

# IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection

Yanhui Li<sup>1</sup>, Yunkang Cao<sup>2</sup>, Chengliang Liu<sup>3</sup>, Yuan Xiong<sup>1</sup>, Xinghui Dong<sup>4</sup>, Chao Huang<sup>1\*</sup>

<sup>1</sup> School of Cyber Science and Technology, Sun Yat-sen University Shenzhen Campus

<sup>2</sup> School of Artificial Intelligence and Robotics, Hunan University

<sup>3</sup> Department of Computer and Information Science, University of Macau

<sup>4</sup> Faculty of Information Science and Engineering, Ocean University of China

liyh665@mail2.sysu.edu.cn, caoyunkang0207@gmail.com, liucl1996@163.com, xiongy89@mail.sysu.edu.cn

xinghui.dong@ouc.edu.cn, huangch253@mail.sysu.edu.cn

## Abstract

Industrial anomaly detection is a critical component of modern manufacturing, yet the scarcity of defective samples restricts traditional detection methods to scenario-specific applications. Although Vision-Language Models (VLMs) demonstrate significant advantages in generalization capabilities, their performance in industrial anomaly detection remains limited. To address this challenge, we propose IAD-R1, a universal post-training framework applicable to VLMs of different architectures and parameter scales, which substantially enhances their anomaly detection capabilities. IAD-R1 employs a two-stage training strategy: the Perception Activation Supervised Fine-Tuning (PA-SFT) stage utilizes a meticulously constructed high-quality Chain-of-Thought dataset (Expert-AD) for training, enhancing anomaly perception capabilities and establishing reasoning-to-answer correlations; the Structured Control Group Relative Policy Optimization (SC-GRPO) stage employs carefully designed reward functions to achieve a capability leap from “Anomaly Perception” to “Anomaly Interpretation”. Experimental results demonstrate that IAD-R1 achieves significant improvements across 7 VLMs, the largest improvement was on the DAGM dataset, with average accuracy 43.3% higher than the 0.5B baseline. Notably, the 0.5B parameter model trained with IAD-R1 surpasses commercial models including GPT-4.1 and Claude-Sonnet-4 in zero-shot settings, demonstrating the effectiveness and superiority of IAD-R1.

## Introduction

Industrial anomaly detection serves as a critical component of modern manufacturing quality control, facing challenges including diverse anomaly types, large inter-class variance, and scarce anomaly samples (Bae, Lee, and Kim 2023; Liu et al. 2024b; Roth et al. 2022; Jeong et al. 2023). The complexity and diversity of anomaly patterns make building universal detection models extremely challenging. Traditional methods primarily rely on hand-crafted feature extractors and domain-specific expert knowledge (You et al. 2022; He et al. 2024; Li et al. 2024d; Zhu and Pang 2024), but these approaches are often limited to single product categories, severely lacking generalization capabilities.

\*Corresponding Author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

In recent years, Vision-Language Models (VLMs) have provided new possibilities for solving industrial anomaly detection problems through their powerful multimodal understanding and generalization capabilities (Li et al. 2025a; Jin et al. 2025; Li et al. 2025b; Zeng et al. 2025; Jiang et al. 2024). Researchers have primarily adopted two strategies to apply VLMs to anomaly detection: (1) Using traditional anomaly detection methods as visual anomaly experts, feeding their generated anomaly localization results along with prompt text and test images into VLMs for judgment (Gu et al. 2024b; Li et al. 2023; Xu et al. 2025; Chen et al. 2025); (2) Based on question-answering anomaly detection datasets, directly adapting VLMs to industrial anomaly detection tasks through post-training techniques such as supervised fine-tuning or reinforcement learning (Chao et al. 2025; Zeng et al. 2025; Li et al. 2025b).

However, both strategies suffer from fundamental limitations. The anomaly expert-assisted approach has overall performance constrained by the capability ceiling of the selected anomaly expert, making it difficult to breakthrough the bottlenecks of traditional methods. End-to-end fine-tuning methods, while avoiding dependence on traditional algorithms, face deeper issues: existing training data lacks high-quality reasoning process annotations, leading models to learn only simple input-output mappings without mastering the intrinsic logic of anomaly analysis. More critically, traditional supervised fine-tuning easily leads to cognitive rigidity, while existing reinforcement learning methods, due to coarse reward design, frequently exhibit inconsistency between reasoning processes and final answers.

Successful experiences in general domains (Guo et al. 2025; Shen et al. 2025) demonstrate that high-quality chain-of-thought (CoT) (Wei et al. 2022) data and reinforcement learning strategies can improve model reasoning capabilities. However, the exploration of the R1-style methods in the anomaly detection domain remains insufficient, lacking high-quality CoT datasets for this task and specifically designed reinforcement learning objectives.

To address these issues, we propose IAD-R1, a two-stage post-training framework specifically designed for industrial anomaly detection, aimed at addressing the scarcity of high-quality CoT data and the lack of reasoning-answer consistency in existing models. As shown in Figure 1, we con-

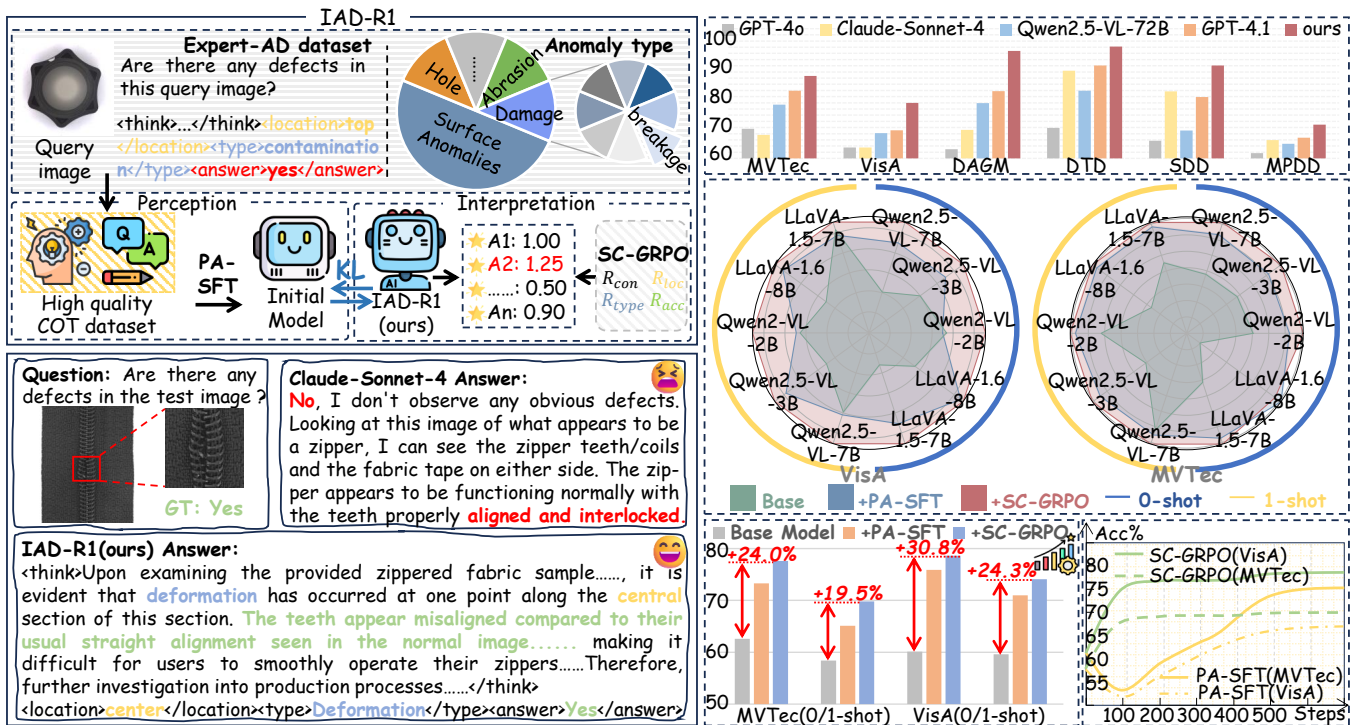


Figure 1: Overview of IAD-R1. The top left panel illustrates the composition of the Expert-AD and the two-stage training framework of IAD-R1. The bottom left panel shows an output example of IAD-R1 for anomaly detection. The right panels present quantitative analyses showcasing the performance of IAD-R1 across different model configurations and datasets.

struct the Expert-AD (totaling 5.9K QA pairs) dataset, the first industrial anomaly detection dataset containing high-quality CoT reasoning. Its CoT template follows a progressive three-layer approach of spatial perception, knowledge-driven analysis, and comprehensive decision-making to systematically simulate the expert anomaly detection process. Inspired by successful two-stage training experiences in DeepSeek-R1 (Guo et al. 2025), we propose the Perception Activation Supervised Fine-Tuning Strategy (PA-SFT) and Structured Control Group Relative Policy Optimization (SC-GRPO) strategies. Targeting the specificity of industrial anomaly detection, we design multi-dimensional reward functions:  $R_{acc}$ ,  $R_{loc}$ ,  $R_{type}$ , and  $R_{con}$  to achieve refined optimization of model reasoning processes, thereby enhancing the model’s anomaly understanding capabilities.

Extensive experimental results demonstrate that IAD-R1 can significantly improve anomaly detection performance across different backbones, achieving up to 43.3% average accuracy. Notably, small parameter models (0.5B) trained with IAD-R1 not only outperform larger baseline models (72B) in zero-shot settings but also surpass advanced commercial models such as GPT-4.1 (OpenAI 2025) and Claude-Sonnet-4 (Anthropic 2025), validating the superior parameter efficiency and performance advantages of our method. IAD-R1 effectively addresses the key problem of insufficient generalization capability in traditional methods, providing an efficient and universal solution for practical applications of VLMs in industrial anomaly detection. Our

main contributions are summarized as follows:

- We construct the first industrial anomaly detection dataset Expert-AD containing high-quality CoT reasoning, providing crucial data support for training.
- We propose IAD-R1, a two-stage post-training framework suitable for industrial anomaly detection, which guides models to achieve a capability leap from “Anomaly Perception” to “Anomaly Interpretation” through PA-SFT and SC-GRPO.
- Extensive experiments validate the effectiveness and universality of IAD-R1, providing an efficient solution for VLMs applications in industrial anomaly detection.

## Related Work

### Traditional Industrial Anomaly Detection

Due to the challenge of scarce anomaly samples in industrial anomaly detection, researchers utilize a small number of normal samples as auxiliary datasets for training and testing. WinCLIP (Jeong et al. 2023) pioneered the application of CLIP (Radford et al. 2021) to industrial anomaly detection by computing the similarity between handcrafted text prompts and test images to achieve defect detection. Addressing the complexity of manual prompt design, PromptAD (Li et al. 2024c) simplified text prompt design through semantic connections to construct negative samples and explicit anomaly boundaries. In contrast, One-to-Normal (Li

et al. 2024e) provided visual support for detection by performing one-to-one normal transformations on query images through an anomaly-free generative model. Considering the unpredictability of anomaly categories in production environments (Deng et al. 2023; Ma et al. 2025; Qu et al. 2024, 2025; Gu et al. 2024a), AnomalyCLIP (Zhou et al. 2023) achieved zero-shot anomaly detection by learning object-agnostic text prompts. AdaCLIP (Cao et al. 2024) further enhanced anomaly detection capabilities by learning dynamic prompts for both text and images.

## VLMs on Industrial Anomaly Detection

With the emergence of VLMs (Zhang et al. 2024; Yin et al. 2024; Wu et al. 2023), researchers have begun exploring the utilization of their powerful generalization capabilities for industrial anomaly detection (Li et al. 2025b; Chen et al. 2025; Jiang et al. 2024; Li et al. 2025a; Jin et al. 2025; Kwon et al. 2025; Zeng et al. 2025). AnomalyGPT (Gu et al. 2024b) adapted to anomaly detection tasks by generating training data through simulated anomaly images and fine-tuning the model using image decoders and prompt learners. However, due to the modality gap between textual and visual domains, Myriad (Li et al. 2023) further introduced anomaly maps generated by visual experts to guide the model’s attention toward anomalous regions. Addressing the requirements of zero-shot scenarios, Anomaly-OV (Xu et al. 2025) achieved more precise zero-shot anomaly detection through feature matching mechanisms that adaptively select and emphasize anomalous visual tokens.

## Methodology

### Problem Definition

In this study, we focus on zero-shot and one-shot scenarios for industrial anomaly detection, where training and testing adopt different data distribution settings. Specifically, our problem formulation is as follows.

**Training Phase.** We adopt a zero-shot training paradigm, where training data comes entirely from our constructed auxiliary dataset Expert-AD, containing no samples from the target test domain.

**Testing Phase.** We conduct evaluations under two settings on the target domain:

- Zero-shot testing: The model receives only text prompts  $p$  and the image to be detected  $I_{test}$  as input, outputting model reasoning and anomaly detection results.
- One-shot testing: In addition to text prompts  $p$  and test images  $I_{test}$ , the model additionally receives a normal reference image  $I_{ref}$  from the same product category as contextual information, with the output being model reasoning and anomaly detection results.

### Expert-AD Dataset

The field of industrial anomaly detection lacks high-quality prompt tuning datasets, severely hindering the application of VLMs in this downstream task. Although Anomaly-Instruct-125k proposed by Anomaly-OV (Xu et al. 2025) contains

a substantial amount of instruction tuning data, the web-crawled data differs significantly from real industrial images in practical scenarios. Moreover, these prompts only include image descriptions and answers, failing to help models develop a complete logic for anomaly inspection. To bridge this gap, we create Expert-AD, a high-quality real industrial scenario tuning dataset containing CoT reasoning.

The core innovation of Expert-AD lies in the construction of a systematic anomaly detection reasoning framework that implements structured anomaly detection reasoning through observation, comparison, identification, evaluation, and decision-making stages. This reasoning process comprises three core layers: the basic perception layer achieves precise localization of anomaly positions through spatial scanning and key component positioning; the knowledge-driven analysis layer combines industrial standard knowledge to identify anomaly types from dimensions of appearance integrity, surface quality, and structural integrity; the comprehensive decision layer provides final judgments on anomaly existence and evaluates their impact. Based on this reasoning framework, we adopt a two-stage training strategy: the PA-SFT stage trains anomaly perception capabilities through QA pairs containing detailed CoT, establishing mappings between structured reasoning processes and detection results; the SC-GRPO stage enhances the model’s analytical and decision-making accuracy in complex scenarios through fine-grained reward function design. In total, our Expert-AD comprises 2.9K QA pairs for PA-SFT stage and 3K QA pairs for SC-GRPO stage.

### IAD-R1

In this paper, we propose IAD-R1, a two-stage post-training framework specifically designed for industrial anomaly detection. As shown in Figure 2, IAD-R1 comprises two progressive stages: (1) PA-SFT stage, which employs supervised fine-tuning on Expert-AD CoT data to enable the model to acquire structured anomaly analysis thinking patterns and establish effective associations between reasoning processes and detection results; (2) SC-GRPO stage, which guides the model to break through simple pattern memorization, achieving more flexible and accurate anomaly analysis decisions through multi-dimensional reward functions.

**Perception Activation Supervised Fine-Tuning** In the PA-SFT stage, we activate the model’s anomaly perception potential through supervised fine-tuning on the Expert-AD dataset. Training samples adopt a triplet format  $(I, p, O)$ , where  $I$  represents the industrial image,  $p$  denotes the text prompt, and  $O$  is the conditional output sequence. Through high-quality CoT data training, the model learns to identify anomaly patterns from visual features and generates outputs with different structures based on image content. For normal images,  $O = (T, A)$ , containing the CoT reasoning process  $T$  and final answer  $A$ ; for anomalous images,  $O = (T, L, t, A)$ , containing the CoT reasoning process  $T$ , anomaly location  $L$ , anomaly type  $t$ , and final answer  $A$ . This training approach ensures that the model can produce logically coherent structured outputs when processing both normal and anomalous images, avoiding contradictions between the reasoning process and final conclusions. The training objective maximizes the conditional probability of gen-

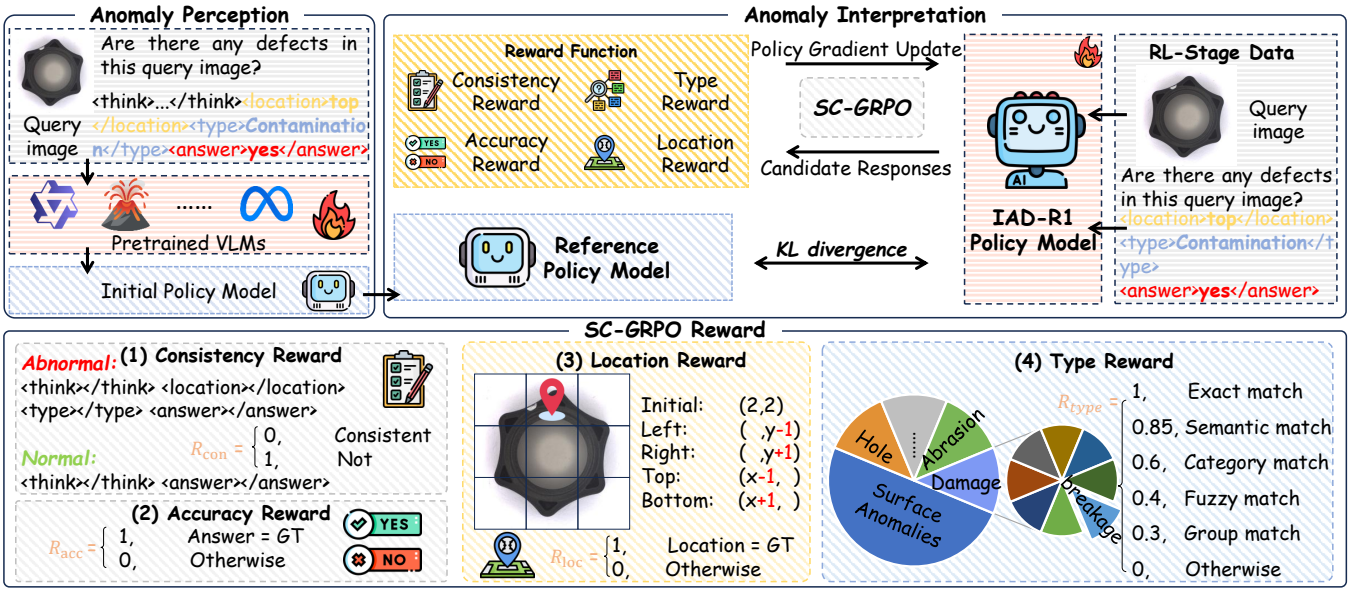


Figure 2: Architecture of IAD-R1. IAD-R1 employs a progressive two-stage training strategy: First, in the PA-SFT stage, supervised fine-tuning is conducted on pre-trained VLMs using CoT reasoning samples from the Expert-AD dataset to enhance the model’s anomaly perception capabilities and establish structured reasoning pathways; Subsequently, in the SC-GRPO stage, reward functions across four dimensions (consistency, accuracy, location, and type) are designed as reinforcement learning objectives to optimize the policy model, thereby improving its anomaly detection and understanding capabilities.

erating the output sequence given the image and prompt:

$$\mathcal{L}_{\text{PA-SFT}} = -\mathbb{E}_{(I,p,O) \sim \mathcal{D}_{\text{Expert-AD}}} \sum_{i=1}^L \log \pi_{\theta}(o_i | I, p, o_{<i>i-1</i>}), \quad (1)$$

where  $\mathcal{D}_{\text{Expert-AD}}$  is the Expert-AD dataset,  $o_i$  is the  $i$ -th token in the output sequence  $O$ ,  $L$  is the sequence length, and  $\pi_{\theta}$  denotes the parameterized model. The trained model serves as the initial policy model for the SC-GRPO stage.

### Structured Control Group Relative Policy Optimization

We utilize the model  $\pi_{\text{PA-SFT}}$ , obtained through PA-SFT training, as the initial policy model. Targeting the unique characteristics of industrial anomaly detection tasks, we develop an optimized SC-GRPO algorithm. Through carefully designed multi-dimensional reward functions, we achieve consistency between model reasoning and answers while enhancing overall detection accuracy.

**Multi-dimensional Reward Function Design.** The core innovation of SC-GRPO lies in the multi-dimensional reward function specifically designed for industrial anomaly detection tasks, comprising four key components.

**Consistency Reward Function.** The consistency reward function  $R_{\text{con}}$  enhances semantic consistency between model reasoning and detection conclusions by enforcing structured output formats that align with anomaly detection results. We define two standard output patterns corresponding to different anomaly states. For normal images, the output follows the Normal Pattern  $P_{\text{normal}}$ , containing only the reasoning process and final answer:

$$P_{\text{normal}} = \langle \text{think} \rangle \dots \langle \text{/think} \rangle \langle \text{answer} \rangle \dots \langle \text{/answer} \rangle.$$

For anomalous images, the output follows the Anomalous Pattern  $P_{\text{abnormal}}$ , which includes complete structured in-

formation to maintain reasoning consistency:

$$P_{\text{abnormal}} = \langle \text{think} \rangle \dots \langle \text{/think} \rangle \langle \text{location} \rangle \dots \langle \text{/location} \rangle \langle \text{type} \rangle \dots \langle \text{/type} \rangle \langle \text{answer} \rangle \dots \langle \text{/answer} \rangle.$$

Given model output  $o_i$  and corresponding ground truth label  $y_i$ , the consistency reward function is defined as:

$$R_{\text{con}}(o_i) = \begin{cases} 1, & \text{if Match}(o_i, P) \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where  $\text{Match}(o_i, P)$  is a pattern matching function that uses regular expressions to verify whether output  $o_i$  strictly conforms to the format requirements of pattern  $P$ .

**Answer Accuracy Reward Function.** The answer accuracy reward function  $R_{\text{acc}}$  is used to evaluate the correctness of the model’s final anomaly judgment and serves as the core reward signal in the reinforcement learning process. This function directly compares the model’s predicted answer and ground truth label. Given the model’s predicted answer  $a_{\text{pred}}$  and the ground truth answer label  $a_{\text{gt}}$ , the answer accuracy reward function is defined as:

$$R_{\text{acc}}(a_{\text{pred}}, a_{\text{gt}}) = \begin{cases} 1, & \text{if } a_{\text{pred}} = a_{\text{gt}} \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

This function ensures that rewards are only obtained when the model provides correct anomaly judgments (“Yes” or “No”), providing a prerequisite condition for subsequent location and type classification rewards.

**Location Accuracy Reward Function.** The location accuracy reward function  $R_{\text{loc}}$  is used to evaluate the accuracy of the model’s anomaly location. This function maps location description text to a standardized  $3 \times 3$  spatial grid we

established for comparison, converting textual descriptions into corresponding grid position numbers by parsing location keywords (such as “left,” “right,” “top,” “bottom,” etc.). Given the model’s predicted location description  $l_{pred}$  and the ground truth location label  $l_{gt}$ , the location accuracy reward function is defined as:

$$R_{loc}(l_{pred}, l_{gt}) = \begin{cases} 1, & \text{if } \Phi(l_{pred}) = \Phi(l_{gt}) \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where  $\Phi(\cdot)$  represents the text-to-grid position mapping function, which provides rewards when the predicted location matches the ground truth location.

**Type Accuracy Reward Function.** The type accuracy reward function  $R_{type}$  is used to evaluate the accuracy of the model’s anomaly type classification. Considering the diversity and semantic complexity of type descriptions in industrial anomaly detection, using only string matching strategies is overly restrictive and may lead to sparse rewards for the model during the reinforcement learning process, affecting training stability. Therefore, we design a multi-level matching mechanism to improve the stability of reinforcement learning and encourage the model to explore diverse output expressions. Given the matching degree between the predicted type  $t_{pred}$  and the ground truth type  $t_{gt}$ , the type accuracy reward function is defined as:

$$R_{type}(t_{pred}, t_{gt}) = \begin{cases} 1.0, & \text{if exact match} \\ 0.85, & \text{if semantic match} \\ 0.6, & \text{if category match} \\ 0.4, & \text{if fuzzy match} \\ 0.3, & \text{if group match} \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

**Relative Advantage Computation.** Our SC-GRPO strategy employs intra-group reward normalization to compute relative advantages between responses. This design avoids the training overhead of value function networks in Proximal Policy Optimization (Schulman et al. 2017) while maintaining optimization effectiveness:

$$A_i = \frac{R_{SC-GRPO}(o_i) - \text{mean}\{R_{SC-GRPO}(o_j)\}_{j=1}^G}{\text{std}\{R_{SC-GRPO}(o_j)\}_{j=1}^G}. \quad (6)$$

## Experiment

### Experiment Setup

**Evaluation Details.** To evaluate the performance of IAD-R1, we select six representative datasets encompassing two major categories: industrial objects (MVTec-AD (Bergmann et al. 2019), VisA (Zou et al. 2022), MPDD (Jia et al. 2022)) and surface textures (DAGM (Wieler and Hahn 2007), DTD (Aota, Tong, and Okatani 2023), SDD (Tabernik et al. 2020)), to simulate complex real-world industrial production scenarios. Considering that overall accuracy is prone to evaluation bias under data imbalance conditions, we adopt balanced accuracy as our evaluation metric. For IAD-R1 models, we extract answers from the structured output tags  $\langle answer \rangle \dots \langle /answer \rangle$ ; for baseline models, we directly obtain their output answers for comparison.

**Implementation Details.** To validate the generalizability of IAD-R1, We employ seven VLMs with diverse architectures

and sizes as backbones, including Qwen2-VL-2B (Wang et al. 2024), Qwen2.5-VL (3B, 7B) (Bai et al. 2025), LLaVA-1.5-7B (Liu et al. 2024a), LLaVA-OneVision-SI (0.5B, 7B) (Li et al. 2024a), and LLaVA-1.6-8B (Li et al. 2024b). All experiments are conducted on  $4 \times A100$  GPUs.

### Main Results

**Model Comparison.** As shown in Table 1, IAD-R1(LLaVA-OneVision-SI-7B) achieved the highest average accuracy of 86.1%, representing improvements of 7.2% and 7.8% over the best open-source model Anomaly-OV (Xu et al. 2025) and commercial model GPT-4.1 (OpenAI 2025), respectively. Notably, small-parameter models trained with IAD-R1 demonstrated exceptional performance: the IAD-R1(Qwen2.5-VL-Instruct-3B) model surpassed the 72B model of the same series by 4.5%, while the IAD-R1(LLaVA-OneVision-SI-0.5B) even outperformed all open-source and commercial models. These results indicate that IAD-R1 is more effective than simply increasing parameters, providing a parameter-efficient solution for industrial applications in resource-constrained environments.

### Ablation Results

**Ablation on Data Activation Method.** To verify the role of CoT responses on model performance, we compare different data preparation methods during the PA-SFT stage. As shown in Table 2, “Expert-AD” refers to fine-tuning with our Expert-AD dataset containing complete structured reasoning processes, and “Original” refers to fine-tuning with the same questions but using direct answers only. Using Expert-AD data in the PA-SFT stage significantly outperforms original data, with this result being validated across all three models. Notably, fine-tuning with original data even exhibits performance degradation in some cases, indicating that simple question-answer training may lead to inadequate learning or overfitting, while the CoT responses in Expert-AD can guide models to learn deeper anomaly analysis logic.

**Ablation on Reward Strategy.** To verify the importance of multi-dimensional reward functions, we compare SC-GRPO and the original reward strategy on three models fine-tuned with PA-SFT, where the original reward strategy employs only the correctness of final answers as the reward signal. As shown in Table 3, experimental results demonstrate that SC-GRPO significantly outperforms the original reward strategy across all models. The original reward strategy exhibits performance degradation primarily because its single reward design easily leads to sparse reward problems and cannot distinguish between “correct answers with clear reasoning” and “correct answers with chaotic reasoning”. In contrast, SC-GRPO provides more refined optimization guidance through multi-dimensional reward mechanisms, achieving more stable reinforcement learning optimization. Table 5 presents an ablation study of the contribution of each reward function in SC-GRPO. Each individual reward function makes a meaningful contribution to the overall performance. Furthermore, the integration of all reward functions achieves optimal performance across different datasets.

**Discussion about Grid Size.** Figure 3 demonstrates the impact of grid partition size in the location accuracy reward

Type	Model	Parameter	Industrial Workpieces			Surface Texture			Average
			MVTec	MPDD	VisA	DAGM	DTD	SDD	
Commercial	GPT-4o-mini	/	71.3	67.9	65.1	72.6	79.5	66.6	70.5
	GPT-4o	/	69.6	60.3	63.5	63.0	69.9	65.7	65.3
	GPT-4.1-nano	/	74.7	61.7	60.5	62.4	78.3	50.0	64.6
	GPT-4.1-mini	/	74.0	69.8	63.4	70.7	82.1	72.1	72.0
	GPT-4.1	/	81.9	66.7	69.1	81.8	90.1	79.9	78.3
	Claude-Sonnet-4	/	67.6	65.9	63.5	69.2	88.4	81.7	72.7
Open Source	LLaVA-OneVision-SI	0.5B	50.0	50.0	50.0	50.0	54.3	50.0	50.7
	Anomaly-OV[CVPR 2025]	0.5B	50.0	50.0	50.0	50.0	53.8	50.0	50.6
	Qwen2.5-VL-Instruct	3B	62.6	52.9	58.4	54.2	64.4	50.3	57.1
	AnomalyR1[arxiv 2025]	3B	69.4	56.0	59.8	56.7	61.0	57.6	60.1
	InternVL-2.5	4B	56.6	59.1	53.7	57.7	81.3	64.1	62.1
	Qwen2.5-VL-Instruct	7B	66.0	56.0	58.4	57.7	59.2	67.4	60.8
	AnomalyGPT[AAAI 2024]	7B	46.6	54.2	57.3	49.6	64.1	49.5	53.6
	LLaVA-OneVision-SI	7B	<u>82.0</u>	57.0	59.6	75.4	76.8	55.1	67.7
	Anomaly-OV[CVPR 2025]	7B	74.3	<u>70.3</u>	74.3	77.5	90.7	<u>88.7</u>	78.9
	LLaVA-1.5	13B	61.4	61.4	67.2	50.4	75.9	50.0	61.1
LLaVA-1.6	34B	53.7	50.0	53.9	50.0	52.7	50.0	51.7	
	Qwen2.5-VL-Instruct	72B	77.4	64.7	68.2	77.9	81.9	69.0	73.2
IAD-R1	IAD-R1(LLaVA-OneVision-SI)	0.5B	81.0	69.4	74.9	<u>93.3</u>	<u>95.5</u>	88.6	<u>83.8</u>
	IAD-R1(Qwen2.5-VL-Instruct)	3B	77.6	59.2	69.8	86.0	89.1	84.5	77.7
	IAD-R1(Qwen2.5-VL-Instruct)	7B	81.9	65.8	<u>75.4</u>	85.2	90.8	83.4	80.4
	IAD-R1(LLaVA-OneVision-SI)	7B	<b>86.7</b>	<b>70.9</b>	<b>78.0</b>	<b>94.8</b>	<b>96.2</b>	<b>90.1</b>	<b>86.1</b>

Table 1: Performance comparison of different models on industrial workpieces and surface texture benchmarks. The best results are highlighted in bold, while the second-best results are underlined.

Model	Data	0-shot	1-shot	Average
LLaVA-OneVision-SI (0.5B)	Base	50.7	50.0	50.4
	Original	50.9	50.0	50.5
	Expert-AD	<b>82.7</b>	<b>69.5</b>	<b>76.1</b>
Qwen2.5-VL-Instruct (3B)	Base	57.1	60.3	58.7
	Original	65.8	59.5	62.7
	Expert-AD	<b>73.1</b>	<b>73.5</b>	<b>73.3</b>
LLaVA-OneVision-SI (7B)	Base	67.7	60.6	64.2
	Original	68.4	62.3	65.4
	Expert-AD	<b>85.9</b>	<b>74.9</b>	<b>80.4</b>

Table 2: Results of different data preparation in PA-SFT.

function on model performance (left: LLaVA-OneVision-SI-0.5B; right: Qwen2-VL-Instruct-2B). The  $3 \times 3$  grid partition achieves optimal performance, providing sufficient anomaly localization details while avoiding noise interference from overly fine-grained partitioning.

**Model Promotion.** Table 4 provides a detailed presentation of the gains brought by IAD-R1 relative to the backbone across different test settings and datasets. The PA-SFT stage significantly activates the anomaly perception capabilities, while the SC-GRPO stage further optimizes output quality. Across the parameter range from 0.5B to 8B, all tested models in the Qwen and LLaVA series achieved significant performance improvements from the two-stage training, validating the broad applicability of IAD-R1 as a general industrial anomaly detection post-training framework. Notably, IAD-R1 demonstrates more pronounced improvement ef-

Model	Strategy	0-shot	1-shot	Average
LLaVA-OneVision-SI (0.5B)	Fine-tuned	82.7	<b>69.5</b>	76.1
	Original	78.1	59.0	68.6
	SC-GRPO	<b>83.8</b>	68.8	<b>76.3</b>
Qwen2.5-VL-Instruct (3B)	Fine-tuned	73.1	73.5	73.3
	Original	70.0	71.9	71.0
	SC-GRPO	<b>77.7</b>	<b>76.4</b>	<b>77.1</b>
LLaVA-OneVision-SI (7B)	Fine-tuned	85.9	74.9	80.4
	Original	78.5	65.8	72.2
	SC-GRPO	<b>86.1</b>	<b>75.8</b>	<b>81.0</b>

Table 3: Results of different reward strategy in SC-GRPO.

fects in 0-shot scenarios, which aligns with our training design using single-image data.

**Analysis of Reasoning and Answer Consistency.** Figure 4 illustrates the performance of different models. Qwen3 (Yang et al. 2025) provides incorrect answers, although Claude-Sonnet-4 (Anthropic 2025) produces the correct answer, it exhibits flawed reasoning processes. IAD-R1 not only delivers the correct answer but also ensures consistency between the reasoning process and the final answer.

## Conclusion

In this paper, we propose IAD-R1 to address key challenges faced by VLMs in industrial anomaly detection. We construct the Expert-AD dataset, contributing the first high-quality CoT reasoning data resource for this domain. Additionally, we design PA-SFT and SC-GRPO methods, achiev-

Model	Parameter	Strategy		0-shot						1-shot	
		PA-SFT	SC-GRPO	MVTec	DAGM	DTD	SDD	MPDD	VisA	MVTec	VisA
LLaVA-OneVision-SI	0.5B	✓		50.0	50.0	54.3	50.0	50.0	50.0	50.0	49.9
		✓	✓	79.4	91.3	<b>96.0</b>	<b>90.7</b>	65.3	73.4	<b>72.6</b>	66.4
Qwen2-VL-Instruct	2B	✓		63.5	54.3	59.3	57.2	55.0	59.6	69.9	59.6
		✓	✓	73.5	65.4	81.6	56.3	59.3	58.8	77.0	60.4
Qwen2.5-VL-Instruct	3B	✓		62.6	54.2	64.4	50.3	52.9	58.4	60.9	59.6
		✓	✓	73.3	82.8	87.7	76.3	53.4	65.1	75.9	71.0
LLaVA-1.5	7B	✓		66.0	57.7	59.2	67.4	56.0	58.4	81.0	69.8
		✓	✓	78.2	83.8	88.0	80.4	63.9	70.4	82.6	70.5
LLaVA-OneVision-SI	7B	✓		67.4	63.7	79.3	58.8	62.0	71.1	59.8	53.6
		✓	✓	78.1	69.6	84.0	<b>83.9</b>	60.9	69.8	76.1	67.2
LLaVA-1.6	8B	✓		82.0	75.4	76.8	55.1	57.0	59.6	67.5	53.7
		✓	✓	<b>86.7</b>	94.3	96.1	89.5	70.5	<b>78.0</b>	79.3	70.4
LLaVA-OneVision-SI	7B	✓		62.7	52.0	73.6	50.0	<b>66.2</b>	56.7	51.5	52.6
		✓	✓	83.1	73.4	80.0	69.4	<b>66.2</b>	68.2	<b>80.1</b>	63.7
LLaVA-1.6	8B	✓		84.7	<b>79.0</b>	<b>88.0</b>	<b>84.5</b>	64.1	<b>70.4</b>	78.0	<b>64.4</b>
		✓	✓	<b>84.7</b>	<b>79.0</b>	<b>88.0</b>	<b>84.5</b>	64.1	<b>70.4</b>	78.0	<b>64.4</b>

Table 4: Performance gains from using IAD-R1 on VLMs with different architectures and parameter scales.

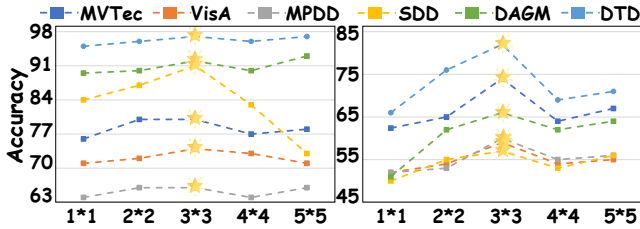


Figure 3: Ablation on grid size.

$R_{con}$	$R_{loc}$	$R_{type}$	MVTec	VisA	DAGM	MPDD	SDD	DTD
			72.7	59.0	72.5	55.5	79.3	81.0
	✓		76.3	66.3	83.9	57.9	77.4	88.4
		✓	77.4	67.0	84.8	57.9	81.6	88.1
✓	✓		75.4	67.0	<b>85.8</b>	58.6	79.5	<b>90.0</b>
✓		✓	76.7	67.5	85.1	56.5	80.8	89.1
✓	✓	✓	<b>77.6</b>	<b>69.8</b>	85.2	<b>59.2</b>	<b>83.4</b>	89.1

Table 5: Ablation of reward function contributions.

ing a leap from “Anomaly Perception” to “Anomaly Interpretation”. Experimental results demonstrate that IAD-R1 achieves significant performance improvements over baseline models across all benchmark datasets, providing a viable and effective technical pathway for advancing VLM applications in industrial anomaly detection scenarios.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No.62301621), Shenzhen Science and Technology Program (No. 20231121172359002,

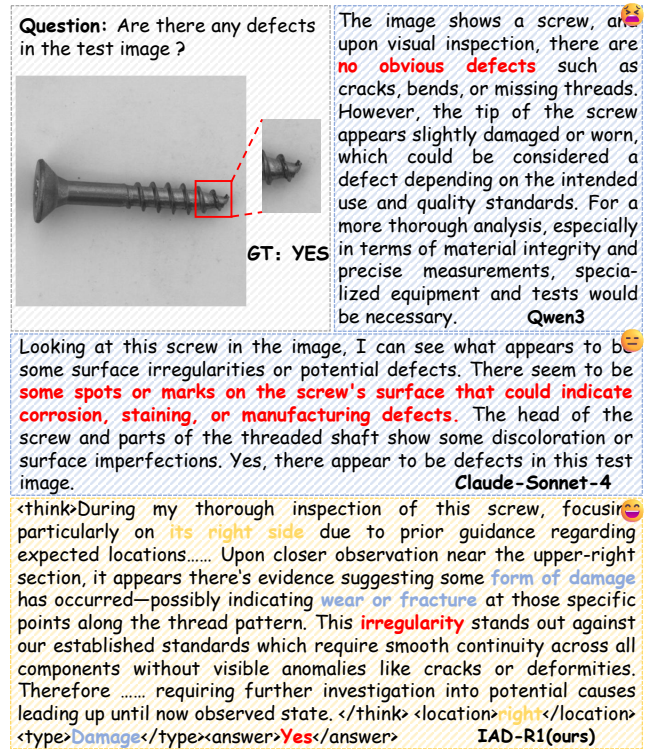


Figure 4: Comparison of model output.

2023A008), Shenzhen General Research Project (No. JCYJ20241202125904007), and Guangdong Basic and Applied Basic Research Foundation (No. 2025A1515011398).

## References

- Anthropic. 2025. Introducing Claude 4. <https://www.anthropic.com/news/claude-4>. Accessed: 2025-07-28.
- Aota, T.; Tong, L. T. T.; and Okatani, T. 2023. Zero-shot versus many-shot: Unsupervised texture anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 5564–5572.
- Bae, J.; Lee, J.-H.; and Kim, S. 2023. Phi: industrial anomaly detection using position and neighborhood information. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6373–6383.
- Bai, S.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; Song, S.; Dang, K.; Wang, P.; Wang, S.; Tang, J.; et al. 2025. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*.
- Bergmann, P.; Fauser, M.; Sattlegger, D.; and Steger, C. 2019. MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9592–9600.
- Cao, Y.; Zhang, J.; Frittoli, L.; Cheng, Y.; Shen, W.; and Boracchi, G. 2024. Adaclip: Adapting clip with hybrid learnable prompts for zero-shot anomaly detection. In *European Conference on Computer Vision*, 55–72. Springer.
- Chao, Y.; Liu, J.; Tang, J.; and Wu, G. 2025. Anomalyr1: A gppo-based end-to-end mllm for industrial anomaly detection. *arXiv preprint arXiv:2504.11914*.
- Chen, Z.; Chen, H.; Imani, M.; and Imani, F. 2025. Can multimodal large language models be guided to improve industrial anomaly detection? *arXiv preprint arXiv:2501.15795*.
- Deng, H.; Zhang, Z.; Bao, J.; and Li, X. 2023. Anovl: Adapting vision-language models for unified zero-shot anomaly localization. *arXiv preprint arXiv:2308.15939*, 2(5).
- Gu, Z.; Zhu, B.; Zhu, G.; Chen, Y.; Li, H.; Tang, M.; and Wang, J. 2024a. Filo: Zero-shot anomaly detection by fine-grained description and high-quality localization. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2041–2049.
- Gu, Z.; Zhu, B.; Zhu, G.; Chen, Y.; Tang, M.; and Wang, J. 2024b. Anomalygpt: Detecting industrial anomalies using large vision-language models. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 1932–1940.
- Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- He, H.; Zhang, J.; Chen, H.; Chen, X.; Li, Z.; Chen, X.; Wang, Y.; Wang, C.; and Xie, L. 2024. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 8472–8480.
- Jeong, J.; Zou, Y.; Kim, T.; Zhang, D.; Ravichandran, A.; and Dabeer, O. 2023. Winclip: Zero/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19606–19616.
- Jia, M.; Tang, L.; Chen, B.-C.; Cardie, C.; Belongie, S.; Hariharan, B.; and Lim, S.-N. 2022. Visual prompt tuning. In *European conference on computer vision*, 709–727. Springer.
- Jiang, X.; Li, J.; Deng, H.; Liu, Y.; Gao, B.-B.; Zhou, Y.; Li, J.; Wang, C.; and Zheng, F. 2024. Mmad: A comprehensive benchmark for multimodal large language models in industrial anomaly detection. *arXiv preprint arXiv:2410.09453*.
- Jin, E.; Feng, Q.; Mou, Y.; Lakemeyer, G.; Decker, S.; Simons, O.; and Stegmaier, J. 2025. Logicad: Explainable anomaly detection via vlm-based text feature extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 4129–4137.
- Kwon, Y.; Moon, D.; Oh, Y.; and Yoon, H. 2025. LogicQA: Logical Anomaly Detection with Vision Language Model Generated Questions. *arXiv preprint arXiv:2503.20252*.
- Li, B.; Zhang, Y.; Guo, D.; Zhang, R.; Li, F.; Zhang, H.; Zhang, K.; Zhang, P.; Li, Y.; Liu, Z.; et al. 2024a. Llava-onevision: Easy visual task transfer. *arXiv preprint arXiv:2408.03326*.
- Li, F.; Zhang, R.; Zhang, H.; Zhang, Y.; Li, B.; Li, W.; Ma, Z.; and Li, C. 2024b. Llava-next-interleave: Tackling multi-image, video, and 3d in large multimodal models. *arXiv preprint arXiv:2407.07895*.
- Li, W.; Chu, G.; Chen, J.; Xie, G.-S.; Shan, C.; and Zhao, F. 2025a. Lad-reasoner: Tiny multimodal models are good reasoners for logical anomaly detection. *arXiv preprint arXiv:2504.12749*.
- Li, X.; Zhang, Z.; Tan, X.; Chen, C.; Qu, Y.; Xie, Y.; and Ma, L. 2024c. 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*, 16838–16848.
- Li, Y.; Feng, Y.; Chen, B.; Chen, W.; Wang, Y.; Hu, X.; Qu, C.; Zhou, M.; et al. 2024d. Vague prototype-oriented diffusion model for multi-class anomaly detection. In *Forty-first International Conference on Machine Learning*.
- Li, Y.; Wang, H.; Yuan, S.; Liu, M.; Zhao, D.; Guo, Y.; Xu, C.; Shi, G.; and Zuo, W. 2023. Myriad: Large multimodal model by applying vision experts for industrial anomaly detection. *arXiv preprint arXiv:2310.19070*.
- Li, Y.; Yuan, S.; Wang, H.; Li, Q.; Liu, M.; Xu, C.; Shi, G.; and Zuo, W. 2025b. Triad: Empowering Imm-based anomaly detection with vision expert-guided visual tokenizer and manufacturing process. *arXiv preprint arXiv:2503.13184*.
- Li, Y.; Zhang, S.; Li, K.; and Lao, Q. 2024e. One-to-normal: Anomaly personalization for few-shot anomaly detection. *Advances in Neural Information Processing Systems*, 37: 78371–78393.
- Liu, H.; Li, C.; Li, Y.; and Lee, Y. J. 2024a. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 26296–26306.
- Liu, J.; Xie, G.; Wang, J.; Li, S.; Wang, C.; Zheng, F.; and Jin, Y. 2024b. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 21(1): 104–135.

- Ma, W.; Zhang, X.; Yao, Q.; Tang, F.; Wu, C.; Li, Y.; Yan, R.; Jiang, Z.; and Zhou, S. K. 2025. Aa-clip: Enhancing zero-shot anomaly detection via anomaly-aware clip. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 4744–4754.
- OpenAI. 2025. Introducing GPT-4.1 in the API. <https://openai.com/index/gpt-4-1/>. Accessed: 2025-07-28.
- Qu, Z.; Tao, X.; Gong, X.; Qu, S.; Chen, Q.; Zhang, Z.; Wang, X.; and Ding, G. 2025. Bayesian Prompt Flow Learning for Zero-Shot Anomaly Detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 30398–30408.
- Qu, Z.; Tao, X.; Prasad, M.; Shen, F.; Zhang, Z.; Gong, X.; and Ding, G. 2024. Vcp-clip: A visual context prompting model for zero-shot anomaly segmentation. In *European Conference on Computer Vision*, 301–317. Springer.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PmLR.
- Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; and Gehler, P. 2022. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14318–14328.
- Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Shen, H.; Liu, P.; Li, J.; Fang, C.; Ma, Y.; Liao, J.; Shen, Q.; Zhang, Z.; Zhao, K.; Zhang, Q.; et al. 2025. Vlm-r1: A stable and generalizable r1-style large vision-language model. *arXiv preprint arXiv:2504.07615*.
- Tabernik, D.; Šela, S.; Skvarč, J.; and Skočaj, D. 2020. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 31(3): 759–776.
- Wang, P.; Bai, S.; Tan, S.; Wang, S.; Fan, Z.; Bai, J.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; et al. 2024. Qwen2-vl: Enhancing vision-language model’s perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*.
- Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.
- Wieler, M.; and Hahn, T. 2007. Weakly supervised learning for industrial optical inspection. In *DAGM symposium in*, volume 6, 11.
- Wu, J.; Gan, W.; Chen, Z.; Wan, S.; and Yu, P. S. 2023. Multimodal large language models: A survey. In *2023 IEEE International Conference on Big Data (BigData)*, 2247–2256. IEEE.
- Xu, J.; Lo, S.-Y.; Safaei, B.; Patel, V. M.; and Dwivedi, I. 2025. Towards zero-shot anomaly detection and reasoning with multimodal large language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 20370–20382.
- Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; et al. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.
- Yin, S.; Fu, C.; Zhao, S.; Li, K.; Sun, X.; Xu, T.; and Chen, E. 2024. A survey on multimodal large language models. *National Science Review*, 11(12): nwae403.
- You, Z.; Cui, L.; Shen, Y.; Yang, K.; Lu, X.; Zheng, Y.; and Le, X. 2022. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35: 4571–4584.
- Zeng, P.; Pang, F.; Wang, Z.; and Yang, A. 2025. LR-IAD: Mask-Free Industrial Anomaly Detection with Logical Reasoning. *arXiv preprint arXiv:2504.19524*.
- Zhang, D.; Yu, Y.; Dong, J.; Li, C.; Su, D.; Chu, C.; and Yu, D. 2024. Mm-llms: Recent advances in multimodal large language models. *arXiv preprint arXiv:2401.13601*.
- Zhou, Q.; Pang, G.; Tian, Y.; He, S.; and Chen, J. 2023. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. *arXiv preprint arXiv:2310.18961*.
- Zhu, J.; and Pang, G. 2024. 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*, 17826–17836.
- Zou, Y.; Jeong, J.; Pemula, L.; Zhang, D.; and Dabeer, O. 2022. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European conference on computer vision*, 392–408. Springer.