ADNet: A Large-Scale and Extensible Multi-Domain Benchmark for Anomaly Detection Across 380 Real-World Categories

Hai Ling[1,2] Jia Guo[3] Zhulin Tao[1*] Yunkang Cao[4] Donglin Di[3]
Hongyan Xu[5] Xiu Su[5] Yang Song[6] Lei Fan[2,6†*]
[1]Communication University of China [2]DZ Matrix [3]Tsinghua University
[4]Hunan University [5]Central South University [6]UNSW Sydney

Abstract

Anomaly detection (AD) aims to identify defects using normal-only training data. Existing anomaly detection benchmarks (e.g., MVTec-AD with 15 categories) cover only a narrow range of categories, limiting the evaluation of cross-context generalization and scalability. We introduce ***ADNet****, a large-scale, multi-domain benchmark comprising* ***380*** *categories aggregated from 49 publicly available datasets across Electronics, Industry, Agrifood, Infrastructure, and Medical domains. The benchmark includes a total of* ***196,294*** *RGB images, consisting of 116,192 normal samples for training and 80,102 test images, of which 60,311 are anomalous. All images are standardized with MVTec-style pixel-level annotations and structured text descriptions spanning both spatial and visual attributes, enabling multimodal anomaly detection tasks. Extensive experiments reveal a clear scalability challenge: existing state-of-the-art methods achieve* 90.6% *I-AUROC in one-for-one settings but drop to* 78.5% *when scaling to all 380 categories in a multi-class setting. To this end, we propose* ***Dinomaly***^{m}*, a context-guided Mixture-of-Experts extension of Dinomaly that expands decoder capacity without added inference cost. It achieves* 83.2% *I-AUROC and* 93.1% *P-AUROC, demonstrating superior performance over existing approaches. We aim to make ADNet a standardized and extensible benchmark, supporting the community in expanding anomaly detection datasets across diverse domains and providing a scalable foundation for AD foundation models. Dataset:* https://grainnet.github.io/ADNet.

1. Introduction

Unsupervised anomaly detection (AD) in visual data aims to identify anomalous regions or defective samples using only anomaly-free samples for training. This topic is critical in safety-sensitive and high-reliability scenarios such as industrial inspection, medical imaging, and infrastructure monitoring. Its development has been largely driven by established benchmarks such as MVTec-AD [7],

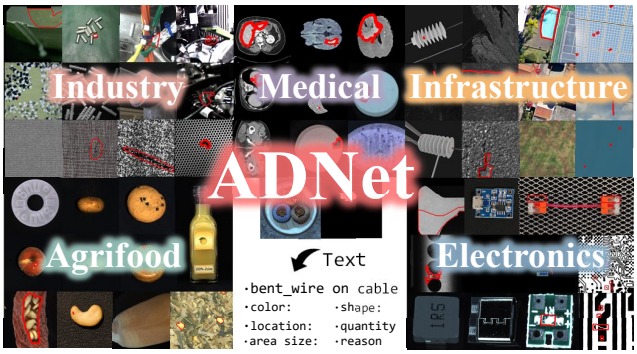

Figure 1. Overview of ADNet, covering 380 categories across five application domains.

VisA [71], and BTAD [32], which provide diverse object categories. These datasets not only enable fair comparison but also drive the exploration of different modeling paradigms, such as reconstruction-based methods [14, 20], data synthesis-based methods [17, 34, 67], embedding-based methods [38], and memory-based methods [51].

Early AD methods focus on training individual models for each category, commonly referred to as the 'one-for-one' setting. However, maintaining separate models for numerous categories is resource-intensive and difficult to manage during deployment. To overcome this limitation, recent efforts [27, 61] have shifted toward 'one-for-all' approaches that aim to build a unified model capable of handling multiple categories simultaneously. Both settings have achieved remarkable performance, consistently attaining near-perfect I-AUROC scores ($> 95\%$) on standard benchmarks. These results highlight the rapid progress in model design but also reveal the inherent limitations of existing datasets. Although recent datasets extend AD to broader domains such as consumer electronics [56, 58], medical imaging [4, 48], tiny objects [19], and new imaging modalities [35, 37], they still contain a limited number of categories (fewer than 40), focus on a single domain, and are collected under controlled imaging conditions (*e.g.,* fixed lighting and viewpoints). As a result, current evaluations remain constrained to isolated, small-scale benchmarks. This mismatch raises fundamental questions: *How well do existing AD methods perform when scaled to a large number of categories? And can we build*

*Corresponding Authors: lei.fan1@unsw.edu.au, taozl@cuc.edu.cn
†Project Leader.

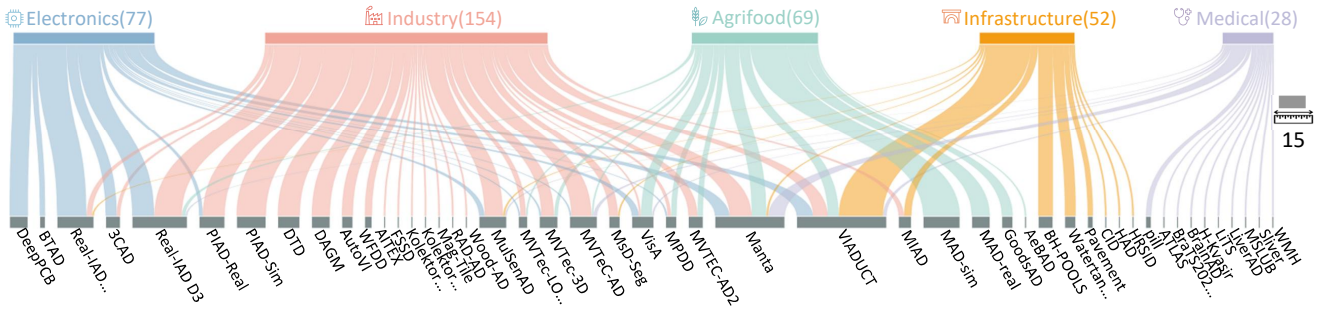

Figure 2. **ADNet is constructed by consolidating 49 publicly available AD datasets**. We validate their composition and quality, and hierarchically integrate all sources into five target domains: Electronics, Industry, Agrifood, Infrastructure, and Medical.

an AD benchmark similar to ImageNet [16]?

In this paper, we introduce ADNet (see Figure 1), a large-scale anomaly detection dataset comprising 380 categories, which is 25× larger in the number of categories than MVTec-AD. ADNet consolidates over 49 publicly available AD datasets into a unified benchmark (see Figure 2), producing 116,192 training and 80,102 test images (including 60,311 anomalous samples). The categories are grouped into five domains: Electronics, Industry, Agrifood, Infrastructure, and Medical. All heterogeneous data sources are standardized into a consistent format with MVTec-style pixel-level annotations, along with expert-provided textual descriptions for each anomalous sample.

Compared to previous datasets, ADNet enables systematic evaluation of two critical capabilities: ***i) Scalability vs. number of categories.*** We evaluate existing methods. While models achieve strong performance in the one-for-one setting (90.6% I-AUROC), performance drops to 78.5% in the multi-class setting. This degradation reveals that current models struggle to maintain robustness and scalability as the number of categories grows, limiting their applicability in real-world environments. ***ii) Anomaly confusion.*** ADNet includes categories spanning diverse domains where identical visual patterns can represent opposite anomaly semantics. For instance, cracks are normal in cookies but defective in printed circuit boards. Models (*e.g.,* PatchCore [51]), which rely on memorizing category-specific appearance priors, are prone to confusion in such cases. This challenges models to move beyond superficial pattern matching.

Our contributions are summarized as follows:

- We introduce **ADNet**, a large-scale AD dataset containing 196,294 images across 380 categories spanning five domains.
- We provide a unified MVTec-style annotation format, and augment each anomalous sample with fine-grained textual descriptions.
- We propose **Dinomaly**^{m}, a strong baseline that integrates a Mixture-of-Experts into Dinomaly [27], improving the I-AUROC from 78.5% to 83.2% on ADNet.

2. Related Work

2.1. Anomaly Detection Datasets

Early studies focused on specific application domains such as industrial manufacturing [7, 32, 71], medical imaging [4, 10, 39], infrastructure inspection [59, 65], and hyperspectral or agricultural imaging [12, 18, 37]. More recent works have expanded along several key directions: i) *Dat a scale and class diversity*. Real-IAD [57] and Kaputt [30] increased both dataset size and category diversity, with Kaputt demonstrating the feasibility of large-scale collection in retail logistics; ii) *Multimodal sensing*. Datasets (*e.g.*, MVTec 3D-AD [8], MulSen-AD [35], Real-IAD D^3 [70]) incorporate depth maps, multi-view imagery, and multi-sensor inputs to enhance robustness beyond RGB data; iii) *Language-grounded annotation*. MANTA [19] provides textual descriptions for anomalies across 38 categories, enabling text-guided anomaly localization; and iv) *Specialized scenarios*. MVTec LOCO [9] focuses on logical rather than visual inconsistencies, while FSSD [21] and MSD [6] explore few-shot settings. Additionally, domain-specific datasets have been developed for textiles [53], electronics [56], and steel [54].

Compared to existing datasets, which are domain-specific, limited in category diversity, and vary in data formats and annotations, we introduce ADNet, which unifies the data structure and annotation protocol across over 49 datasets, aggregating 380 categories spanning five domains.

2.2. Anomaly Detection Methods

Existing AD approaches have shifted from one-for-one models to more general frameworks to support scalable deployment across diverse scenarios.

One-for-One. This paradigm trains an individual model for each category. Existing methods may be grouped into: i) *Reconstruction-based methods*, which detect anomalies through reconstruction errors from autoencoders [14, 20] or diffusion models [40, 50]; ii) *Synthesis-based methods*, which generate pseudo anomalies to enable supervised training [17, 34, 43, 62, 67]; iii) *Embedding-based meth-*

ods, which learn compact representations of normal data to distinguish anomalies [13, 38, 52]; and iv) *Memory-based methods*, which store normal features and detect anomalies based on deviations during inference [3, 51]. However, this paradigm faces inherent scalability challenges when extended to multiple classes and prevents cross-category knowledge transfer.

One-for-All. Recent studies [25, 26, 61] have shifted towards training a unified model that can handle multiple categories simultaneously. For example, transformer-based methods (*e.g.*, UniAD [61], IUF [55], Dinomaly [27]) leverage attention mechanisms to capture inter-category relationships. Data generation approaches, including OmniAL [68] and DiAD [29], synthesize diverse anomaly patterns, while density-based methods, such as CRAD [33] and HGAD [60], learn distributional representations shared across categories. However, multi-class models often underperform compared to one-for-one counterparts, and existing evaluations remain limited on relatively small-scale benchmarks (< 40 classes), leaving open questions regarding scalability, cross-domain knowledge transfer, and the feasibility of foundation models.

Few-Shot. Vision-language models have enabled new paradigms for anomaly detection with limited training samples. WinCLIP [31] and AnomalyCLIP [69] adapt CLIP [49] for zero-shot detection through designed text prompts, while PromptAD [36] introduces learnable visual prompts for task adaptation. Recent advances further explore multimodal fusion and alignment [41, 42]. For example, AA-CLIP [46] proposes anomaly-aware adaptation mechanisms, FE-CLIP [23] incorporates frequency-domain information, UniVAD [24] presents a training-free approach with Contextual Component Clustering for cross-domain detection, and ReMP-AD [45] addresses intra-class variation through retrieval-based token selection and Vision-Language Prior Fusion. While these methods demonstrate strong data efficiency, their performance and robustness under domain shifts and generalization across diverse application contexts remain limited.

3. ADNet

The motivation behind ADNet is to establish a large-scale anomaly detection dataset that spans broader real-world scenarios and can be easily scaled through a standardized data processing pipeline. We first introduce the data sources and standardization pipeline (Sec. 3.1). We then present the data distribution and structure, including both images and textual annotations (Sec. 3.2). Finally, we discuss the potential challenges associated with using ADNet (Sec. 3.3).

3.1. Data Sources and Standardization Pipeline

We construct ADNet following a standardized pipeline that aggregates 49 publicly available datasets into a unified benchmark, as illustrated in Figure 3.

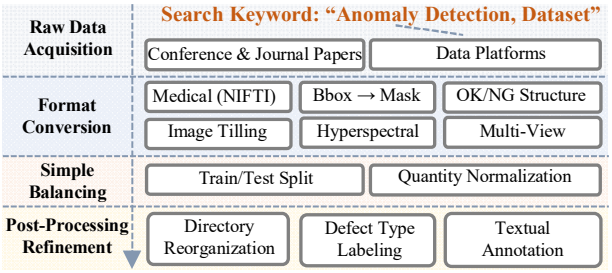

Figure 3. **ADNet construction pipeline.** We collect datasets from diverse sources, standardize formats, balance category-wise samples, and refine annotations with fine-grained defect type labels.

Data Sources. Existing benchmarks are often constructed from isolated domains or constrained scenarios (*e.g.*, single-device setups), limiting their applicability across diverse real-world settings. To ensure comprehensive domain diversity and mitigate single-source bias, we systematically collect publicly available datasets from three types of sources: academic publications (*e.g.*, conference papers), public data platforms (*e.g.*, Kaggle), and research collaborations [1]. For each candidate source, we apply a set of selection criteria to ensure data quality and usability:

- *Data Completeness.* Each dataset must contain normal samples (> 50 images), anomalous samples (> 20 images), and corresponding pixel-level masks indicating anomalous regions.
- *Background Consistency.* Normal and anomalous samples should be acquired under consistent imaging conditions to avoid shortcut learning [22], ensuring that models focus on anomalies rather than background variation.
- *Category Labels.* Each sample must be clearly associated with an explicit category (*e.g.*, hazelnut) to enable standardized organization and downstream usage.
- *Accessibility.* The dataset must be publicly available with clear licensing for academic use.

By adhering to these criteria, we gather 49 datasets that cover a broad spectrum of AD scenarios. To better organize them, we categorize all images into five application domains based on their semantic attributes: Electronics, Industry, Agrifood, Infrastructure, and Medical. More importantly, the principled selection process facilitates scalable and continual expansion of the dataset in the future.

Standardization Pipeline. To ensure fair comparison and compatibility with existing methods, we design an automated standardization pipeline that converts and unifies all source datasets into the MVTec-style format.

- *Format Conversion.* We implement specialized converters to handle heterogeneous data formats, includ-

[1] The complete list with sources and licenses is provided in the *Supp*.

Table 1. Comparison with existing benchmarks.

Dataset	#Cats	#Images (Train/Test)	#Anomalies
MVTec-AD [7]	15	5,354 (3,629/1,725)	1,258
VisA [71]	12	10,821 (8,659/2,162)	1,200
Real-IAD [57]	30	151,050 (36,465/114,585)	51,329
Manta-tiny [19]	38	66,068 (21,483/44,585)	11,025
MulSen-AD [35]	15	2,035 (1,391/644)	491
Uni-Medical [11]	3	20,352 (13,339/7,013)	4,499
MIAD [5]	7	105,000 (70,000/35,000)	17,500
ADNet (Ours)	**380**	**196,294 (116,192/80,102)**	**60,311**

ing tiling large-size images (*e.g.*, AITEX textile fabrics, hyperspectral remote sensing) into fixed-size images (256×256 pixels), converting medical imaging formats (NIFTI/DICOM) into slice-level RGB images, and extracting single-view images from multi-view images (*e.g.*, MANTA).

- *Sample Balancing.* To prevent domain biases and ensure computational tractability, we first adopt a 9:1 train/test split for normal samples, then normalize per-category sample quantities with capping limits (a maximum of 500 training samples and 100 normal test samples after splitting, plus up to 100 anomaly samples per defect type).

This unified format ensures consistent preprocessing while preserving domain-specific semantics, resulting in balanced categories and reliable statistical evaluation [2].

Text Annotation. We follow previous studies [1, 2, 19] and provide textual descriptions through structured text-visual correspondences. The annotations consist of two components: *spatial localization* and *semantic attributes*. For spatial localization, each image is divided into a 3×3 grid, yielding nine regions. Each defect is assigned a positional descriptor (*e.g.*, "top-left", "middle-center") to support natural language-based spatial queries. For semantic attributes, we employ a Large Language Model (*i.e.*, GPT-4o-mini) with domain-specific prompts to generate five visual properties for each anomaly: color, area size, shape, quantity, and reason. Combined with location, all six annotations undergo automated validation followed by manual verification to ensure consistency and accuracy.

3.2. Dataset Statistics

ADNet encompasses 380 categories derived from 49 public datasets. Figure 2 illustrates the hierarchical integration of these source datasets into five target domains: Electronics (77 categories), Industry (154 categories), Agrifood (69 categories), Infrastructure (52 categories), and Medical (28 categories).

Data Scale. The dataset comprises 196,294 images across 380 categories, including 116,192 training samples and

[2]Scripts are provided to support community contributions and extension with new datasets.

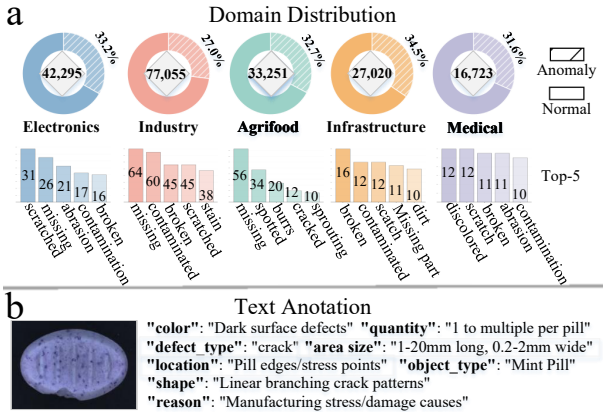

Figure 4. **ADNet domain summary and text annotations. (a)** Total image counts of the five domains with normal vs. anomaly proportions. **(b)** Top-5 defect types and a text-annotation example (image, mask, and JSON fields).

80,102 test images, among which 19,791 are anomaly-free and 60,311 contain anomalies. On average, each category contains 516 images, providing sufficient data while maintaining computational efficiency. As shown in Table 1, ADNet offers not only a significantly larger data volume, about $46\times$ more than MVTec-AD, but also substantially greater diversity, with a $25\times$ increase in category count over MVTec-AD (15 categories) and a $10\times$ increase over Manta (38 categories). This reflects broader coverage of real-world scenarios and richer visual variations.

Text Information. ADNet provides comprehensive textual descriptions, with all annotations stored in a unified JSON format following consistent schemas across diverse industrial domains. The dataset includes fine-grained defect type annotations covering 420 unique defect types across 60,311 anomalous samples. Unlike prior datasets that typically use only generic "defect" labels, ADNet captures detailed defect semantics. The most prevalent types include *scratch* (2,393 samples, 83 categories), *fragmentation* (2,373 samples, 31 categories), and *missing* (2,227 samples, 30 categories). This dual-level scale (380 categories and 420 defect types) offers finer granularity for evaluation.

Specifically, each textual description characterizes an anomaly through six attributes: *Location*, indicating the approximate position within the image (*e.g.*, "top-left"), *Color*, describing the most salient visual appearance (*e.g.*, "copper-colored residue"), *Shape*, outlining the geometry of the anomalous region (*e.g.*, "semicircular notches"), *Area Size*, capturing the relative or physical size (*e.g.*, "0.2–10 mm linear breaks"), *Quantity*, denoting the number of defect instances (*e.g.*, "2"), and *Reason*, describing the underlying cause (*e.g.*, "over-etching" or "incomplete separation"). Representative examples are shown in Figure 4.

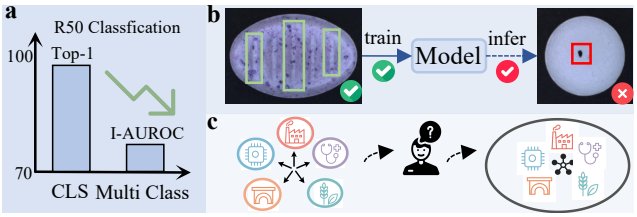

Figure 5. Illustration of the key challenges introduced by ADNet. **a)** Category Catastrophe: existing methods exhibit significant performance degradation on ADNet. **b)** Context-Dependent Anomalies: "spots" are good on mint tablets but bad contamination in a pill. c) Can Anomalies Transfer?, ADNet spans multiple domains, requiring models to learn transferable feature representations.

3.3. Challenges

ADNet introduces several fundamental challenges for existing AD methods, highlighting limitations in generalization, robustness, and adaptability across diverse domains, as shown in Figure 5.

Category Catastrophe. Unlike general image classification datasets (*e.g.*, ImageNet), which aim to recognize semantic content, AD datasets focus on identifying deviations from normal samples. To highlight this difference, we conducted a toy experiment by training a ResNet-50 classifier on all 380 categories in ADNet. Normal and anomalous samples were merged within each category, and the data were split into 8:2 train/test ratio. The model achieved 97.9% top-1 accuracy after 30 epochs. However, when we evaluated advanced multi-class AD methods, their results were notably lower, with I-AUROC scores below 80%.

These results indicate that category separation in ADNet is sufficient with minimal inter-class similarity. In contrast, advanced multi-class AD methods, mainly following the reconstruction-based paradigm, show moderate performance, likely due to class confusion [20, 61]. This evidence suggests that the main challenges in ADNet stem from the intrinsic complexity of anomaly detection rather than from data quality or class-level ambiguity.

Context-Dependent Anomalies. Given ADNet's scale of 380 categories, models must maintain robust performance across a wide range of scenarios. Unlike one-to-one AD approaches, where each model specializes in a single category, multi-class AD methods must detect anomalies across hundreds of diverse categories simultaneously. In many cases, identical visual patterns may carry opposite semantic meanings depending on the context.

For example, "surface cracks" are natural textures in baked cookies (Agrifood) but represent critical defects in printed circuit boards (Electronics). Similarly, "decorative spots" are expected on mint tablets (Agrifood) yet indicate contamination in pharmaceutical pills (Medical). Therefore, models evaluated on ADNet must acquire context-aware anomaly understanding rather than relying on universal pattern matching. This scalability requirement imposes a rigorous test on whether existing methods can generalize to large-scale, real-world deployment scenarios.

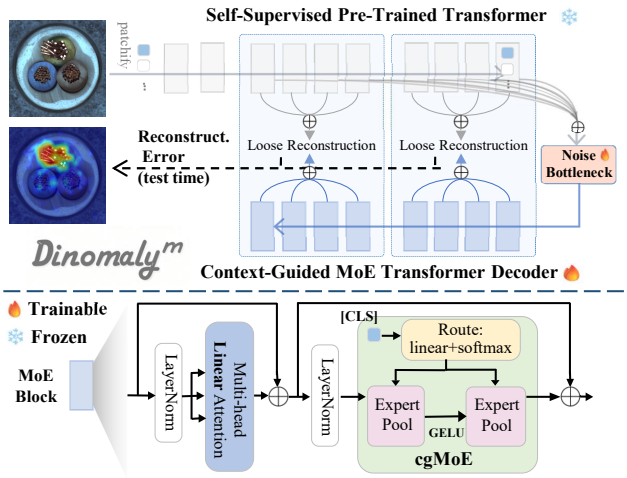

Figure 6. **Overview of Dinomaly^m.** We extend Dinomaly by augmenting the decoder with context-guided MoE modules.

Can Anomalies Transfer? The rich textual information and extensive category coverage of ADNet support a range of downstream applications. A central question arises: *Can we build foundation models for anomaly detection that generalize across diverse categories and contexts simultaneously*? In particular, cross-domain transfer, such as whether models trained on industrial defects can generalize to medical or agrifood anomalies, remains largely unexplored. These questions are fundamental to real-world deployment but cannot be adequately addressed using existing single-domain or small-scale benchmarks. ADNet provides the necessary scale and diversity to investigate this direction.

4. Baseline

To validate and explore the potential of ADNet, we introduce a baseline model, **Dinomaly^m**, a context-guided extension of Dinomaly [27] based on a mixture-of-experts framework. The vanilla Dinomaly employs a frozen DINOv2 [47] encoder, a noisy bottleneck, and a lightweight Transformer decoder to reconstruct clean features. Only the bottleneck and decoder are trained, and anomalies are detected via reconstruction errors. While effective on existing benchmarks, scaling to ADNet results in only moderate performance (78.5% I-AUROC). We attribute this degradation to the fixed-capacity decoder, which struggles to model the diverse visual patterns across hundreds of categories and to the strong context dependence of anomaly semantics.

We augment Dinomaly by introducing a context-guided mixture-of-experts (cgMoE) module into the decoder, enhancing its representational capacity with minimal runtime

Table 2. Results of Multi-class methods on ADNet. We report image-level (I-) AUROC, pixel-level (P-) AUROC, and pixel-level (P-AP) Average Precision (%) for each domain and the overall average.

Method	Electronics			Industry			Agrifood			Infrastructure			Medical			Overall		
	I	P	P-AP	I	P	P-AP	I	P	P-AP	I	P	P-AP	I	P	P-AP	I	P	P-AP
PatchCore [51]	77.5	89.0	25.2	80.4	90.4	25.8	**65.6**	85.0	15.8	75.7	83.5	15.9	76.0	84.6	28.6	76.2	87.7	22.7
UniAD [61]	59.9	86.2	8.2	66.1	87.5	14.4	58.6	86.1	9.1	64.4	87.1	10.8	68.2	83.7	17.1	63.4	86.7	11.8
RD [15]	68.3	90.4	14.7	70.6	89.6	19.3	59.4	82.1	9.5	63.9	83.8	9.9	67.9	87.6	26.6	66.9	87.4	15.7
SimpleNet [44]	51.9	63.8	2.4	53.6	65.3	4.3	51.5	66.9	4.4	53.5	63.4	3.3	51.7	65.2	6.2	52.7	65.0	3.9
ViTAD [64]	77.1	91.7	18.3	78.0	91.2	24.4	63.0	82.6	11.9	63.3	80.9	12.2	71.6	89.6	31.7	72.4	88.1	19.6
DeSTSeg [66]	72.6	86.1	15.3	66.1	83.3	17.9	58.0	79.7	11.1	56.4	72.1	6.8	59.1	79.0	14.0	63.9	81.0	14.0
MambaAD [28]	74.0	91.8	20.7	77.3	91.0	25.7	64.0	83.1	12.0	69.4	85.8	13.4	72.5	89.4	29.9	72.7	88.8	20.7
Dinomaly [27]	82.7	92.8	23.3	85.2	93.8	30.3	63.4	85.7	13.3	73.0	87.0	15.4	77.6	90.3	32.1	78.5	91.0	23.9
LGC-AD [20]	73.7	92.4	19.0	74.8	91.8	23.6	63.7	83.2	11.3	69.3	87.4	13.0	74.0	89.6	31.8	71.8	89.6	19.6
Dinomaly^m	**88.9**	**95.6**	**32.6**	**90.6**	**95.8**	**38.8**	65.3	**88.4**	**16.2**	**78.8**	**88.9**	**21.6**	**81.4**	**91.6**	**36.0**	**83.2**	**93.1**	**30.6**

overhead and enabling context-specific reconstruction. As illustrated in Figure 6, given an input image I, the frozen encoder extracts features f_E and produces a final-layer [CLS] token $\mathbf{z}_{\text{cls}} \in \mathbb{R}^d$ that captures the global context. In each decoder block, a gating network uses $\mathbf{z}_{\text{cls}}$ to compute soft routing weights over K experts:

$$\mathbf{g} = [g_1, \ldots, g_K] = \text{softmax}(\mathbf{W}_g \mathbf{z}_{\text{cls}}) \in \mathbb{R}^K, \quad (1)$$

where $\mathbf{W}_g \in \mathbb{R}^{K \times d}$ is a learnable parameter matrix, and $g_k \in \mathbb{R}$ is the scalar weight for expert k. Each expert k is a two-layer feed-forward network with parameters $\mathbf{W}_k^{(1)} \in \mathbb{R}^{h \times d}$ and $\mathbf{W}_k^{(2)} \in \mathbb{R}^{d \times h}$, where h is the hidden dimension. These are adaptively mixed based on the routing weights:

$$\text{cgMoE}(\mathbf{x}) = \sum_{k=1}^{K} g_k \mathbf{W}_k^{(2)} \text{GELU}(\sum_{k=1}^{K} g_k \mathbf{W}_k^{(1)} \mathbf{x}). \quad (2)$$

All decoder blocks share the same $\mathbf{z}_{\text{cls}}$ to ensure consistent routing, and the decoder outputs layer-wise features f_D for reconstruction. Training is performed using a reconstruction loss defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{D}_{\cos}(\mathcal{F}(f_E), \mathcal{F}(sg(f_D))), \quad (3)$$

where $\mathcal{D}_{\cos}$ is cosine distance, $\mathcal{F}(\cdot)$ flattens the feature maps, and $sg(\cdot)$ denotes the hard-mining mechanism that shrinks gradients of well-restored feature points [27]. During training, the encoder is frozen, while the bottleneck, expert weights, and gating network are optimized jointly. At inference, anomaly scores are computed from reconstruction errors, with routing performed once per image.

5. Experiments

5.1. Experimental Setup

Evaluation Settings. We evaluate methods under two settings: *(1) Single-Class Evaluation:* Following the standard protocol [51], we train 380 independent category-specific models, one per category. This setting provides an upper-bound performance. *(2) Multi-Class Evaluation:* We train a single unified model on all 380 categories and evaluate it across all test sets. This setting directly assesses a model's ability to handle diverse contexts.

Metrics and Implementation Details. We report standard image-level (I-) and pixel-level (P-) Area Under the Receiver Operating Characteristic curve (AUROC) scores [7, 51] to evaluate model performance in anomaly detection and localization, respectively. We implemented baseline methods using the ADer toolbox [63]. For multi-class evaluation, we trained models for 30 epochs following the default setups in ADer. Dinomaly variants were trained for 100k iterations (27.5 epochs) with ViT-B and 280^2 input. Complete details, additional metrics, and comprehensive experimental results are provided in the *Supp*.

5.2. Multi-Class Results

We evaluated multi-class performance on ADNet by comparing our Dinomaly^m against nine representative baselines across three major paradigms: memory-based (PatchCore), reconstruction-based (UniAD, RD, ViTAD, MambaAD, Dinomaly, LGC-AD), and synthesis-based methods (SimpleNet, DeSTSeg). These baselines consistently achieve over 95% I-AUROC on MVTec-AD [7], making them strong candidates for assessing the challenges posed by ADNet. Results are shown in Table 2.

Dinomaly^m achieved an overall 83.2% I-AUROC, outperforming the baseline (Dinomaly at 78.5%) by a margin of 4.7% I-AUROC. This improvement validates the effectiveness of incorporating context-adaptive capacity. The mixture-of-experts decoder enables specialized subnetworks for different semantic contexts, directly addressing context-dependent anomaly semantics without incurring additional inference costs. This architectural choice proves particularly effective for ADNet's multi-domain.

However, this performance remains far below the 95-99% typically reported on MVTec-AD, highlighting the substantial scalability challenges introduced by ADNet. Al-

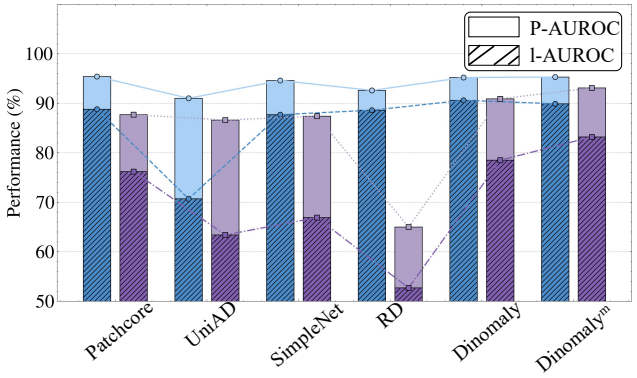

Figure 7. Comparison between single-class and multi-class settings. We report results averaged across all 380 categories.

though P-AUROC remains relatively high (93.1%), pixel-level Average Precision (P-AP) remains considerably low across all methods (3.9%-30.6%). This gap reveals that establishing precise anomaly boundaries under diverse semantic contexts remains a fundamental challenge for current AD approaches.

Single-Class *vs.* Multi-Class.

To quantify the gap between single-class models and multi-class models, we conducted five representative methods under both settings. Figure 7 illustrates the stark contrast between single-class (average 380 models) and training a single model on all categories. In the single-class setting, all methods achieved 87-90% I-AUROC, demonstrating that individual ADNet categories are inherently learnable. This performance confirms that ADNet's challenge stems not from ambiguity within categories but from the complexity of unified multi-category modeling. Under the multi-class setting, however, we observe substantial and method-dependent performance degradation. SimpleNet collapsed with a 35.9% absolute drop in I-AUROC, and RD degraded by 20.8%. In contrast, PatchCore and Dinomaly exhibited more moderate degradation (approximately 12% each).

These degradation patterns reveal that architectural design is critical for multi-class robustness. Methods employing context-invariant strategies (SimpleNet's fixed augmentation templates, RD's single student-teacher architecture) collapse under context diversity. Conversely, incorporating implicit context-awareness (PatchCore's memory-based matching) degrades more gracefully. These observations motivate our Dinomaly^m's explicit context-adaptive design through mixture-of-experts, which enables specialized subnetworks to handle diverse semantic contexts effectively.

5.3. Cross-Domain Generalization

Real-world deployment frequently demands that models generalize to new application domains with limited or no domain-specific training data. To evaluate zero-shot cross-domain transferability, we train models exclusively on the Industry domain (154 categories) and evaluate their performance on the remaining four domains without any fine-tuning, as shown in Table 3.

Table 3. Zero-shot cross-domain transfer. Models trained solely on Industry are tested on other domains. Numbers show performance drop from in-domain training.

Method	In-Domain	Cross-Domain (Zero-Shot Transfer)				Avg Transfer
	Industry	Electronics	Agrifood	Infrastructure	Medical	Gap
PatchCore	81.8	48.8 -33.0	55.6 -26.2	60.1 -21.7	58.4 -23.4	-26.1%
UniAD	73.4	48.3 -25.1	56.1 -17.3	58.7 -14.7	64.1 -9.3	-16.6%
RD	71.9	55.3 -16.6	57.0 -14.9	57.8 -14.1	61.1 -10.8	-14.1%
VITAD	82.2	58.5 -23.7	55.5 -26.7	52.8 -29.4	58.8 -23.4	-25.8%
Dinomaly	90.1	54.3 -35.8	55.2 -34.9	59.0 -31.1	55.5 -34.6	-34.1%

Table 4. Zero-shot and few-shot anomaly detection performance on ADNet. We report I-AUROC and P-AUROC (%) across different shot settings.

Method	0-shot		1-shot		4-shot	
	I	P	I	P	I	P
WinCLIP [31]	60.3	74.7	63.9	85.9	66.4	86.5
Dinomaly	-	-	68.6	89.9	72.6	90.6
Dinomaly^m	-	-	66.8	88.6	72.3	91.1

All methods experience substantial performance degradation under zero-shot cross-domain evaluation, with average drops ranging from 14.1% (RD) to 34.1% (Dinomaly). Notably, high in-domain performance does not guarantee cross-domain transferability: Dinomaly achieves the highest in-domain performance (90.1% on Industry) yet exhibits the most severe transfer gap, whereas RD with moderate source domain performance (71.9%) transfers most robustly across domains.

5.4. Few-shot Multi-Class Anomaly Detection

We evaluated the few-shot capability of multi-class models (Dinomaly and Dinomaly^m) under 1-shot and 4-shot settings, with WinCLIP [31] serving as the single-class few-shot baseline (Table 4). Both multi-class models demonstrated strong sample efficiency. Dinomaly yielded 68.6% and 72.6% I-AUROC with 1-shot and 4-shot, while Dinomaly^m exhibited similar results of 66.8% and 72.3%. More remarkably, at the 4-shot level, Dinomaly and Dinomaly^m not only surpassed one-for-one few-shot methods (WinCLIP) but also outperformed several full-shot multi-class baselines, including UniAD (63.4%), RD (66.9%), and ViTAD (72.4%, as shown in Table 2).

5.5. Performance Across Dataset Scales

To systematically understand how methods scale with increasing category diversity, we evaluated performance on ADNet subsets of progressively increasing sizes: 50, 100, 200, and 380 categories, as shown in Table 5. To isolate the degeneration from varied categories, we also evaluated a fixed 50-category subset across different training scales.

Table 5. Performance degradation as dataset scale increases. All methods use multi-class evaluation on ADNet subsets of varying sizes. For ADNet-100/200/380, we also report performance on the fixed ADNet-50 subset to demonstrate negative transfer effects.

Method	ADNet-50		ADNet-100				ADNet-200				ADNet-380			
	All		All		ADNet-50		All		ADNet-50		All		ADNet-50	
	I	P	I	P	I	P	I	P	I	P	I	P	I	P
UniAD	68.5	88.5	67.9	88.6	67.2	87.6	67.3	88.2	65.6	86.9	63.4	86.6	61.9	85.1
RD	76.7	89.7	73.4	89.7	73.5	88.5	67.7	88.0	67.8	87.3	66.9	87.4	68.6	86.9
SimpleNet	56.9	76.5	54.3	77.5	56.5	77.7	52.6	67.9	52.6	67.3	52.7	65.0	53.2	65.4
Dinomaly	86.0	93.7	83.2	93.2	82.8	92.4	82.3	92.4	81.1	91.8	78.5	91.0	78.4	90.9
Dinomaly^m	89.0	94.9	86.7	94.6	86.0	93.8	85.9	94.0	84.9	93.4	83.2	93.1	83.0	92.9

Most methods exhibited non-linear performance degradation characterized by a rapid decline from 50 to 200 categories, followed by a plateau phase. For instance, RD experienced a sharp 9.0% drop in 200 categories but only an additional 0.8% decline when scaling to 380 categories. This pattern suggests that methods encounter a critical complexity threshold around 100-200 categories, beyond which additional categories introduce diminishing marginal difficulty. Our Dinomaly^m achieved the optimal results, exhibiting -6.0% degradation on the fixed subset ADNet-50. This result demonstrates that MoE-based context-adaptive capacity effectively mitigates negative transfer at scale.

5.6. Ablation Study

We conducted comprehensive ablation studies on Dinomaly^m using the full ADNet benchmark. Table 6 systematically validates our key design choices.

Impact of Expert Count. Our cgMoE replaces standard feed-forward layers (FFN) with a weighted ensemble of expert networks, where routing weights are dynamically computed from the encoder's [CLS] token. Starting from a single FFN baseline (78.5% I-AUROC), we observed consistent improvements with increasing expert count. Notably, scaling to 12 experts yielded only marginal gains (83.3%), indicating that 8 experts sufficiently capture ADNet's context diversity without introducing redundant capacity.

Routing Strategy. We compared two routing strategies: using the encoder's [CLS] token versus the [CLS] token in each MoE layer in the decoder. Encoder-based routing (83.2%) outperformed decoder-based routing (82.7%), validating the hypothesis that pre-encoded representations contain richer semantic context for expert selection.

Backbone Scaling. We further evaluated Dinomaly^m with ViT-Large encoder. The larger backbone achieved 84.3% I-AUROC, representing a 1.1% absolute improvement over ViT-Base (83.2%). This result demonstrates that cgMoE and backbone scaling provide complementary benefits. However, we adopted ViT-Base as our default configuration, as it offers the optimal accuracy-efficiency trade-off for practical deployment scenarios where computational resources are constrained.

Visualization. We further visualize anomaly heatmaps by comparing our Dinomaly^m with the baseline Dinomaly, as shown in Figure 8. The results show that Dinomaly^m attends more precisely to anomalous regions, producing clearer and more focused localization.

Table 6. Ablation of Dinomaly^m design on ADNet.

Configuration	I-AUROC	P-AUROC
Single FFN (baseline)	78.5	89.4
cgMoE-2 experts	79.6	91.5
cgMoE-4 experts	81.6	92.1
cgMoE-8 experts (ours)	83.2	93.1
cgMoE-12 experts	83.3	92.8
cgMoE-8, decoder [CLS] routing	82.7	92.8
cgMoE-8, ViT-Large	84.3	93.6

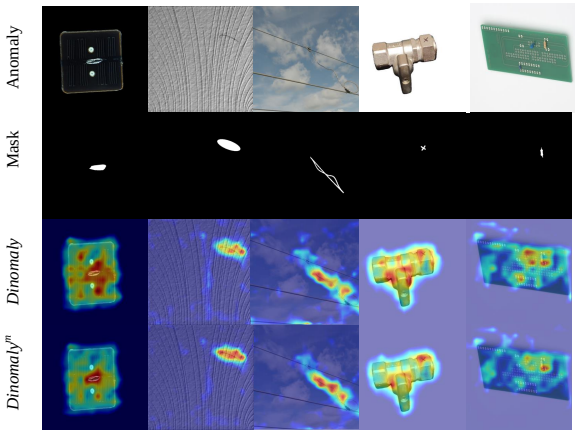

Figure 8. Result Visualization of Dinomaly and Dinomaly^m.

6. Conclusion

We introduced ADNet, a large-scale benchmark with 380 categories and 196,294 images spanning five domains, and proposed Dinomaly^m, a context-guided mixture-of-experts extension that achieves 83.2% I-AUROC. Our evaluation reveals that current AD methods struggle with both scalability (78.5% on 380 categories vs. 92–99% on small benchmarks) and cross-domain transfer (14–34% degradation), highlighting the substantial challenges posed by ADNet.

Limitation and Outlook. Despite this progress, ADNet's scale remains limited, with an imbalanced domain distribution (*e.g.*, Industry with 154 categories vs. Medical with only 28 categories). In future work, we plan to continuously expand ADNet by incorporating more diverse datasets to achieve better domain balance and larger category coverage. We will also explore stronger baseline methods and investigate the feasibility of building foundation models for anomaly detection that can effectively handle the two core challenges: scaling to thousands of categories while maintaining performance, and understanding context-dependent anomalies where identical visual patterns exhibit opposite meanings across different application domains.

References

[1] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, et al. Flamingo: a visual language model for few-shot learning. *NeurIPS*, 35:23716–23736, 2022. 4

[2] Krzysztof R Apt, Howard A Blair, and Adrian Walker. Towards a theory of declarative knowledge. In *Foundations of deductive databases and logic programming*, pages 89–148. Elsevier, 1988. 4

[3] Jaehyeok Bae, Jae-Han Lee, and Seyun Kim. Pni: Industrial anomaly detection using position and neighborhood information. In *ICCV*, pages 6373–6383, 2023. 3

[4] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021. 1, 2

[5] Tianpeng Bao, Jiadong Chen, Wei Li, Xiang Wang, Jingjing Fei, et al. Miad: A maintenance inspection dataset for unsupervised anomaly detection. In *ICCV*, pages 993–1002, 2023. 4

[6] Yanqi Bao, Kechen Song, Jie Liu, Yanyan Wang, Yunhui Yan, et al. Triplet-graph reasoning network for few-shot metal generic surface defect segmentation. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2021. 2

[7] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, pages 9592–9600, 2019. 1, 2, 4, 6

[8] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045*, 2021. 2

[9] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *IJCV*, 130(4):947–969, 2022. 2

[10] Hanna Borgli, Vajira Thambawita, Pia H Smedsrud, Steven Hicks, Debesh Jha, et al. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data*, 7(1):283, 2020. 2

[11] Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. *arXiv preprint arXiv:2309.16588*, 2023. 4

[12] ITI-Instituto Tecnológico de Informática. Wood anomaly detection one class classification, 2024. 2

[13] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021. 3

[14] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, pages 9737–9746, 2022. 1, 2

[15] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, pages 9737–9746, 2022. 6

[16] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255. Ieee, 2009. 2

[17] Lei Fan, Yiwen Ding, Maurice Pagnucco, and Yang Song. Patch-wise augmentation for anomaly detection and localization. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5425–5429. IEEE, 2024. 1, 2

[18] Lei Fan, Dongdong Fan, Yiwen Ding, Yong Wu, Donglin Di, et al. Grainbrain: Multiview identification and stratification of defective grain kernels. *IEEE Transactions on Industrial Informatics*, 2025. 2

[19] Lei Fan, Dongdong Fan, Zhiguang Hu, Yiwen Ding, Donglin Di, Kai Yi, et al. Manta: A large-scale multi-view and visual-text anomaly detection dataset for tiny objects. In *CVPR*, pages 25518–25527, 2025. 1, 2, 4

[20] Lei Fan, Junjie Huang, Donglin Di, Anyang Su, Tianyou Song, et al. Salvaging the overlooked: Leveraging class-aware contrastive learning for multi-class anomaly detection. In *ICCV*, pages 21419–21428, 2025. 1, 2, 5, 6

[21] Hu Feng, Kechen Song, Wenqi Cui, Yiming Zhang, and Yunhui Yan. Cross position aggregation network for few-shot strip steel surface defect segmentation. *IEEE Transactions on Instrumentation and Measurement*, 72:1–10, 2023. 2

[22] Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, et al. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11): 665–673, 2020. 3

[23] Tao Gong, Qi Chu, Bin Liu, Wei Zhou, and Nenghai Yu. Feclip: Frequency enhanced clip model for zero-shot anomaly detection and segmentation. In *ICCV*, pages 21220–21230, 2025. 3

[24] Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. Univad: A training-free unified model for few-shot visual anomaly detection. In *CVPR*, pages 15194–15203, 2025. 3

[25] Jia Guo, Shuai Lu, Lize Jia, Weihang Zhang, and Huiqi Li. Recontrast: Domain-specific anomaly detection via contrastive reconstruction. *NeurIPS*, 36:10721–10740, 2023. 3

[26] Jia Guo, Shuai Lu, Lei Fan, Zelin Li, Donglin Di, et al. One dinomaly2 detect them all: A unified framework for full-spectrum unsupervised anomaly detection. *arXiv preprint arXiv:2510.17611*, 2025. 3

[27] Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, and Hongen Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *CVPR*, pages 20405–20415, 2025. 1, 2, 3, 5, 6

[28] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, et al. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *NeurIPS*, 37:71162–71187, 2024. 6

[29] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, et al. A diffusion-based framework for multi-class anomaly detection. In *AAAI*, pages 8472–8480, 2024. 3

[30] Sebastian Höfer, Dorian Henning, Artemij Amiranashvili, Douglas Morrison, Mariliza Tzes, et al. Kaputt: A large-

scale dataset for visual defect detection. In *ICCV*. IEEE, 2025. 2

[31] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *CVPR*, pages 19606–19616, 2023. 3, 7

[32] Xi Jiang, Jianlin Liu, Jinbao Wang, Qiang Nie, Kai Wu, et al. Softpatch: Unsupervised anomaly detection with noisy data. *NeurIPS*, 35:15433–15445, 2022. 1, 2

[33] Joo Chan Lee, Taejune Kim, Eunbyung Park, Simon S Woo, and Jong Hwan Ko. Continuous memory representation for anomaly detection. In *ECCV*, pages 438–454. Springer, 2024. 3

[34] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, pages 9664–9674, 2021. 1, 2

[35] Wenqiao Li, Bozhong Zheng, Xiaohao Xu, Jinye Gan, Fading Lu, et al. Multi-sensor object anomaly detection: Unifying appearance, geometry, and internal properties. In *CVPR*, pages 9984–9993, 2025. 1, 2, 4

[36] Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, et al. Promptad: Learning prompts with only normal samples for few-shot anomaly detection. In *CVPR*, pages 16838–16848, 2024. 3

[37] Zhaoxu Li, Yingqian Wang, Chao Xiao, Qiang Ling, Zaiping Lin, and Wei An. You only train once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–18, 2023. 1, 2

[38] Yun Liang, Zhiguang Hu, Junjie Huang, Donglin Di, Anyang Su, et al. Tocoad: Two-stage contrastive learning for industrial anomaly detection. *IEEE Transactions on Instrumentation and Measurement*, 74:1–9, 2025. 1, 3

[39] Sook-Lei Liew, Bethany P Lo, Miranda R Donnelly, Artemis Zavaliangos-Petropulu, Jessica N Jeong, et al. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. *Scientific data*, 9(1):320, 2022. 2

[40] Wenrui Liu, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Diversity-measurable anomaly detection. In *CVPR*, pages 12147–12156, 2023. 2

[41] Xiaohao Liu, Xiaobo Xia, See-Kiong Ng, and Tat-Seng Chua. Continual multimodal contrastive learning. *arXiv preprint arXiv:2503.14963*, 2025. 3

[42] Xiaohao Liu, Xiaobo Xia, See-Kiong Ng, and Tat-Seng Chua. Principled multimodal representation learning. *arXiv preprint arXiv:2507.17343*, 2025. 3

[43] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *CVPR*, pages 20402–20411, 2023. 2

[44] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *CVPR*, pages 20402–20411, 2023. 6

[45] Hongchi Ma, Guanglei Yang, Debin Zhao, Yanli Ji, and Wangmeng Zuo. Remp-ad: Retrieval-enhanced multi-modal prompt fusion for few-shot industrial visual anomaly detection. In *ICCV*, pages 20425–20434, 2025. 3

[46] Wenxin Ma, Xu Zhang, Qingsong Yao, Fenghe Tang, Chenxu Wu, Yingtai Li, Rui Yan, Zihang Jiang, and S Kevin Zhou. Aa-clip: Enhancing zero-shot anomaly detection via anomaly-aware clip. In *CVPR*, pages 4744–4754, 2025. 3

[47] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 5

[48] Yanjun Peng and Jindong Sun. The multimodal mri brain tumor segmentation based on ad-net. *Biomedical Signal Processing and Control*, 80:104336, 2023. 1

[49] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763. PMLR, 2021. 3

[50] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, et al. Self-supervised predictive convolutional attentive block for anomaly detection. In *CVPR*, pages 13576–13586, 2022. 2

[51] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, pages 14318–14328, 2022. 1, 2, 3, 6

[52] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, et al. Deep one-class classification. In *ICML*, pages 4393–4402. PMLR, 2018. 3

[53] Javier Silvestre-Blanes, Jorge Moreno, Ignacio Miralles, Teresa Albero-Albero, and Rubén Pérez-Llorens. Afid: A public fabric image database for defect detection. *AUTEX Research Journal*, pages 1–10, 2019. Early access. 2

[54] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-Based Deep-Learning Approach for Surface-Defect Detection. *Journal of Intelligent Manufacturing*, 2019. 2

[55] Jiaqi Tang, Hao Lu, Xiaogang Xu, Ruizheng Wu, Sixing Hu, et al. An incremental unified framework for small defect inspection. In *ECCV*, pages 307–324. Springer, 2024. 3

[56] Sanli Tang, Fan He, Xiaolin Huang, and Jie Yang. Online pcb defect detector on a new pcb defect dataset. *arXiv preprint arXiv:1902.06197*, 2019. 1, 2

[57] Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, et al. Real-iad: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In *CVPR*, pages 22883–22892, 2024. 2, 4

[58] Enquan Yang, Peng Xing, Hanyang Sun, Wenbo Guo, Yuanwei Ma, et al. 3cad: A large-scale real-world 3c product dataset for unsupervised anomaly detection. In *AAAI*, pages 9175–9183, 2025. 1

[59] Fan Yang, Lei Zhang, Sijia Yu, Danil Prokhorov, Xue Mei, and Haibin Ling. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Transactions on Intelligent Transportation Systems*, 21(4):1525–1535, 2019. 2

[60] Xincheng Yao, Ruoqi Li, Zefeng Qian, Lu Wang, and Chongyang Zhang. Hierarchical gaussian mixture normalizing flow modeling for unified anomaly detection. In *ECCV*, pages 92–108. Springer, 2024. 3

[61] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, et al. A unified model for multi-class anomaly detection. *NeurIPS*, 35:4571–4584, 2022. 1, 3, 5, 6

[62] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8330–8339, 2021. 2

[63] Jiangning Zhang, Haoyang He, Zhenye Gan, Qingdong He, Yuxuan Cai, et al. Ader: A comprehensive benchmark for multi-class visual anomaly detection. *arXiv preprint arXiv:2406.03262*, 2024. 6

[64] Jiangning Zhang, Xuhai Chen, Yabiao Wang, Chengjie Wang, Yong Liu, et al. Exploring plain vit features for multi-class unsupervised visual anomaly detection. *Computer Vision and Image Understanding*, 253:104308, 2025. 6

[65] Tao Zhang, Sheng Zhong, Wenhui Xu, Luxin Yan, and Xu Zou. Catenary insulator defect detection: A dataset and an unsupervised baseline. *IEEE Transactions on Instrumentation and Measurement*, 73:1–15, 2024. 2

[66] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, et al. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *CVPR*, pages 3914–3923, 2023. 6

[67] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *CVPR*, pages 16699–16708, 2024. 1, 2

[68] Ying Zhao. Omnial: A unified cnn framework for unsupervised anomaly localization. In *CVPR*, pages 3924–3933, 2023. 3

[69] Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. *arXiv preprint arXiv:2310.18961*, 2023. 3

[70] Wenbing Zhu, Lidong Wang, Ziqing Zhou, Chengjie Wang, Yurui Pan, et al. Real-iad d3: A real-world 2d/pseudo-3d/3d dataset for industrial anomaly detection. In *CVPR*, pages 15214–15223, 2025. 2

[71] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *ECCV*, pages 392–408. Springer, 2022. 1, 2, 4