. https://doi.org/10.1109/CVPR52734.2025.01744.
153. Jin, Y.; Peng, J.; He, Q.; Hu, T.; Wu, J.; Chen, H.; Wang, H.; Zhu, W.; Chi, M.; Liu, J.; et al. Dual-Interrelated Diffusion Model for Few-Shot Anomaly Image Generation. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 30420–30429. https://doi.org/10.1109/CVPR52734.2025.02832.
154. Sun, H.; Cao, Y.; Dong, H.; Fink, O. Unseen Visual Anomaly Generation. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 25508–25517.
155. Sakai, S.; He, X.; Gu, C.; Sigal, L.; Hasegawa, T. InvAD: Inversion-based Reconstruction-Free Anomaly Detection with Diffusion Models. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026.
156. Zhu, W.; Wang, C.; Gao, B.B.; Zhang, J.; Jiang, G.; Hu, J.; Gan, Z.; Wang, L.; Zhou, Z.; Cheng, L.; et al. Real-IAD Variety: Pushing Industrial Anomaly Detection Dataset to a Modern Era, 2025, [arXiv:cs.CV/2511.00540].

 doi:10.20944/preprints202606.0719.v1

157. Zhu, W.; Wang, L.; Zhou, Z.; Wang, C.; Pan, Y.; Zhang, R.; Chen, Z.; Cheng, L.; Gao, B.B.; Zhang, J.; et al. Real-IAD D3: A Real-World 2D/Pseudo-3D/3D Dataset for Industrial Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 15214–15223.
158. Yang, E.; Xing, P.; Sun, H.; Guo, W.; Ma, Y.; Li, Z.; Zeng, D. 3CAD: A Large-Scale Real-World 3C Product Dataset for Unsupervised Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence* **2025**, *39*, 9175–9183. https://doi.org/10.1609/aaai.v39i9.32993.
159. Zhang, J.; Ding, R.; Ban, M.; Dai, L. PKU-GoodsAD: A Supermarket Goods Dataset for Unsupervised Anomaly Detection and Segmentation. *IEEE Robotics and Automation Letters* **2024**, *9*, 2008–2015. https://doi.org/10.1109/LRA.2024.3352358.
160. Fan, L.; Fan, D.; Hu, Z.; Ding, Y.; Di, D.; Yi, K.; Pagnucco, M.; Song, Y. MANTA: A Large-Scale Multi-View and Visual-Text Anomaly Detection Dataset for Tiny Objects. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 25518–25527.
161. Shi, D.; He, C.; Zhang, S.; Qian, B.; Quan, X.; Zhang, W.; Wei, X. Omni-AD: A Large-scale and Versatile Benchmark for Industrial Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2026, pp. 14157–14166.
162. McHard, P.M.; Audonnet, F.P.; Summerell, O.; Andraos, S.; Henderson, P.; Aragon-Camarasa, G. 3D-ADAM: A Dataset for 3D Anomaly Detection in Additive Manufacturing. In Proceedings of the Proceedings of the 2026 IEEE International Conference on Robotics and Automation (ICRA), 2026.
163. Guo, B.; Li, H.; Yu, R.; Liang, H.; Wang, J. IEC3D-AD: A 3D Dataset of Industrial Equipment Components for Unsupervised Point Cloud Anomaly Detection, 2025, [arXiv:cs.CV/2511.03267].
164. Li, W.; Zheng, B.; Xu, X.; Gan, J.; Lu, F.; Li, X.; Ni, N.; Tian, Z.; Huang, X.; Gao, S.; et al. Multi-Sensor Object Anomaly Detection: Unifying Appearance, Geometry, and Internal Properties. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 9984–9993.
165. Cao, Y.; Cheng, Y.; Zhang, Y.; Xu, X.; Zhang, Y.; Sun, Y.; Tan, Y.; Huang, X.; Huang, C.; Shen, W. Visual Anomaly Detection under Complex View-Illumination Interplay: A Large-Scale Benchmark. *Pattern Recognition* **2026**, *179*, 113666. https://doi.org/10.1016/j.patcog.2026.113666.
166. Heckler-Kram, L.; Neudeck, J.H.; Scheler, U.; König, R.; Steger, C. The MVTec AD 2 Dataset: Advanced Scenarios for Unsupervised Anomaly Detection. *International Journal of Computer Vision* **2026**, *134*, 175. https://doi.org/10.1007/s11263-026-02743-0.
167. Bai, H.; Mou, S.; Likhomanenko, T.; Cinbis, R.G.; Tuzel, O.; Huang, P.; Shan, J.; Shi, J.; Cao, M. VISION Datasets: A Benchmark for Vision-based InduStrial InspectiON, 2023, [arXiv:cs.CV/2306.07890]. Presented at the CVPR 2023 Workshop on Vision-Based Industrial Inspection.
168. Lei, T.; Wang, B.; Chen, S.; Cao, S.; Zou, N. Texture-AD: An Anomaly Detection Dataset and Benchmark for Real Algorithm Development, 2024, [arXiv:cs.CV/2409.06367].
169. Wang, Q.; Gao, S.; Hu, J.; Yu, J.; Tong, X.; Li, Y.; Zhang, W. HSS-IAD: A Heterogeneous Same-Sort Industrial Anomaly Detection Dataset. In Proceedings of the 2025 IEEE International Conference on Multimedia and Expo (ICME), 2025, pp. 1–6. https://doi.org/10.1109/ICME59968.2025.11208914.
170. Yu, R.; Guo, B.; Li, H. Anomaly Detection of Integrated Circuits Package Substrates Using the Large Vision Model SAIC: Dataset Construction, Methodology, and Application. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 22563–22574.
171. Chen, Q.; Luo, H.; Lv, C.; Zhang, Z. A Unified Anomaly Synthesis Strategy with Gradient Ascent for Industrial Anomaly Detection and Localization. In Proceedings of the Computer Vision – ECCV 2024. Springer, 2024, Vol. 15125, *Lecture Notes in Computer Science*, pp. 37–54. https://doi.org/10.1007/978-3-031-72855-6_3.
172. Vieira e Silva, A.L.B.; Felix, H.d.C.; Simões, F.P.M.a.; Teichrieb, V.; dos Santos, M.; Santiago, H.; Sgotti, V.; Lott Neto, H. InsPLAD: A Dataset and Benchmark for Power Line Asset Inspection in UAV Images. *International Journal of Remote Sensing* **2023**, *44*, 7294–7320. https://doi.org/10.1080/01431161.2023.2283900.
173. Arodi, A.; Luck, M.; Bedwani, J.L.; Zaimi, A.; Li, G.; Pouliot, N.; Beaudry, J.; Marceau Caron, G. CableInspect-AD: An Expert-Annotated Anomaly Detection Dataset. In Proceedings of the Advances in Neural Information Processing Systems, 2024, Vol. 37, pp. 64703–64716. Datasets and Benchmarks Track.
174. Yang, S.; Chen, Z.; Chen, P.; Fang, X.; Liang, Y.; Liu, S.; Chen, Y. Defect Spectrum: A Granular Look of Large-Scale Defect Datasets with Rich Semantics. In Proceedings of the Computer Vision – ECCV 2024.

 doi:10.20944/preprints202606.0719.v1

Springer, 2025, Vol. 15065, *Lecture Notes in Computer Science*, pp. 187–203. https://doi.org/10.1007/978-3-031-72667-5_11.

175. Yang, C.A.; Peng, K.C.; Yeh, R.A. Toward Long-Tailed Online Anomaly Detection through Class-Agnostic Concepts. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2025, pp. 23419–23430.
176. Hu, L.; Gan, Z.; Deng, L.; Liang, J.; Liang, L.; Huang, S.; Chen, T. ReplayCAD: Generative Diffusion Replay for Continual Anomaly Detection. In Proceedings of the Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI), 2025, pp. 2946–2954. https://doi.org/10.24963/ijcai.2025/328.
177. Safarov, S.; Park, J.; Jung, Y.G.; Peng, K.C.; Kim, W.; Bang, S.; Camps, O. Memory-Distilled Selection for Noise-Robust Anomaly Detection. In Proceedings of the Proceedings of the 43rd International Conference on Machine Learning (ICML), 2026.
178. Horwitz, E.; Hoshen, Y. Back to the Feature: Classical 3D Features Are (Almost) All You Need for 3D Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2023, pp. 2968–2977. https://doi.org/10.1109/CVPRW59228.2023.00298.
179. Jiang, X.; Zhao, Y.; Yang, Z.; Zheng, F. AnomalyClaw: A Universal Visual Anomaly Detection Agent via Tool-Grounded Refutation, 2026, [arXiv:cs.CV/2605.10397].
180. Liu, J.; Yan, Y.; Li, J.; Zhao, W.; Chu, P.; Sheng, X.; Liu, Y.; Yang, X. IPAD: Industrial Process Anomaly Detection Dataset. *IEEE Transactions on Circuits and Systems for Video Technology* **2025**, *35*, 380–393. https://doi.org/10.1109/TCSVT.2024.3465517.
181. Li, W.; Gu, Y.; Chen, X.; Xu, X.; Hu, M.; Huang, X.; Wu, Y. Towards Visual Discrimination and Reasoning of Real-World Physical Dynamics: Physics-Grounded Anomaly Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 30409–30419. https://doi.org/10.1109/CVPR52734.2025.02831.
182. Dabouei, A.; Parayil Shibu, J.; Dalal, V.; Cao, C.; MacWilliams, A.; Kangas, J.; Xu, M. Deep video anomaly detection in automated laboratory setting. *Expert Systems with Applications* **2025**, *271*, 126581. https://doi.org/10.1016/j.eswa.2025.126581.