

UniADC: A Unified Framework for Anomaly Detection and Classification

Ximiao Zhang¹ Min Xu² Zheng Zhang¹ Junlin Hu³ Xiuzhuang Zhou¹*

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

² College of Information and Engineering, Capital Normal University

³ School of Software, Beihang University

2024010482@bupt.cn¹, {zhangzheng, xiuzhuang.zhou}@bupt.edu.cn¹

xumin@cnu.edu.cn², hujunlin@buaa.edu.cn³

Abstract

In this paper, we introduce the task of unified anomaly detection and classification, which aims to simultaneously detect anomalous regions in images and identify their specific categories. Existing methods typically treat anomaly detection and classification as separate tasks, thereby neglecting their inherent correlation, limiting information sharing, and resulting in suboptimal performance. To address this, we propose UniADC, a **Unified Anomaly Detection and Classification** model that can effectively perform both tasks with only a few or even no anomaly images. Specifically, UniADC consists of two key components: a training-free controllable inpainting network and a multi-task discriminator. The inpainting network can synthesize anomaly images of specific categories by repainting normal regions guided by anomaly priors, and can also repaint few-shot anomaly samples to augment the available anomaly data. The multi-task discriminator is then trained on these synthesized samples, enabling precise anomaly detection and classification by aligning fine-grained image features with anomaly-category embeddings. We conduct extensive experiments on three anomaly detection and classification datasets, including MVTEC-FS, MTD, and WFDD, and the results demonstrate that UniADC consistently outperforms existing methods in anomaly detection, localization, and classification. The code is available at <https://github.com/cnulab/UniADC>.

1. Introduction

Image anomaly detection aims to train models to detect and localize anomalous regions within images. It has numerous practical applications and has attracted growing research interest in recent years [5, 22, 37, 40]. Anomaly classification further categorizes detected anomalies, which

*Corresponding author.

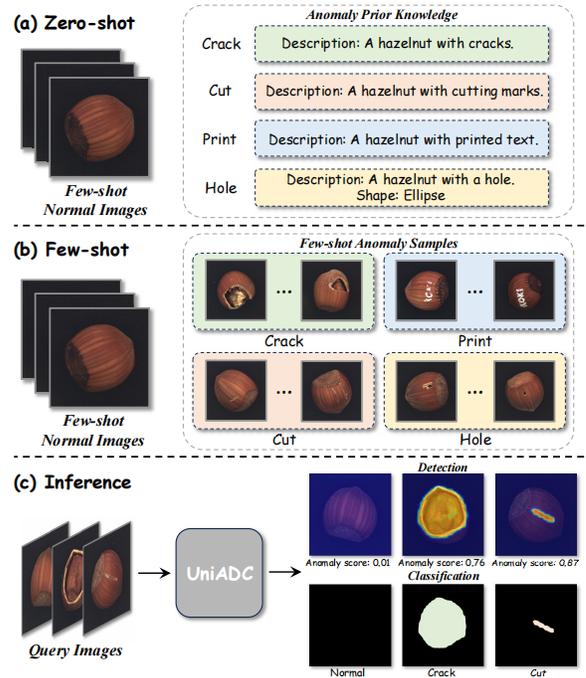


Figure 1. Task settings for unified anomaly detection and classification. (a) Zero-shot: training with a few normal samples and prior knowledge for each anomaly category. (b) Few-shot: training with a few normal samples and few-shot anomaly samples per category. (c) Inference: UniADC predicts anomaly scores and categories for query images.

aids in assessing anomaly severity and facilitating root-cause analysis. Existing studies [13, 23] typically treat anomaly classification as a downstream task of anomaly detection, in which the detected anomalous regions are cropped into sub-images and then fed into a separate classification model. This two-stage pipeline ignores the inherent connection between anomaly detection and classification, and hinders information sharing across the two tasks. Moreover, missed and over detections during the de-

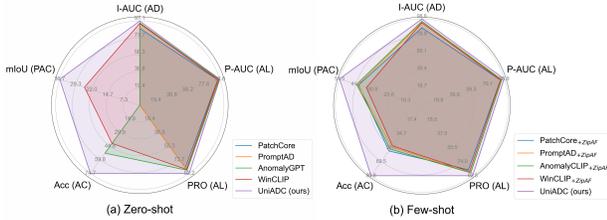


Figure 2. Comparison of UniADC with alternative methods on the MVTec-FS dataset. UniADC demonstrates superior performance over existing methods in Anomaly Detection (AD), Anomaly Localization (AL), Anomaly Classification (AC), and Pixel-level Anomaly Classification (PAC).

tection stage, as well as the difficulty in selecting appropriate anomaly score thresholds, pose a series of challenges for subsequent classification. As a result, such methods often suffer from high implementation complexity and suboptimal performance, limiting their practicality in real-world scenarios.

In this paper, we integrate anomaly detection and classification tasks and address them with a unified model. We propose zero-shot and few-shot settings for unified anomaly detection and classification, both extending the widely studied few-shot anomaly detection task [8, 14, 20, 36]. Few-shot anomaly detection provides only a small number of normal samples for model training, thereby reducing data collection costs, and has been widely applied in industrial [8, 14, 20] and medical [11, 45] fields. This setting is highly aligned with anomaly classification tasks in which available data is limited. Building upon this, the zero-shot anomaly detection and classification task does not provide additional anomaly samples for model training, but instead offers prior knowledge about anomaly categories, such as textual descriptions or shape information. This setting simulates real-world scenarios where collecting anomalous data is challenging. Differently, the few-shot anomaly detection and classification task provides a small number of anomaly samples for each category during model training, simulating practical scenarios where anomalous data is limited. Further explanation of the two settings is provided in Fig. 1.

To address the above tasks, we propose UniADC, a unified model for anomaly detection and classification, which tackles the challenge of scarce anomalous data through controllable image inpainting. It consists of two key components: a training-free controllable inpainting network and a multi-task discriminator. The controllable inpainting network consists of a pre-trained latent diffusion model [26] and an inpainting control network [44], enabling controllable inpainting guided by either anomaly priors or anomaly samples. Specifically, anomaly prior-guided controllable inpainting synthesizes category-specific anomalous images by repainting normal regions according to the provided prior knowledge of anomalies, whereas anomaly sample-

guided controllable inpainting refines and repaints few-shot anomalous samples to enhance data diversity. In addition, we propose a category consistency selection strategy to filter synthetic anomaly samples that are highly consistent with the target category. The multi-task discriminator is trained on these synthetic anomaly samples and leverages the powerful generalization capability of the existing vision-language models such as CLIP [25] or `dino.txt` [16, 30] to achieve end-to-end anomaly detection and classification. As shown in Fig. 2, UniADC comprehensively outperforms existing methods in anomaly detection, localization, and classification, particularly strengthening the reliability of defect categorization. Our main contributions are summarized as follows:

- We introduce the task of unified anomaly detection and classification, which has broad application prospects yet remains underexplored. To this end, we propose UniADC, which enables accurate anomaly detection and classification under both zero-shot and few-shot settings.
- We propose a controllable inpainting network that can generate category-specific anomaly samples conditioned on either anomaly prior or few-shot anomaly images, without additional training. This enables broad applicability across various anomaly detection and classification tasks, serving as an effective alternative to existing anomaly synthesis methods.
- Extensive experiments on the MVTec-FS [23], MTD [12], and WFDD [3] demonstrate the effectiveness of UniADC and highlight its potential for real-world anomaly detection and classification applications.

2. Related Work

Few-shot Anomaly Detection. Few-shot anomaly detection methods train models using only a small number of normal samples and often rely on the general knowledge of foundation models to overcome the challenges posed by limited data. WinCLIP [14] employs the CLIP model [25] and aligns image patches with normal and anomalous prompts via sliding windows, enabling anomaly detection and fine-grained anomaly localization. PromptAD [20] introduces prompt learning to enhance the alignment process and refine decision boundaries. AnomalyGPT [8] feeds the query images and their anomaly scores into an LLM for anomaly inspection, eliminating the dependency on decision thresholds. It is worth noting that some existing methods also incorporate certain prior knowledge of anomalies to improve anomaly detection. For example, PromptAD [20] employs textual descriptions of each anomaly category to improve detection accuracy. AnomalyGPT [8] pre-defines the expected appearance of query images and potential anomaly types. AnoGen [9] constrains the size range of synthesized anomalies according to their types. In contrast, the proposed UniADC leverages anomaly prior knowledge

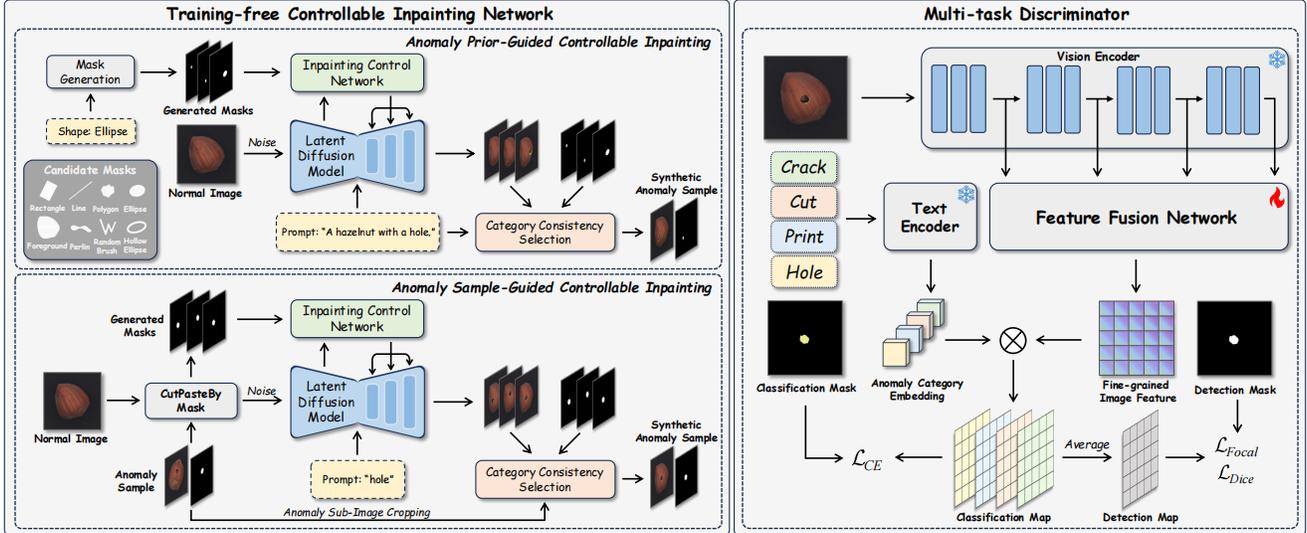


Figure 3. Overview of the proposed UniADC pipeline, which consists of a training-free controllable inpainting network and a multi-task discriminator. The controllable inpainting network supports two modes: anomaly prior-guided controllable inpainting and anomaly sample-guided controllable inpainting, enabling the generation of category-specific anomaly samples under different settings. The multi-task discriminator aligns fine-grained image features with anomaly category embeddings for accurate anomaly detection and classification.

for both anomaly detection and classification, providing a functional extension to existing methods.

Anomaly Classification. Anomaly classification aims to identify the categories of detected anomalous regions, which remains highly challenging due to the scarcity of anomaly samples. ZipAF [23] employs the AlphaCLIP model [35] to extract region-contextual anomaly features and adopts a zero-initialized projection to align query features with cached anomaly features for few-shot anomaly classification. However, it does not incorporate the anomaly detection process, which may result in degraded performance in practical scenarios. AnomalyNCD [13] integrates novel class discovery methods into the existing anomaly detection pipeline for anomaly detection and clustering. Recent work [28] proposes a zero-shot multi-type anomaly detection method, MultiADS, which trains the model to learn cross-domain anomaly priors for anomaly classification. However, due to the lack of targeted training, this method struggles to identify domain-specific anomaly types, such as identifying highly similar defects within a specific product [12, 33, 43].

Anomaly Synthesis. Existing anomaly synthesis methods can generally be classified into zero-shot and few-shot approaches, depending on the availability of real anomaly data during the synthesis process [18]. Zero-shot approaches generate anomaly samples relying on predefined data augmentation rules [19, 29], noise injection [38, 46], or textual descriptions [34], without access to real anomaly samples. However, the synthesized anomalies may exhibit a distribution shift from real-world cases, limiting their ef-

fectiveness. In contrast, few-shot approaches [7, 10, 15, 32] aim to enhance data diversity by augmenting a limited number of real anomaly samples. To the best of our knowledge, UniADC is the first method capable of synthesizing anomalies under both zero-shot and few-shot settings, thereby significantly enhancing its applicability across a wide range of anomaly detection and classification tasks.

3. Method

In this section, we present our proposed anomaly detection and classification model, UniADC, with its overall pipeline illustrated in Fig. 3. We first provide the problem definition, then describe the main functionalities and implementation details of the controllable inpainting network, including anomaly prior-guided controllable inpainting and anomaly sample-guided controllable inpainting. Finally, we introduce the multi-task discriminator of UniADC.

3.1. Problem Definition

In few-shot anomaly detection, the model is provided with a support set of normal samples $\mathcal{D}^n = \{X_1, X_2, \dots, X_{K_n}\}$, where K_n denotes the number of normal samples per image class. Given a query image, the model predicts an image-level anomaly detection score $I_d \in [0, 1]$ and a pixel-level anomaly map $S_d \in [0, 1]^{H \times W}$, where H and W denote the image height and width, respectively. The zero-shot and few-shot anomaly detection and classification tasks respectively provide additional anomaly prior knowledge and an anomaly support set for model training. Specif-

ically, the anomaly support set is represented as $\mathcal{D}^a = \bigcup_{y \in \{1, 2, \dots, Y\}} \{(X_1^y, M_1^y), (X_2^y, M_2^y), \dots, (X_{K_a}^y, M_{K_a}^y)\}$, where Y denotes the number of anomaly categories, K_a is the number of samples per anomaly category, and M denotes the binary anomaly mask. The model is further trained to predict both an image-level classification result $I_c \in \{0, 1, \dots, Y\}$ and a pixel-level classification result $S_c \in \{0, 1, \dots, Y\}^{H \times W}$, where 0 indicates the normal class, and 1 to Y represent the Y anomaly categories.

3.2. Anomaly Prior-Guided Controllable Inpainting

This module synthesizes category-specific anomaly samples based on provided anomaly prior, making it suitable for zero-shot anomaly detection and classification tasks. The anomaly prior includes the shape, size, and textual description of each anomaly category. Shape and size are used to generate the anomaly mask, while the textual description serves as a prompt to guide the synthesis of the desired anomalous appearance. Specifically, we define eight types of candidate masks to simulate various anomaly shapes, including *Rectangle*, *Line*, *Polygon*, *Ellipse*, *Hollow Ellipse*, *Random Brush*, *Perlin Noise*, and *Foreground Mask*. Except for *Foreground Mask*, each type is generated at three scales: *Large*, *Medium*, and *Small*. We use specific mask types for mask generation based on the given anomaly shape and size. For example, we specify the shape of the anomaly category ‘‘Hole’’ as *Ellipse*, which will generate elliptical masks of arbitrary sizes. When both shape and size priors are unavailable for an anomaly category, all mask types are used for mask generation.

For a given normal image X and prior knowledge from anomaly category y , we first generate an anomaly mask M^y under the shape-size constraints. Then, the latent diffusion model encodes the image X and the text prompt p^y into latent variables $z = \mathbb{E}_v(X)$ and $z^p = \mathbb{E}_t(p^y)$, where $\mathbb{E}_v(\cdot)$ and $\mathbb{E}_t(\cdot)$ denote the vision and text encoders, respectively. We define the number of forward diffusion steps as $T' = T \cdot \gamma$, where T denotes the total number of diffusion steps in the original latent diffusion model, and $\gamma \in (0, 1]$ controls the noise strength. A larger γ injects more noise into the latent variable, resulting in synthesized anomalies that deviate further from the original image distribution and exhibit more visually prominent anomalous patterns. To ensure spatial consistency between synthesized anomalous and the anomaly mask, we use an inpainting control network to precisely control the locations of the synthesized anomalies. Specifically, at each time step $t \in \{1, 2, \dots, T'\}$, the inpainting control network ϕ predicts the conditioning variable $z^m = \phi(z_t, t, z, M^y)$ of noisy latent z_t for controlling the denoising process:

$$z_{t-1} = \Psi(z_t, t, z^p, z^m) \quad (1)$$

where Ψ denotes the latent diffusion model, and the conditioning variables z^p and z^m respectively control the semantics and spatial positions of the synthesized anomalies. The inpainting control network ensures that the denoised latent representation z_0 remains consistent with the original latent z outside the masked region, while repainting only the masked region to align with the prompt embedding z^p . Finally, the denoised latent representation z_0 is decoded by decoder $\mathbb{D}(\cdot)$ into the anomaly image $X^y = \mathbb{D}(z_0)$. In our implementation, we adopt Stable Diffusion v1.5 [26] as the latent diffusion model and BrushNet [17] as the inpainting control network. Both models are trained on high-quality and diverse data, enabling UniADC to generalize effectively across diverse anomaly synthesis scenarios.

However, the above synthesis process does not ensure the semantic alignment between anomaly masks and their corresponding textual descriptions, which may lead to inconsistencies between the synthesized anomaly samples and their intended categories. For example, in the case of ‘‘a broken transistor lead’’, the synthesized anomaly should be located in the lead region, rather than randomly elsewhere. To this end, we propose a category consistency selection strategy to further filter synthetic anomalous images with high category consistency, thereby eliminating noisy samples introduced by anomaly synthesis. Given an anomaly category y , we first generate a mini-batch of synthesized samples using diverse anomaly masks, denoted as $\mathcal{S}^y = \{(X_b^y, M_b^y)\}_{b=1}^B$, where B is the mini-batch size. We then adopt the AlphaCLIP model [35] to evaluate the semantic consistency between the synthesized image-mask pairs and the textual description of the anomaly category. For the synthetic anomaly sample (X_b^y, M_b^y) , we calculate its category matching score as:

$$\mathcal{P}_b = \frac{\exp(\langle \psi_v(X_b^y, M_b^y), \psi_t(p^y) \rangle)}{\sum_{y \in \{1, 2, \dots, Y\}} \exp(\langle \psi_v(X_b^y, M_b^y), \psi_t(p^y) \rangle)} \quad (2)$$

where ψ_v and ψ_t denote the vision and text encoders of AlphaCLIP, respectively, and $\langle \cdot, \cdot \rangle$ denotes the inner product. The AlphaCLIP model aligns the image region features guided by the masks with the text embeddings. If the synthetic anomalous region fails to match the corresponding anomaly description, such as being placed in the wrong location, the matching score decreases. In the end, we select the synthetic anomaly sample with the highest matching score in the mini-batch \mathcal{S}^y for discriminator training.

3.3. Anomaly Sample-Guided Controllable Inpainting

Anomaly sample-guided controllable inpainting aims to enrich the diversity of anomalies by repainting few-shot anomaly samples, making it well-suited for few-shot anomaly detection and classification tasks. Given a normal image X and an anomalous sample (X^y, M^y) , we first

Table 1. Comparison of UniADC with alternative methods under zero-shot anomaly detection and classification settings. † denotes methods that use anomaly priors.

K_n	Method	MVTec-FS [23]					MTD [12]					WFDD [3]				
		I-AUC	P-AUC	PRO	Acc	mIoU	I-AUC	P-AUC	PRO	Acc	mIoU	I-AUC	P-AUC	PRO	Acc	mIoU
1	InCTRL [49]	92.36	-	-	-	-	70.34	-	-	-	-	96.66	-	-	-	-
	PatchCore [27]	84.76	93.77	82.44	-	-	68.22	73.44	59.50	-	-	84.81	96.07	70.62	-	-
	† PromptAD [20]	91.80	95.09	87.06	-	-	86.27	71.70	70.61	-	-	96.90	97.10	86.72	-	-
	† AnomalyGPT [8]	93.48	95.92	86.60	45.85	-	71.91	68.41	58.67	27.57	-	96.89	97.72	85.69	45.76	-
	† WinCLIP [14]	93.20	94.43	86.47	40.75	25.17	77.53	69.26	57.58	28.69	15.10	95.72	93.94	78.98	34.64	27.53
	† UniADC (CLIP)	95.03	96.33	88.75	66.07	32.90	86.19	80.80	77.69	55.82	27.92	97.01	99.00	87.03	86.70	45.55
	† UniADC (DINO)	96.37	96.11	89.16	68.30	35.06	90.09	82.65	79.98	59.22	29.58	98.08	99.46	92.13	88.88	52.17
2	InCTRL [49]	93.01	-	-	-	-	72.07	-	-	-	-	97.36	-	-	-	-
	PatchCore [27]	88.49	94.43	84.65	-	-	69.90	75.07	60.52	-	-	88.06	96.35	71.45	-	-
	† PromptAD [20]	93.95	95.42	87.93	-	-	87.06	73.53	69.91	-	-	97.15	97.20	86.81	-	-
	† AnomalyGPT [8]	94.91	96.24	87.97	52.65	-	72.57	70.20	60.12	32.56	-	97.48	97.78	85.75	47.39	-
	† WinCLIP [14]	94.37	94.60	86.95	41.93	25.34	78.07	71.57	57.70	29.90	15.28	96.50	94.16	80.19	35.83	27.62
	† UniADC (CLIP)	95.27	96.53	88.52	71.64	35.12	90.31	82.92	78.99	62.87	30.56	97.55	98.98	87.14	89.48	48.15
	† UniADC (DINO)	97.09	97.04	92.15	74.74	36.66	92.30	83.52	81.22	64.15	31.85	98.21	99.46	92.35	89.22	53.38
4	InCTRL [49]	93.62	-	-	-	-	73.23	-	-	-	-	97.38	-	-	-	-
	PatchCore [27]	90.75	95.28	86.56	-	-	71.40	75.35	59.67	-	-	88.26	97.21	71.41	-	-
	† PromptAD [20]	94.91	95.92	89.86	-	-	88.16	73.82	71.49	-	-	96.79	97.53	87.80	-	-
	† AnomalyGPT [8]	96.10	96.42	91.09	56.59	-	74.38	68.68	59.92	42.48	-	98.00	97.92	86.35	48.24	-
	† WinCLIP [14]	95.17	94.98	87.67	42.69	25.70	78.90	69.84	58.11	29.91	16.17	96.74	94.58	80.28	34.99	27.79
	† UniADC (CLIP)	96.18	96.47	91.17	73.33	35.14	90.88	84.88	79.81	63.98	31.29	98.15	99.09	88.29	90.85	48.24
	† UniADC (DINO)	97.65	97.36	92.68	76.35	37.23	93.04	85.82	81.43	66.03	33.75	98.60	99.39	94.76	91.06	55.44

crop the anomalous region from X^y based on the mask M^y , and paste it at a random location on the normal image X , yielding a preliminary synthetic anomaly image \hat{X}^y and its corresponding mask M^y . To increase the shape variation, we apply identical data augmentations, such as affine transformations, to both the anomalous region and its mask. However, at this stage, the synthesized anomalies still suffer from limited diversity and suboptimal visual coherence. To improve diversity and image quality, we use the latent diffusion model along with an inpainting control network to repaint the pasted regions in \hat{X}^y . The repainting process is guided by simple text prompts (e.g., the name of the anomaly category) and follows the same diffusion and denoising procedures as described in anomaly prior-guided inpainting. The degree of diversity and deviation from the original anomaly is controlled by the noise factor γ . To ensure semantic consistency, we perform category consistency selection by computing the Structural Similarity Index (SSIM) [39] between the original and repainted anomalous sub-images. These sub-images include both the anomalous region and a fixed amount of surrounding normal context. We compute the SSIM-based category matching score for each sample and retain the highest-scoring synthetic anomaly sample for discriminator training.

3.4. Multi-task Discriminator

We train the multi-task discriminator on synthetic anomalous samples so that it learns to align fine-grained image features with anomaly category embeddings, enabling unified anomaly detection and classification. First, we represent each synthetic anomalous sample as a 4-tuple (X, y, M_d, M_c) , where X is the synthesized anomalous im-

age, y is the anomaly category label, and M_d and M_c respectively denote the anomaly detection mask and classification mask, for a pixel location (i, j) satisfying:

$$M_c(i, j) = \begin{cases} y, & M_d(i, j) = 1 \\ 0, & M_d(i, j) = 0 \end{cases} \quad (3)$$

We employ a pre-trained vision encoder to extract multi-scale image features $\{x_1, x_2, \dots, x_N\} = \varphi(X)$, and fuse these features into a fine-grained representation through a feature fusion network $f = \theta(x_1, x_2, \dots, x_N)$, where $f \in \mathbb{R}^{H' \times W' \times C}$. The fusion network follows a fully convolutional architecture that progressively fuses intermediate-layer features from high to low levels. We then employ a text encoder to embed the names of anomaly categories to obtain $\{g_1, g_2, \dots, g_Y\}$, where each $g_y \in \mathbb{R}^C$ denotes the embedding of the y -th anomaly category. The similarity between the anomaly category embedding g_y and the feature map f at a spatial location (h, w) is computed as:

$$s_y(h, w) = \sigma(\langle f(h, w), g_y \rangle / \epsilon) \quad (4)$$

where $\sigma(\cdot)$ denotes the Sigmoid function, and ϵ is a learnable scaling parameter. We compute the similarity score at each spatial location and upsample the resulting similarity matrix to the original image resolution, yielding a classification map $S_y \in (0, 1)^{H \times W}$, which indicates the likelihood of anomaly category y at each pixel. By computing classification maps for all categories, we obtain a set of classification maps $\{S_1, S_2, \dots, S_Y\}$. The pixel-level anomaly detection map S_d is then computed as the average of these classification maps: $S_d = \frac{1}{Y} \sum_{y=1}^Y S_y$. The maximum value of S_d is used as the image-level anomaly detection score I_d .

Table 2. Comparison of UniADC with alternative methods under few-shot anomaly detection and classification settings.

K_n	K_a	Method	MVTec-FS [23]					MTD [12]					WFDD [3]				
			I-AUC	P-AUC	PRO	Acc	mIoU	I-AUC	P-AUC	PRO	Acc	mIoU	I-AUC	P-AUC	PRO	Acc	mIoU
1	1	PatchCore+ZipAF [27]	84.76	93.77	82.44	52.10	37.30	68.22	73.44	59.50	28.26	20.33	84.81	96.07	70.62	61.75	39.10
		PromptAD+ZipAF [20]	92.74	95.72	88.06	49.12	39.57	86.85	72.30	70.87	31.79	28.92	97.47	97.51	86.23	64.11	44.60
		WinCLIP+ZipAF [14]	93.20	94.43	86.47	49.60	34.39	77.53	69.26	57.58	33.78	24.23	95.72	93.94	78.98	59.59	37.30
		AnomalyCLIP+ZipAF [48]	95.59	96.00	87.98	52.60	40.25	76.18	75.98	69.54	25.61	21.78	93.76	98.59	87.58	76.48	45.99
		UniADC (CLIP)	97.69	98.53	89.72	84.72	48.62	88.49	82.41	77.41	60.81	34.80	98.50	98.83	92.96	93.51	48.76
		UniADC (DINO)	98.42	98.96	92.26	86.74	51.28	91.41	82.88	80.07	62.47	32.21	99.85	99.37	94.07	96.10	50.53
2	1	PatchCore+ZipAF [27]	88.49	94.43	84.65	56.16	39.63	69.90	75.07	60.52	30.15	21.03	88.06	96.35	71.45	63.82	39.71
		PromptAD+ZipAF [20]	94.58	95.66	88.93	51.40	40.09	87.41	74.10	71.12	36.09	29.78	97.52	97.74	87.01	67.17	44.51
		WinCLIP+ZipAF [14]	94.37	94.60	86.95	49.99	34.77	78.07	71.57	57.70	33.31	24.54	96.50	94.16	80.19	60.21	37.28
		AnomalyCLIP+ZipAF [48]	95.94	96.12	88.77	53.80	40.90	76.42	76.11	70.62	26.57	21.17	94.96	98.75	87.80	77.99	46.49
		UniADC (CLIP)	98.35	98.63	90.64	85.20	48.93	90.70	84.49	80.71	63.78	35.58	98.97	99.16	94.64	93.66	49.71
		UniADC (DINO)	98.56	98.90	92.48	86.85	51.49	92.57	86.79	83.17	65.10	34.33	99.87	99.48	94.12	97.22	51.78
	2	PatchCore+ZipAF [27]	88.49	94.43	84.65	58.25	40.32	69.90	75.07	60.52	34.71	21.78	88.06	96.35	71.45	64.52	41.29
		PromptAD+ZipAF [20]	95.16	95.60	89.03	52.66	40.93	88.46	74.53	71.91	43.89	31.16	97.75	98.20	86.81	67.91	45.25
		WinCLIP+ZipAF [14]	94.37	94.60	86.95	51.76	35.25	78.07	71.57	57.70	37.36	27.14	96.50	94.16	80.19	61.64	37.69
		AnomalyCLIP+ZipAF [48]	95.93	96.47	89.12	52.93	42.49	76.72	76.32	73.90	35.57	24.81	95.25	98.73	90.38	81.17	47.74
		UniADC (CLIP)	98.75	98.88	93.31	89.88	54.93	93.15	86.20	83.36	71.34	37.05	99.14	99.34	94.85	94.90	51.06
		UniADC (DINO)	99.05	99.10	93.55	88.72	54.21	93.85	89.31	85.78	70.87	35.46	99.93	99.49	94.19	97.76	51.91
4	1	PatchCore+ZipAF [27]	90.75	95.28	86.56	60.32	39.80	71.40	75.35	59.67	31.58	21.92	88.26	97.21	71.41	64.17	42.93
		PromptAD+ZipAF [20]	96.49	96.10	90.02	55.21	41.66	87.78	73.82	71.49	38.36	30.13	97.57	97.83	87.80	70.59	44.87
		WinCLIP+ZipAF [14]	95.17	94.98	87.67	53.75	35.98	78.90	69.84	58.11	33.89	28.32	96.74	94.58	80.28	63.59	37.59
		AnomalyCLIP+ZipAF [48]	95.98	96.67	89.77	55.07	40.83	76.84	76.22	71.11	29.58	22.37	95.22	98.74	90.17	78.88	46.97
		UniADC (CLIP)	98.57	98.55	91.42	85.57	50.01	92.74	84.98	82.15	64.40	35.86	99.35	99.19	94.48	93.91	49.94
		UniADC (DINO)	98.70	98.91	93.35	86.83	52.37	93.32	87.12	83.83	66.59	34.87	99.87	99.45	94.59	97.67	51.80

Then, the pixel-level anomaly classification result at a pixel location (i, j) is:

$$S_c(i, j) = \begin{cases} \operatorname{argmax}_{y \in \{1, 2, \dots, Y\}} S_y(i, j), & S_d(i, j) \geq \tau \\ 0, & S_d(i, j) < \tau \end{cases} \quad (5)$$

where τ represents the anomaly score threshold. We use the anomaly category with the largest pixel area in S_c as the image-level anomaly classification result I_c . If all pixels in S_c are classified as normal, then I_c is set to 0. We optimize the feature fusion network using Binary Focal loss [21] and Dice loss [24] for anomaly detection, and Cross-Entropy loss for anomaly classification. The overall loss function is:

$$\mathcal{L} = \mathcal{L}_{Focal}(S_d, M_d) + \mathcal{L}_{Dice}(S_d, M_d) + \lambda \mathcal{L}_{CE}([S_1, S_2, \dots, S_Y], M_c) \quad (6)$$

where $[\cdot, \cdot]$ denotes the concatenation operation, and λ is a weight hyperparameter. The normal regions in M_c are ignored during the computation of \mathcal{L}_{CE} . Unlike prior works [14, 20], we discard normal prompts to avoid performance degradation caused by the severe class imbalance between normal pixels and per-category anomalous pixels.

4. Experiments

4.1. Experimental Setup

Datasets. We conduct extensive evaluations on three anomaly detection and classification datasets, including MVTec-FS [23], MTD [12], and WFDD [3]. The MVTec-FS dataset is introduced by ZipAF [23] as an extension of the MVTec-AD benchmark [1] for anomaly classification. It contains 15 types of industrial products, each with an average of four anomaly categories. The MTD dataset [12]

consists of 1,344 images of magnetic tiles, including five types of anomalies. The subtle differences among these anomalies make the classification task particularly challenging. The WFDD dataset [3] contains 4,101 images across four types of fabrics. We reclassified the dataset by grouping defects of the same type, resulting in an average of three anomaly types per fabric. More details of the WFDD dataset [3] are provided in the **Appendix**. Some popular anomaly detection datasets, such as VisA [50], are excluded from this work due to the lack of anomaly category labels.

Implementation Details. We adopt a consistent experimental setup for both anomaly prior-guided and anomaly sample-guided inpainting. Specifically, we use the DDIM sampler [31] with 1,000 original diffusion steps. The noise factor γ is uniformly sampled between 0.4 and 0.6, and the number of accelerated sampling steps is set to 10. For mask generation, we leverage BiRefNet [47] for binary segmentation to ensure that the mask is located within the foreground region. We set the mini-batch size to 32 for category consistency selection, and generate 16 samples with a resolution of 512×512 per anomaly category for discriminator training. For the multi-task discriminator, we adopt two experimental settings, namely **UniADC (CLIP)** and **UniADC (DINO)**, which use the CLIP ViT-L/14 [25] and the DINOv3-based `dino.txt` [30] as the vision-language backbone of the discriminator, respectively. We set both the anomaly score threshold τ and the loss weight λ to 0.5. The list of anomaly prior is provided in the **Appendix**.

Metrics. We use the Area Under the Receiver Operating Characteristic Curve (I-AUC) metric to evaluate image-level anomaly detection performance. For pixel-level anomaly location, we use Pixel-AUROC (P-AUC) and Per Region Overlap (PRO) [2] metrics. In addition, we

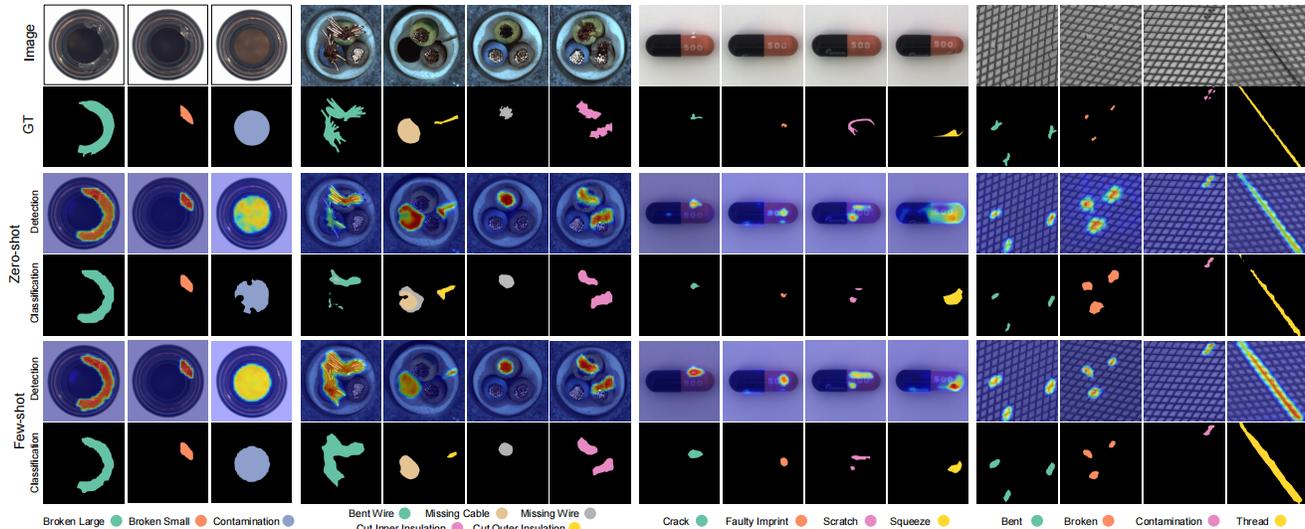


Figure 4. Qualitative results of UniADC on the MVTec-FS dataset under zero-shot ($K_n=2$) and few-shot ($K_n=2, K_a=1$) settings.

report the Top-1 classification accuracy (Acc) and Mean Intersection over Union (mIoU) to evaluate image-level and pixel-level anomaly classification performance, respectively. For the “combined” category in MVTec-FS dataset [23], we only evaluate its anomaly detection performance and exclude it from the anomaly classification metrics, as in previous works [13, 23].

Baselines. In the zero-shot setting, we adopt InCTRL [49], PatchCore [27], PromptAD [20], AnomalyGPT [8], and WinCLIP [14] as baseline methods. For AnomalyGPT [8], we cast classification as a single-choice QA task by prompting the model with candidate anomaly types. For WinCLIP [14], we provide it with anomaly category descriptions and compute the similarity between patch features and anomaly category embeddings, obtaining the anomaly classification results. In the few-shot setting, we combine anomaly detection methods such as PatchCore [27], PromptAD [20], WinCLIP [14], and AnomalyCLIP [48] with the anomaly classification method ZipAF [23] as our baseline methods. We employ the threshold selection strategy proposed in AnomalyNCD [13] to determine the anomaly score threshold, and fine-tune PromptAD [20] and AnomalyCLIP [48] using few-shot anomaly samples to enhance their anomaly detection performance.

4.2. Experimental Results

Tables 1 and 2 report the anomaly detection and classification results under zero-shot and few-shot settings, respectively. In the zero-shot setting, most baselines struggle to identify specific anomaly categories due to the absence of real anomaly samples. Although some methods incorporate Vision-Language Models (VLMs) or Large Language Models (LLMs) with anomaly priors, they still fail to achieve satisfactory anomaly detection and classification perfor-

mance. In contrast, UniADC fully exploits the relevance between anomaly detection and classification, thereby achieving promising performance in both tasks. Compared to other methods, it achieves approximately a 20% improvement in classification accuracy and a 10% improvement in mIoU. Moreover, **UniADC (DINO)** achieves superior detection and classification accuracy over **UniADC (CLIP)**, highlighting the remarkable performance of DINOv3 [30] in extracting fine-grained image features. In the few-shot setting, UniADC delivers substantial performance gains, even with only one anomaly sample per category, demonstrating its low cost and effective data utilization. Furthermore, it markedly outperforms other two-stage methods, showcasing the benefit of unifying anomaly detection and classification. Qualitative results in Fig. 4 further illustrate its ability to accurately localize and classify anomalous regions under various settings, emphasizing its strong generalization capability and practical applicability.

In addition, UniADC can be extended to the setting of full-shot normal samples, with the results provided in Table 3. We use the unsupervised anomaly detection methods PatchCore [27], RD4AD [4], and RealNet [46], as well as the semi-supervised method BGAD [41], in combination with the anomaly classification method ZipAF [23] as our baselines. When sufficient normal samples are available, UniADC can achieve anomaly detection performance comparable to mainstream full-shot methods. Moreover, it significantly improves anomaly classification performance, addressing the limitations of existing approaches in this aspect and demonstrating its unique practical value.

Fig. 5 presents synthetic anomaly samples generated by UniADC. By incorporating an inpainting control network and reliable category consistency selection strategies,

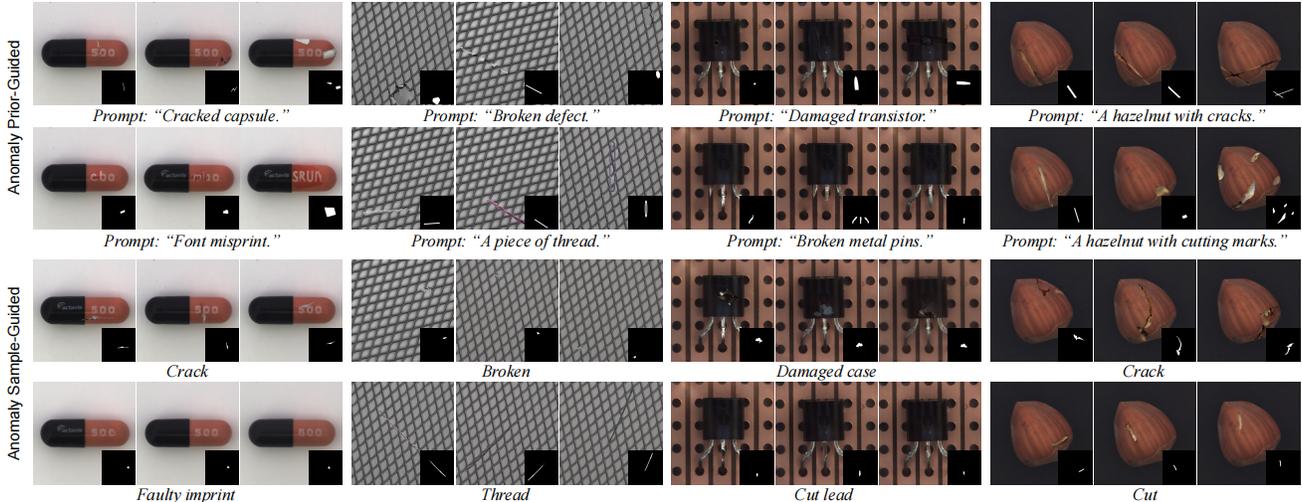


Figure 5. Examples of synthetic anomaly samples generated by UniADC on the MVTec-FS dataset.

Table 3. Comparison of UniADC with alternative methods on the MVTec-FS dataset using full-shot normal samples, with $K_a = 1$.

Method	Detection			Classification	
	I-AUC	P-AUC	PRO	Acc	mIoU
PatchCore+ZipAF [27]	98.77	98.76	95.17	69.50	39.60
RD4AD+ZipAF [4]	98.89	98.54	94.23	62.23	39.76
RealNet+ZipAF [46]	99.60	99.02	93.71	67.20	43.24
BGAD+ZipAF [41]	98.80	98.24	92.33	65.62	41.93
UniADC (CLIP)	98.85	98.73	93.76	87.01	51.59
UniADC (DINO)	99.41	98.95	94.06	87.86	51.93

UniADC can generate high-quality anomaly samples that are mask-consistent, category-aligned, and visually diverse. These properties enable UniADC to generalize well across a wide range of anomaly detection and classification scenarios. A comparison with alternative anomaly synthesis methods is provided in the **Appendix**.

4.3. Ablation Study

We conduct extensive experiments to evaluate the effectiveness of the UniADC components. By default, all ablation experiments use DINOv3 [30] as the backbone of the discriminator. Table 4 investigates the impact of anomaly classification loss and category consistency selection on the performance of UniADC. When the anomaly classification loss is removed, UniADC degenerates into a standard anomaly detection model. In the zero-shot setting, image-level anomaly detection performance improves slightly, while anomaly localization performance degrades. In the few-shot setting, all metrics decline, indicating that incorporating the anomaly classification task does not compromise anomaly detection but rather enhances the robustness and accuracy of anomaly detection. Moreover, Uni-

Table 4. Ablation results of UniADC on the MVTec-FS dataset. \mathcal{L}_{CE} denotes the Cross-Entropy loss, and CCS refers to the category consistency selection strategy.

Setting	Component	I-AUC	P-AUC	PRO	Acc	mIoU
Zero-shot ($K_n = 2$)	w/o \mathcal{L}_{CE}	97.76	96.23	87.10	-	-
	w/o CCS	95.19	96.02	89.26	66.16	33.15
	UniADC	97.09	97.04	92.15	74.74	36.66
Few-shot ($K_n = 2, K_a = 1$)	w/o \mathcal{L}_{CE}	97.76	98.33	86.26	-	-
	w/o CCS	96.84	98.08	88.56	79.91	45.37
	UniADC	98.56	98.90	92.48	86.85	51.49

ADC exhibits consistent performance improvement across all settings with the integration of category consistency selection, highlighting its critical role in enhancing the quality of synthesized anomaly samples. More ablation experiments and analysis can be found in the **Appendix**.

5. Conclusion

This paper integrates image anomaly detection and classification into a unified task and proposes UniADC, an innovative framework capable of anomaly synthesis, detection, and classification. UniADC consists of a training-free controllable inpainting network and a multi-task discriminator. The former can synthesize high-quality anomaly samples of specific categories guided by anomaly priors or anomaly samples, while the latter is trained on these synthesized anomaly samples, achieving unified anomaly detection and classification. Extensive experiments across multiple benchmarks validate the effectiveness and superiority of UniADC. In future work, we aim to further enhance the versatility of our approach by extending its application to diverse domains, such as medical image anomaly detection and classification.

References

- [1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec-ad: A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 6
- [2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4183–4192, 2020. 6
- [3] Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization. In *European Conference on Computer Vision*, pages 37–54. Springer, 2024. 2, 5, 6, 12, 16
- [4] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9737–9746, 2022. 7, 8
- [5] Jan Diers and Christian Pigorsch. A survey of methods for automated quality control based on images. *International Journal of Computer Vision*, 131(10):2553–2581, 2023. 1
- [6] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7388–7398, 2022. 15
- [7] Yuxuan Duan, Yan Hong, Li Niu, and Liqing Zhang. Few-shot defect image generation via defect-aware feature manipulation. In *Proceedings of the AAAI conference on artificial intelligence*, pages 571–578, 2023. 3
- [8] Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. Anomalygpt: Detecting industrial anomalies using large vision-language models. In *Proceedings of the AAAI conference on artificial intelligence*, pages 1932–1940, 2024. 2, 5, 7
- [9] Guan Gui, Bin-Bin Gao, Jun Liu, Chengjie Wang, and Yunsheng Wu. Few-shot anomaly-driven generation for anomaly classification and segmentation. In *European Conference on Computer Vision*, pages 210–226. Springer, 2024. 2, 12, 14
- [10] 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 *Proceedings of the AAAI conference on artificial intelligence*, pages 8526–8534, 2024. 3, 12, 14
- [11] Chaoqin Huang, Aofan Jiang, Jinghao Feng, Ya Zhang, Xinchao Wang, and Yanfeng Wang. Adapting visual-language models for generalizable anomaly detection in medical images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11375–11385, 2024. 2
- [12] Yibin Huang, Congying Qiu, and Kui Yuan. Surface defect saliency of magnetic tile. *The Visual Computer*, 36(1):85–96, 2020. 2, 3, 5, 6, 16
- [13] Ziming Huang, Xurui Li, Haotian Liu, Feng Xue, Yuzhe Wang, and Yu Zhou. Anomalyncd: Towards novel anomaly class discovery in industrial scenarios. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4755–4765, 2025. 1, 3, 7
- [14] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19606–19616, 2023. 2, 5, 6, 7, 15
- [15] Ying Jin, Jinlong Peng, Qingdong He, Teng Hu, Jiafu Wu, Hao Chen, Haoxuan Wang, Wenbing Zhu, Mingmin Chi, Jun Liu, et al. Dual-interrelated diffusion model for few-shot anomaly image generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 30420–30429, 2025. 3, 12, 14
- [16] Cijo Jose, Théo Moutakanni, Dahyun Kang, Federico Baldassarre, Timothée Darcet, Hu Xu, Daniel Li, Marc Szafraniec, Michaël Ramamonjisoa, Maxime Oquab, et al. Dinov2 meets text: A unified framework for image-and pixel-level vision-language alignment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 24905–24916, 2025. 2
- [17] Xuan Ju, Xian Liu, Xintao Wang, Yuxuan Bian, Ying Shan, and Qiang Xu. Brushnet: A plug-and-play image inpainting model with decomposed dual-branch diffusion. In *European Conference on Computer Vision*, pages 150–168. Springer, 2024. 4
- [18] Zhangyu Lai, Yilin Lu, Xinyang Li, Jianghang Lin, Yansong Qu, Liujuan Cao, Ming Li, and Rongrong Ji. Anomalypainter: Vision-language-diffusion synergy for zero-shot realistic and diverse industrial anomaly synthesis. *arXiv preprint arXiv:2503.07253*, 2025. 3
- [19] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9664–9674, 2021. 3, 12, 14
- [20] Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Promptad: Learning prompts with only normal samples for few-shot anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16838–16848, 2024. 2, 5, 6, 7, 15
- [21] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 6
- [22] 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. 1
- [23] Shuai Lyu, Rongchen Zhang, Zeqi Ma, Fangjian Liao, Dongmei Mo, and Waikung Wong. Mvrec: A general few-shot defect classification model using multi-view region-context. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5937–5945, 2025. 1, 2, 3, 5, 6, 7, 16
- [24] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric

- medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016. 6
- [25] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 2, 6
- [26] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 4
- [27] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2022. 5, 6, 7, 8
- [28] Ylli Sadikaj, Hongkuan Zhou, Lavdim Halilaj, Stefan Schmid, Steffen Staab, and Claudia Plant. Multiads: Defect-aware supervision for multi-type anomaly detection and segmentation in zero-shot learning. *arXiv preprint arXiv:2504.06740*, 2025. 3
- [29] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *European Conference on Computer Vision*, pages 474–489. Springer, 2022. 3, 12, 14
- [30] Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025. 2, 6, 7, 8
- [31] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. 6
- [32] Jaewoo Song, Daemin Park, Kanghyun Baek, Sangyub Lee, Jooyoung Choi, Eunji Kim, and Sungroh Yoon. Defectfill: Realistic defect generation with inpainting diffusion model for visual inspection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 18718–18727, 2025. 3
- [33] Kechen Song and Yunhui Yan. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, 285:858–864, 2013. 3
- [34] Han Sun, Yunkang Cao, Hao Dong, and Olga Fink. Unseen visual anomaly generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 25508–25517, 2025. 3, 12, 14
- [35] Zeyi Sun, Ye Fang, Tong Wu, Pan Zhang, Yuhang Zang, Shu Kong, Yuanjun Xiong, Dahua Lin, and Jiaqi Wang. Alphaclip: A clip model focusing on wherever you want. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13019–13029, 2024. 3, 4
- [36] Fenfang Tao, Guo-Sen Xie, Fang Zhao, and Xiangbo Shu. Kernel-aware graph prompt learning for few-shot anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7347–7355, 2025. 2
- [37] Xian Tao, Xinyi Gong, Xin Zhang, Shaohua Yan, and Chandranath Adak. Deep learning for unsupervised anomaly localization in industrial images: A survey. *IEEE Transactions on Instrumentation and Measurement*, 71:1–21, 2022. 1
- [38] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan Duong, Chanh D Tr Nguyen, and Steven QH Truong. Revisiting reverse distillation for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24511–24520, 2023. 3
- [39] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5
- [40] Xuan Xia, Xizhou Pan, Nan Li, Xing He, Lin Ma, Xiaoguang Zhang, and Ning Ding. Gan-based anomaly detection: A review. *Neurocomputing*, 493:497–535, 2022. 1
- [41] Xincheng Yao, Ruoqi Li, Jing Zhang, Jun Sun, and Chongyang Zhang. Explicit boundary guided semi-push-pull contrastive learning for supervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24490–24499, 2023. 7, 8
- [42] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draema: a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8330–8339, 2021. 12, 14, 16
- [43] Zhu Zhan, Jinfeng Zhou, and Bugao Xu. Fabric defect classification using prototypical network of few-shot learning algorithm. *Computers in Industry*, 138:103628, 2022. 3
- [44] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3836–3847, 2023. 2
- [45] Ximiao Zhang, Min Xu, Dehui Qiu, Ruixin Yan, Ning Lang, and Xiuzhuang Zhou. Medclip: Adapting clip for few-shot medical image anomaly detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 458–468. Springer, 2024. 2
- [46] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. 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*, pages 16699–16708, 2024. 3, 7, 8, 12, 14
- [47] Peng Zheng, Dehong Gao, Deng-Ping Fan, Li Liu, Jorma Laaksonen, Wanli Ouyang, and Nicu Sebe. Bilateral reference for high-resolution dichotomous image segmentation. *arXiv preprint arXiv:2401.03407*, 2024. 6
- [48] Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection. In *The Twelfth International Conference on Learning Representations*, 2024. 6, 7, 15

- [49] Jiawen Zhu and Guansong Pang. Toward generalist anomaly detection via in-context residual learning with few-shot sample prompts. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17826–17836, 2024. [5](#), [7](#)
- [50] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*, pages 392–408. Springer, 2022. [6](#)