

LogiCo: A Unified Framework for Logical and Structural Anomaly Detection

Ximiao Zhang¹, Min Xu², and Xiuzhuang Zhou^{1✉}

¹ School of Intelligent Engineering and Automation, Beijing University of Posts and Telecommunications, Beijing, China

xiuzhuang.zhou@bupt.edu.cn

² College of Information and Engineering, Capital Normal University, Beijing, China

Abstract. Current anomaly detection methods primarily focus on structural anomalies, while paying insufficient attention to anomalies that violate logical constraints. Conversely, top-performing logical anomaly detection approaches address this by modeling global semantic consistency, but perform poorly on subtle structural anomalies due to inadequate detection granularity. In this paper, we propose **LogiCo**, a unified framework for **Logical** and structural anomaly detection via **Component-level** feature reconstruction. Unlike existing methods that rely on explicit global semantic modeling, LogiCo employs a novel component-level feature reconstruction technique to capture inter-component logical constraints. Specifically, LogiCo maps pre-trained image features into a discrete component-level feature space and performs collaborative feature reconstruction at both component and patch levels, enabling it to effectively detect both logical and structural anomalies. Furthermore, to address the specific challenge of count-related logical anomalies, we integrate a segmentation-map discriminator that extends the model's capability to identify quantitative inconsistencies. LogiCo achieves state-of-the-art performance on both logical and structural anomaly detection across four benchmarks, including MVTec-LOCO, MVTec-AD, VisA, and Real-IAD, demonstrating its superiority and practical feasibility. The code is available at <https://github.com/cnulab/LogiCo>.

Keywords: Logical Anomaly Detection · Surface Anomaly Detection · Component-level Feature Reconstruction · Component Segmentation

1 Introduction

Image anomaly detection has emerged as a fundamental problem in industrial vision. Given the inherent scarcity of defective samples in real-world scenarios, this task is typically formulated as an unsupervised learning paradigm [12, 28, 37]. Anomalies are generally classified into two types: structural anomalies and logical anomalies. Structural anomalies are characterized by unseen local visual patterns, such as scratches, breakage, and stains. In contrast, logical anomalies represent invalid configurations that deviate from production flows or assembly requirements, such as missing components, incorrect locations, and count

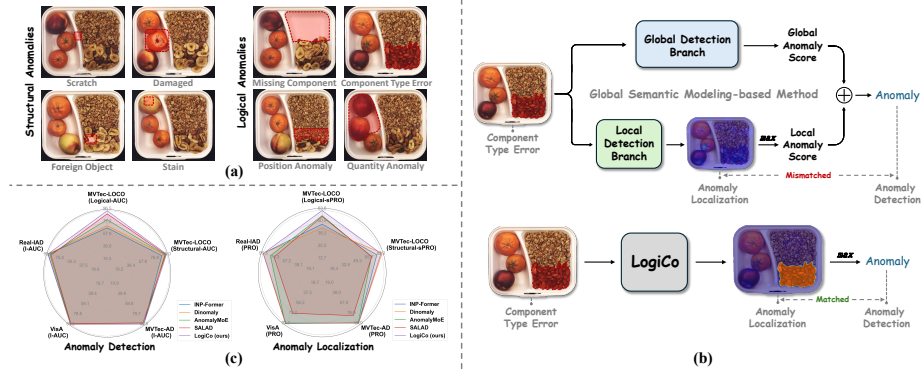


Fig. 1: (a): Examples of structural and logical anomalies. (b): Comparison of the overall pipelines between LogiCo and global semantic modeling-based methods. (c): Comparison of LogiCo with other methods on anomaly detection and localization performance.

anomalies, as illustrated in Fig. 1(a). Prevalent unsupervised anomaly detection approaches [10, 11, 17, 33, 52, 53] typically model the feature distribution of image patches. While effective for structural defects, these methods are intrinsically constrained by their local modeling schemes, failing to capture long-range logical dependencies. To mitigate this, recent state-of-the-art (SOTA) logical anomaly detection methods [14, 18, 21, 31, 36] usually employ an additional global detection branch to learn the global semantic consistency, forming a global-local architecture, as shown in Fig. 1(b). The global branch discards spatial structure, allowing it to effectively capture overall appearance patterns and thereby help identify logical anomalies. Nevertheless, due to the lack of fine-grained discriminative information, the global branch struggles to detect subtle structural anomalies, which may impair the model’s capability for structural anomaly detection. Furthermore, the global branch is typically unable to predict pixel-level anomaly maps, leading to mismatched anomaly detection and localization results. Currently, the performance of logical anomaly detection methods in detecting structural anomalies is still notably inferior to that of specialized structural detection methods.

To address the above issues, we propose LogiCo, a unified framework for logical and structural anomaly detection. Unlike previous works limited by coarse-grained global consistency modeling, LogiCo consistently preserves the image’s spatial layout, thereby facilitating precise anomaly detection and localization, as illustrated in Fig. 1(b). LogiCo consists of three key components: a Component-level Reconstruction Network (CRN), a Structural Reconstruction Network (SRN), and a Segmentation Map Discriminator (SMD). Specifically, the CRN is trained to restore discretized component-level feature maps to their normal pre-trained counterparts, thereby capturing inter-component logical inconsistencies. The SRN reconstructs fine-grained local features under cross-attention guidance, enabling precise detection of structural anomalies. The SMD models quantity con-

straints by learning the normal patterns of component segmentation maps, facilitating the identification of count-related anomalies. Compared with existing methods, LogiCo significantly extends the scope of detectable anomalies while preserving high detection accuracy. As illustrated in Fig. 1(c), LogiCo surpasses existing methods in logical anomaly detection, achieving SOTA performance in both anomaly detection and localization. For structural anomalies, it achieves performance comparable to or exceeding that of top-tier methods. Our main contributions are summarized as follows:

- We propose LogiCo, a unified framework for logical and structural anomaly detection. LogiCo captures local structural consistency and global logical constraints in a fine-grained discriminative space, effectively addressing the limitations of existing methods regarding detection scope and granularity.
- We introduce a novel component-level feature reconstruction technique, which transforms the complex modeling of multiple logical constraints into a unified reconstruction paradigm. This allows for the precise detection and pixel-level localization of logical anomalies without compromising spatial structures.
- LogiCo achieves SOTA logical and structural anomaly detection performance on four benchmarks: MVTec-LOCO, MVTec-AD, VisA, and Real-IAD, demonstrating its effectiveness and potential for real-world applications.

2 Related Work

Structural Anomaly Detection. Structural anomalies are prevalent in fields such as industrial manufacturing [3, 5, 20, 40, 55] and medical imaging [7, 42, 43, 51], prompting extensive research. Since such anomalies typically manifest as localized abnormal patterns, existing methods usually employ local feature modeling schemes. Among them, reconstruction-based methods [13, 45, 47–49] train models to learn the normal patterns of images, and identify anomalous regions via pixel-wise or patch-wise reconstruction. Knowledge distillation-based methods [2, 6, 11, 17, 38] train a student network to learn the normal feature distribution of the teacher network. In the inference phase, they localize anomalous regions via patch-level feature comparison. Deep feature embedding-based methods learn the normal feature distribution during training, for instance, by constructing memory banks [1, 8, 33] or utilizing normalizing flows [16, 24, 46]. Subsequently, anomaly detection is achieved by measuring the deviation between query features and the learned feature distribution. Despite these advances, the above methods lack global perception capabilities, and thus often fail in detecting logical anomalies. Anomaly synthesis-based methods [19, 25, 34, 52] train anomaly detection models using synthetic anomalous samples. These approaches emphasize that synthetic anomalies can effectively simulate real-world anomalies. However, for logical anomaly detection, synthesizing realistic logical anomalous samples in either the pixel space or the continuous feature space remains a significant challenge [39, 54].

Logical Anomaly Detection. Logical anomaly detection focuses on capturing global semantic constraints and long-range dependencies; it has broad application prospects yet remains underexplored. Bergmann et al. [4] released MVTEC-LOCO, the first dataset for logical anomaly detection, and proposed GCAD, a logical anomaly detection framework based on global-local knowledge distillation. GCAD employs a global network branch with a low-dimensional bottleneck to learn global context and overall semantic constraints. EfficientAD [2] inherits the core idea of GCAD and identifies logical anomalies by comparing the reconstruction error between two reconstruction networks with different model capacities. Sugawara et al. [36] observed that EfficientAD remains ineffective at detecting certain types of logical anomalies, such as missing components or count-related anomalies. To address this, they introduced a global detection branch that disregards inter-component spatial relationships and instead models global semantic consistency using a Gaussian probability model [32] to predict image-level anomaly scores. Recent works such as CSAD [18], SALAD [14], and PSAD [21] have integrated global semantic modeling with component segmentation techniques, achieving SOTA performance in logical anomaly detection. However, due to the limited detection granularity of the global branch, these methods struggle with subtle structural anomalies, thereby limiting their practical application.

3 Method

The overall pipeline of the proposed LogiCo is illustrated in Fig. 2. Given an input image, LogiCo first employs open-vocabulary semantic segmentation to generate component segmentation maps. Guided by these maps, pre-trained image features are discretized and mapped into a component-level feature space. During the training phase, LogiCo synthesizes pseudo-anomalous features by manipulating the segmentation maps. A Component-level Reconstruction Network (CRN) is then trained to restore these features back to their normal, pre-trained counterparts, thereby capturing underlying logical inconsistencies. For structural anomalies, LogiCo employs cross-attention guided reconstruction to compel the Structural Reconstruction Network (SRN) to restore the anomaly features, ensuring comprehensive recall of the anomalous regions. Additionally, a Segmentation Map Discriminator (SMD) is integrated to identify irregularities in segmentation maps, enhancing the detection of quantity-related anomalies.

3.1 Component-level Feature Reconstruction

We employ DINOv3 [35] for feature extraction and open-vocabulary semantic segmentation, which aligns image and text embeddings into a joint embedding space. Specifically, for an input image $I \in \mathbb{R}^{H \times W \times 3}$, we define a set of text prompts $\{P_1, P_2, \dots, P_K\}$ to describe its component categories, such as $\{\text{"Peach"}, \text{"Orange"}, \dots\}$, where H and W represent the height and width of the image, respectively, and K is the number of component types. We then extract

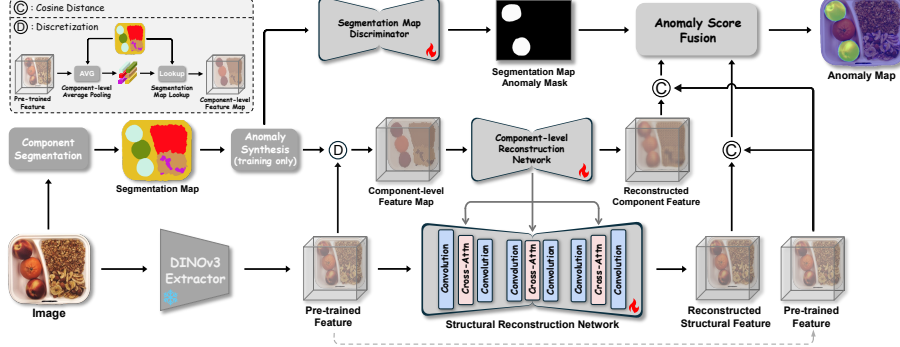


Fig. 2: Overview of the proposed **LogiCo** pipeline, which consists of a Component-level Reconstruction Network (CRN), a Structural Reconstruction Network (SRN), and a Segmentation Map Discriminator (SMD). The CRN and SRN learn global logical constraints and local structural consistency to detect logical and structural anomalies, respectively, whereas the SMD learns quantity rules to identify count-related anomalies.

the pre-trained feature $X = \phi_v(I) \in \mathbb{R}^{h \times w \times c}$ and project it into the joint embedding space, yielding $X_p = \phi_p(X) \in \mathbb{R}^{h \times w \times c_p}$, where $\phi_v(\cdot)$ and $\phi_p(\cdot)$ denote the DINOv3 vision encoder and the projection head, respectively, while c and c_p represent the feature dimensions before and after projection. For text prompts, we use the DINOv3 text encoder $\phi_t(\cdot)$ to embed them into the embedding space to obtain $\{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_K\}$, where $\mathcal{P}_k = \phi_t(P_k) \in \mathbb{R}^{c_p}$ denotes the embedding of the k -th prompt. Utilizing the shared embedding space between text embeddings and projected image features, we compute their similarity at each spatial location and then generate the segmentation map $G \in \{1, 2, \dots, K\}^{h \times w}$ by assigning each patch to the component category with the highest similarity:

$$G(i, j) = \arg \max_{k \in \{1, 2, \dots, K\}} \cos(X_p(i, j), \mathcal{P}_k), \quad (1)$$

where (i, j) denotes the spatial coordinates. Finally, we apply a simple post-processing procedure to refine the segmentation maps, for example by filtering out very small connected components. In contrast to recent works [14, 18, 21, 30], we do not rely on overly refined segmentation maps to avoid additional computational overhead.

Based on the generated segmentation map, we transform the pre-trained feature into a discretized component-level representation $X_c \in \mathbb{R}^{h \times w \times c}$. Specifically, we first compute the component prototypes $\{f_1, f_2, \dots, f_K\}$ via component-level average pooling:

$$f_k = \frac{1}{|\Omega_k|} \sum_{(i, j) \in \Omega_k} X(i, j), \quad (2)$$

where $\Omega_k = \{(i, j) | G(i, j) = k\}$ denotes the spatial indices of the k -th component. Then, we construct the discretized feature map X_c via a lookup operation

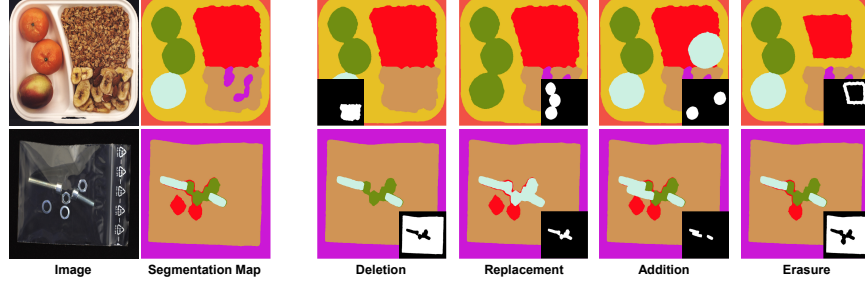


Fig. 3: Examples of segmentation maps generated by various anomaly synthesis methods.

by assigning the corresponding component prototype to each spatial location:

$$X_c(i, j) = f_{G(i, j)}. \quad (3)$$

It offers two key advantages: (1) The component-level feature map preserves the image’s spatial layout and essential information for logical anomaly detection while filtering out irrelevant textural details, thereby facilitating the modeling of logical consistency. (2) The discretized feature space facilitates anomaly synthesis; specifically, by manipulating the segmentation maps during the lookup phase, we can generate diverse logical anomalies to enhance model training.

We design four anomaly synthesis strategies for LogiCo: component deletion, replacement, addition, and erasure, as shown in Fig. 3. **Deletion:** Replacing the component with its neighboring background to remove it. **Replacement:** Replacing one component with another. **Addition:** Employing CutPaste [25] to duplicate a component at a different location, thereby increasing the component count. **Erasure:** Distinct from Deletion, Erasure only removes partial regions of a component. When a component consists of a single connected region, Erasure performs an erosion operation to alter its shape. When it comprises multiple connected regions, Erasure deletes parts of regions to reduce the component count. By integrating these methods, we generate diverse anomalous segmentation maps, which are then used to construct the anomalous component-level feature maps X'_c . Subsequently, a Component-level Reconstruction Network (CRN) ψ_c is trained to restore these features into their normal pre-trained representations: $X_L = \psi_c(X'_c) \in \mathbb{R}^{h \times w \times c}$. To this end, we optimize ψ_c with the cosine distance loss:

$$\mathcal{L}_{\text{Log}} = 1 - \cos(X_L, X). \quad (4)$$

The logical anomaly map $\mathcal{S}_L \in \mathbb{R}^{h \times w}$ is defined as the patch-wise cosine distance between the reconstructed features and the original pre-trained features. Our key insight lies in reconstructing synthesized anomalous features back to pre-trained features rather than their normal discretized counterparts. This provides a continuous learning target that facilitates model convergence and prevents degeneration into trivial solutions. Furthermore, this reconstruction paradigm

enhances LogiCo’s robustness against segmentation noise; even in the presence of noisy segmentation maps, the CRN can still generate normal reconstructed features to avoid false positive detections.

3.2 Cross-attentive Structural Reconstruction

Component-level reconstruction preserves the components’ spatial layout, enhancing the granularity of logical anomaly detection. However, due to the lack of intra-component structural information, its capability for structural anomaly detection remains limited. To this end, we introduce a Structural Reconstruction Network (SRN) to reconstruct the fine-grained pre-trained features for structural anomaly detection: $X_s = \psi_s(X) \in \mathbb{R}^{h \times w \times c}$. Similar to the CRN, the loss function is defined as $\mathcal{L}_{\text{str}} = 1 - \cos(X_s, X)$, and the structural anomaly map $\mathcal{S}_s \in \mathbb{R}^{h \times w}$ is computed as the patch-wise cosine distance. The key challenge in the above process is controlling the reconstruction capability of the SRN to prevent it from preserving anomaly cues, which would result in an “identical shortcut” [17, 45, 49]. To address this, we propose Cross-attentive Structural Reconstruction, a novel method bridging logical and structural anomaly detection. It utilizes the CRN to guide the structural reconstruction process, compelling the SRN to restore anomalous features to their normal representations. Recall that local structural information is discarded during the feature discretization; consequently, the feature space of CRN is free of structural anomalies. Meanwhile, it retains the overall spatial layout of the image, providing a valuable reference for structural reconstruction. Specifically, for the h -th intermediate-layer feature $X_s^h \in \mathbb{R}^{h \times w \times c}$ from the SRN, we extract the corresponding layer feature $X_L^h \in \mathbb{R}^{h \times w \times c}$ from the CRN to refine the reconstruction process via a cross-attention mechanism:

$$\begin{aligned} Q_h &= X_s^h W_Q^h, \quad K_h = X_L^h W_K^h, \quad V_h = X_L^h W_V^h, \\ \hat{X}_s^h &= FFN(\text{Attention}(Q_h, K_h, V_h)), \end{aligned} \quad (5)$$

where $Q_h, K_h, V_h \in \mathbb{R}^{h \times w \times c}$ denote the query, key, and value matrices, respectively, and $W_Q^h, W_K^h, W_V^h \in \mathbb{R}^{c \times c}$ are the learnable projection weights. $\text{Attention}(\cdot)$ and $FFN(\cdot)$ represent the attention operation and the feed-forward network, while $\hat{X}_s^h$ is the reconstructed feature map. The above cross-attention operation uses the anomaly-free X_L^h as the key-value pair, explicitly mapping X_s^h to its normal representation, thereby enhancing the reliability of reconstruction while avoiding missed detections. In our implementation, the CRN and SRN adopt Autoencoder and U-Net architectures of the same depth, respectively. Specifically, we insert a self-attention layer into each CRN block to capture global semantics and generate guiding features, and place a cross-attention layer at the corresponding location in the SRN to enable cross-attentive reconstruction. We omit the residual connection in the standard cross-attention module to prevent the leakage of anomalous information.

3.3 Segmentation Map Discriminator

The Segmentation Map Discriminator (SMD) is specifically designed to detect quantity-related logical anomalies. We find that learning constraints related to component quantity is challenging in the feature space, whereas segmentation maps offer a more intuitive representation. The SMD directly predicts the anomaly mask based on the segmentation map: $\mathcal{S}_D = \psi_D(G) \in [0, 1]^{h \times w}$, where 0 and 1 denote normal and anomalous pixels, respectively. We employ a lightweight fully convolutional U-Net model as our SMD, trained using synthetic segmentation map-mask pairs $(G', \mathcal{M})$, as illustrated in Fig. 3. We optimize the SMD using Binary Focal Loss [27], denoted as:

$$\mathcal{L}_{\text{Dis}} = \text{Focal}(\psi_D(G'), \mathcal{M}). \quad (6)$$

3.4 Training and Inference

We jointly train the different branches within LogiCo, with the overall loss function defined as:

$$\mathcal{L} = \mathcal{L}_{\text{Log}} + \mathcal{L}_{\text{Str}} + \mathcal{L}_{\text{Dis}}. \quad (7)$$

In the inference phase, we omit the anomaly synthesis and perform the same segmentation and reconstruction steps to obtain anomaly maps $\mathcal{S}_L$, $\mathcal{S}_s$, and $\mathcal{S}_D$ for the query image. Subsequently, we compute the means and standard deviations of these anomaly maps on the validation set, utilizing these statistics for normalization and anomaly score fusion:

$$\mathcal{S} = \max\left\{\frac{\mathcal{S}_L - \mu(\mathcal{S}_L)}{\sigma(\mathcal{S}_L)}, \frac{\mathcal{S}_s - \mu(\mathcal{S}_s)}{\sigma(\mathcal{S}_s)}, \frac{\mathcal{S}_D - \mu(\mathcal{S}_D)}{\sigma(\mathcal{S}_D)}\right\}. \quad (8)$$

To comprehensively capture diverse types of anomalies, we derive the overall anomaly map $\mathcal{S} \in \mathbb{R}^{h \times w}$ by taking the element-wise maximum across these maps. Finally, we upsample $\mathcal{S}$ to the image resolution to obtain the pixel-level anomaly score, while its maximum is used as the image-level anomaly score.

4 Experiments

4.1 Experimental Setup

Datasets. We conduct extensive evaluations on four anomaly detection benchmarks: MVTec-LOCO [4], MVTec-AD [5], VisA [55], and Real-IAD [40]. MVTec-LOCO [4] is a benchmark specifically designed for logical anomaly detection, comprising 3,644 images across five categories, each featuring diverse logical and structural anomalies. The MVTec-AD benchmark [5] comprises 5,354 images across 15 industrial products, where certain categories such as “cable” and “transistor” contain logical anomalies like missing or misplaced components. The VisA benchmark [55] comprises 10,821 images across 12 products. Certain categories present significant challenges due to their intricate structures and the

Table 1: Comparison of LogiCo with other anomaly detection methods. [†] denotes methods specifically designed for logical anomaly detection.

Method	MVTec-LOCO		MVTec-AD				VisA				Real-IAD			
	I-AUC	sPRO	I-AUC	P-AUC	P-AP	PRO	I-AUC	P-AUC	P-AP	PRO	I-AUC	P-AUC	P-AP	PRO
PatchCore [33]	75.6	34.4	99.1	98.4	65.8	94.4	94.9	98.5	48.0	91.0	91.7	98.9	39.6	92.8
UniNet [41]	83.0	65.5	99.2	98.6	64.3	<u>95.4</u>	97.9	98.7	45.6	93.0	91.7	99.2	40.9	95.8
INP-Former [29]	79.3	63.0	<u>99.5</u>	98.4	69.6	95.5	98.4	98.7	49.2	93.9	93.2	<u>99.3</u>	<u>48.4</u>	<u>95.6</u>
Dinomaly [17]	83.4	64.8	99.6	98.8	71.2	95.5	98.9	99.3	<u>57.7</u>	<u>94.8</u>	<u>93.7</u>	99.4	48.1	95.7
[†] GLCF [44]	78.1	63.7	93.9	96.4	49.0	80.5	87.1	96.1	28.5	73.2	85.4	95.1	26.1	75.3
[†] AnomalyMoE [15]	88.2	63.6	<u>99.5</u>	98.1	67.0	94.4	98.1	<u>99.0</u>	54.6	94.9	90.5	97.8	35.8	83.9
[†] CSAD [18]	92.9	<u>70.6</u>	95.8	97.7	56.3	86.6	90.6	97.0	31.1	73.7	83.4	96.3	29.8	78.9
[†] SALAD [14]	<u>93.2</u>	63.3	97.2	96.9	52.6	81.1	97.4	97.4	32.7	75.7	88.8	96.2	30.2	77.9
[†] LogiCo	96.3	72.9	99.6	<u>98.7</u>	<u>70.2</u>	95.0	<u>98.5</u>	99.3	60.1	93.7	93.8	99.4	53.3	95.3

Table 2: Comparison between LogiCo and other logical anomaly detection methods on the MVTec-LOCO dataset. [†] indicates that the results are taken from the original paper. LogiCo+ denotes LogiCo using SAM-refined [22] component segmentation maps.

Method	Fine-grained Segmentation	Detection (I-AUC)			Localization (sPRO)		
		Logical	Structural	AVG	Logical	Structural	AVG
GLCF [44]	X	72.9	83.2	78.1	58.7	68.6	63.7
[†] GCAD [4]	X	86.0	80.6	83.3	71.1	<u>69.2</u>	<u>70.1</u>
[†] EfficientAD [2]	X	86.8	<u>94.7</u>	90.7	-	-	-
SALAD [14]	X	<u>91.7</u>	<u>94.7</u>	<u>93.2</u>	57.4	<u>69.2</u>	63.3
LogiCo	X	96.5	96.0	96.3	<u>63.6</u>	82.2	72.9
PSAD [21]	✓	97.7	91.3	<u>94.5</u>	-	-	-
LogSAD [50]	✓	89.3	91.6	90.4	58.3	78.8	68.6
CSAD [18]	✓	92.6	<u>93.1</u>	92.9	65.5	75.6	70.6
[†] SAM-LAD [30]	✓	95.8	85.5	90.7	85.6	<u>80.7</u>	83.2
LogiCo+	✓	<u>97.1</u>	96.2	96.7	<u>76.2</u>	84.6	<u>80.4</u>

presence of multiple target objects. The Real-IAD benchmark [40] comprises 151,050 multi-view images across 30 different products, with some categories exhibiting subtle structural anomalies that are difficult to detect. We adopt the single-view experimental setting, following recent works [23, 26, 53].

Implementation Details. We utilize DINOv3 ViT-B/16 [35] as our visual backbone and extract feature maps from the final (12th) layer. All images are resized to a resolution of 512×512 for both semantic segmentation and anomaly detection. For the texture classes in the MVTec-AD dataset [5], we treat them as a single component and disable the anomaly synthesis and SMD. Otherwise, we uniformly sample from the four anomaly synthesis strategies and apply the same configuration across all other image classes. For the MVTec-AD [5], VisA [55], and Real-IAD [40] datasets, we hold out 20% of the training images to compute the means and standard deviations of the anomaly maps for normalization.

Metrics. For the MVTec-AD [5], VisA [55], and Real-IAD [40] datasets, we evaluate anomaly detection performance using image-level AUROC (I-AUC) and anomaly localization performance using pixel-level AUROC (P-AUC), average precision (P-AP), and per-region overlap (PRO) [6]. For the MVTec-LOCO dataset [4], we employ I-AUC and saturated PRO (sPRO) [4] to evaluate

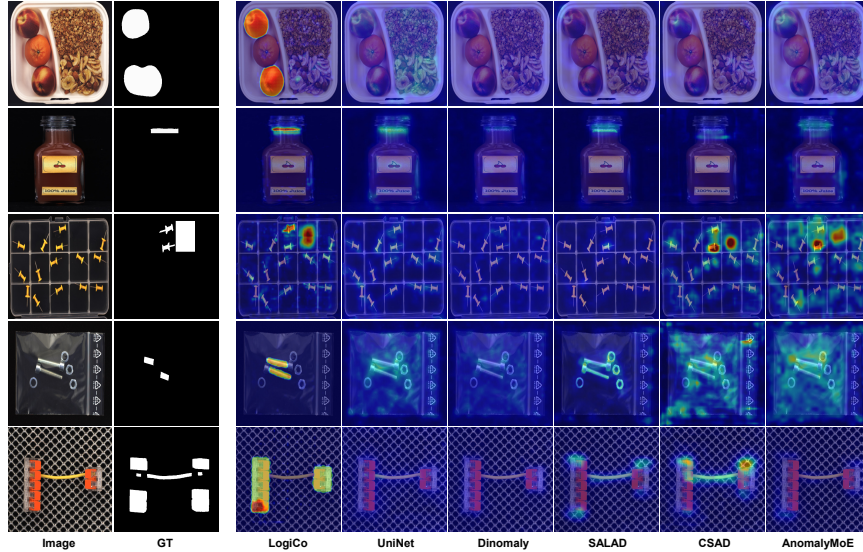


Fig. 4: Qualitative comparison of LogiCo and other methods on logical anomaly detection.

anomaly detection and localization performance, respectively, noting that sPRO is specifically designed for logical anomalies. Following previous work [4, 14, 18], metrics for logical and structural anomalies are computed separately, and their average is reported as the overall metric.

Baselines. For structural anomaly detection, we use PatchCore [33], UniNet [41], INP-Former [29], and Dinomally [17] as our baseline methods. For logical anomaly detection, we use GLCF [44], AnomalyMoE [15], CSAD [18], SALAD [14], GCAD [4], EfficientAD [2], PSAD [21], LogSAD [50], and SAM-LAD [30] as baseline methods. For INP-Former [29], Dinomally [17], and AnomalyMoE [15], we use the same backbone and image size as LogiCo to ensure a fair comparison. To enable a comprehensive and unified evaluation with consistent metrics and more challenging datasets, we rigorously reproduce the baselines using their official code, with detailed reproduction protocols provided in the **Appendix**. Unless otherwise specified, all reported results are reproduced results.

4.2 Experimental Results

Tab. 1 presents the performance of LogiCo compared to other methods across different benchmarks. Current SOTA methods like Dinomally [17] and INP-Former [29] excel on benchmarks focusing on structural anomalies, such as MVTec-AD [5], VisA [55], and Real-IAD [40]. However, they struggle to generalize to the logical anomalies present in the MVTec-LOCO dataset [4]. Conversely, methods explicitly designed for logical anomalies, such as CSAD [18], SALAD [14],

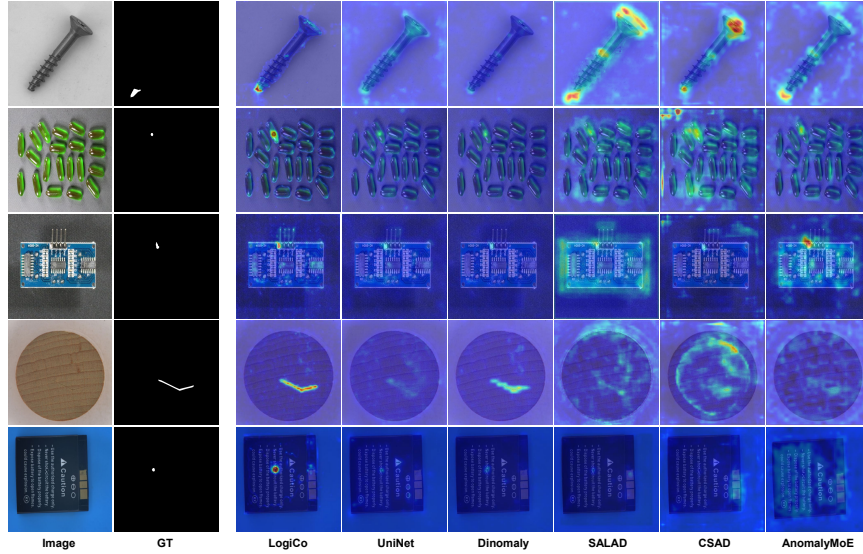


Fig. 5: Qualitative comparison of LogiCo and other methods on structural anomaly detection.



Fig. 6: Examples of component segmentation on the MVTec-LOCO dataset.

and AnomalyMoE [15], detect logical defects more effectively but lag significantly behind on benchmarks characterized by subtle structural anomalies, particularly on the challenging VisA [55] and Real-IAD [40] datasets. In contrast, LogiCo demonstrates superior performance across all benchmarks. Specifically, it achieves an impressive 96.3% I-AUC and 72.9% sPRO on the MVTec-LOCO benchmark [4], while performing on par with or even surpassing top-performing methods on other structural anomaly detection benchmarks. Tab. 2 further reports the logical and structural anomaly detection and localization performance of different methods on the MVTec-LOCO dataset [4]. Among them, some approaches [14, 21] achieve strong anomaly detection performance due to their global semantic modeling capabilities, but perform poorly in anomaly localization. In contrast, some methods [18, 30] excel at anomaly localization by leveraging fine-grained segmentation maps, yet show limited performance in anomaly detection. By comparison, LogiCo achieves a favorable trade-off between these

Table 3: Comparison of computational efficiency between LogiCo and other methods.

Method	Training Time (h) ↓	Images/s ↑	Performance on MVTec-LOCO ↑			
			(Log.-AUC/Str.-AUC/Log.-sPRO/Str.-sPRO)			
Dinomaly [17]	0.7	30.2	73.71	93.04	45.83	83.74
AnomalyMoE [15]	18.0	15.2	82.79	93.57	58.36	68.78
SALAD [14]	1.7	37.6	91.73	94.69	57.39	69.22
LogSAD [50]	0.2	0.3	89.29	91.58	58.31	78.81
LogiCo	1.0	17.4	96.53	96.03	63.63	82.16

Table 4: Ablation study results of LogiCo components on the MVTec-LOCO dataset.

CRN	SRN	SMD	Cross Attn.	Detection (I-AUC)		Localization (sPRO)	
				Logical	Structural	Logical	Structural
✓	-	-	-	90.88	84.59	67.51	70.07
-	✓	-	-	75.34	94.86	53.00	81.55
-	-	✓	-	85.63	61.20	58.05	35.21
✓	-	✓	-	96.87	82.47	64.22	62.84
-	✓	✓	-	86.85	93.84	56.79	80.00
✓	✓	-	-	89.43	94.94	<u>67.71</u>	81.98
✓	✓	-	✓	90.05	<u>95.95</u>	68.08	84.71
✓	✓	✓	-	95.81	94.82	63.04	81.05
✓	✓	✓	✓	<u>96.53</u>	96.03	63.63	<u>82.16</u>

two aspects. Moreover, when using finer segmentation maps, LogiCo obtains a substantial performance improvement. Figs. 4 and 5 present the qualitative results of LogiCo and other methods on logical and structural anomalies, respectively. The results demonstrate that LogiCo precisely localizes various types of anomaly regions, significantly extending the detectable range of existing methods and emphasizing its feasibility for real-world scenarios.

Fig. 6 displays the component segmentation results of LogiCo on the MVTec-LOCO dataset. Unlike prior works that rely on elaborate segmentation pipelines [18, 21, 30, 50], LogiCo adopts a simple open-vocabulary segmentation scheme, significantly reducing computational overhead. Moreover, LogiCo demonstrates robustness against segmentation noise. Specifically, in the **Juice Bottle** category, it yields an I-AUC of 98.93% and an sPRO of 82.67% (refer to the **Appendix** for per-category metrics), despite the segmentation module struggling to distinguish between specific juice types. We attribute this to LogiCo’s robust reconstruction capability, which effectively restores normal feature representations from noisy inputs, thereby mitigating false positives. More experimental results and analysis on component segmentation are provided in the **Appendix**.

Tab. 3 compares the computational efficiency of different methods. The experiments are implemented in PyTorch and conducted on a single Nvidia RTX 5090 GPU. At inference time, we set the batch size to 1 to mimic practical deployment scenarios. Benefiting from a simplified component segmentation pipeline, LogiCo eliminates the need to train an additional component segmentation model [14, 18, 21], which shortens its training time and enables it to meet the practical demands of rapid product iteration and updates. Furthermore, LogiCo

Table 5: Ablation study results of anomaly synthesis on the MVTec-LOCO dataset.

Anomaly Synthesis Method	Detection (I-AUC)		Localization (sPRO)	
	Logical	Structural	Logical	Structural
Deletion	85.70	94.64	46.03	81.13
Replacement	88.47	95.44	53.08	82.04
Erasure	90.29	95.81	52.81	<u>82.37</u>
Addition	93.41	95.01	61.45	81.99
Addition+Deletion	93.73	95.24	61.90	82.12
Addition+Replacement	94.58	95.35	62.68	82.09
Addition+Erasure	95.73	95.59	62.29	82.39
Addition+Replacement+Deletion	94.88	95.55	62.84	81.89
Addition+Erasure+Deletion	95.99	95.72	63.36	82.08
Addition+Erasure+Replacement	<u>96.26</u>	<u>95.91</u>	63.65	82.13
All	96.53	96.03	<u>63.63</u>	82.16

Table 6: Ablation study results on different backbones and image sizes.

Backbone	Image Size	MVTec-LOCO		MVTec-AD			
		I-AUC	sPRO	I-AUC	P-AUC	P-AP	PRO
DINOv2-R [9]	224 ²	94.42	69.48	99.30	98.14	65.98	93.19
	448 ²	95.67	71.85	99.55	98.58	68.09	94.27
DINOv3 _{VIT-S} [35]	512 ²	95.50	70.20	99.49	98.43	68.81	94.37
DINOv3 _{VIT-B} [35]	256 ²	94.78	69.65	99.04	98.33	66.81	93.75
	384 ²	95.59	71.38	99.40	98.47	68.30	94.26
	512 ²	96.28	<u>72.90</u>	<u>99.59</u>	<u>98.68</u>	<u>70.19</u>	<u>95.00</u>
DINOv3 _{VIT-L} [35]	512 ²	<u>95.80</u>	73.88	99.70	98.73	70.98	95.29

achieves a high inference speed of 17.4 FPS at a resolution of 512×512 , thereby effectively balancing anomaly detection performance and efficiency to meet the requirements of real-world scenarios.

4.3 Ablation Study

Ablation Study on LogiCo Components. Tab. 4 validates the effectiveness of each component within LogiCo. As a key module of LogiCo, the CRN provides the primary capability for logical anomaly detection. When used independently, it achieves a logical-AUC of 90.88% and a logical-sPRO of 67.51%, which outperforms a wide range of logical anomaly detection methods [2, 4, 44, 50]. In contrast, the SRN is mainly responsible for detecting structural anomalies, compensating for CRN’s limitations in this aspect. The introduction of the SRN improves the structural anomaly detection scores by approximately 10%. Furthermore, by employing cross-attentive reconstruction, SRN’s anomaly detection and localization performance is further enhanced, with structural AUC and sPRO increasing by 1.01% and 2.73%, respectively, thereby demonstrating its indispensability. The SMD module is primarily designed to identify quantity-related logical anomalies, thereby expanding the detection scope of LogiCo. However, incorporating SMD may compromise anomaly localization performance. Since it operates on segmentation maps devoid of appearance details, it can effectively identify the

Table 7: Ablation study results on different intermediate layers.

Layers	MVTec-LOCO		MVTec-AD			
	I-AUC	sPRO	I-AUC	P-AUC	P-AP	PRO
Early Layers _{1,2,3,4,5,6}	93.33	67.15	99.19	98.00	65.18	94.14
Middle Layers _{4,5,6,7,8,9}	93.51	68.83	99.40	98.20	68.84	95.38
Late Layers _{7,8,9,10,11,12}	<u>95.70</u>	<u>71.24</u>	<u>99.57</u>	98.36	69.03	<u>95.26</u>
Even-numbered Layers _{2,4,6,8,10,12}	95.28	70.68	99.56	<u>98.52</u>	70.34	<u>95.26</u>
Final Layer _{12}	96.28	72.90	99.59	98.68	<u>70.19</u>	95.00

anomalous component but fails to pinpoint the precise location of the anomaly within it. Despite this limitation, we advocate using SMD for products subject to quantity constraints to improve overall detection performance.

Ablation Study on Anomaly Synthesis. Tab. 5 investigates the impact of different anomaly synthesis methods. Compared with using a single anomaly synthesis method, using multiple strategies simultaneously leads to significant performance improvements. Furthermore, for the MVTec-LOCO dataset, the **Addition** and **Erase** operations demonstrate superior performance, which may be attributed to their ability to generate diverse anomalies. Our anomaly synthesis process operates within a discrete pixel space, facilitating straightforward customization tailored to the logical constraints of specific products. This flexibility significantly enhances LogiCo’s adaptability and scalability across diverse application scenarios.

Ablation Study on Feature Extraction. Tab. 6 reports the anomaly detection performance of LogiCo using different backbones and image sizes. When using the popular DINOv2-R backbone [9], LogiCo consistently delivers excellent results, confirming its broad effectiveness across various pre-trained models. Furthermore, we observe that performance improves notably as the backbone size and input image resolution increase; however, this incurs a higher computational overhead. In this paper, we employ DINOv3 ViT-B [35] with a resolution of 512×512 as the default configuration to balance performance and computational efficiency. Tab. 7 investigates the impact of extracting pre-trained features from different intermediate layers, using DINOv3 ViT-B as the backbone. For multi-layer settings, we reconstruct the averaged feature representation. We observe that LogiCo benefits from feature maps at deeper layers; specifically, utilizing late layers improves anomaly detection performance by 2.37% on MVTec-LOCO and 0.38% on MVTec-AD compared to early layers. It is worth noting that LogiCo achieves the best overall performance when using only the final layer. This can be ascribed to the Gram Anchoring technique [35] employed by DINOv3, which imposes additional consistency constraints on the final layer during pre-training, enabling it to generate higher-quality feature maps.

5 Conclusion

In this paper, we proposed LogiCo, a unified framework for logical and structural anomaly detection. LogiCo learns diverse logical constraints via component-level

feature reconstruction, and extends its detection capability by integrating cross-attentive structural reconstruction with segmentation map discrimination, effectively addressing the limited detection scope and granularity of existing methods. Specifically, LogiCo comprises a component-level reconstruction network, a structural reconstruction network, and a segmentation map discriminator, all operating in a fine-grained discriminative space, enabling effective detection of diverse anomalies. Extensive experiments on multiple benchmarks demonstrate that LogiCo achieves SOTA performance in both logical and structural anomaly detection. In future work, we aim to further enhance LogiCo’s adaptability across diverse application scenarios, specifically by developing learnable anomaly synthesis strategies and optimization mechanisms.

6 Acknowledgements

This work was funded in part by the National Natural Science Foundation of China under Grants 62576049 and 61972046.

References

1. Bae, J., Lee, J.H., Kim, S.: Pni: industrial anomaly detection using position and neighborhood information. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6373–6383 (2023)
2. Batzner, K., Heckler, L., König, R.: Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 128–138 (2024)
3. Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C.: The mvtec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision* **129**(4), 1038–1059 (2021)
4. Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C.: Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision* **130**(4), 947–969 (2022)
5. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9592–9600 (2019)
6. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4183–4192 (2020)
7. Cai, Y., Chen, H., Yang, X., Zhou, Y., Cheng, K.T.: Dual-distribution discrepancy for anomaly detection in chest x-rays. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 584–593. Springer (2022)
8. Damm, S., Laszkiewicz, M., Lederer, J., Fischer, A.: Anomalydino: Boosting patch-based few-shot anomaly detection with dinov2. In: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 1319–1329. IEEE (2025)

9. Darcet, T., Oquab, M., Mairal, J., Bojanowski, P.: Vision transformers need registers. In: The Twelfth International Conference on Learning Representations (2024)
10. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: International conference on pattern recognition. pp. 475–489. Springer (2021)
11. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9737–9746 (2022)
12. Diers, J., Pigorsch, C.: A survey of methods for automated quality control based on images. *International Journal of Computer Vision* **131**(10), 2553–2581 (2023)
13. Fučka, M., Zavrtnik, V., Skočaj, D.: Transfusion—a transparency-based diffusion model for anomaly detection. In: European conference on computer vision. pp. 91–108. Springer (2024)
14. Fučka, M., Zavrtnik, V., Skočaj, D.: Salad—semantics-aware logical anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21843–21852 (2025)
15. Gu, Z., Zhu, B., Zhu, G., Chen, Y., Ge, W., Tang, M., Wang, J.: Anomalymoe: Towards a language-free generalist model for unified visual anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 4348–4356 (2026)
16. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 98–107 (2022)
17. Guo, J., Lu, S., Zhang, W., Chen, F., Li, H., Liao, H.: Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20405–20415 (2025)
18. Hsieh, Y.H., Lai, S.H.: Csad: Unsupervised component segmentation for logical anomaly detection. In: 35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25–28, 2024. BMVA (2024)
19. Hu, T., Zhang, J., Yi, R., Du, Y., Chen, X., Liu, L., Wang, Y., Wang, C.: Anomaly-diffusion: Few-shot anomaly image generation with diffusion model. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 8526–8534 (2024)
20. Huang, Y., Qiu, C., Yuan, K.: Surface defect saliency of magnetic tile. *The Visual Computer* **36**(1), 85–96 (2020)
21. Kim, S., An, S., Chikontwe, P., Kang, M., Adeli, E., Pohl, K.M., Park, S.H.: Few shot part segmentation reveals compositional logic for industrial anomaly detection. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 8591–8599 (2024)
22. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
23. Lee, J.C., Kim, T., Park, E., Woo, S.S., Ko, J.H.: Continuous memory representation for anomaly detection. In: European Conference on Computer Vision. pp. 438–454. Springer (2024)
24. Lei, J., Hu, X., Wang, Y., Liu, D.: Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14143–14152 (2023)
25. Li, C.L., Sohn, K., Yoon, J., Pfister, T.: Cutpaste: Self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9664–9674 (2021)

26. Liao, J., Xu, X., Su, Y., Tu, R.C., Liu, Y., Tao, D., Yang, X.: Robust distribution alignment for industrial anomaly detection under distribution shift. *arXiv preprint arXiv:2503.14910* (2025)
27. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2980–2988 (2017)
28. Liu, J., Xie, G., Wang, J., Li, S., Wang, C., Zheng, F., Jin, Y.: Deep industrial image anomaly detection: A survey. *Machine Intelligence Research* **21**(1), 104–135 (2024)
29. Luo, W., Cao, Y., Yao, H., Zhang, X., Lou, J., Cheng, Y., Shen, W., Yu, W.: Exploring intrinsic normal prototypes within a single image for universal anomaly detection. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 9974–9983 (2025)
30. Peng, Y., Lin, X., Ma, N., Du, J., Liu, C., Liu, C., Chen, Q.: Sam-lad: Segment anything model meets zero-shot logic anomaly detection. *Knowledge-Based Systems* **314**, 113176 (2025)
31. Peng, Y., Lin, X., Ma, N., Liu, C., Chen, Q.: Vllm-lad: Visual large language model for zero-shot logical anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology* (2026)
32. Rippel, O., Mertens, P., Merhof, D.: Modeling the distribution of normal data in pre-trained deep features for anomaly detection. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. pp. 6726–6733. IEEE (2021)
33. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 14318–14328 (2022)
34. Schlüter, H.M., Tan, J., Hou, B., Kainz, B.: Natural synthetic anomalies for self-supervised anomaly detection and localization. In: *European Conference on Computer Vision*. pp. 474–489. Springer (2022)
35. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Kholidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. *arXiv preprint arXiv:2508.10104* (2025)
36. Sugawara, S., Imamura, R.: Puad: Frustratingly simple method for robust anomaly detection. In: *2024 IEEE International Conference on Image Processing (ICIP)*. pp. 842–848. IEEE (2024)
37. Tao, X., Gong, X., Zhang, X., Yan, S., Adak, C.: Deep learning for unsupervised anomaly localization in industrial images: A survey. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–21 (2022)
38. Tien, T.D., Nguyen, A.T., Tran, N.H., Huy, T.D., Duong, S., Nguyen, C.D.T., Truong, S.Q.: Revisiting reverse distillation for anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 24511–24520 (2023)
39. Tong, X., Chang, Y., Zhao, Q., Yu, J., Wang, B., Lin, J., Lin, Y., Mai, X., Wang, H., Tao, Z., et al.: Component-aware unsupervised logical anomaly generation for industrial anomaly detection. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 16722–16729. IEEE (2025)
40. Wang, C., Zhu, W., Gao, B.B., Gan, Z., Zhang, J., Gu, Z., Qian, S., Chen, M., Ma, L.: Real-iad: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 22883–22892 (2024)

41. Wei, S., Jiang, J., Xu, X.: Uninet: A contrastive learning-guided unified framework with feature selection for anomaly detection. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 9994–10003 (2025)
42. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 650–656 (2022)
43. Xiang, T., Zhang, Y., Lu, Y., Yuille, A.L., Zhang, C., Cai, W., Zhou, Z.: Squid: Deep feature in-painting for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 23890–23901 (2023)
44. Yao, H., Yu, W., Luo, W., Qiang, Z., Luo, D., Zhang, X.: Learning global-local correspondence with semantic bottleneck for logical anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology* **34**(5), 3589–3605 (2023)
45. You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., Le, X.: A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems* **35**, 4571–4584 (2022)
46. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677* (2021)
47. Zavrtanik, V., Kristan, M., Skočaj, D.: Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 8330–8339 (2021)
48. Zavrtanik, V., Kristan, M., Skočaj, D.: Dsr—a dual subspace re-projection network for surface anomaly detection. In: *European conference on computer vision*. pp. 539–554. Springer (2022)
49. Zhang, J., Chen, X., Wang, Y., Wang, C., Liu, Y., Li, X., Yang, M.H., Tao, D.: Exploring plain vit features for multi-class unsupervised visual anomaly detection. *Computer Vision and Image Understanding* **253**, 104308 (2025)
50. Zhang, J., Wang, G., Jin, Y., Huang, D.: Towards training-free anomaly detection with vision and language foundation models. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 15204–15213 (2025)
51. Zhang, X., Xu, M., Qiu, D., Yan, R., Lang, N., Zhou, X.: Mediclip: Adapting clip for few-shot medical image anomaly detection. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 458–468. Springer (2024)
52. Zhang, X., Xu, M., Zhou, X.: Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 16699–16708 (2024)
53. Zhang, X., Xu, M., Zhou, X.: Towards high-resolution industrial image anomaly detection. *arXiv preprint arXiv:2508.12931* (2025)
54. Zhao, Y.: Logical: Towards logical anomaly synthesis for unsupervised anomaly localization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4022–4031 (2024)
55. Zou, Y., Jeong, J., Pemula, L., Zhang, D., Dabeer, O.: Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In: *European conference on computer vision*. pp. 392–408. Springer (2022)