

# ReplayCAD: Generative Diffusion Replay for Continual Anomaly Detection

Lei Hu<sup>1</sup>, Zhiyong Gan<sup>1,2</sup>, Ling Deng<sup>2</sup>, Jinglin Liang<sup>1</sup>, Lingyu Liang<sup>1</sup>,  
Shuangping Huang<sup>1,3,\*</sup>, Tianshui Chen<sup>4</sup>

<sup>1</sup>South China University of Technology

<sup>2</sup>China United Network Communications Corporation Limited Guangdong Branch

<sup>3</sup>Pazhou Laboratory

<sup>4</sup>Guangdong University of Technology  
eehulei@mail.scut.edu.cn, eehsp@scut.edu.cn

## Abstract

Continual Anomaly Detection (CAD) enables anomaly detection models in learning new classes while preserving knowledge of historical classes. CAD faces two key challenges: catastrophic forgetting and segmentation of small anomalous regions. Existing CAD methods store image distributions or patch features to mitigate catastrophic forgetting, but they fail to preserve pixel-level detailed features for accurate segmentation. To overcome this limitation, we propose **ReplayCAD**, a novel diffusion-driven generative replay framework that replay high-quality historical data, thus effectively preserving pixel-level detailed features. Specifically, we compress historical data by searching for a class semantic embedding in the conditional space of the pre-trained diffusion model, which can guide the model to replay data with fine-grained pixel details, thus improving the segmentation performance. However, relying solely on semantic features results in limited spatial diversity. Hence, we further use spatial features to guide data compression, achieving precise control of sample space, thereby generating more diverse data. Our method achieves state-of-the-art performance in both classification and segmentation, with notable improvements in segmentation: 11.5% on VisA and 8.1% on MVTec. Our source code is available at <https://github.com/HULEI7/ReplayCAD>.

## 1 Introduction

Nowadays, unsupervised Anomaly Detection (AD) is crucial in modern manufacturing [Liu *et al.*, 2024b]. It automatically identifies defects in samples at image and pixel levels, enhancing production efficiency and product quality while reducing reliance on manual inspections. However, in practical industrial production, changes in the generation plan continuously produce new class samples, requiring the model to learn knowledge from new samples while retaining knowledge of previous ones, i.e., Continual Anomaly Detection (CAD).

\*Corresponding author.

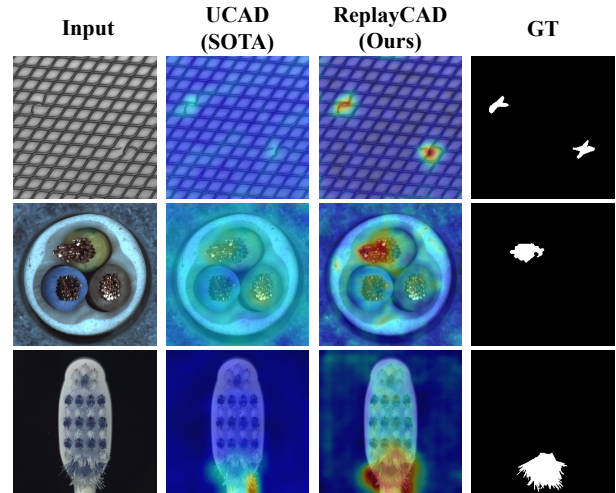


Figure 1: Compared to the SOTA method UCAD, our approach achieves more accurate anomaly segmentation, since we better preserves pixel-level detailed features through high-quality data replay.

The key challenge of CAD lies in two aspects: First, catastrophic forgetting, where models tend to catastrophically forget historical class knowledge when learning new class data. Second, CAD requires not only the identification of image-level anomalous samples but also the segmentation of pixel-level anomalous regions. Since anomaly regions are usually small (As illustrated in Figure 3), this requires the model to remember more detailed pixel-level features from historical samples to ensure accurate anomaly segmentation.

A straightforward approach to addressing these two challenges is to apply existing Continual Learning (CL) methods (e.g. EWC [Kirkpatrick *et al.*, 2017], MAS [Aljundi *et al.*, 2018]) to anomaly detection models. However, existing CL methods do not focus on preserving the detailed features of historical samples, making them ineffective for CAD [Tang *et al.*, 2025]. Advanced CAD methods [Li *et al.*, 2022; Liu *et al.*, 2024a] aim to mitigate catastrophic forgetting by compressing and storing features of historical samples. Specifically, DNE [Li *et al.*, 2022] stores the mean and variance of image features. However, DNE remembers only the overall features of the distribution while ignoring the detailed features of individual samples. UCAD [Liu *et al.*, 2024a] di-

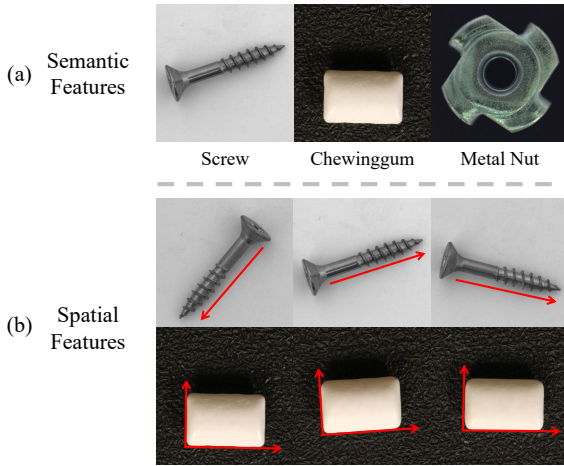


Figure 2: Visualization of semantic and spatial features. (a) Semantic features, samples from different classes exhibit distinct texture and shape characteristics. (b) Spatial features, samples within the same class are distributed across different spatial locations.

vides the image into multiple patches and encodes them into features for storage, resulting in the loss of pixel-level details from historical samples. In summary, these methods fail to preserve the pixel-level detailed features of samples, which negatively impacts segmentation performance.

To address this, we propose **ReplayCAD**, a novel CAD framework that can replay real and diverse historical data, thus preserving pixel-level detailed features. ReplayCAD consists of two stages: (1) compressing historical data of each class into conditional features; and (2) utilizing these features to replay historical data, and using the replayed data together with new class data for model training. Our key contribution lies in enhancing first-stage data compression to generate more diverse and authentic historical data. Specifically, in the first stage, we employ reverse engineering to learn a class semantic embedding in the conditional space of a pre-trained diffusion model for each class’s historical data. This class semantic embedding guides the diffusion model to generate historical data with pixel-level detailed features, significantly enhancing the segmentation performance of anomalous regions. However, as illustrated in Figure 2, samples contain not only semantic features but also rich spatial features. Compressing data using semantic features alone leads to replayed data lacking spatial diversity. To address this issue, we utilize the sample’s mask as its spatial features to guide the data compression process. This approach enables controlled generation of spatial positions, resulting in the replay of more diverse historical data. Our method offers several advantages: Firstly, compared to current CAD methods based on feature replay [Li *et al.*, 2022; Liu *et al.*, 2024a] and regularization [Tang *et al.*, 2025; Li *et al.*, 2025], the replay-based approach is more effective in preserving the detailed features of samples, which leads to better performance in anomaly segmentation. Secondly, by considering both semantic and spatial features, we generate high-quality and diverse samples, which significantly mitigates catastrophic forgetting. Finally, by guiding the pre-

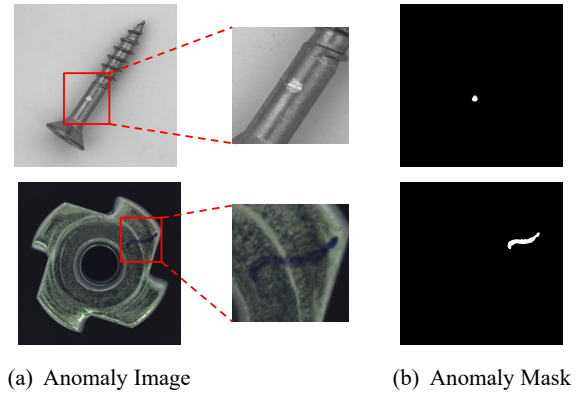


Figure 3: Visualization of anomalous regions, illustrating their typically small areas.

trained diffusion model to perform reverse engineering, we significantly reduce training time and cost compared to existing replay methods [Shin *et al.*, 2017; Wang *et al.*, 2021; Gao and Liu, 2023; Li *et al.*, 2024] that require training the entire generator.

To sum up, our contributions are as follows:

- We propose a novel CAD framework ReplayCAD, which effectively preserves fine-grained pixel details through high-quality data replay, significantly enhancing anomaly segmentation performance. To our best knowledge, this is the first attempt to apply diffusion-driven generative replay approach in CAD.
- To generate more diverse historical data, we employ spatial features of samples to guide data compression, thereby achieving spatially controlled data replay.
- Extensive experiments demonstrate that our method significantly outperforms SOTA approaches, particularly in segmentation performance, achieving improvements of 11.5% and 8.1% on VisA and MVTec, respectively.

## 2 Related Work

### 2.1 Unsupervised Anomaly Detection

Currently, unsupervised anomaly detection methods can be classified into the following three types: **Embedding-based methods** [Roth *et al.*, 2022; Deng and Li, 2022; Liu *et al.*, 2023] model the feature distribution of normal samples and detect anomalies by identifying deviations of anomalous samples from this distribution. **Reconstruction-based methods** [You *et al.*, 2022; He *et al.*, 2024; Zhang *et al.*, 2024a; Wu *et al.*, 2024] train a model to reconstruct normal images and detect anomalies by comparing the differences between the input image and the reconstructed image. **Synthesizing-based methods** [Hu *et al.*, 2024; Li *et al.*, 2021; Zavrtnik *et al.*, 2021a; Zuo *et al.*, 2024] train a model by adding artificially synthesized anomalies to normal images, thereby improving anomaly detection capability. However, in real-world industrial scenarios where new classes of data frequently emerge, and existing AD methods lack the capability for continual learning, our goal is to equip existing anomaly detection methods with continual learning capability.

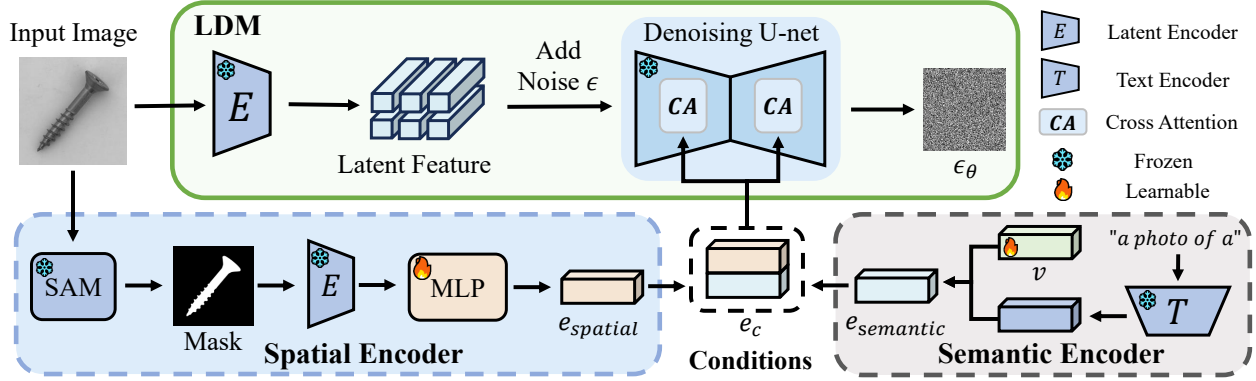


Figure 4: Overview of **Data Compression**. We compress the data from both semantic and spatial dimensions: (1) **Semantic**. In the semantic encoder, we get the semantic feature by initializing a class semantic embedding  $v$  and combining it with a prompt  $p$  encoded by the text encoder. (2) **Spatial**. In the spatial encoder, we first obtain the mask of the sample using SAM and learn the mapping of mask to spatial features by optimizing a MLP layer. Finally, we use spatial and semantic features as conditions to guide the denoising process of LDM.

## 2.2 Continual Learning

The current methods for continual learning can be broadly categorized into the following three types: **Regularization-based methods** [Kirkpatrick *et al.*, 2017; Pan *et al.*, 2020; Aljundi *et al.*, 2018] prevent forgetting of previous tasks by imposing constraints on model parameters. **Replay-based methods** [Rebuffi *et al.*, 2017; Liu *et al.*, 2020; Riemer *et al.*, 2019] consolidate the retention of prior knowledge by replaying data or features from past tasks. **Architecture-based methods** [Mallya and Lazebnik, 2018; Mallya *et al.*, 2018; Yoon *et al.*, 2017] assign specific parameters to each task, isolating them to minimize interference. However, existing CL methods are not specifically designed to preserve detailed features of historical samples in CAD, making them unsuitable for direct application to AD models.

## 2.3 Continual Anomaly Detection

Current CAD research can be broadly categorized into two approaches: **Feature replay-based** [Li *et al.*, 2022; Liu *et al.*, 2024a] and **Regularization-based** [Tang *et al.*, 2025; Li *et al.*, 2025]. DNE [Li *et al.*, 2022] and UCAD [Liu *et al.*, 2024a] store the feature embeddings of samples from previous classes and locate anomalies by comparing the test sample features with the stored embeddings. However, these methods cannot save the pixel-level detailed features, resulting in poor performance of anomalous region segmentation. IUF [Tang *et al.*, 2025] and CDAD [Li *et al.*, 2025] prevent knowledge forgetting by restricting parameter updates for previous classes, but its performance deteriorates in long class sequences. To address these issues, we focus on the data itself to remember detailed features, we propose a novel CAD framework that can replay high-quality historical data.

## 3 Problem Definition

In this section, we provide a detailed introduction to the specific setting of the CAD. Unlike traditional AD task where all classes are known during training, CAD requires the model to continuously learn new classes, essentially falling under the category of class-incremental learning. Specifically,

for an anomaly detection dataset with  $\mathcal{N}$  classes, its training and testing sets are represented as  $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N\}$  and  $\{\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_N\}$ , respectively. When learning the  $i$ -th class, we can only access the training data  $\mathcal{X}_i$  and cannot access data from previous classes  $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_{i-1}\}$ , during testing, we need to perform testing on  $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_i\}$ . Our goal is to sequentially train on the training set of each class and ultimately evaluate on the testing set  $\{\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_N\}$ .

## 4 Methods

### 4.1 Overview

ReplayCAD consists of two stages: feature-guided data compression stage and replay-enhanced anomaly detection stage. In the first stage, we compress the data of historical classes from both semantic and spatial dimensions, thereby extracting corresponding conditional features, as shown in Figure 4. Specifically, for semantic features, we search for a class semantic embedding within the conditional space of the pre-trained LDM to guide the model in generating samples for the corresponding class. For spatial features, we learn the mapping relationship between the mask and the spatial features, where the mask is obtained through SAM [Kirillov *et al.*, 2023]. In the second stage, we use the conditional features learned in the first stage to replay the historical class data, and train the anomaly detection model together with the new class data. The main contribution of this paper is to enhance the data compression of the first stage from both semantic and spatial dimensions, so as to generate more diversified and authentic historical data. Therefore, in the following, we will focus on how to use semantic and spatial features to guide LDM to compress samples in the first stage.

### 4.2 Semantic-Aware Generative Replay

In this section, we will investigate how to accomplish the two stages of generative replay using only semantic features.

Latent Diffusion Model (LDM) [Rombach *et al.*, 2022] generates data by denoising in a low-dimensional latent space. Compared to directly working in high-dimensional



pixel space, this approach significantly improves computational efficiency while preserving essential semantic information, ensuring high-quality outputs. To further enhance controllability, LDM incorporates a conditional guidance mechanism. This allows conditions such as text descriptions, image masks, or class labels to guide the generation process, ensuring that the results closely align with the given conditions while maintaining diversity and flexibility. Specifically, for the input image  $x$ , the LDM first encodes it using a latent encoder  $\mathcal{E}$  and then adds random noise  $\epsilon$  to generate the noisy image  $z_t$  for denoising. Subsequently, the noisy image  $z_t$  and conditions  $c$  are fed into the denoising U-Net, where the noise  $\epsilon_\theta$  is predicted using a cross-attention mechanism [Dai *et al.*, 2023; Dai *et al.*, 2024; Peng *et al.*, 2024; Hu and Huang, 2025]. The optimization function of LDM is as follows:

$$\mathcal{L}_{LDM} = \mathbb{E}_{\mathcal{E}(x), \epsilon \sim \mathcal{N}(0,1), t, c} [\|\epsilon - \epsilon_\theta(\mathcal{E}(x), t, \mathcal{E}_c(c))\|_2^2], \quad (1)$$

Instead of directly training the LDM, we learn the conditional feature to guide it in generating samples for the corresponding class. Specifically, to represent semantic features, we first randomly initialize a learnable embedding  $v \in \mathbb{R}^{K \times C}$ , which serves as the foundational representation. In addition, we take a text prompt  $p$  (e.g., “a photo of a”), encode it using a frozen text encoder  $\mathcal{T}$ , and then concatenate it with the semantic embedding  $v$  to form the final semantic features  $e_{semantic}$ , guiding the LDM’s denoising process. The optimization objective is as follows:

$$e_{semantic} = \{\mathcal{T}(p), v\}, \quad (2)$$

$$v^* = \arg \min_v \mathcal{L}_{LDM}(x, t, e_{semantic}), \quad (3)$$

We store the learned semantic embedding  $v$  for each class. When new class data appears, we use these semantic embeddings to guide LDM to replay the historical class data. After replaying data from historical classes, we integrate this data with new class data to train the AD model, effectively mitigating the catastrophic forgetting of the historical classes. Our approach imposes no restrictions on anomaly detection models, in this study, we select the multi-class anomaly detection model InvAD [Zhang *et al.*, 2024b] as the baseline. The multi-class AD model combines samples from all classes during training, eliminating the need to know sample classes beforehand during inference, thus addressing class uncertainty in CAD. InvAD detects multiple types of anomalies by introducing a feature inversion mechanism and leveraging high-quality feature reconstruction capabilities. Specifically, InvAD consists of an encoder  $E$  and a generator  $G$ . For a given normal input sample  $x$ , its original feature  $F$  is first extracted using a frozen encoder, and then its reconstructed feature  $F^*$  are generated by the generator:

$$F = E(x), F^* = G(F). \quad (4)$$

Then, the abnormal region is obtained by calculating the difference between the original feature  $F$  and the generated feature  $F^*$ . The loss function is defined as follows:

$$\mathcal{L}_{InvAD} = \mathbb{E}_{E, G, x} [\|F - F^*\|_2^2], \quad (5)$$

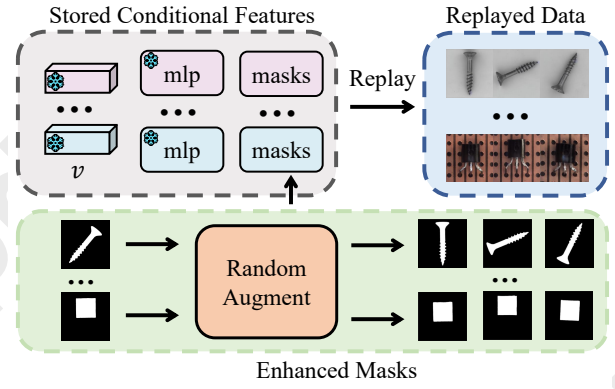


Figure 5: Overview of **Data Replay**. We use stored semantic embeddings  $v$ , the mlp layer, and mask images to guide the LDM in replaying data from historical classes. To enhance the diversity of replay samples, we apply random augmentations to the masks.

### 4.3 Spatially-Enhanced Diverse Generation

However, the samples not only present semantic features, but also various spatial features, as shown in Figure 2, focusing solely on learning semantic features may cause the generated samples to overfit certain specific examples [Yang *et al.*, 2025; Hu *et al.*, 2024; Liang *et al.*, 2024], resulting in a lack of spatial diversity. Therefore, we introduce spatial features to guide the data compression process.

The spatial features of a sample include its coordinates and angle, which are difficult to represent directly. To address this, we use a mask image to denote the sample’s spatial features. However, due to the lack of mask labels in the training set, we utilize SAM’s [Kirillov *et al.*, 2023] powerful zero-shot segmentation capability to generate the mask for each sample. As the diffusion model cannot directly interpret mask information, we optimize a mlp layer to learn the mapping from the mask to spatial features, thus guiding the diffusion model in denoising. Specifically, for an input image  $x \in \mathbb{R}^{H \times W \times 3}$ , we first generate its corresponding object mask image by SAM. The mask image is then encoded by a latent encoder  $\mathcal{E}$  to produce feature representations  $x_{mask} \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 4}$ , which are then reshaped to match the input size required by the mlp layer, resulting in  $x'_{mask} \in \mathbb{R}^{\frac{HW}{16n} \times n}$ . Subsequently, a mlp layer  $\mathcal{W}$  with a size of  $(n, m)$  is applied to compute the spatial feature  $x_{spatial} \in \mathbb{R}^{\frac{HW}{16n} \times m}$ . Finally,  $x_{spatial}$  is reshaped again to align with the size of the semantic features, yielding the final spatial features  $e_{spatial} \in \mathbb{R}^{M \times C}$ .

We concatenate the extracted semantic features  $e_{semantic}$  and spatial features  $e_{spatial}$  to obtain the condition  $e_c$ , which is used to guide the denoising process of the LDM:

$$e_c = \{e_{semantic}, e_{spatial}\} \in \mathbb{R}^{(K+M) \times C}, \quad (6)$$

Our goal is to find the best semantic embedding  $v$  as well as the mlp layer  $\mathcal{W}$  for each class, the optimization objectives are outlined as follows:

$$v^*, \mathcal{W}^* = \arg \min_{v, \mathcal{W}} \mathcal{L}_{LDM}(x, t, e_c), \quad (7)$$

Method	Venue	VisA				MVTec			
		Image-AUROC		Pixel-AP		Image-AUROC		Pixel-AP	
		Avg ( $\uparrow$ )	FM ( $\downarrow$ )	Avg ( $\uparrow$ )	FM ( $\downarrow$ )	Avg ( $\uparrow$ )	FM ( $\downarrow$ )	Avg ( $\uparrow$ )	FM ( $\downarrow$ )
PatchCore	CVPR 22	57.6	38.0	10.3	32.4	60.7	40.7	20.9	43.1
SimpleNet	CVPR 23	58.9	33.4	7.3	30.2	62.4	36.9	17.7	32.6
MambaAD	NeurIPS 24	59.5	32.4	8.8	27.6	60.6	37.6	12.7	43.8
InvAD	arXiv 24	67.5	28.1	9.8	33.1	66.8	31.7	21.5	38.2
SimpleNet+EWC	CVPR 23	59.6	35.3	9.0	28.2	69.0	29.8	21.6	28.1
MambaAD+EWC	NeurIPS 24	60.6	32.0	8.2	29.3	65.1	33.0	13.5	43.4
SimpleNet+MAS	CVPR 23	60.6	34.4	9.3	28.1	70.3	28.8	25.4	24.3
MambaAD+MAS	NeurIPS 24	65.2	25.9	9.5	25.6	64.5	33.9	12.2	45.2
DNE	MM 22	60.1	17.8	–	–	87.6	<u>4.7</u>	–	–
DNE + PANDA	CVPR 21	46.0	36.8	–	–	66.3	21.8	–	–
DNE + CutPaste	CVPR 21	61.3	18.9	–	–	83.9	9.8	–	–
UCAD	AAAI 24	<u>87.4</u>	–	<u>30.0</u>	–	<u>93.0</u>	–	<u>45.6</u>	–
IUF	ECCV 24	68.1	<u>8.5</u>	3.4	<b>0.3</b>	76.2	6.7	17.1	<u>5.9</u>
CDAD	CVPR 25	62.8	22.2	8.3	21.7	74.3	20.1	28.0	26.1
<b>ReplayCAD</b>	<b>Ours</b>	<b>90.3</b>	<b>5.5</b>	<b>41.5</b>	<u>5.0</u>	<b>94.8</b>	<b>4.5</b>	<b>53.7</b>	<b>5.5</b>

Table 1: Comparison with state-of-the-art methods. **Bold** indicates the best performance, while underline indicates the second-best performance. The term "Avg" refers to the mean value obtained by testing across all classes after completing training on the final class.

After training is completed for each class, we save the semantic embedding  $v$  and the mlp layer  $\mathcal{W}$  for subsequent data replay. In addition,  $M(M \in [1, 10])$  randomly selected mask images are stored to control the spatial positioning of the generated images. These stored components constitute the conditional features for each class, requiring minimal storage space while enabling precise control over both semantic and spatial aspects during data replay.

In the second stage, to achieve real and diverse data replay, we use stored conditional features to achieve controllable generation in both semantic and spatial aspects, as illustrated in Figure 5. Semantically, we use semantic embedding  $v$  to control the class of generated samples. Spatially, we control the spatial position of samples through image masks and the mlp layer  $\mathcal{W}$  in the spatial encoder. In addition, we randomly rotate and shift the stored masks to enhance their spatial diversity. By achieving semantic and spatial controllable generation, we can generate real and diverse historical data, thus preserving pixel-level details more efficiently.

## 5 Experiments

### 5.1 Experiments Setup

**Datasets.** We conducted experiments on two commonly used industrial anomaly detection datasets: VisA [Zou *et al.*, 2022] and MVTec [Bergmann *et al.*, 2019]. VisA comprises 12 sample classes, with 8659 training samples (all normal) and 2162 test samples (962 normal, 1200 anomalous). MVTec includes 15 sample classes, with 3629 training samples (all normal) and 1725 test samples (467 normal, 1258 anomalous).

**Compared Methods.** We compare our approach with three methods: (1) **CAD Methods.** We compare our method

with SOTA CAD methods, including CDAD [Li *et al.*, 2025], IUF [Tang *et al.*, 2025], UCAD [Liu *et al.*, 2024a], and DNE [Li *et al.*, 2022]. Additionally, we migrate the framework of DNE to PANDA [Reiss *et al.*, 2021] and CutPaste [Li *et al.*, 2021]. (2) **AD Methods.** We apply current SOTA AD methods directly to CAD, including PatchCore [Roth *et al.*, 2022], SimpleNet [Liu *et al.*, 2023], MambaAD [He *et al.*, 2024] and InvAD [Zhang *et al.*, 2024b], we fine-tune the old model as new class data comes in. (3) **AD+CL Methods.** We adapt CL methods, such as EWC [Kirkpatrick *et al.*, 2017] and MAS [Aljundi *et al.*, 2018], to SimpleNet and MambaAD to enhance their adaptability in CAD.

**Metrics.** Similar to previous methods, we assess the performance of anomaly detection from both classification and segmentation perspectives. For classification performance, we use Area Under the Receiver Operating Characteristics (AUROC) [Zhou and Paffenroth, 2017], while for segmentation performance, we use Area Under Precision-Recall (AP) [Zavrtanik *et al.*, 2021b]. Moreover, we use Forgetting Measure (FM) [Chaudhry *et al.*, 2018] to evaluate the extent to which the model forgets historical classes:

$$FM = \frac{1}{\mathcal{N}-1} \sum_{i=1}^{\mathcal{N}-1} \max_{\gamma \in \{1, \dots, \mathcal{N}-1\}} (a_{\gamma,i} - a_{\mathcal{N},i}), \quad (8)$$

Here,  $\mathcal{N}$  represents the total number of classes,  $i$  denotes the class number, and  $a$  indicates the classification or segmentation score for that class.

**Implementation Details.** We set the number of tokens  $K$  for the semantic embedding  $v$  to 20. For VisA, we use pre-trained Stable Diffusion 1.5 weights with a resolution of

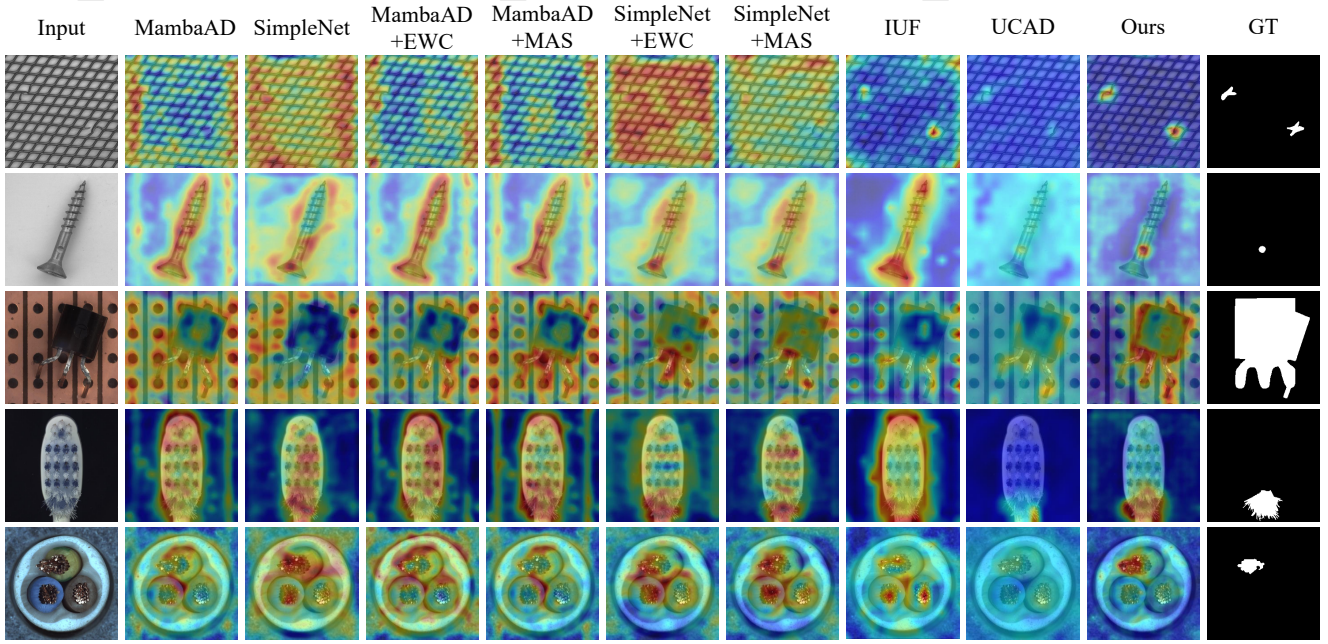


Figure 6: Visualization of anomalous region segmentation. From left to right are the test image, the heatmaps predicted by the comparison methods, the heatmap predicted by our method, and the Ground Truth (GT). Areas with high anomaly scores are shown in red, while areas with low anomaly scores are shown in blue.

Method	VisA	MVTec
UCAD	18.6 MB	23.3 MB
CDAD	2.7 GB	3.5 GB
ReplayCAD (Ours)	<b>1.9 MB</b>	<b>3.0 MB</b>

Table 2: Comparison of storage space with the SOTA method.

Model	Semantic	Spatial	I-AUROC	P-AP
T1	✗	✗	66.8	21.5
T2	✓	✗	92.4	49.5
T3	✗	✓	86.9	42.5
T4	✓	✓	<b>94.8</b>	<b>53.7</b>

Table 3: Effect of semantic and spatial features on MVTEC.

(512, 512), where the dimension  $C$  of the semantic and spatial features is 768, and the MLP layer size is (128, 196). For MVTEC, we use pre-trained LDM weights with a resolution of (256, 256), where the dimension  $C$  of the semantic and spatial features is 1280, and the MLP layer size is (128, 200). We generate 800 samples for each class for training the anomaly detection model. We use Invad [Zhang *et al.*, 2024b] as the baseline model for anomaly detection, keeping all settings consistent with the original paper.

## 5.2 Experiment Performance

**Quantitative Evaluation.** As shown in Table 1, our method achieves state-of-the-art performance on both image-level and pixel-level metrics across two datasets. Specifically, we

Method	Venue	I-AUROC	P-AP
PatchCore	CVPR 22	93.7	<b>54.7</b>
RLR	ECCV 24	94.0	50.6
MambaAD	NeurIPS 24	<b>95.3</b>	53.1
InvAD	arXiv 24	94.8	53.7

Table 4: Transferring to different AD models on MVTEC.

improved pixel-level metrics by 11.5% on VisA and 8.1% on MVTEC compared to the previous state-of-the-art method. Existing anomaly detection methods often suffer from catastrophic forgetting in continual learning scenarios. Although some continual learning methods can mitigate this issue, there remains a significant gap compared to our approach. We observe that although IUF performs poorly in terms of pixel-level and image-level metrics, it demonstrates relatively low FM scores. This is because IUF experiences severe feature conflicts between new and historical classes when learning data from new classes. This conflict causes the historical best performance  $a_{\gamma,i}$  (Eq. 8) of new classes to fall below their actual potential, thus resulting in lower FM scores. Additionally, although UCAD reported the FM metric in the original paper, its historical best score  $a_{\gamma,i}$  was achieved by storing a larger number of patch features, which fails to reflect the true FM metric. Therefore, we marked it as “-”.

Additionally, we compare our method’s storage space with the SOTA methods. As shown in Table 2, different from UCAD storage patch-level features and CDAD storage gradient projection, our approach significantly reduces storage space by storing semantic and spatial features of samples,



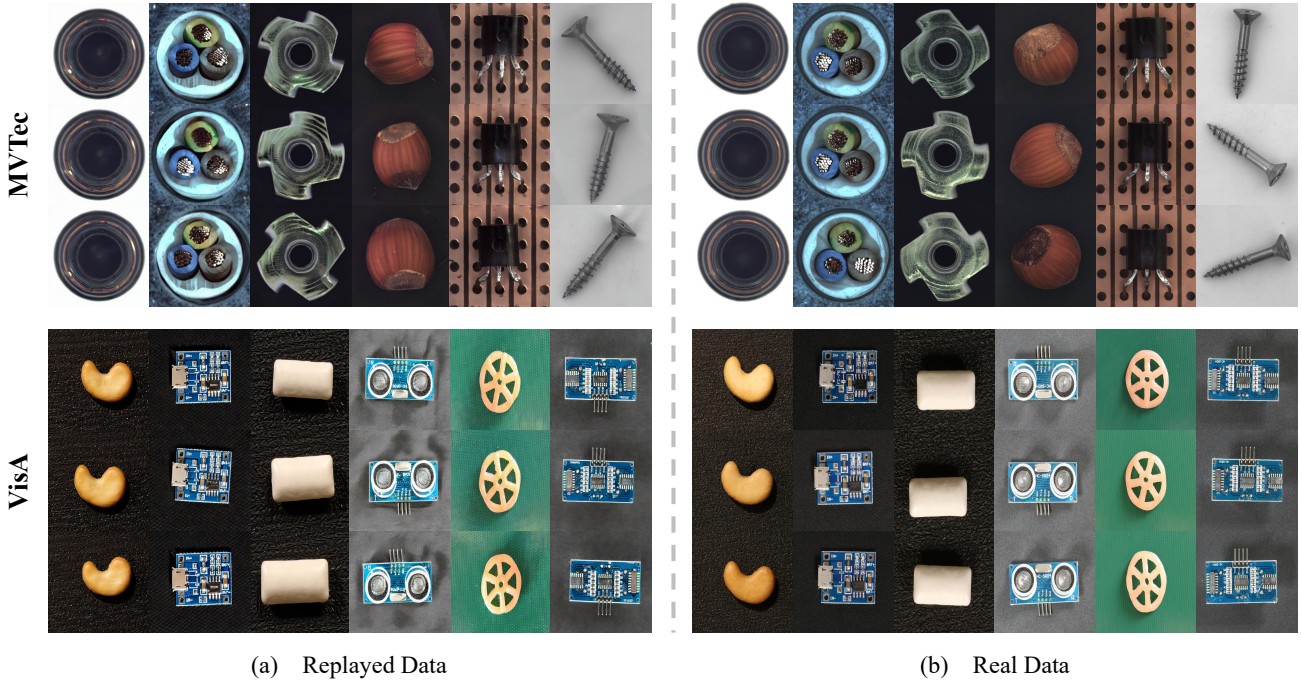


Figure 7: The visual comparison of replayed data and real data on the MVTec and VisA datasets.

while also achieving superior experimental performance.

**Qualitative Evaluation.** Figure 6 shows the heatmaps of anomaly regions predicted by our method and other SOTA methods. We found that the AD methods, AD + CL methods, and IUF often result in a large area of error prediction. This occurs because, when learning new classes, these methods conflict with the features of historical classes, leading to catastrophic forgetting of previously acquired knowledge. Although UCAD shows significant improvement compared to previous methods, it still suffers from missed detections in anomalous regions. Visualization results demonstrate that our method achieves higher accuracy in anomaly segmentation.

**Visualization of Replayed Data.** As shown in Figure 7, we present a comparison between the replayed data and the real data. Since AD is conducted in an unsupervised setting, the replayed data should closely resemble the real normal data. The results indicate that the replayed data exhibit a high degree of similarity to the real data, validating the effectiveness of our framework based on generative replay.

### 5.3 Analysis

In this section, we analyze the impact of the two proposed features and different AD models on MVTec.

**Effect of semantic and spatial features.** As shown in Table 3, we analyze the effectiveness of the proposed semantic features  $e_{semantic}$  and spatial features  $e_{spatial}$ . **T1** indicates that we do not retain any features of historical samples but instead directly train the AD model sequentially on samples from each class. In this case, the model suffers from catastrophic forgetting, resulting in a sharp decline in performance. **T2** and **T3** represent the processes of learning and storing spatial features and semantic features of the sam-

ples, respectively, followed by replaying historical data using the learned features. The results indicate that both types of features can significantly alleviate catastrophic forgetting. **T4** simultaneously learns both spatial and semantic features. Compared to learning them individually, this approach significantly improves classification and segmentation performance, further demonstrating the critical role of these two types of features in replaying historical data.

**Transferring to different AD models.** To validate the generalization capability of the replayed data, we transfer it to various anomaly detection models for training, including Patchcore [Roth *et al.*, 2022], RLR [He *et al.*, 2025], MambaAD [He *et al.*, 2024], and InvAD [Zhang *et al.*, 2024b], as shown in Table 4. Our replayed data demonstrated strong performance across four methods, confirming the superior generalization ability of the proposed method. Moreover, this indicates that our method can be readily adapted to SOTA AD methods at any time. Compared to other CAD methods that can only adapt to a single AD method, our approach demonstrates greater adaptability and sustainability.

### 6 Conclusion

In this paper, we introduce ReplayCAD, marking the first use of generative replay based on a pre-trained diffusion model in CAD. Compared to current state-of-the-art methods, our approach more effectively preserves the detailed features of historical samples, leading to significant improvements in segmenting anomalous regions. Additionally, by incorporating semantic and spatial feature guidance into data compression, we achieve high-quality data generation. Experimental results demonstrate that our method achieves SOTA performance on both image-level and pixel-level metrics.

## Acknowledgments

The research is partially supported by National Natural Science Foundation of China (No.62176093, 61673182), National Key Research and Development Program of China (No.2023YFC3502900), Key Realm Research and Development Program of Guangzhou (No.202206030001), Guangdong-Hong Kong-Macao Joint Innovation Project (No.2023A0505030016), Guangdong Emergency Management Science and Technology Program (No.2025YJKY001).

## References

- [Aljundi *et al.*, 2018] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *ECCV*, pages 139–154, 2018.
- [Bergmann *et al.*, 2019] 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.
- [Chaudhry *et al.*, 2018] Arslan Chaudhry, Puneet K Dokania, Thalaiyasingam Ajanthan, and Philip HS Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *ECCV*, pages 532–547, 2018.
- [Dai *et al.*, 2023] Gang Dai, Yifan Zhang, Qingfeng Wang, Qing Du, Zhuliang Yu, Zhuoman Liu, and Shuangping Huang. Disentangling writer and character styles for handwriting generation. In *CVPR*, pages 5977–5986, 2023.
- [Dai *et al.*, 2024] Gang Dai, Yifan Zhang, Quhui Ke, Qiangya Guo, and Shuangping Huang. One-dm: One-shot diffusion mimicker for handwritten text generation. In *ECCV*, pages 410–427. Springer, 2024.
- [Deng and Li, 2022] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, pages 9737–9746, 2022.
- [Gao and Liu, 2023] Rui Gao and Weiwei Liu. Ddgr: Continual learning with deep diffusion-based generative replay. In *ICML*, pages 10744–10763. PMLR, 2023.
- [He *et al.*, 2024] 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. *arXiv preprint arXiv:2404.06564*, 2024.
- [He *et al.*, 2025] Liren He, Zhengkai Jiang, Jinlong Peng, Wenbing Zhu, Liang Liu, Qiangang Du, Xiaobin Hu, Mingmin Chi, Yabiao Wang, and Chengjie Wang. Learning unified reference representation for unsupervised multi-class anomaly detection. In *ECCV*, pages 216–232. Springer, 2025.
- [Hu and Huang, 2025] Lei Hu and Shuangping Huang. Enhancing table structure recognition via bounding box guidance. In *ICPR*, pages 209–225. Springer, 2025.
- [Hu *et al.*, 2024] Teng Hu, Jiangning Zhang, Ran Yi, Yuzhen Du, Xu Chen, Liang Liu, Yabiao Wang, and Chengjie Wang. Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In *AAAI*, volume 38, pages 8526–8534, 2024.
- [Kirillov *et al.*, 2023] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *ICCV*, pages 4015–4026, 2023.
- [Kirkpatrick *et al.*, 2017] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.
- [Li *et al.*, 2021] 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.
- [Li *et al.*, 2022] 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 *ACMMM*, pages 2871–2880, 2022.
- [Li *et al.*, 2024] Wei Li, Jingyang Zhang, Pheng-Ann Heng, and Lixu Gu. Comprehensive generative replay for task-incremental segmentation with concurrent appearance and semantic forgetting. In *MICCAI*, pages 80–90. Springer, 2024.
- [Li *et al.*, 2025] 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. *arXiv preprint arXiv:2502.19848*, 2025.
- [Liang *et al.*, 2024] Jinglin Liang, Jin Zhong, Hanlin Gu, Zhongqi Lu, Xingxing Tang, Gang Dai, Shuangping Huang, Lixin Fan, and Qiang Yang. Diffusion-driven data replay: A novel approach to combat forgetting in federated class continual learning. In *ECCV*, pages 303–319. Springer, 2024.
- [Liu *et al.*, 2020] Xialei Liu, Chenshen Wu, Mikel Menta, Luis Herranz, Bogdan Raducanu, Andrew D Bagdanov, Shangling Jui, and Joost van de Weijer. Generative feature replay for class-incremental learning. In *CVPRW*, pages 226–227, 2020.
- [Liu *et al.*, 2023] 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.
- [Liu *et al.*, 2024a] 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*, volume 38, pages 3639–3647, 2024.



- [Liu *et al.*, 2024b] Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 21(1):104–135, 2024.
- [Mallya and Lazebnik, 2018] Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *CVPR*, pages 7765–7773, 2018.
- [Mallya *et al.*, 2018] Arun Mallya, Dillon Davis, and Svetlana Lazebnik. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *ECCV*, pages 67–82, 2018.
- [Pan *et al.*, 2020] Pingbo Pan, Siddharth Swaroop, Alexander Immer, Runa Eschenhagen, Richard Turner, and Mohammad Emtiyaz E Khan. Continual deep learning by functional regularisation of memorable past. *NIPS*, 33:4453–4464, 2020.
- [Peng *et al.*, 2024] Wenjie Peng, Hongxiang Huang, Tianshui Chen, Quhui Ke, Gang Dai, and Shuangping Huang. Globally correlation-aware hard negative generation. *IJCV*, pages 1–22, 2024.
- [Rebuffi *et al.*, 2017] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *CVPR*, pages 2001–2010, 2017.
- [Reiss *et al.*, 2021] Tal Reiss, Niv Cohen, Liron Bergman, and Yedid Hoshen. Panda: Adapting pretrained features for anomaly detection and segmentation. In *CVPR*, pages 2806–2814, 2021.
- [Riemer *et al.*, 2019] Matthew Riemer, Tim Klinger, Djallel Bouneffouf, and Michele Franceschini. Scalable recollections for continual lifelong learning. In *AAAI*, volume 33, pages 1352–1359, 2019.
- [Rombach *et al.*, 2022] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022.
- [Roth *et al.*, 2022] 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.
- [Shin *et al.*, 2017] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. *NIPS*, 30, 2017.
- [Tang *et al.*, 2025] Jiaqi Tang, Hao Lu, Xiaogang Xu, Ruizheng Wu, Sixing Hu, Tong Zhang, Tsz Wa Cheng, Ming Ge, Ying-Cong Chen, and Fuguee Tsung. An incremental unified framework for small defect inspection. In *ECCV*, pages 307–324. Springer, 2025.
- [Wang *et al.*, 2021] Liyuan Wang, Kuo Yang, Chongxuan Li, Lanqing Hong, Zhenguo Li, and Jun Zhu. Ordisco: Effective and efficient usage of incremental unlabeled data for semi-supervised continual learning. In *CVPR*, pages 5383–5392, 2021.
- [Wu *et al.*, 2024] Di Wu, Shicai Fan, Xue Zhou, Li Yu, Yuzhong Deng, Jianxiao Zou, and Baihong Lin. Unsupervised anomaly detection via masked diffusion posterior sampling. In *IJCAI*, pages 2442–2450, 2024.
- [Yang *et al.*, 2025] Shuai Yang, Zhifei Chen, Pengguang Chen, Xi Fang, Yixun Liang, Shu Liu, and Yingcong Chen. Defect spectrum: a granular look of large-scale defect datasets with rich semantics. In *ECCV*, pages 187–203. Springer, 2025.
- [Yoon *et al.*, 2017] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. *arXiv preprint arXiv:1708.01547*, 2017.
- [You *et al.*, 2022] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *NIPS*, 35:4571–4584, 2022.
- [Zavrtanik *et al.*, 2021a] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8330–8339, 2021.
- [Zavrtanik *et al.*, 2021b] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8330–8339, 2021.
- [Zhang *et al.*, 2024a] Jiangning Zhang, Chengjie Wang, Xi-angtai 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.
- [Zhang *et al.*, 2024b] Jiangning Zhang, Chengjie Wang, Xi-angtai 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.
- [Zhou and Paffenroth, 2017] Chong Zhou and Randy C Paffenroth. Anomaly detection with robust deep autoencoders. In *SIGKDD*, pages 665–674, 2017.
- [Zou *et al.*, 2022] 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.
- [Zuo *et al.*, 2024] Zuo Zuo, Yao Wu, Baoqiang Li, Jiahao Dong, You Zhou, Lei Zhou, Yanyun Qu, and Zongze Wu. Clip-fsac: Boosting clip for few-shot anomaly classification with synthetic anomalies. In *IJCAI*, pages 1834–1842, 2024.