Large Vision-Language Models for Industrial Anomaly Detection: A Comprehensive Survey

Ze Gao, Mengxue Wang

Abstract— **Industrial Anomaly Detection (IAD) plays a critical role in quality control and manufacturing efficiency across various industries. Recent advancements in Large Vision-Language Models (LVLMs) have introduced new paradigms for addressing the challenges in IAD, overcoming limitations of traditional approaches. This survey provides a comprehensive review of the integration of LVLMs with IAD, analyzing their evolution, methodologies, and applications. We systematically categorize existing approaches into three main frameworks: traditional anomaly detection, zero-shot methods, and LVLM-based solutions. We examine how these models leverage multimodal capabilities to enhance anomaly detection, reasoning, and explanation in industrial settings. Furthermore, we compare the performance of various methods across standard benchmarks, discuss current challenges, and highlight promising future research directions. Our findings indicate that LVLM-based approaches offer significant advantages in terms of flexibility, interpretability, and generalization capabilities, particularly in scenarios with limited anomaly samples and complex industrial environments.**

Index Terms—**Anomaly Detection, Large Multimodal Model, Vision Expert**

I. INTRODUCTION

Industrial Anomaly Detection (IAD) is a fundamental task in manufacturing that aims to identify defects or irregularities in industrial products, ensuring quality control and operational safety. Traditional IAD approaches typically rely on unsupervised or self-supervised learning techniques that model the normal data distribution and identify deviations as anomalies [1, 2]. These methods have been effective in controlled environments but face significant challenges in dynamic production settings where anomalies can manifest in various forms and severities.

The scarcity of anomaly samples in industrial settings presents a fundamental challenge for conventional approaches. Since manufacturing processes are designed to minimize defects, collecting sufficient anomalous data for training becomes impractical. Furthermore, traditional methods often struggle with the "one-class-one-model" paradigm, requiring separate models for each object category and limiting their applicability in flexible production environments [3].

Recent years have witnessed the emergence of Large Vision-Language Models (LVLMs) as a revolutionary force in artificial intelligence. Models such as GPT-4V, LLaVA, and MiniGPT-4 have demonstrated remarkable capabilities in understanding and reasoning across visual and textual modalities [4, 5]. These models are trained on massive datasets, endowing them with extensive knowledge and generalization abilities that can be leveraged for specialized tasks like IAD.

The integration of LVLMs with IAD represents a paradigm shift in addressing industrial quality control challenges. By combining the visual understanding capabilities of computer vision models with the reasoning and language generation abilities of large language models, LVLMs offer several advantages for IAD:

1,**Zero-shot and few-shot capabilities**: LVLMs can generalize to new object categories and anomaly types with minimal or no task-specific training.

2,**Interpretability**: Unlike black-box approaches, LVLMs can provide textual explanations for detected anomalies, enhancing human understanding and trust.

3,**Multimodal** reasoning: LVLMs can leverage both visual cues and domain knowledge represented in text to make more informed decisions.

4,**Flexibility**: The same model can handle multiple tasks beyond binary anomaly detection, such as severity assessment, root cause analysis, and recommendation generation.

Despite these advantages, applying LVLMs to IAD presents unique challenges. Industrial anomalies often manifest as subtle visual differences that require fine-grained perception and domain-specific knowledge. Additionally, industrial settings demand high precision and robustness, which may be challenging for general-purpose LVLMs without domain adaptation.

In this survey, we provide a comprehensive review of LVLM-based approaches for IAD, tracing their evolution from traditional methods to recent advancements. We systematically categorize existing approaches, analyze their strengths and limitations, and discuss future research directions. Our contributions include:

1,A systematic categorization of LVLM-based IAD approaches, highlighting their architectural designs and key innovations.

2,A comparative analysis of different methods across standard benchmarks and evaluation metrics.

3,An examination of current challenges and potential solutions in applying LVLMs to industrial settings.

4,A discussion of emerging trends and future research directions in this rapidly evolving field.

The remainder of this survey is organized as follows: Section 2 reviews related work in anomaly detection and LVLMs. Section 3 presents the evolution of IAD methods, from traditional approaches to LVLM-based solutions. Section 4 provides a detailed comparison of different methods across

Manuscript received May 07, 2025
Ze Gao, School of Software, Tiangong University, Tianjin, China
Mengxue Wang, School of Software, Tiangong University, Tianjin, China

multiple dimensions. Section 5 discusses challenges and future directions, and Section 6 concludes the survey.

II. Related Work

A. Traditional Industrial Anomaly Detection

Traditional IAD methods can be broadly categorized into reconstruction-based and feature embedding-based approaches. Reconstruction-based methods [6, 7] aim to reconstruct input images through encoder-decoder architectures and identify anomalies by measuring the reconstruction error. These approaches assume that the model will struggle to reconstruct anomalous regions, resulting in higher reconstruction errors for these areas.

Feature embedding-based methods [8, 9] extract representative features from normal samples and detect anomalies by measuring deviations in the feature space. These include one-class classification approaches [10], memory bank methods [11], and knowledge distillation techniques [12]. PatchCore [13], for instance, establishes a memory bank of patch embeddings from normal samples and detects anomalies by measuring the distance between test sample embeddings and their nearest normal embeddings in the memory bank.

While these methods have shown promising results on benchmark datasets like MVTec-AD [14], they typically follow a "one-class-one-model" paradigm, requiring abundant normal samples for each object category. This limits their applicability in dynamic production environments where new object categories frequently emerge.

B. Large Vision-Language Models

The development of Large Vision-Language Models (LVLMs) represents a significant advancement in multimodal learning. These models integrate visual and textual understanding capabilities, enabling them to perform tasks that require reasoning across modalities. Early works like CLIP [15] demonstrated the power of contrastive learning between images and text, aligning visual and textual representations in a shared embedding space.

More recent models such as BLIP-2 [16], MiniGPT-4 [17], and LLaVA [18] have further enhanced this integration by connecting visual encoders with large language models (LLMs) through trainable adapters. These models leverage the extensive knowledge and reasoning capabilities of LLMs while providing them with visual grounding.

The capabilities of LVLMs extend beyond simple image captioning to complex tasks such as visual question answering, image reasoning, and following visual instructions. Their ability to understand and generate human-like text based on visual inputs makes them particularly suitable for tasks requiring interpretation and explanation, such as industrial anomaly detection.

C. Zero-Shot and Few-Shot Learning for IAD

Zero-shot and few-shot learning approaches for IAD have gained significant attention as they address the fundamental challenge of limited anomaly samples. These methods leverage pre-trained models and transfer knowledge from related tasks or domains to detect anomalies with minimal or no task-specific training.

WinCLIP [19] pioneered the use of CLIP for zero-shot IAD, comparing image features to textual descriptions representing normal and anomalous states. AnomalyCLIP [20] further improved this approach by substituting manual templates with object-agnostic text vectors for more generic representation. These methods demonstrate that with appropriate prompting, pre-trained vision-language models can effectively discriminate between normal and anomalous samples without seeing anomalies during training.

Few-shot learning approaches for IAD aim to adapt models to new object categories or anomaly types with a minimal number of examples. Methods like RegAD [21] and Meta-AD [22] have shown promising results in this direction, enabling more flexible deployment in dynamic industrial environments.

III. Evolution of IAD Methods

The landscape of industrial anomaly detection has evolved significantly over the past decade, transitioning from traditional computer vision approaches to advanced multimodal methods leveraging large vision-language models. Figure 1 illustrates this evolution along a timeline, highlighting key milestones and technological advancements.

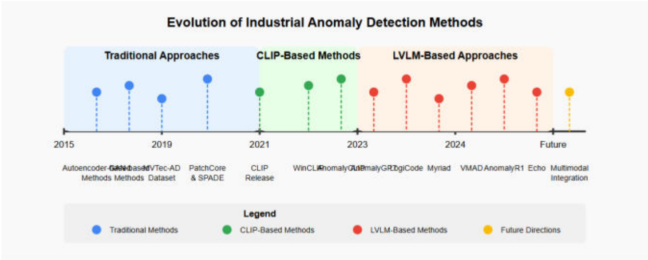

Figure 1: Evolution of Industrial Anomaly Detection Methods

A. Traditional Approaches (pre-2021)

Traditional IAD methods primarily relied on unsupervised or self-supervised learning techniques, focusing on modeling the distribution of normal samples and identifying deviations as anomalies. These methods can be categorized into two main streams:

Reconstruction-based methods like AE-SSIM [23], VAE [24], and GAN-based approaches [25] learn to reconstruct normal samples and identify anomalies through reconstruction errors. These methods typically employ encoder-decoder architectures to compress the input into a latent representation and then reconstruct it. The assumption is that the model will struggle to reconstruct anomalous regions, resulting in higher reconstruction errors for these areas.

Feature embedding-based methods such as PatchSVDD [26], SPADE [27], and PatchCore [13] extract representative features from normal samples and detect anomalies by measuring deviations in the feature space. These methods often leverage pre-trained backbone networks to extract discriminative features and employ various techniques to model the distribution of normal features.

While these approaches demonstrated effectiveness on benchmark datasets, they faced significant limitations in practical industrial settings. Most notably, they follow a "one-class-one-model" paradigm, requiring separate models for each object category and limiting their scalability in dynamic production environments.

B. CLIP-Based Zero-Shot Methods (2021-2023)

The introduction of CLIP [15] in 2021 marked a significant turning point in IAD, enabling zero-shot anomaly detection without requiring normal samples for each object category. CLIP-based approaches leverage the alignment between visual and textual representations learned through contrastive learning on large-scale image-text pairs.

WinCLIP [19] pioneered this approach by comparing image features to textual prompts describing normal and anomalous states. By computing the similarity between image features and these prompts, the model can determine whether a sample contains anomalies without seeing any anomalous examples during training.

AnomalyCLIP [20] further improved this approach by employing object-agnostic text vectors and implementing a combined global-local context optimization strategy. This enhanced the model's ability to capture normal and abnormal semantics across different object categories.

ClipSAM [28] integrated CLIP with the Segment Anything Model (SAM) to improve anomaly localization, demonstrating the potential of combining foundation models for more precise IAD.

These CLIP-based approaches marked a significant advancement in IAD, offering flexibility and generalization capabilities previously unattainable with traditional methods. However, they still operated within a relatively closed-world setting, performing binary classification with predefined prompts and struggling with novel or complex anomaly types.

C. LVLM-Based Approaches (2023-Present)

The most recent evolution in IAD is the integration of Large Vision-Language Models (LVLMs), which combine the visual understanding capabilities of computer vision models with the reasoning and language generation abilities of large language models. These approaches have further pushed the boundaries of IAD, enabling more complex reasoning, interpretation, and explanation of anomalies.

AnomalyGPT [29] was among the first to apply LVLMs to IAD, integrating an image decoder to provide fine-grained semantic understanding and employing a prompt learner to fine-tune the LVLM. This approach not only detected anomalies but also provided textual descriptions and supported multi-turn dialogues.

Myriad [30] enhanced this concept by introducing a modulation module and interactive corpus training strategy, embedding domain knowledge into the pre-trained model and effectively balancing query types related to defect knowledge and other information.

VMAD [31] further advanced LVLM-based IAD by incorporating a defect-sensitive structure learning scheme and a locality-enhanced token compression mechanism, improving the model's ability to discriminate anomalies and perceive fine-grained details.

More recent approaches like LogiCode [32], FabGPT [33], and AnomalyR1 [34] have explored different aspects of LVLM-based IAD, focusing on logical anomaly detection, domain-specific knowledge integration, and end-to-end training, respectively.

These LVLM-based approaches represent the current state-of-the-art in IAD, offering unprecedented flexibility, interpretability, and generalization capabilities. By leveraging the extensive knowledge and reasoning abilities of LVLMs, these methods can handle complex anomaly detection tasks and provide detailed explanations that enhance human understanding and trust.

IV. Comparison of LVLM-Based IAD Methods

In this section, we provide a comprehensive comparison of various LVLM-based IAD methods across multiple dimensions, including their architectural design, training approach, capabilities, and performance on standard benchmarks. Table 1 summarizes this comparison.

Method	Architecture	Training Approach	Zero-Shot	Anomaly Localization	Explanation
WinCLIP	CLIP + Window-based	Zero-shot	✓	✓	✗
Anomaly CLIP	CLIP + Object-agnostic prompts	Zero-shot	✓	✓	✗
Anomaly GPT	LVLM + Image decoder + Prompt learner	Fine-tuning	✓	✓	✓
LogiCode	LLM + Code generation	Instruction tuning	✓	✗	✓
Myriad	LVLM + Modulation module	Fine-tuning	✓	✓	✓
FabGPT	LVLM + Modal enhancement + Detection head	Fine-tuning	✓	✓	✓
VMAD	LVLM + DSSL + LTC	Fine-tuning	✓	✓	✓
EIAD	LVLM + Multi-modal defect localization	Fine-tuning	✓	✓	✓
Anomaly R1	VLM-R1 + GRPO + ROAM	GRPO fine-tuning	✓	✓	✓
Echo	Multi-expert LVLM framework	Retrieval-augmented	✓	✓	✓

Table 1: Comparison of LVLM-based IAD methods across various dimensions.

A. Architectural Design

LVLM-based IAD methods employ diverse architectural designs to integrate visual perception with language understanding and reasoning. We identify three main architectural patterns:

1,**End-to-end LVLM frameworks** like AnomalyR1 [34] directly leverage the multimodal capabilities of LVLMs for anomaly detection, fine-tuning the entire model for IAD tasks. These approaches benefit from simplified architectures but may require substantial computational resources for training.

2,**Hybrid frameworks** like AnomalyGPT [29], Myriad [30], and VMAD [31] combine LVLMs with specialized modules for anomaly detection and localization. These modules enhance the LVLM's ability to perceive fine-grained visual details and discriminate anomalies, addressing the limitations of general-purpose LVLMs in specialized industrial tasks.

3,**Multi-expert frameworks** like Echo [36] employ multiple specialized modules that work collaboratively to enhance the LVLM's performance in IAD. By separating different functionalities into distinct expert modules, these approaches achieve greater flexibility and can be adapted to various industrial settings with minimal modifications.

The choice of architecture significantly impacts the model's capabilities, computational requirements, and applicability to different industrial scenarios. End-to-end frameworks offer simplicity and potentially better integration between modalities, while hybrid and multi-expert frameworks provide greater flexibility and can leverage specialized domain knowledge.

B. *Training Approaches*

LVLM-based IAD methods employ various training approaches to adapt general-purpose LVLMs to industrial anomaly detection tasks. These approaches can be categorized into four main types:

1,**Zero-shot approaches** like WinCLIP [19] and AnomalyCLIP [20] leverage pre-trained models without any task-specific fine-tuning, relying on carefully designed prompts to guide the model's behavior. These approaches offer immediate applicability but may lack the precision required for specialized industrial tasks.

2,**Fine-tuning approaches** like AnomalyGPT [29], Myriad [30], and VMAD [31] adapt pre-trained LVLMs to IAD tasks through supervised learning on task-specific data. These approaches can achieve higher performance but require curated datasets and may struggle with generalization to unseen anomaly types.

3,**Reinforcement learning approaches** like AnomalyR1 [34] employ techniques such as Group Relative Policy Optimization (GRPO) and Reasoned Outcome Alignment Metric (ROAM) to optimize the model's performance on IAD tasks. These approaches can potentially achieve better alignment with human preferences and require fewer labeled examples.

4,**Retrieval-augmented approaches** like Echo [36] enhance the LVLM's reasoning by retrieving relevant information from external sources, such as normal reference images or domain-specific knowledge bases. These approaches can improve the model's context awareness and adaptability without requiring extensive retraining.

The choice of training approach depends on factors such as the availability of labeled data, computational resources, and the specific requirements of the industrial application. Zero-shot approaches offer immediate applicability but may lack precision, while fine-tuning and reinforcement learning approaches can achieve higher performance but require more resources and data.

C. *Capabilities and Performance*

LVLM-based IAD methods offer a wide range of capabilities beyond binary anomaly detection, differentiating them from traditional approaches. We identify five key capabilities that characterize these methods:

1,**Zero-shot detection**: The ability to detect anomalies in previously unseen object categories without task-specific training.

2,**Multi-turn dialogue**: The ability to engage in interactive conversations with users, answering follow-up questions and providing additional information upon request.

3,**Anomaly localization**: The ability to precisely locate anomalous regions within an image, enabling more targeted inspection and remediation.

4,**Explanation generation**: The ability to provide textual explanations for detected anomalies, enhancing human understanding and trust.

5,**Severity assessment**: The ability to evaluate the severity of detected anomalies, enabling more informed decision-making in industrial settings.

The performance of LVLM-based IAD methods on standard benchmarks like MVTec-AD [14] has shown significant improvement over time. Recent approaches like VMAD [31] and AnomalyR1 [34] achieve image-level AUC scores exceeding 96%, surpassing both traditional methods and earlier LVLM-based approaches. This improvement demonstrates the rapid advancement of the field and the potential of LVLMs for industrial anomaly detection.

V. Challenges and Future Directions

Despite the significant progress in LVLM-based IAD, several challenges remain to be addressed for their widespread adoption in industrial settings. In this section, we discuss these challenges and highlight promising future research directions.

A. *Domain-Specific Knowledge Integration*

While LVLMs possess extensive general knowledge, they often lack the specialized domain knowledge required for industrial anomaly detection. Future research should explore effective methods for integrating domain-specific knowledge into LVLMs, such as:

1,**Domain-specific fine-tuning datasets**: Developing comprehensive datasets that capture the diversity of industrial anomalies and their contextual information.

2,**Knowledge distillation from domain experts**: Extracting and transferring knowledge from human experts or specialized models to enhance the LVLM's understanding of industrial processes.

3,**Modular domain adaptation**: Creating adaptable components that can be integrated with LVLMs to provide domain-specific capabilities without compromising their general knowledge.

B. *Fine-Grained Visual Perception*

Industrial anomalies often manifest as subtle visual differences that require fine-grained perception. Enhancing the visual perception capabilities of LVLMs for industrial

settings is a crucial research direction, including:
1,**High-resolution image processing**: Developing efficient methods for LVLMs to process high-resolution industrial images without compromising performance.
2,**Attention mechanisms for fine details**: Designing specialized attention mechanisms that focus on subtle visual cues relevant to anomaly detection.
3,**Multi-scale feature integration**: Incorporating features from multiple spatial scales to capture both global context and local details.

C. Interpretability and Trustworthiness

In industrial settings, the interpretability and trustworthiness of anomaly detection systems are paramount. Future research should focus on:
1,**Explainable decision-making**: Enhancing the ability of LVLMs to provide clear, consistent, and accurate explanations for their anomaly detection decisions.
2,**Uncertainty quantification**: Developing methods for LVLMs to express uncertainty in their predictions, enabling more reliable decision-making.
3,**Human-AI collaboration**: Designing interfaces and interaction paradigms that facilitate effective collaboration between human operators and LVLM-based systems.

D. Efficiency and Resource Requirements

The computational resources required by LVLMs can be prohibitive for deployment in industrial settings with limited hardware capabilities. Future research should address:
1,**Model compression**: Developing techniques to reduce the size and computational requirements of LVLMs without significant performance degradation.
2,**Edge deployment**: Adapting LVLM-based IAD methods for deployment on edge devices with constrained resources.
3,**Incremental learning**: Enabling LVLMs to continuously learn and adapt to new anomaly types and object categories with minimal computational overhead.

E. Multimodal and Temporal Integration

Industrial anomalies may manifest across multiple modalities and time scales. Future research should explore:
1,**Multimodal fusion**: Integrating information from various sensors beyond visual data, such as thermal imaging, spectroscopy, and acoustic signals.
2,**Temporal reasoning**: Enhancing the ability of LVLMs to reason about anomalies that evolve over time or exhibit temporal patterns.
3,**Causal understanding**: Developing methods for LVLMs to understand and explain the causal factors underlying industrial anomalies.

VI. CONCLUSION

This survey has provided a comprehensive review of the integration of Large Vision-Language Models (LVLMs) with Industrial Anomaly Detection (IAD), tracing its evolution from traditional methods to recent advancements. We have systematically categorized existing approaches, analyzed their strengths and limitations, and discussed future research directions.

The emergence of LVLM-based IAD represents a paradigm shift in addressing industrial quality control challenges. By combining the visual understanding capabilities of computer vision models with the reasoning and language generation abilities of large language models, these approaches offer unprecedented flexibility, interpretability, and generalization capabilities. They can detect anomalies in previously unseen object categories, provide detailed explanations, engage in interactive dialogues, and adapt to dynamic industrial environments.

Despite these advancements, several challenges remain to be addressed for the widespread adoption of LVLM-based IAD in industrial settings. These include integrating domain-specific knowledge, enhancing fine-grained visual perception, ensuring interpretability and trustworthiness, reducing computational requirements, and incorporating multimodal and temporal information.

As the field continues to evolve, we anticipate further advancements that will bridge these gaps and enable more effective and efficient anomaly detection in industrial settings. The integration of LVLMs with IAD represents a promising direction for enhancing manufacturing quality control and operational safety, ultimately contributing to more reliable and efficient industrial processes.

REFERENCES

[1] Roth, K., et al. "PatchCore: Towards total recall in industrial anomaly detection." CVPR, 2022.
[2] Defard, T., et al. "PaDiM: a patch distribution modeling framework for anomaly detection and localization." ICPR, 2020.
[3] Bergmann, P., et al. "The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection." IJCV, 2019.
[4] Liu, H., et al. "Visual instruction tuning." NeurIPS, 2023.
[5] Zhu, Y., et al. "MiniGPT-4: Enhancing vision-language understanding with advanced large language models." arXiv preprint, 2023.
[6] Zimmerer, D., et al. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery." IPMI, 2017.
[7] Schlegl, T., et al. "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks." MIA, 2019.
[8] Rudolph, M., et al. "Same same but DifferNet: Semi-supervised defect detection with normalizing flows." WACV, 2021.
[9] Yi, J., et al. "Patch SVDD: Patch-level SVDD for anomaly detection and segmentation." ACCV, 2020.
[10] Cohen, N., and Hoshen, Y. "Sub-image anomaly detection with deep pyramid correspondences." arXiv preprint, 2020.
[11] Deng, C., and Li, Z. "RD4AD: Real difference for anomaly detection." ECCV, 2022.
[12] Bergmann, P., et al. "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings." CVPR, 2020.
[13] Roth, K., et al. "Towards total recall in industrial anomaly detection." CVPR, 2022.
[14] Bergmann, P., et al. "MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection." CVPR, 2019.
[15] Radford, A., et al. "Learning transferable visual models from natural language supervision." ICML, 2021.
[16] Li, J., et al. "BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models." ICML, 2023.
[17] Zhu, Y., et al. "MiniGPT-4: Enhancing vision-language understanding with advanced large language models." arXiv preprint, 2023.
[18] Liu, H., et al. "Visual instruction tuning." NeurIPS, 2023.
[19] Jeong, J., et al. "WinCLIP: Zero-/few-shot anomaly classification and segmentation." CVPR, 2023.
[20] Zhou, D., et al. "AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection." ICCV, 2023.
[21] Huang, S., et al. "Registration based few-shot anomaly detection." ECCV, 2022.
[22] Yang, Y., et al. "Meta-AD: Meta-learning for industrial anomaly detection." NIPS Workshop, 2021.
[23] Bergmann, P., et al. "MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection." CVPR, 2019.
[24] Kingma, D. P., and Welling, M. "Auto-encoding variational Bayes." ICLR, 2014.

[25] Akcay, S., et al. "GANomaly: Semi-supervised anomaly detection via adversarial training." ACCV, 2018.
[26] Yi, J., and Yoon, S. "Patch SVDD: Patch-level SVDD for anomaly detection and segmentation." ACCV, 2020.
[27] Cohen, N., and Hoshen, Y. "Sub-image anomaly detection with deep pyramid correspondences." arXiv preprint, 2020.
[28] Cao, Y., et al. "ClipSAM: CLIP and SAM collaboration for zero-shot anomaly segmentation." arXiv preprint, 2023.
[29] Gu, Q., et al. "AnomalyGPT: Detecting industrial anomalies using large vision-language models." ICCV, 2023.
[30] Li, M., et al. "Myriad: Large language models meet industrial anomaly detection." arXiv preprint, 2023.
[31] Ge, C., et al. "VMAD: Visual-enhanced multimodal large language model for zero-shot anomaly detection." arXiv preprint, 2024.
[32] Zhang, S., et al. "LogiCode: an LLM-driven framework for logical anomaly detection." IEEE TASE, 2023.
[33] Jiang, W., et al. "FabGPT: An efficient large multimodal model for complex wafer defect knowledge queries." arXiv preprint, 2024.
[34] Zhang, J., et al. "AnomalyR1: A GRPO-based end-to-end MLLM for industrial anomaly detection." arXiv preprint, 2024.
[35] Lin, Y., et al. "EIAD: Explainable industrial anomaly detection via multi-modal large language models." arXiv preprint, 2024.
[36] Wang, H., et al. "Can multimodal large language models be guided to improve industrial anomaly detection?" arXiv preprint, 2024.