This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Towards an Incremental Unified Multimodal Anomaly Detection: Augmenting Multimodal Denoising From an Information Bottleneck Perspective

Kaifang Long[1], Lianbo Ma[1*], Jiaqi Liu[2], Liming Liu[1], Guoyang Xie[3]
[1]Software College, Northeastern University, China, [2]Independent Researcher, [3]CATL, China
longkf@stumail.neu.edu.cn, malb@swc.neu.edu.cn, guoyang.xie@ieee.org

Abstract

The quest for incremental unified multimodal anomaly detection seeks to empower a single model with the ability to systematically detect anomalies across all categories and support incremental learning to accommodate emerging objects/categories. Central to this pursuit is resolving the catastrophic forgetting dilemma, which involves acquiring new knowledge while preserving prior learned knowledge. Despite some efforts to address this dilemma, a key oversight persists: ignoring the potential impact of spurious and redundant features on catastrophic forgetting. In this paper, we delve into the negative effect of spurious and redundant features on this dilemma in incremental unified frameworks, and reveal that under similar conditions, the multimodal framework developed by naive aggregation of unimodal architectures is more prone to forgetting. To address this issue, we introduce a novel denoising framework called IB-IUMAD, which exploits the complementary benefits of the Mamba decoder and information bottleneck fusion module: the former dedicated to disentangle inter-object feature coupling, preventing spurious feature interference between objects; the latter serves to filter out redundant features from the fused features, thus explicitly preserving discriminative information. A series of theoretical analyses and experiments on MVTec 3D-AD and Eyecandies datasets demonstrates the effectiveness and competitive performance of IB-IUMAD. Code Link.

1. Introduction

Motivation. Multimodal anomaly detection (MAD) has emerged as a pivotal task in industrial quality inspection, with the aim of recognizing and locating product surface defects using RGB and depth images [16, 27, 35]. Most existing MAD methods follow an *N-objects-N-models* principle [8], where a separate model is designed for each object/category (i.e., an industrial product). However, this paradigm suffers from significant limitations of high computational cost, substantial memory consumption, and weak generalization. Then, the *N-objects-One-model* has become a promising schema, enabling a single model to detect anomalies between multiple categories [5, 17]. Especially, in real industrial scenarios, the continuous emergence of new objects is a prevalent situation [32], which requires MAD models based on the *N-objects-One-model* to support incremental learning for the adaptation to new ones. Thus, this study focuses on the exploration of the incremental unified multimodal anomaly detection (IUMAD) task.

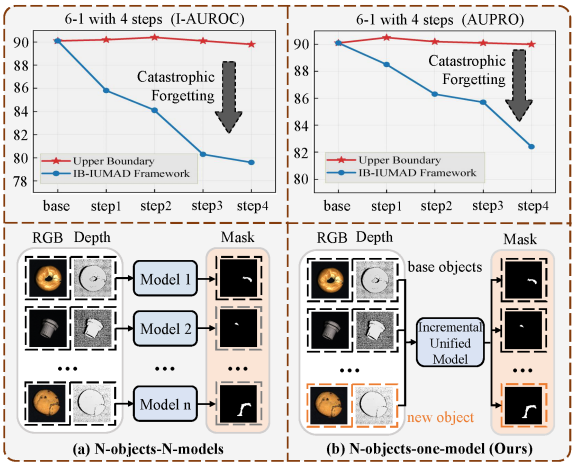


Figure 1. **Top** shows the performance of IUMAD task on the MVTec 3D-AD dataset, where catastrophic forgetting occurs significantly. ***Bottom*** is the comparison of our incremental unified multimodal framework with previous paradigms.

Challenges. In fact, existing works about *N-objects-One-model* focus on unimodal anomaly detection (e.g., for RGB images) [25] rather than IUMAD. As shown in Figure 1, the performance on the base unified model, trained with 6 objects in MVTec 3D-AD, is excellent but decreases sharply as the model trains on new objects incrementally. Thus, central to this quest is how to effectively preserve acquired knowledge as new objects are continuously introduced, i.e.,

*Corresponding author.

how to overcome catastrophic forgetting dilemma. For this purpose, the object-aware self-attention and semantic compression loss are used to capture semantic boundaries, improving adaptability to new objects [43]. Then, the gradient projection method is devised for stable incremental learning using diffusion models [29]. However, these works neglect the potential impacts of inter-object spurious features and redundant information in fused features towards catastrophic forgetting, which has been demonstrated in conventional machine learning scenarios [14]. More critically, we observe that compared to unimodal detection scenarios, the aforementioned impacts exert substantially amplified influence on MAD due to the inherent complexity of cross-modal feature fusion. Consequently, the incremental unified design for MAD presents significantly greater technical challenges, particularly in maintaining inter-modal interactions during sequential learning phases.

To investigate the above issue, we conducted a series of preliminary experiments on IUF [43] and its variant to examine the impact of spurious and redundant features on catastrophic forgetting in incremental unified unimodal and multimodal frameworks, evaluating accuracy and average forgetting when new objects are learned incrementally. As shown in Figure 2, introducing spurious and redundant feature perturbations leads to substantial performance degradation in both frameworks, along with a marked increase in average forgetting. Especially for multimodal frameworks, the feature fusion process increases the model's susceptibility to capturing spurious and redundant features from different modalities, resulting in more severe deterioration than unimodal and even causing performance collapse.

How to mitigate catastrophic forgetting triggered by spurious and redundant features in IUMAD? A straightforward and intuitive solution is to devise two independent modules with distinct functions: one dedicated to mitigating interference from inter-object spurious features, and the other focused on filtering out redundant information in multimodal fusion features. Building on these insights, we extend the Mamba technique for IUMAD and enhance multimodal fusion from the information bottleneck perspective, thereby proposing a novel denoising framework, called IB-IUMAD, which leverages the complementary strengths of the Mamba decoder and information bottleneck fusion module. Specifically, the former focuses on integrating the fine-grained features with label information into the multimodal reconstruction network, effectively disentangling inter-object feature coupling and preventing spurious feature interference. The latter is designed to filter out redundant information from the fused multimodal features, explicitly preserving discriminative features. Extensive experiments show that IB-IUMAD consistently improved performance compared to baselines.

Contributions. In summary, the main contributions of this work are as follows:

- We empirically validate the impact of spurious and redundant features on catastrophic forgetting and then identify the critical denoising module designs that boost IUMAD performance from an information bottleneck perspective.
- We propose a novel framework, called IB-IUMAD, which leverages the complementary strengths of the Mamba decoder and information bottleneck regularization to mitigate the interference of spurious and redundant features.
- A series of theoretical analyses and experiments demonstrates that the proposed method consistently improves performance in terms of accuracy, memory usage, and frame rate.

To the best of our knowledge, this is the first work to address MAD in an incremental and unified manner. Our method outperforms state-of-the-art baselines across four incremental settings, achieving higher accuracy and frame rates with lower forgetting and memory costs. Specifically, under the 6-1 with 4-step setting, IB-IUMAD boosts I-AUROC/AUPRO by 3.5%/2.9% and reduces forgetting by 5.8%/1.5% on MVTec 3D-AD. In the 10-0 with 0-step setting, it cuts memory usage by 44× and delivers at least 41× faster inference than the N-objects-N-models approach, while maintaining comparable performance.

2. Spurious and Redundant Features Impact

This section investigates the impact of spurious and redundant features on catastrophic forgetting in incremental unified frameworks, with a focus on multimodal settings.

2.1. Problem Formulation

Figure 1-(b) outlines the formal definition of the problem setting for IUMAD tasks, which aim to develop a single model capable of: comprehensive anomaly identification across heterogeneous objects, and continuous knowledge integration through incremental learning to accommodate emerging object types. Specifically, we divide the dataset containing N objects into T training steps ($T \leq N$). In the t-th training step, the model updates its weight parameters solely based on the current object, without accessing any data from previous steps. During testing at the t-th step, we evaluate anomaly scores for all seen objects so far and employ the forgetting metric (FM) to quantify the degree of forgetting of previously learned knowledge. For example, in the 6-1 with 4 steps or 6-4 with 1 step setting, we first train a base model with six objects, and then incrementally learn one or four additional objects per step across four or one incremental step(s), respectively, to obtain the final IUMAD model. Other settings follow the same principle.

2.2. Empirical Study Settings

Benchmark. IUF [43] is selected as the foundational framework, given its widespread application in existing re-

search [23]. Our primary goal is to examine how spurious and redundant features exacerbate catastrophic forgetting in incremental settings, evaluating three input patterns: RGB-only, depth-only, and their combination.

Evaluation Metrics. We evaluate model performance using several standard metrics (I-AUROC, P-AUROC, and AUPRO) and introduce the forgetting metric (FM) to quantify the degree of forgetting of previously learned knowledge when learning new objects incrementally; the lower the FM, the better the performance. FM as defined by:

$$FM = \frac{1}{N-1}\sum_{o}^{N-1} max_{s\in[1,N-1]}(I_{s,o}^{Acc} - I_{N,o}^{Acc}), \quad (1)$$

where N denotes the total incremental steps. $I_{s,o}^{Acc}/I_{N,o}^{Acc}$ represents the I-AUROC, P-AUROC, or AUPRO score for the o-th object at the s-th/N-th step. $max_s(\cdot)$ denotes the maximum score drop for each object across all steps.

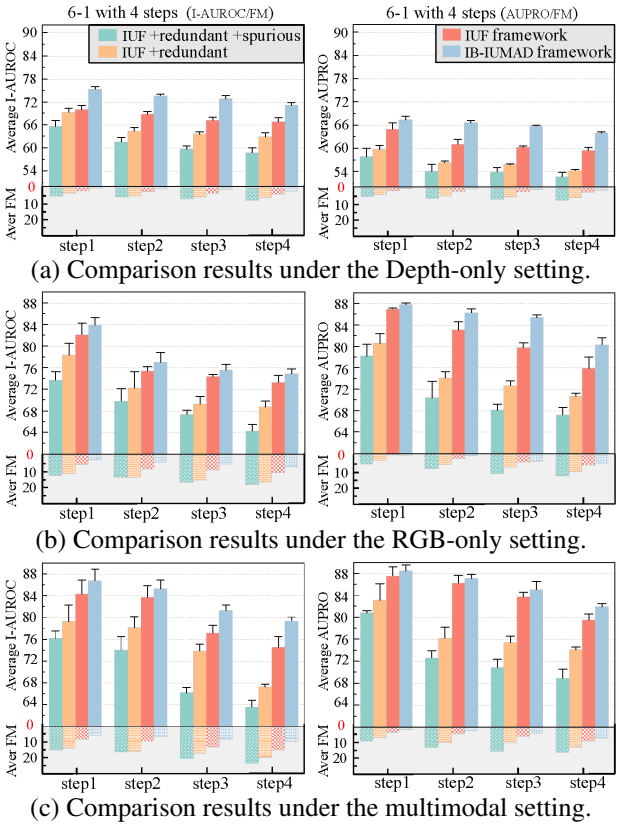
Implementation Details. In our empirical study protocol, we train a unified model on 6 basic objects from MVTec 3D-AD, followed by incrementally introducing the remaining objects over 4 steps (i.e., 6-1 with 4 steps), and simulate the interference of spurious and redundant features by injecting background information (acquired from the other objects, as detailed in Supplementary Material) and Berlin noise with different intensities into the model. To ensure statistical robustness, we perform four runs with different random seeds and report the mean and standard deviation of results. Each run consists of: (i) initial training for 1,000 epochs on basic objects, and (ii) training for 800 epochs per incremental step.

2.3. Empirical Observations

To evaluate the impact of spurious and redundant features on catastrophic forgetting in the incremental unified unimodal and multimodal framework, under the 6-1 with 4 steps setting, we report the performance at each incremental step, involving I-AUROC, AUPRO, and FM. In brief, as previewed in Figure 2, we derive the following **observations**:

❶ Under the unimodal setting, injecting redundant features degrades model performance and accelerates forgetting, while introducing additional spurious features further intensifies this decline. This indicates that the model becomes increasingly sensitive to such features during incremental learning, thereby amplifying their negative impact on overall performance. (Figure 2-a and b)

❷ Under the multimodal setting, we observe a performance degradation trend similar to the unimodal case, but with a more pronounced decline. A key contributing factor is the inherent complexity of cross-modal fusion, increases the likelihood of capturing spurious or redundant features, thereby exacerbating catastrophic forgetting. (Figure 2-c)



(a) Comparison results under the Depth-only setting.

(b) Comparison results under the RGB-only setting.

(c) Comparison results under the multimodal setting.

Figure 2. The impact of spurious and redundant features on catastrophic forgetting in incremental unified frameworks.

❸ As shown in Figure 2 (blue bars), employing a denoising constraint for IUMAD, such as information bottleneck regularization, can effectively alleviate the negative impact of spurious and redundant features on overall performance.

❹ In summary, how to effectively mitigate the catastrophic forgetting problem triggered by spurious and redundant features is crucial to improving model performance.

Based on the above insights, we identify that spurious and redundant features pose a considerable challenge to IUMAD tasks. This motivates us to design an incremental unified denoising framework for improving the model's performance.

3. Method

In this section, we present IB-IUMAD, a denoising method that mitigates catastrophic forgetting arising from spurious and redundant features. We further outline its framework, detailing the denoising module and theoretical analysis.

3.1. Overview

Figure 3 shows the skeleton of IB-IUMAD, which comprises Multimodal Feature Extraction Networks (MFEN), Mamba decoders, Multimodal Reconstruction Networks

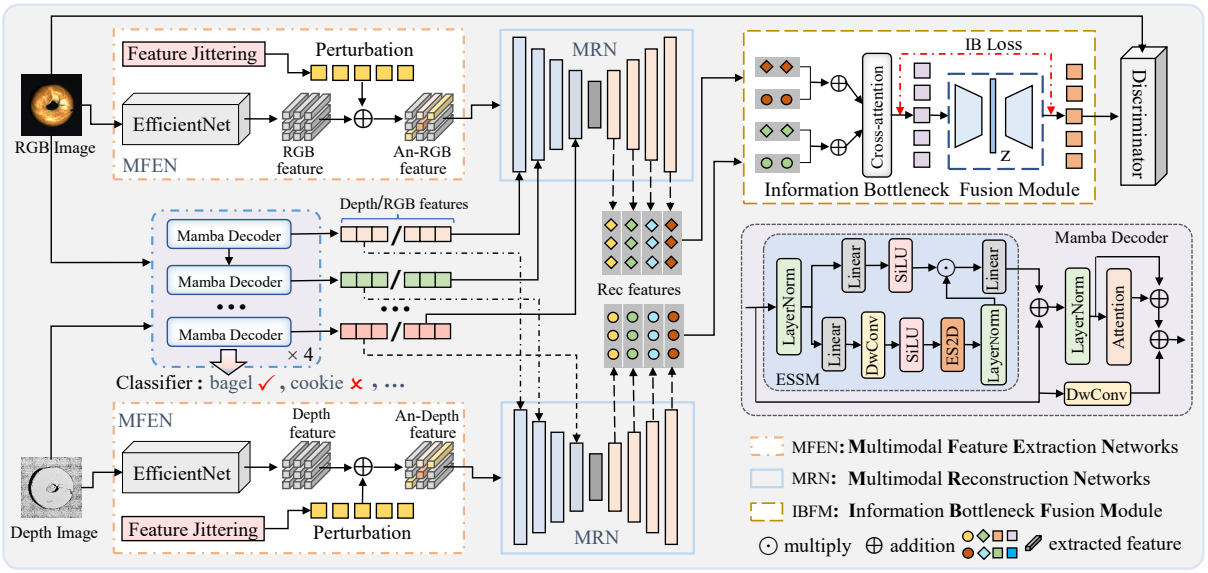


Figure 3. The overall framework of IB-IUMAD, where the Mamba decoder and information bottleneck fusion module (IBFM) are the core designs of this work, aims to mitigate the effects of spurious and redundant features on catastrophic forgetting.

(MRN), an Information Bottleneck Fusion Module (IBFM), and a Discriminator. MFEN aims to extract RGB and depth image features via EfficientNet and synthesize them into abnormal features using feature jitters. Then, MRN restores these abnormal features to normal ones while acquiring multi-scale multimodal representations. To mitigate interference of spurious features between objects during normal feature reconstruction within the MRN, we employ the Mamba decoder to disentangle inter-object feature coupling. Subsequently, IBFM serves to fuse multi-scale multimodal features from MRN and imposes information bottleneck regularization to filter out redundant information from the multimodal fusion features, which effectively alleviates catastrophic forgetting. Particularly, the Mamba decoders and IBFM constitute the core components of the IB-IUMAD framework. The following subsections detail their designs.

3.2. Mamba Decoder

The above experiments demonstrate that spurious and redundant features significantly exacerbate catastrophic forgetting in IUMAD tasks. To improve its performance, it is crucial to enhance the model's denoising capability to mitigate the interference of spurious and redundant features.

Specifically, the generation of spurious features primarily stems from the tight coupling of features between objects that share the same feature space. Especially, as the model incrementally learns new objects, it tends to indiscriminately update the feature space of previously learned ones, ultimately making the MRN more vulnerable to inter-object spurious features during reconstruction. To address this issue, inspired by [39], we incorporate the Mamba decoders and a label classifier to disentangle inter-object feature coupling. As shown in Figure 3, each Mamba decoder comprises an efficient state space module (ESSM), deep separable convolution (DwConv), and an attention mechanism. Firstly, the ESSM employs DwConv and Efficient 2D Scanning (ES2D) [34] to efficiently sample each visual patch, extracting fine-grained information from RGB and depth images. Then, the extracted features are processed by the attention mechanism, forming the Mamba decoder's output features, as follows:

$$\begin{aligned} \hat{X}_{R/D}^{i+1} &= DwConv(X_{R/D}^{i}), \\ \widetilde{X}_{R/D}^{i+1} &= ESSM(LN(X_{R/D}^{i})), \\ X_{R/D}^{i+1} &= Attention(LN(\widetilde{X}_{R/D}^{i+1})) + \hat{X}_{R/D}^{i+1}, \end{aligned} \tag{2}$$

where R/D denotes RGB and depth images, respectively; $X_{R/D}^{i}$ represents the output of the i-th mamba decoder, $i \in [0, 4]$; specifically, when $i = 0$, $X_{R/D}^{i}$ denotes the original image features. LN indicates layer normalization.

Subsequently, the output of each Mamba decoder is independently integrated into the MRN to assist feature reconstruction, while the output of the final block is also fed into the classifier for label prediction. This design aims to guide the model to leverage label information to disentangle inter-object features, thereby minimizing the interference of spurious features. Thus, the classification loss function is

$$\mathcal{L}_{\mathrm{R/D}} = min \ \ \mathcal{L}_{\mathrm{CE}}(Y_{mab}^{\mathrm{RGB/Depth}}, Y), \tag{3}$$

where $\mathcal{L}_{\mathrm{CE}}(\cdot)$ denotes cross-entropy loss; Y_{mab}^{RGB} and

Y_{mab}^{Depth} represent the outputs of the RGB and depth image classifiers, respectively; Y is the ground-truth label.

3.3. Information Bottleneck Fusion Module

Once the MRN reconstructs the abnormal RGB and depth images synthesized by the MFEN into normal features using the additional features provided by the Mamba encoders, the cascade and cross-attention mechanisms are employed to fuse both modalities, as detailed follows:

$$\begin{aligned} F_{fusion}^{1} = f_{R}^{rec2} \oplus f_{D}^{rec2}, F_{fusion}^{2} = f_{R}^{rec4} \oplus f_{D}^{rec4}, \\ F_{fu} = Cross_Att(F_{fusion}^{1}, F_{fusion}^{2}). \end{aligned} \tag{4}$$

Then, we use an information bottleneck regularization module to filter redundant information from the initial fusion feature (i.e., F_{fu}), obtaining a more predictive feature (i.e., F_{fu}^{g}). This module comprises two linear projection layers, dropout, and ReLU activation. Specifically, z denotes the reprojection of F_{fu}^{g} back to the dimension of F_{fu}, as:

$$F_{fu} \xrightarrow[\text{Linear Projection}]{\text{Dropout + ReLU}} z \xrightarrow[\text{Linear Projection}]{\text{Dropout + ReLU}} F_{fu}^{g}. \tag{5}$$

To ensure that F_{fu}^{g} sufficiently retains the predictive information of F_{fu} while discarding redundant features, we quantify this relationship using the mutual information between F_{fu}^{g} and F_{fu}, defined as follows:

$$I(F_{fu}; F_{fu}^{g}) = \mathbb{E}_{p(F_{fu}, F_{fu}^{g})} \left[\log \frac{p(F_{fu}, F_{fu}^{g})}{p(F_{fu}) p(F_{fu}^{g})} \right]. \tag{6}$$

Moreover, $I(F_{fu}; F_{fu}^{g})$ can be futher divide into two parts, $I(F_{fu}; F_{fu}^{g}|Y)$ and $I(F_{fu}^{g}; Y)$, via the mutual information chain rule (See ***Corollary* 1** in **Theoretical Analysis Section** for details), where $I(F_{fu}; F_{fu}^{g}|Y)$ represents redundant information and $I(F_{fu}^{g}; Y)$ indicates prediction-related information for the current object. Thus, to effectively eliminate redundant features, we only need to maximize $I(F_{fu}^{g}; Y)$ while minimizing $I(F_{fu}; F_{fu}^{g}|Y)$. Given F_{fu}^{g} derived from F_{fu}, the information contained in F_{fu}^{g} cannot exceed that in F_{fu}, i.e., $I(F_{fu}^{g}; Y) \leq I(F_{fu}; Y)$. Consequently, the above information bottleneck objective function is equivalent to:

$$min \quad I(F_{fu}; Y) - I(F_{fu}^{g}; Y). \tag{7}$$

Finally, to optimize the above objective function Eq. 7, we adopt Kullback-Leibler divergence as the loss function to optimize feature F_{fu}^{g} (See ***Corollary* 2** in **Theoretical Analysis Section** for details), defined as:

$$\mathcal{L}_{\text{IB}} = \text{KL}[P(Y|F_{fu})||P(Y|F_{fu}^{g})]. \tag{8}$$

In summary, for the IUMAD task, we first employ the classic MSE loss for reconstruction, followed by the cross-entropy loss for classification on both RGB and depth images to prevent interference from spurious features. To further mitigate redundant information in the fused features, we use the KL divergence loss. Thus, the overall loss can be formulated as:

$$\mathcal{L}_{\text{Fusion}} = \frac{1}{W \times H} \left\| F_{\text{org}}^{\text{RGB}} - F_{fusion}^{g} \right\|_{2}^{2}, \tag{9}$$

$$\begin{aligned} \mathcal{L}_{\text{All}} = \lambda_1 \mathcal{L}_{\text{CE}}(Y_{mab}^{\text{RGB}}, Y) + \lambda_2 \mathcal{L}_{\text{CE}}(Y_{mab}^{\text{Depth}}, Y) \\ \underbrace{+ \lambda_3 \mathcal{L}_{\text{Fusion}} + \lambda_4 \mathcal{L}_{\text{IB}}(Y_{F_{fu}}, Y_{F_{fu}^{g}})}_{\text{Information Bottleneck Fusion Module}}, \end{aligned} \tag{10}$$

where λ_1, λ_2, λ_3, and λ_4 are loss balance hyperparameters, respectively, and all are set to 1 in our experiments.

3.4. Theoretical Analysis

Further, we theoretically analyze the effectiveness of the information bottleneck regularization in filtering redundant information within the multimodal fusion feature. For this purpose, we follow the information theory [14, 15, 44] and prepare several corollaries.

Corollary 1.

Given a random variable Y, the mutual information between features F_{fu} and F_{fu}^{g}, expressed in Eq. 6, can be equivalently rewritten as $I(F_{fu}; F_{fu}^{g}) = I(F_{fu}; F_{fu}^{g}|Y) + I(F_{fu}^{g}; Y)$ based on the chain rule.

Proof. Due to the limitation of pages, the detailed proof is provided in the Appendix A.

Corollary 2.

If the KL divergence between $p(Y|F_{fu})$ and $p(Y|F_{fu}^{g})$ is 0, that is, $\text{KL}[P(Y|F_{fu})||P(Y|F_{fu}^{g})] = 0$, we have $I(Y; F_{fu}) - I(Y; F_{fu}^{g}) = I(F_{fu}; Y) - I(F_{fu}^{g}; Y) = 0$.

Proof. $I(Y; F_{fu}) - I(Y; F_{fu}^{g})$

$$\begin{aligned} &= \iint P(F_{fu}) P(Y|F_{fu}) \log P(Y|F_{fu}) dF_{fu} dY \\ &- \iint P(F_{fu}^{g}) P(Y|F_{fu}^{g}) \log P(Y|F_{fu}^{g}) dF_{fu}^{g} dY \\ &= \mathbb{E}_{F_{fu}}[\text{KL}[P(Y|F_{fu})||P(Y|F_{fu}^{g})]] - \mathbb{E}_{F_{fu}^{g}}[\text{KL}[P \\ &(Y|F_{fu}^{g})||P(Y|F_{fu})]] + \int P(Y) \log \frac{P(Y|F_{fu}^{g})}{p(Y|F_{fu})} dY. \end{aligned}$$

By Jensen's inequality and the strict convexity of $-\log$, we conclude that the KL divergence is non-negative. When $\text{KL}[P(Y|F_{fu})||P(Y|F_{fu}^{g})] = 0$, it follows that $P(Y|F_{fu}^{g}) = P(Y|F_{fu})$ almost everywhere, which means that

Table 1. I-AUROC/AUPRO results of our approach on MVTec 3D-AD under four incremental unified anomaly detection settings.

	Method	10-0 with 0 step (setting 1)			9-1 with 1 step (setting 2)			6-4 with 1 step (setting 3)			6-1 with 4 steps (setting 4)		
		I-AUROC	AUPRO	FM (⇓)	I-AUROC	AUPRO	FM (⇓)	I-AUROC	AUPRO	FM (⇓)	I-AUROC	AUPRO	FM (⇓)
RGB	IUF (ECCV24)	87.1	88.5	– / –	82.4	85.1	2.9 / 1.5	76.7	78.6	8.3 / 6.9	73.9	76.3	12.5 / 7.9
	CDAD (CVPR25)	76.4	87.5	– / –	74.2	83.9	2.1 / 1.9	68.9	77.5	7.4 / 6.3	67.4	73.9	**8.0** / 6.7
	IB-IUMAD	**88.9**	**89.3**	– / –	**85.3**	**87.6**	**1.8** / **1.3**	**79.8**	**81.2**	**6.5** / **5.8**	**75.7**	**80.5**	8.7 / **6.3**
3D	IUF (ECCV24)	73.7	65.4	– / –	70.4	62.1	1.6 / 1.4	67.4	59.2	4.4 / 3.6	66.8	59.7	4.7 / 3.0
	CDAD (CVPR25)	60.6	40.6	– / –	58.8	39.2	1.1 / **0.2**	55.5	37.9	1.7 / 0.9	56.1	38.5	2.5 / **1.4**
	IB-IUMAD	**75.8**	**67.6**	– / –	**74.9**	**65.5**	**-0.3** / 0.7	**72.6**	**64.3**	**1.3** / **0.2**	**71.2**	**64.1**	**2.2** / **1.4**
RGB+3D	IUF (ECCV24)	88.7	89.2	– / –	84.2	86.6	3.7 / 2.2	79.2	80.2	10.6 / 7.3	75.1	79.5	15.1 / 8.4
	CDAD (CVPR25)	79.1	88.1	– / –	75.1	85.3	3.4 / 2.1	70.1	78.5	8.5 / 7.1	69.5	75.7	**8.9** / 7.7
	IB-IUMAD	**91.0**	**90.4**	– / –	**87.5**	**89.1**	**2.3** / **0.8**	**82.4**	**84.7**	**6.8** / **6.4**	**78.6**	**82.4**	9.3 / **6.9**

Table 2. I-AUROC/AUPRO results of our approach on Eyecandies under four incremental unified anomaly detection settings.

	Method	10-0 with 0 step (setting 1)			9-1 with 1 step (setting 2)			6-4 with 1 step (setting 3)			6-1 with 4 steps (setting 4)		
		I-AUROC	AUPRO	FM (⇓)	I-AUROC	AUPRO	FM (⇓)	I-AUROC	AUPRO	FM (⇓)	I-AUROC	AUPRO	FM (⇓)
RGB	IUF (ECCV24)	77.2	84.0	– / –	73.9	79.6	3.8 / 3.5	62.9	68.6	12.4 / 9.1	63.6	69.1	10.6 / 8.8
	CDAD (CVPR25)	78.1	82.1	– / –	73.2	76.5	4.2 / 3.1	63.7	70.4	9.5 / 7.7	65.5	72.3	8.4 / 6.3
	IB-IUMAD	**78.5**	**85.3**	– / –	**74.1**	**83.2**	**2.7** / **1.1**	**65.4**	**71.2**	**8.1** / **6.5**	**66.5**	**72.7**	**7.3** / **5.7**
3D	IUF (ECCV24)	57.2	39.4	– / –	55.6	36.4	2.1 / 0.8	51.5	37.4	3.8 / 1.2	51.9	38.2	3.0 / 1.1
	CDAD (CVPR25)	58.4	30.2	– / –	57.1	29.3	**0.9** / **0.1**	**53.1**	29.6	4.2 / **-0.4**	52.9	30.1	4.4 / **0.3**
	IB-IUMAD	**59.2**	**43.7**	– / –	**57.5**	**39.8**	1.4 / 1.3	52.9	**41.3**	**3.5** / 1.4	**53.3**	**41.9**	**2.6** / 0.9
RGB+3D	IUF (ECCV24)	78.4	85.5	– / –	74.4	80.3	4.3 / 3.8	64.1	69.8	14.3 / 8.6	65.8	71.5	12.5 / 8.8
	CDAD (CVPR25)	79.2	83.7	– / –	74.1	78.8	4.6 / 3.5	66.5	72.3	10.1 / 7.9	68.2	73.8	9.2 / 6.6
	IB-IUMAD	**80.6**	**86.1**	– / –	**76.7**	**85.6**	**3.2** / **1.6**	**67.8**	**73.1**	**8.4** / **6.7**	**69.4**	**75.1**	**7.7** / **6.3**

$\int P(Y) \log \frac{P(Y|F^g_{fu})}{p(Y|F_{fu})} dY = 0$, and we have $I(Y|F_{fu}) - I(Y|F^g_{fu}) = 0$. Therefore, based on ***Corollary* 1** and ***Corollary* 2**, we conclude that using KL divergence as the target loss function between F_{fu} and F^g_{fu} effectively eliminates redundant information from the multimodal fusion feature.

4. Evaluation

We then conduct empirical comparisons to validate the effectiveness of IB-IUMAD.

4.1. Dataset and Experimental Setting

Dataset. We evaluate IB-IUMAD on MVTec 3D-AD [2] and Eyecandies [3] datasets. Both contain 10 objects, where the former is collected from realistic scenes, and the latter is synthetic.

Setup. Firstly, we conducted comparison experiments of IB-IUMAD with incremental unified methods such as IUF [43] and CDAD [29] in four different settings: 10-0 with 0 step, 9-1 with 1 step, 6-4 with 1 step, and 6-1 with 4 steps. Then, for comparison with unified MAD methods, we report the results of IB-IUMAD (i.e., 10-0 with 0 step setting) with UniAD [51], SimpleNet [33], DeSTSeg [58], DiAD [19], and MambaAD [18]. For implementation details, please refer to Appendix F.

4.2. Quantitative Evaluation

Comparison of Incremental Settings. Tables 1 and 2 present performance comparisons between IB-IUMAD and other incremental unified methods on the MVTec 3D-AD and Eyecandies datasets, respectively. The results show that IB-IUMAD consistently outperforms the baselines across different methods, incremental settings, and datasets. Specifically, when trained with RGB images on the MVTec 3D-AD under setting 2, IB-IUMAD surpasses CDAD by 11.1% and 3.7% in I-AUROC and AUPRO, respectively, while significantly dropping the FM metric by 0.3% and 0.6%. When trained with RGB and depth images on the MVTec 3D-AD (Eyecandies) under setting 4, IB-IUMAD outperforms IUF by 3.5% and 2.9% (3.6% and 3.6%) in I-AUROC and AUPRO, respectively, and significantly reduces the FM metric by 5.8% and 1.5% (4.8% and 2.5%). These results highlight the effectiveness of our devised denoising module in mitigating the impact of spurious and redundant features on model performance and catastrophic forgetting.

Table 3. I-AUROC scores on MVTec 3D-AD dataset (10-0 with 0 step). The red / blue indicates the best/second-best results.

	Method	Year	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	UniAD	NIPS22	82.7	89.8	76.8	77.3	96.7	70.5	70.0	51.6	**97.4**	75.7	78.9
	SimpleNet	CVPR23	76.2	70.3	71.4	66.7	83.7	77.4	62.0	56.7	95.6	65.3	72.5
	DeSTSeg	CVPR23	89.7	84.8	79.1	67.4	77.3	77.9	82.2	62.9	93.5	80.8	79.6
	MambaAD	NIPS24	87.7	**94.3**	90.7	61.2	**97.6**	**84.0**	**92.8**	66.8	**97.4**	**90.0**	86.2
	DiAD	AAAI24	**100.0**	68.1	**94.4**	69.4	**98.0**	**100.0**	58.0	76.3	89.2	**92.7**	84.6
	IUF	ECCV24	88.3	**95.0**	**100.0**	**89.7**	93.7	73.6	83.3	**93.1**	91.1	63.5	**87.1**
	CDAD	CVPR25	87.4	88.4	74.9	67.6	91.2	70.7	64.2	59.9	86.2	73.8	76.4
	IB-IUMAD	-	**95.4**	93.1	90.0	**82.1**	92.9	68.4	**94.6**	**90.4**	**96.7**	85.6	**88.9**
3D	DiAD	AAAI24	**79.4**	59.2	**71.7**	65.2	**78.4**	74.5	**56.3**	65.0	60.1	**64.2**	67.4
	IUF	ECCV24	68.9	**96.5**	**68.4**	**73.5**	**75.4**	**76.7**	55.0	**72.3**	**89.1**	61.4	**73.7**
	CDAD	CVPR25	64.7	65.2	54.7	68.9	56.3	62.7	55.1	51.9	58.6	**67.7**	60.6
	IB-IUMAD	-	**69.1**	**99.3**	63.6	**74.8**	65.5	**82.2**	**69.7**	**96.4**	**76.6**	60.4	**75.8**
RGB + 3D	DiAD	AAAI24	**100.0**	73.9	**97.2**	71.6	**97.6**	**98.7**	69.4	**78.3**	**94.3**	**85.6**	86.7
	IUF	ECCV24	97.2	**96.9**	**89.1**	**75.4**	**93.4**	**89.7**	**95.6**	74.2	**98.4**	77.3	**88.7**
	CDAD	CVPR25	90.3	89.7	78.2	68.4	92.5	73.4	65.9	63.3	89.4	79.7	79.1
	IB-IUMAD	-	**98.3**	**97.5**	86.3	**84.6**	90.5	69.4	**98.7**	**100.0**	**98.4**	**86.4**	**91.0**

Table 4. I-AUROC scores on Eyecandies dataset (10-0 with 0 step). The red / blue indicates the best/second-best results.

	Method	Year	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwish	Lollipop	Marsh-mallow	Peppermint Candy	Mean
RGB	DiAD	AAAI24	56.6	83.5	77.3	**89.7**	59.8	**59.4**	**84.3**	66.7	97.4	86.8	76.2
	IUF	ECCV24	56.3	86.2	**78.7**	**90.8**	59.5	**61.0**	84.0	**68.8**	**98.1**	**88.3**	77.2
	CDAD	CVPR25	**62.2**	**92.1**	78.2	87.4	**68.3**	56.1	**85.4**	**70.6**	95.8	85.2	**78.1**
	IB-IUMAD	-	**63.7**	**88.6**	**82.4**	87.8	**64.4**	56.9	82.7	67.7	**98.7**	**92.4**	**78.5**
3D	DiAD	AAAI24	51.8	57.1	55.0	58.1	56.6	50.4	52.9	**56.8**	54.9	54.1	54.8
	IUF	ECCV24	**52.5**	53.1	**65.8**	55.8	**58.6**	**65.8**	52.2	53.9	53.9	**63.7**	57.5
	CDAD	CVPR25	**55.9**	**62.8**	50.9	**65.9**	52.2	54.4	**61.9**	**66.7**	**55.4**	58.3	**58.4**
	IB-IUMAD	-	51.2	**57.4**	**65.6**	**59.8**	**57.7**	**65.6**	**53.2**	56.1	**60.9**	**64.7**	**59.2**
RGB + 3D	DiAD	AAAI24	54.1	87.5	78.2	**91.7**	58.2	55.7	**85.0**	71.9	**98.8**	88.3	76.9
	IUF	ECCV24	**67.3**	87.2	**81.6**	85.6	61.7	**57.9**	81.6	66.3	98.2	**96.3**	78.4
	CDAD	CVPR25	**66.9**	**91.4**	75.3	**88.9**	**74.2**	55.8	**86.1**	**77.1**	94.0	82.5	**79.2**
	IB-IUMAD	-	66.5	**90.1**	**84.6**	87.7	**62.3**	**59.7**	84.2	**75.4**	**99.4**	**95.6**	**80.6**

Comparison of Unified MAD. To further demonstrate the effectiveness of the proposed approach, Tables 3 and 4 present the experimental results of IB-IUMAD (setting 1) in comparison with the unified MAD baselines across both datasets. As shown, IB-IUMAD achieves superior detection results in most test cases under unimodal and multimodal settings. Notably, when both RGB and depth images are utilized for training, IB-IUMAD improves the I-AUROC by 4.3% and 2.3% on MVTec 3D-AD, and by 3.7% and 2.2% on Eyecandies, compared to DiAD and IUF, respectively. These results, along with the visualizations in Figure 4, further validate the effectiveness of our proposed approach in improving unified MAD. Due to page limitations, more experimental results are provided in Appendices B and C[1].

4.3. Ablation Study

Impact of Key Components. In this part, we scrutinize the contribution of each component within IB-IUMAD to incremental unified tasks. As shown in Table 5, IB-IUMAD obtains consistently better results when both the Mamba decoder and IBFM module are used, compared to using only one of them or neither. For example, on 6-1 with 4 steps, I-AUROC increases by 2.5% and FM decreases by 1.9%; on 9-1 with 1 step, I-AUROC improves by 2.4% and FM decreases by 2.3%, compared to the variant without both Mamba and IBFM. It is clear that our proposed method is extremely competitive in incremental unified framework.

Impact of Fusion Operations. To reveal the impact of various fusion operations in the IBFM module on model performance, we conducted several groups of tests with different fusion operations on the MVTec 3D-AD dataset with setting 1. As shown in Table 6, the cross-attention mechanism achieves better results with 91.0%, 97.6%, and 90.4% on the I-AUROC, P-AUROC, and AUPRO metrics, respectively. This further shows that well-designed multimodal fusion operations equipped with information bottleneck constraints can effectively improve the IUMAD's overall performance.

Computing Efficiency. To further show the advantages of IB-IUMAD, we report the results in terms of mem-

[1] Appendix released at https://github.com/longkaifang/IB-IUMAD

Table 5. The ablation study results on the MVTec 3D-AD dataset.

Ablation		6-1 with 4		6-4 with 1		9-1 with 1		10-0 with 0
Mamba	IBFM	I-AUROC	FM	I-AUROC	FM	I-AUROC	FM	I-AUROC
✗	✗	75.3	12.7	79.9	8.7	85.1	4.6	86.7
✓	✗	76.0	11.0	80.8	7.6	86.0	4.0	88.8
✗	✓	76.9	10.2	81.5	7.1	86.8	3.2	89.2
✓	✓	**78.6**	**9.3**	**82.4**	**6.8**	**87.5**	**2.3**	**91.0**

Table 6. The impact of multimodal fusion operations.

Fusion types	MVTec 3D-AD		
	I-AUROC	P-AUROC	AUPRO
Addition	89.8	96.7	88.6
ConcatFC	88.2	96.4	89.5
LinearGLU	86.7	95.8	88.3
Cross-attention	**91.0**	**97.6**	**90.4**

Table 7. Comparison results in terms of the computing efficiency.

Method	MVTec 3D-AD		
	Frame Rate	Memory	I-AUROC
BTF	3.197	3810.6	86.5
AST	4.966	4639.4	93.7
M3DM	0.514	65261.2	94.5
IB-IUMAD	**21.427**	**1483.7**	**91.0**

ory usage, frame rate and I-AUROC on MVTec 3D-AD. As shown in Table 7, compared to M3DM, IB-IUMAD gained a 41x increase in frame rate and 44x shrinkage in memory usage while maintaining comparable performance. This once again reveals the enormous benefits of our method in IUMAD tasks.

5. Related Work

N-objects-N-models is a typical anomaly detection paradigm, which refers to training a separate model for each object to identify anomalies [10, 22, 46, 52]. Specifically, early efforts primarily focused on flaw inspection in RGB images, utilizing feature embedding [1, 11, 12, 24, 28, 31, 40] and reconstruction-based methods [13, 42, 53–55, 59]. Recently, with the emergence of multimodal anomaly detection (MAD) [20, 26, 41, 45, 61], which leverages the complementary information of RGB and depth images to enhance detection performance, has attracted considerable attention [7]. For example, methods like M3DM [47], CFM [8], and FIND [30] have substantially advanced defect detection through different fusion strategies. However, despite its near-saturated performance in single-category tasks, the N-objects-N-models paradigm faces substantial challenges in scaling to multi-category anomaly detection.

N-objects-one-model paradigm is an emerging and promising approach that aims to train a single model capable of detecting anomalies across all objects [9, 36, 38, 50, 57, 60]. A pioneering approach, UniAD [51], employs a Transformer-based architecture for feature reconstruction, sparking a wave of subsequent research [4, 6, 48, 49, 56]. Such as, DiAD [19] introduces diffusion models into RGB-based multi-class anomaly detection via a semantics-guided network, while He *et al.* [18] achieve efficient and lightweight flaw detection by integrating Mamba. However, existing studies have largely overlooked the development of unified MAD models.

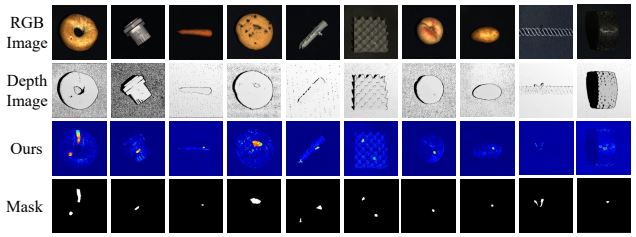


Figure 4. Visualizations results on MVTec 3D-AD.

Moreover, in industrial scenarios, the frequent emergence of new objects requires MAD models within the N-objects-one-model framework to support incremental learning (IL) for effective adaptation. Although IL has shown promise in RGB-based anomaly detection (e.g., IUF [43], CDAD [29]), existing approaches [21, 23, 37] often neglect the influence of spurious and redundant features, which exacerbate catastrophic forgetting during incremental updates. This issue becomes more serious in the MAD due to the inherent complexity of cross-modal fusion. To address this issue, we propose a novel incremental unified multimodal denoising framework for MAD.

6. Conclusion

This paper investigates the impact of spurious and redundant features on catastrophic forgetting in IUMAD tasks. To address this challenge, we propose IB-IUMAD, an incremental unified multimodal denoising framework that integrates the complementary strengths of the Mamba decoder and IBFM module. This design effectively mitigates inter-object spurious feature interference and filters out redundant information in multimodal fusion features, thereby substantially alleviating catastrophic forgetting and achieving consistent performance gains. Extensive experiments on multiple datasets validate the effectiveness and superiority of our proposed method. We hope this work will inspire further research on MAD from an incrementally unified perspective, opening up new possibilities.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 62472079.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the CVPR*, pages 4182–4191, 2020. 8

[2] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *VISIGRAPP*, 2022. 6

[3] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the ACCV*, 2022. 6

[4] Qiyu Chen, Huiyuan Luo, Haiming Yao, Wei Luo, Zhen Qu, Chengkan Lv, and Zhengtao Zhang. Center-aware residual anomaly synthesis for multiclass industrial anomaly detection. *IEEE TII*, 21(9):7276–7286, 2025. 8

[5] Jiayi Cheng, Can Gao, Jie Zhou, Jiajun Wen, Tao Dai, and Jinbao Wang. MC3D-AD: A unified geometry-aware reconstruction model for multi-category 3d anomaly detection. In *Proceedings of the IJCAI*, pages 837–845, 2025. 1

[6] Yuqi Cheng, Yunkang Cao, Dongfang Wang, Weiming Shen, and Wenlong Li. Boosting global-local feature matching via anomaly synthesis for multi-class point cloud anomaly detection. *IEEE TASE*, 22:12560–12571, 2025. 8

[7] Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. In *ICML*, pages 6185–6194, 2023. 8

[8] Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, and Luigi Di Stefano. Multimodal industrial anomaly detection by crossmodal feature mapping. In *CVPR*, 2024. 1, 8

[9] Zhewei Dai, Shilei Zeng, Haotian Liu, Xurui Li, Feng Xue, and Yu Zhou. Seas: few-shot industrial anomaly image generation with separation and sharing fine-tuning. In *Proceedings of the ICCV*, pages 23135–23144, 2025. 8

[10] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *Proceedings of the ICPR*, pages 475–489, 2021. 8

[11] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, 2022. 8

[12] Juan Du, Chengyu Tao, Xuanming Cao, and Fugee Tsung. 3d vision-based anomaly detection in manufacturing: A survey. *Frontiers of Engineering Management*, 2025. 8

[13] Qingqing Fang, Qinliang Su, Wenxi Lv, Wenchao Xu, and Jianxing Yu. Boosting fine-grained visual anomaly detection with coarse-knowledge-aware adversarial learning. In *Proceedings of the AAAI*, pages 16532–16540, 2025. 8

[14] Yingying Fang, Shuang Wu, Sheng Zhang, Chaoyan Huang, Tieyong Zeng, Xiaodan Xing, Simon Walsh, and Guang Yang. Dynamic multimodal information bottleneck for multimodality classification. In *WACV*, 2024. 2, 5

[15] Marco Federici, Anjan Dutta, Patrick Forré, Nate Kushman, and Zeynep Akata. Learning robust representations via multi-view information bottleneck. In *ICLR*, 2020. 5

[16] Zhihao Gu, Jiangning Zhang, Liang Liu, Xu Chen, Jinlong Peng, Zhenye Gan, Guannan Jiang, Annan Shu, Yabiao Wang, and Lizhuang Ma. Rethinking reverse distillation for multi-modal anomaly detection. In *AAAI*, 2024. 1

[17] 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 *Proceedings of the CVPR*, pages 20405–20415, 2025. 1

[18] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. In *Advances in NIPS*, 2024. 6, 8

[19] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI*, 2024. 6, 8

[20] Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *Proceedings of the CVPR*, 2023. 8

[21] Lei Hu, Zhiyong Gan, Ling Deng, Jinglin Liang, Lingyu Liang, Shuangping Huang, and Tianshui Chen. Replaycad: Generative diffusion replay for continual anomaly detection. In *Proceedings of the IJCAI*, pages 2946–2954, 2025. 8

[22] Xi Jiang, Jianlin Liu, Jinbao Wang, Qiang Nie, Kai Wu, Yong Liu, Chengjie Wang, and Feng Zheng. Softpatch: Unsupervised anomaly detection with noisy data. *Advances in NIPS*, pages 15433–15445, 2022. 8

[23] Geonu Lee, Yujeong Oh, Geonhui Jang, Soyoung Lee, Jeonghyo Song, Sungmin Cha, and YoungJoon Yoo. Continual-mega: A large-scale benchmark for generalizable continual anomaly detection. *arXiv*, 2025. 3, 8

[24] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, 2021. 8

[25] Wujin Li, Jiawei Zhan, Jinbao Wang, Bizhong Xia, Bin-Bin Gao, Jun Liu, Chengjie Wang, and Feng Zheng. Towards continual adaptation in industrial anomaly detection. In *Proceedings of the ACM MM*, pages 2871–2880, 2022. 1

[26] Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, and Yingna Wu. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. In *Proceedings of the CVPR*, pages 22207–22216, 2024. 8

[27] Wenqiao Li, Bozhong Zheng, Xiaohao Xu, Jinye Gan, Fading Lu, Xiang Li, Na Ni, Zheng Tian, Xiaonan Huang, Shenghua Gao, et al. Multi-sensor object anomaly detection: Unifying appearance, geometry, and internal properties. In *Proceedings of the CVPR*, pages 9984–9993, 2025. 1

[28] Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Promptad: Learning prompts with only normal samples for few-shot anomaly detection. In *Proceedings of the CVPR*, 2024. 8

[29] Xiaofan Li, Xin Tan, Zhuo Chen, Zhizhong Zhang, Ruixin Zhang, Rizen Guo, Guanna Jiang, Yulong Chen, Yanyun Qu, Lizhuang Ma, et al. One-for-more: Continual diffusion model for anomaly detection. In *CVPR*, 2025. 2, 6, 8

[30] Yiting Li, Fayao Liu, Jingyi Liao, Sichao Tian, Chuan-Sheng Foo, and Xulei Yang. Find: Few-shot anomaly inspection with normal-only multi-modal data. In *ICCV*, 2025. 8

[31] Hanzhe Liang, Guoyang Xie, Chengbin Hou, Bingshu Wang, Can Gao, and Jinbao Wang. Look inside for more: Internal spatial modality perception for 3d anomaly detection. In *Proceedings of the AAAI*, 2025. 8

[32] Jiaqi Liu, Kai Wu, Qiang Nie, Ying Chen, Bin-Bin Gao, Yong Liu, Jinbao Wang, Chengjie Wang, and Feng Zheng. Unsupervised continual anomaly detection with contrastively-learned prompt. In *AAAI*, 2024. 1

[33] 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

[34] Kaifang Long, Guoyang Xie, Lianbo Ma, Qing Li, Min Huang, Jianhui Lv, and Zhichao Lu. Enhancing multimodal learning via hierarchical fusion architecture search with inconsistency mitigation. *IEEE TIP*, 2025. 4

[35] Kaifang Long, Guoyang Xie, Lianbo Ma, Jiaqi Liu, and Zhichao Lu. Revisiting multimodal fusion for 3d anomaly detection from an architectural perspective. In *Proceedings of the AAAI*, 2025. 1

[36] Fanbin Lu, Xufeng Yao, Chi-Wing Fu, and Jiaya Jia. Removing anomalies as noises for industrial defect localization. In *Proceedings of the ICCV*, pages 16166–16175, 2023. 8

[37] Haoquan Lu, Hanzhe Liang, Jie Zhang, Chenxi Hu, Jinbao Wang, and Can Gao. C3d-ad: Toward continual 3d anomaly detection via kernel attention with learnable advisor. *arXiv preprint arXiv:2508.01311*, 2025. 8

[38] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. *Advances in NIPS*, 36:8487–8500, 2023. 8

[39] Xiaohuan Pei, Tao Huang, and Chang Xu. Efficientvmamba: Atrous selective scan for light weight visual mamba. In *Proceedings of the AAAI*, pages 6443–6451, 2025. 4

[40] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for image-based defect detection. In *WACV*, 2022. 8

[41] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *WACV*, 2023. 8

[42] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *ECCV*, 2022. 8

[43] Jiaqi Tang, Hao Lu, Xiaogang Xu, Ruizheng Wu, Sixing Hu, Tong Zhang, Tsz Wa Cheng, Ming Ge, Ying-Cong Chen, and Fugee Tsung. An incremental unified framework for small defect inspection. In *ECCV*, pages 307–324, 2024. 2, 6, 8

[44] Xudong Tian, Zhizhong Zhang, Shaohui Lin, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Farewell to mutual information: Variational distillation for cross-modal person re-identification. In *Proceedings of the CVPR*, 2021. 5

[45] Yuanpeng Tu, Boshen Zhang, Liang Liu, Yuxi Li, Jiangning Zhang, Yabiao Wang, Chengjie Wang, and Cairong Zhao. Self-supervised feature adaptation for 3d industrial anomaly detection. In *ECCV*, pages 75–91, 2024. 8

[46] Fuyun Wang, Tong Zhang, Yuanzhi Wang, Yide Qiu, Xin Liu, Xu Guo, and Zhen Cui. Distribution prototype diffusion learning for open-set supervised anomaly detection. In *Proceedings of the CVPR*, 2025. 8

[47] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *CVPR*, 2023. 8

[48] Shun Wei, Jielin Jiang, and Xiaolong Xu. Uninet: A contrastive learning-guided unified framework with feature selection for anomaly detection. In *CVPR*, 2025. 8

[49] Haiming Yao, Yunkang Cao, Wei Luo, Weihang Zhang, Wenyong Yu, and Weiming Shen. Prior normality prompt transformer for multiclass industrial image anomaly detection. *IEEE TII*, 20(10):11866–11876, 2024. 8

[50] Xincheng Yao, Ruoqi Li, Zefeng Qian, Lu Wang, and Chongyang Zhang. Hierarchical gaussian mixture normalizing flow modeling for unified anomaly detection. In *Proceedings of the ECCV*, 2024. 8

[51] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in NIPS*, 2022. 6, 8

[52] Ruiyun Yu, Haoyuan Li, Bingyang Guo, and Ziming Zhao. Background-weaken generalization network for few-shot industrial metal defect segmentation. *IEEE TIM*, 2025. 8

[53] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draema discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, 2021. 8

[54] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr–a dual subspace re-projection network for surface anomaly detection. In *Proceedings of the ECCV*, 2022.

[55] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Cheating depth: Enhancing 3d surface anomaly detection via depth simulation. In *WACV*, pages 2164–2172, 2024. 8

[56] Jiangning Zhang, Xuhai Chen, Yabiao Wang, Chengjie Wang, Yong Liu, Xiangtai Li, Ming-Hsuan Yang, and Dacheng Tao. Exploring plain vit reconstruction for multi-class unsupervised anomaly detection. *arXiv*, 2023. 8

[57] Jiangning Zhang, Chengjie Wang, Xiangtai Li, Guanzhong Tian, Zhucun Xue, Yong Liu, Guansong Pang, and Dacheng Tao. Learning feature inversion for multi-class anomaly detection under general-purpose coco-ad benchmark. *arXiv preprint arXiv:2404.10760*, 2024. 8

[58] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *CVPR*, 2023. 6

[59] Zhe Zhang, Mingxiu Cai, Hanxiao Wang, Gaochang Wu, Tianyou Chai, and Xiatian Zhu. Costfilter-ad: Enhancing anomaly detection through matching cost filtering. *In Proceedings of the ICML*, 2025. 8

[60] Ying Zhao. Omnial: A unified cnn framework for unsupervised anomaly localization. In *CVPR*, 2023. 8

[61] Yuan Zhao, Youwei Pang, Lihe Zhang, Hanqi Liu, Jiaming Zuo, Huchuan Lu, and Xiaoqi Zhao. Unimmad: Unified multi-modal and multi-class anomaly detection via moe-driven feature decompression. *arXiv preprint arXiv:2509.25934*, 2025. 8