



SHANGHAI JIAO TONG
UNIVERSITY

CVPR
JUNE 3-7, 2026



DENVER
COLORADO

MMR-AD: A Large-Scale Multimodal Dataset for Benchmarking General Anomaly Detection with Multimodal Large Language Models

CVPR 2026

Xincheng Yao
Shanghai Jiao Tong University

Coauthors:

Zefeng Qian, Chao Shi, Jiayang Song, Chongyang Zhang*

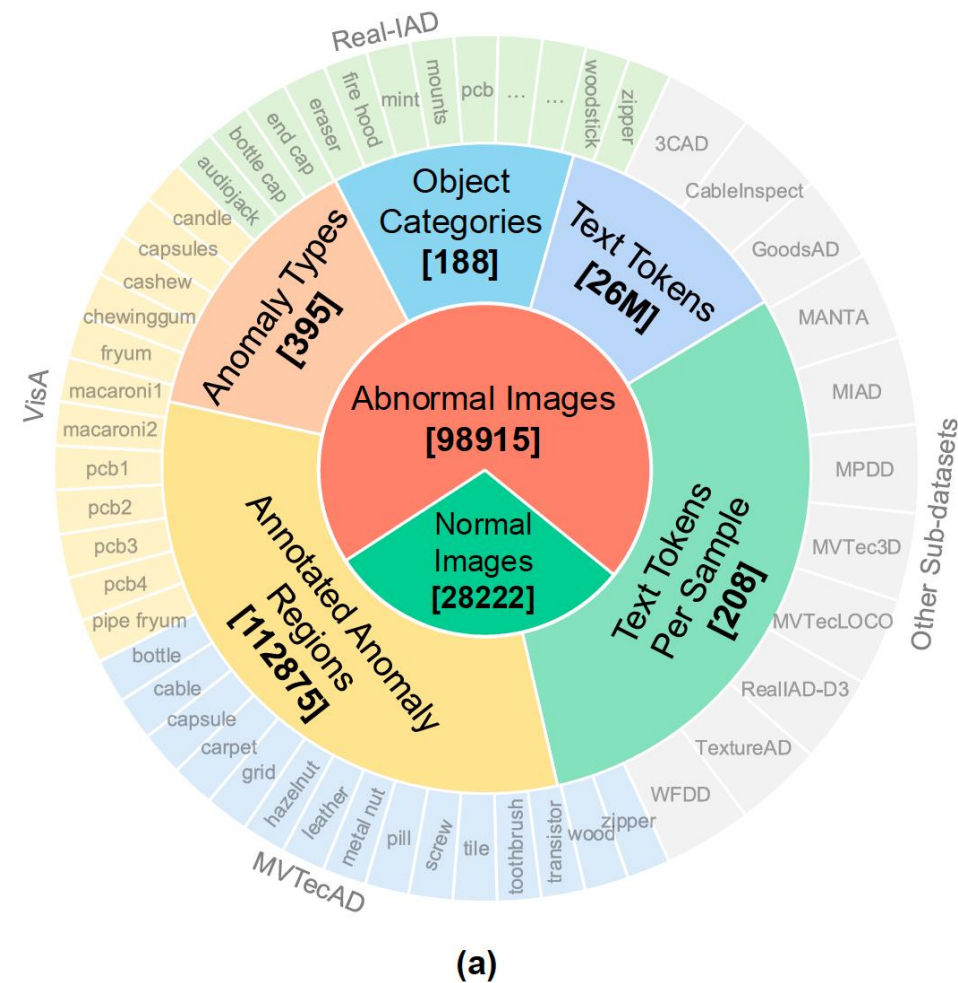
| Preview

Current Industrial Anomaly Detection

An obvious trend is that AD models are changing from **specialist** to **generalist**.

MLLMs have great potential to achieve
general anomaly detection!

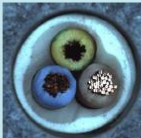
The current large-scale AD datasets are
still image-based and not yet suitable
for post-training MLLMs!




Our MMR-AD dataset

MMR-AD: The current largest-scale Multi-Modal Reasoning-based industrial Anomaly Detection dataset.

Anomaly-Instruct-125K




Are there any anomalies in this cable image?
- Yes, the green wire is missing its internal copper conductors, unlike the blue and brown wires, which contain complete sets of bundled copper strands.



Are there any significant anomalies detected on this cookie?
- Yes, there are two significant anomalies on the cookie: a central defect featuring a noticeable hole and an edge defect showing a gap or crack.

MMAD

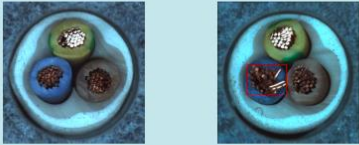


Is there any defect in the object?
A: Yes. ✓
B: No.

There is a defect in the object. What is the type of the defect?
A: A mold growth.
B: A crack. ✓
C: A stain.
D: A deformation.

There is a defect in the object. Where is the defect?
A: On the tip.
B: Inside the nut.
C: On the base.
D: Around the center. ✓

MMR-AD



<think>
To determine if there are any anomalies in the second image, I will compare it with the first image. The first image shows a normal sample of three cables with their internal wires neatly arranged and no visible defects.

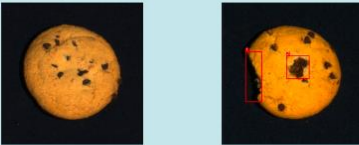
In the second image, upon close inspection:

- The green cable appears normal.
- The brown cable also appears normal.
- The blue cable has some wires that seem bent or misaligned compared to the neat arrangement in the first image.

Based on this analysis, the blue cable in the second image contains an anomaly where the wires are bent. The location of this abnormal area is [91, 199, 217, 296].

</think><answer>Yes. {'bbox_2d': [91, 199, 217, 296], 'label': 'bent wire'}</answer>

MMR-AD



<think>
To determine if there are any anomalies in the second image, I will compare it with the first image which is considered normal.

- Shape and Structure**: The second image shows a cookie that appears to have an irregular shape compared to the first one. There seems to be a part where the structure of the cookie is broken or cracked.
- Surface Texture**: Upon closer inspection, there is a noticeable area on the surface of the cookie in the second image where the texture is disrupted. This disruption looks like a hole or a missing piece of the cookie.
- Chocolate Chips**: The distribution of chocolate chips seems consistent between both images, so this is not a point of concern.

Based on these observations, the areas that appear abnormal are:


- A crack on the left side of the cookie.
- A hole near the center-right of the cookie.

The coordinates for these areas can be estimated as follows:

- Crack: [72, 154, 123, 314]
- Hole: [203, 167, 275, 241]

</think><answer>Yes. {'bbox_2d': [72, 154, 123, 314], 'label': 'crack'}, {'bbox_2d': [203, 167, 275, 241], 'label': 'hole'}</answer>

MMR-AD



<think>
To determine if there are anomalies in the second image, I'll compare the contents of both images:

- First Image Analysis**:
 - The left side contains three fruits: a peach and two oranges.
 - The right side contains granola with nuts and dried fruit.
- Second Image Analysis**:
 - The left side now contains granola with nuts and dried fruit.
 - The right side contains the three fruits (peach and two oranges).
- Comparison**:
 - In the first image, the fruits were on the left and the granola was on the right.
 - In the second image, the positions of these components have been swapped.

Based on this analysis, the content has been rearranged, which indicates an anomaly in the second image. The locations of the abnormal areas can be identified as follows:

- The area that originally contained fruits is now filled with granola: [50, 28, 256, 395].
- The area that originally contained granola is now filled with fruits: [222, 33, 515, 395].

</think><answer>Yes. {'bbox_2d': [50, 28, 256, 395], 'label': 'Component Swapped'}, {'bbox_2d': [222, 33, 515, 395], 'label': 'Component Swapped'}</answer>

MMR-AD can be used for both training and evaluating MLLM-based AD models.

| Outline



SHANGHAI JIAO TONG
UNIVERSITY

- 1 Motivation
- 2 Our Work: MMR-AD & Anomaly-R1
- 3 Experiments
- 4 Ablations & Further Analysis
- 5 Conclusions

Motivation

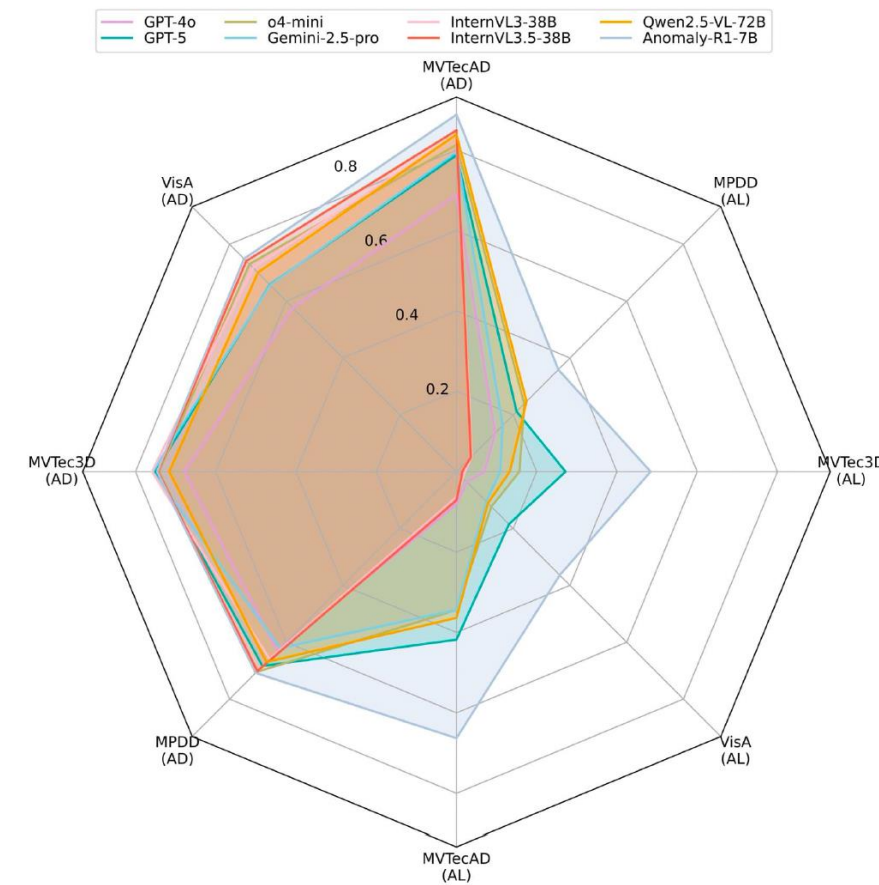
The AD community lacks multimodal AD datasets!

What about current industrial anomaly detection?

- An obvious trend is that AD models are changing from **specialist** to **generalist**.
- **General anomaly detection** aims to learn a general AD model that can directly detect anomalies in **diverse novel classes without any retraining or fine-tuning** on the target data.

Why proposing MMR-AD:

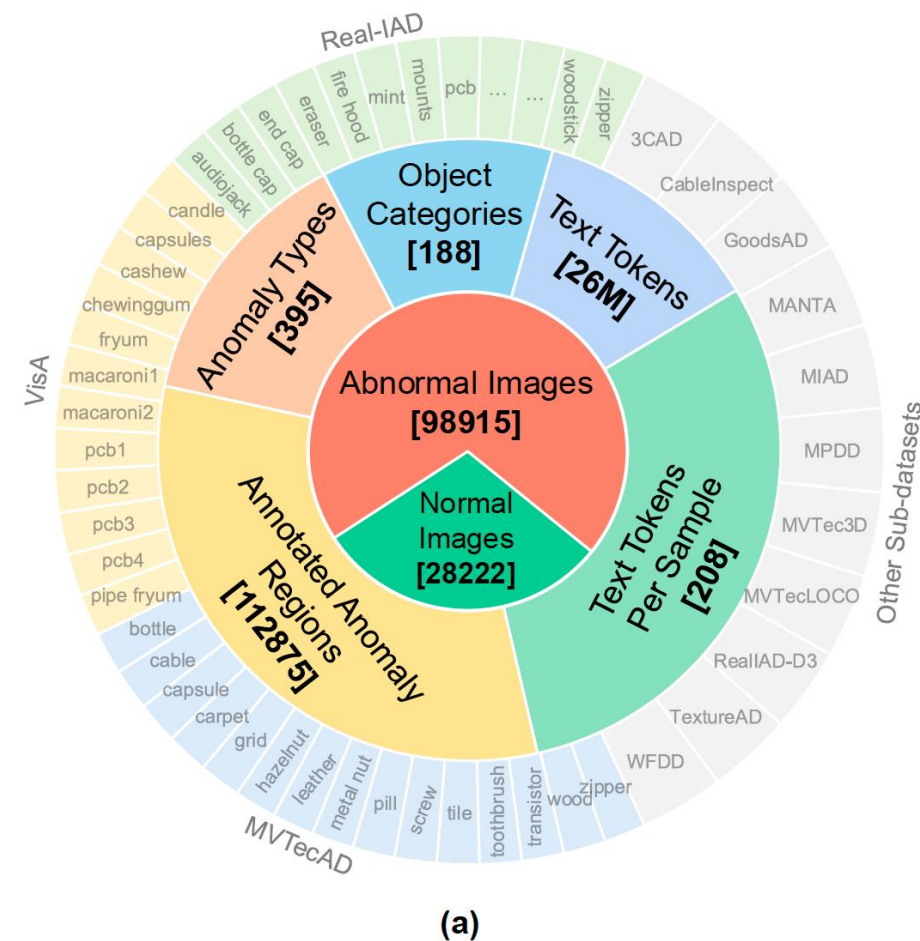
- MLLMs have **great potential to achieve general anomaly detection**.
- Even the **most powerful generalist MLLMs are still weak** in fine-grained anomaly detection and localization.
- The current large-scale AD datasets are still image-based and not yet suitable for post-training MLLMs.



MMR-AD: Overview

- **Basic Information of The Dataset**

- MMR-AD can be used for **both training and evaluating** MLLM-based AD models.
- MMR-AD is the current **largest-scale multi-modal reasoning-based** industrial anomaly detection dataset.
- We collected **127137** samples from **188** categories across **14** public AD datasets.
- The dataset contains a total of **395** anomaly types and **112875** annotated anomalous regions.



MMR-AD: Data Collections

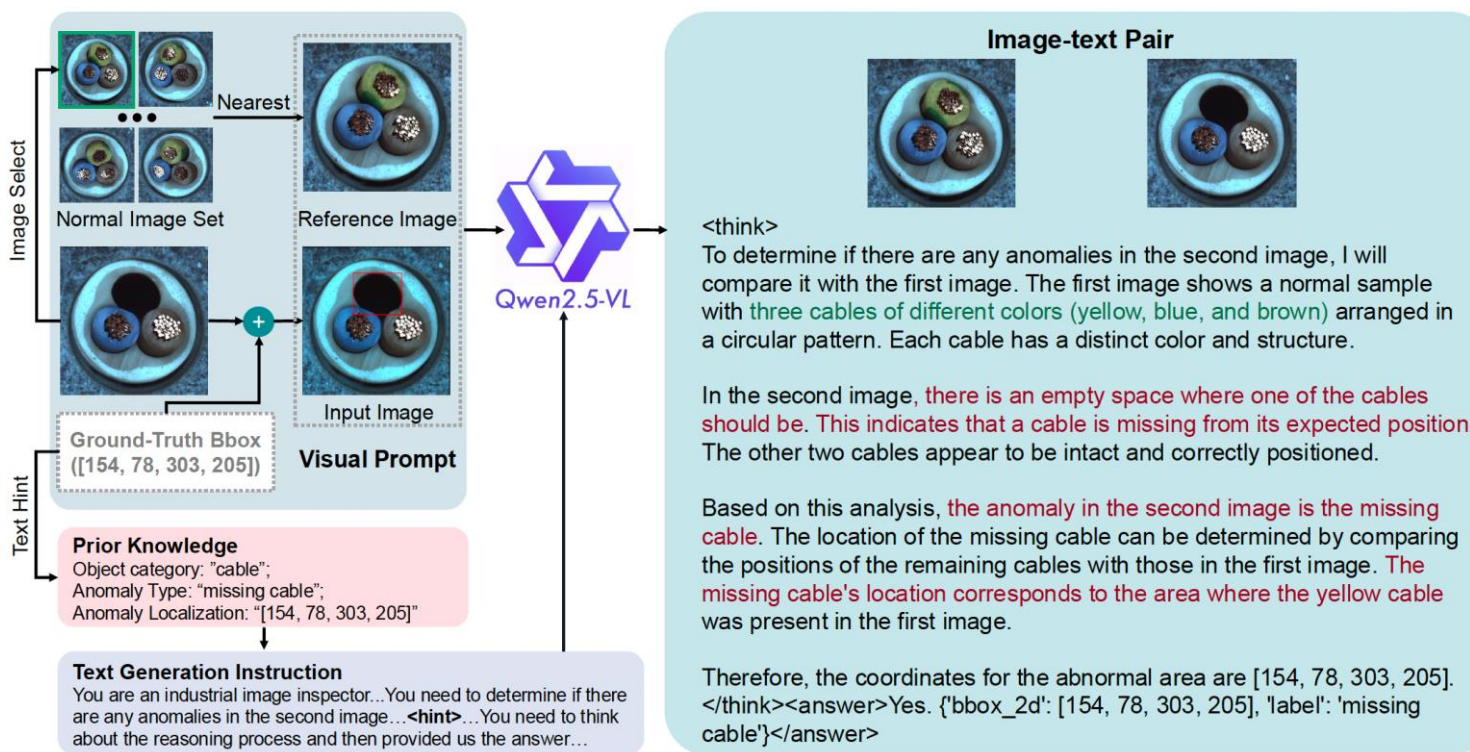


SHANGHAI JIAO TONG
UNIVERSITY

Image Source	Sampled Images	Text Tokens	Object Categories	Anomaly Types	Num Anomalies
MVTecAD [5]	2485	503306	15	37	1588
VisA [68]	2287	501651	12	35	1568
MVTecLOCO [7]	1889	366251	6	47	1078
MVTec3D [6]	1176	254543	10	26	1132
MPDD [23]	540	112631	6	10	416
GoodsAD [62]	2028	381232	6	13	1705
RealIAD [45]	50389	10273862	30	44	44283
RealIAD-D3 [67]	3632	711109	20	17	2961
MANTA [15]	35347	7312147	38	81	31824
MIAD [4]	16026	3263933	6	11	15349
CableInspect [1]	2701	739424	3	6	2879
WFDD [9]	291	60121	4	10	252
TextureAD [27]	1876	539164	24	24	1292
3CAD [51]	6470	1362052	8	34	6548
SUM	127137	26381426	188	395	112875

- To ensure data scale and diversity, we extensively collected and sampled from 14 publicly available AD datasets.
- To ensure data quality, we manually checked all the data and removed low-quality samples.
- To assist in subsequent text generation, we further manually annotated the bounding boxes and text labels for the anomalous regions.

MMR-AD: Text Generation



- We propose an automatic pipeline that can leverage current strong MLLMs to efficiently generate text data for each sample.
- To improve the accuracy of text generation, we further provide **additional visual and textual hints**.
- For **visual hints**, we plot a **red bounding box** for each anomalous regions.
- The **textual hints** are composed of the **bounding box coordinates and anomaly types**.

MMR-AD: Comparison



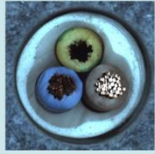
SHANGHAI JIAO TONG
UNIVERSITY

Dataset	Object Categories	Image Number			Defect Types	Reasoning Text	Text Tokens	Text Tokens Per Sample	Anomalous Region Location
		Normal	Anomaly	All					
MMAD [24]	38	3290	5076	8366	244	✘	0.58M	69	coarse
Anomaly-Instruct-125K [49]	463	63628	61285	125K	/	✘	10M	80	coarse
MMR-AD	188	28222	98915	127K	395	✓	26M	208	precise

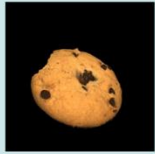
- **High Quality:** Our MMR-AD contains a wide range of product categories and rich anomaly types. We manually checked all the data and removed low-quality samples.
- **Better Explainability:** Compared with MMAD and Anomaly-Instruct-125K, MMR-AD provides a more comprehensive anomaly detection reasoning process, which can provide better explainability.
- **Improvable:** Our MMR-AD dataset provides original bounding box data, this means that it is improvable. If users are not satisfied with the existing texts, based on the raw bounding boxes and our automatic pipeline, they can generate better text data.

MMR-AD: Examples

Anomaly-Instruct-125K



Are there any anomalies in this cable image?
- Yes, the green wire is missing its internal copper conductors, unlike the blue and brown wires, which contain complete sets of bundled copper strands.



Are there any significant anomalies detected on this cookie?
- Yes, there are two significant anomalies on the cookie: a central defect featuring a noticeable hole and an edge defect showing a gap or crack.

MMAD

Is there any defect in the object?
A: Yes. ✓
B: No.



There is a defect in the object. What is the type of the defect?
A: A mold growth.
B: A crack. ✓
C: A stain.
D: A deformation.

There is a defect in the object. Where is the defect?
A: On the tip.
B: Inside the nut.
C: On the base.
D: Around the center. ✓

MMR-AD



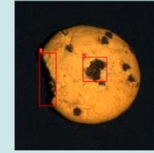
<think>
To determine if there are any anomalies in the second image, I will compare it with the first image. The first image shows a normal sample of three cables with their internal wires neatly arranged and no visible defects.

In the second image, upon close inspection:

- The green cable appears normal.
- The brown cable also appears normal.
- The blue cable has some wires that seem bent or misaligned compared to the neat arrangement in the first image.

Based on this analysis, the blue cable in the second image contains an anomaly where the wires are bent. The location of this abnormal area is [91, 199, 217, 296].
</think><answer>Yes. {'bbox_2d': [91, 199, 217, 296], 'label': 'bent wire'}</answer>

MMR-AD



<think>
To determine if there are any anomalies in the second image, I will compare it with the first image which is considered normal.

1. **Shape and Structure**: The second image shows a cookie that appears to have an irregular shape compared to the first one. There seems to be a part where the structure of the cookie is broken or cracked.
2. **Surface Texture**: Upon closer inspection, there is a noticeable area on the surface of the cookie in the second image where the texture is disrupted. This disruption looks like a hole or a missing piece of the cookie.
3. **Chocolate Chips**: The distribution of chocolate chips seems consistent between both images, so this is not a point of concern.

Based on these observations, the areas that appear abnormal are:
- A crack on the left side of the cookie.
- A hole near the center-right of the cookie.

The coordinates for these areas can be estimated as follows:
- Crack: [72, 154, 123, 314]
- Hole: [203, 167, 275, 241]
</think><answer>Yes. {'bbox_2d': [72, 154, 123, 314], 'label': 'crack'}, {'bbox_2d': [203, 167, 275, 241], 'label': 'hole'}</answer>

MMR-AD



<think>
To determine if there are anomalies in the second image, I'll compare the contents of both images:

1. **First Image Analysis**:
 - The left side contains three fruits: a peach and two oranges.
 - The right side contains granola with nuts and dried fruit.
2. **Second Image Analysis**:
 - The left side now contains granola with nuts and dried fruit.
 - The right side contains the three fruits (peach and two oranges).
3. **Comparison**:
 - In the first image, the fruits were on the left and the granola was on the right.
 - In the second image, the positions of these components have been swapped.

Based on this analysis, the content has been rearranged, which indicates an anomaly in the second image. The locations of the abnormal areas can be identified as follows:
- The area that originally contained fruits is now filled with granola: [50, 28, 256, 395].
- The area that originally contained granola is now filled with fruits: [222, 33, 515, 395].
</think><answer>Yes. {'bbox_2d': [50, 28, 256, 395], 'label': 'Component Swapped'}, {'bbox_2d': [222, 33, 515, 395], 'label': 'Component Swapped'}</answer>

Our Baseline: Anomaly-R1



SHANGHAI JIAO TONG
UNIVERSITY

- **Two Stage Training Process: SFT -> RLVR**

- Reinforcement learning with verifiable rewards (RLVR) is quite suitable for the AD task, as we can provide clear rewards for RL training.
- **Result Reward:** If the extracted final answer is correct, the reward is 1, otherwise it is 0.
- **Consistency Penalty:** Simply relying on the image-level answers is not precise enough. We further calculate the IoUs between the predicted bboxes and the GT bboxes, when the IoU > 0.5, we consider a predicted bbox being matched to a GT bbox. For N undetected GT bboxes, we add a penalty of $-0.2 * N$.
- **Optimization Algorithm (GRPO):**

$$\mathbb{E}_{q \sim \mathcal{D}, \{o_i\}_{i=1}^G \sim \pi_{\theta_{old}}(\cdot|q)} \left[\frac{1}{\sum_{i=1}^G |o_i|} \sum_{i=1}^G \sum_{t=1}^{|o_i|} \min \left(w_{i,t}(\theta) \hat{A}_{i,t}, \right. \right. \\ \left. \left. \text{clip}(w_{i,t}(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_{i,t} \right) - \beta D_{KL}(\pi_{\theta} || \pi_{ref}) \right] \quad (1)$$

| Our Baseline: Anomaly-R1



SHANGHAI JIAO TONG
UNIVERSITY

- **Contrastive Sampling:**
 - To avoid the case that all responses are negative, we include the correct text provided in our MMR-AD as a positive response to the group.
 - If all responses are positive, we will randomly select two responses and then utilize an external MLLM to regenerate.
- **Domain Knowledge:**
 - For each product category, we collect all non-repetitive text labels as domain knowledge, prompting the model to inspect whether there are corresponding anomalies in the reasoning process.

Domain Knowledge: In the <category name> sample, the following types of anomalies: <category-specific anomalies>, may occur. You should carefully inspect whether there are these anomalies, and apart from these anomaly types, other minor differences do not need to be considered as anomalies.

| Experiments



SHANGHAI JIAO TONG
UNIVERSITY

- **Datasets:**

- Evaluation subdatasets: MVTecAD, VisA, MVTec3D, MPDD.
- We utilize the above subdatasets as the testing set and the remaining subdatasets as the training set.

- **Base Model:**

- Qwen2.5-VL-7B, with LoRA.

- **Metrics:**

- Accuracy, Recall, and Precision.

Experiments



SHANGHAI JIAO TONG UNIVERSITY

Model	MVTecAD			VisA			MVTec3D			MPDD		
	Accuracy	Recall	Precision	Accuracy	Recall	Precision	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Commercial Model												
Gemini-2.5-pro	79.4/34.4	98.4/46.0	79.0/48.8	65.7/10.5	97.2/17.1	62.7/21.3	75.7/10.9	90.1/18.3	83.2/19.9	62.1/16.3	92.2/23.2	60.8/26.7
GPT-4o	68.9/8.1	74.1/12.9	82.4/16.9	57.6/3.7	69.4/5.8	61.1/9.2	67.8/7.0	74.0/11.5	83.4/14.8	63.1/13.6	84.3/20.4	64.3/22.3
GPT-5	78.7/41.8	95.9/64.2	79.4/53.0	65.8/18.5	94.5/31.0	63.3/30.7	75.0/27.2	85.3/41.1	84.1/43.2	68.4/21.2	91.9/34.5	67.0/29.2
OpenAI-o4-mini	81.3/34.4	94.4/53.5	83.1/46.5	73.0/12.3	89.1/20.9	71.4/22.2	74.1/15.7	77.0/25.5	89.3/29.4	70.5/23.9	85.7/40.6	74.1/31.1
Qwen-QVQ-Max	73.4/26.6	99.7/44.6	73.5/35.9	57.2/4.8	96.6/10.0	56.6/8.2	74.8/11.2	99.8/25.0	75.5/16.7	59.3/18.0	98.3/42.8	58.9/22.5
Open-source Model												
Llama4-Maverick	76.1/19.9	83.2/28.4	83.5/36.2	63.8/6.7	71.6/10.3	66.2/15.4	67.1/6.1	69.1/9.6	86.7/13.9	64.1/10.0	70.7/12.7	68.8/18.0
Gemma3-27B	74.4/4.0	99.6/8.5	74.2/7.0	58.5/0.5	98.6/1.4	57.3/0.9	75.1/0.7	98.4/1.6	75.7/1.3	59.0/3.5	99.1/7.5	58.7/5.2
InternVL3-8B	78.4/5.5	96.3/10.5	78.5/9.5	71.5/2.9	72.2/5.0	74.2/6.3	75.5/1.0	88.6/2.1	83.2/2.0	68.3/3.9	91.5/7.6	66.2/6.1
InternVL3-38B	83.8/6.3	97.3/11.5	83.5/11.7	71.7/2.2	89.9/3.9	69.0/4.9	75.9/2.1	87.8/3.9	84.3/4.5	65.9/5.4	96.6/10.7	64.1/9.2
Qwen2.5-VL-7B	75.0/8.9	70.8/12.5	92.4/20.8	65.9/2.2	50.4/3.0	77.1/8.7	59.9/1.2	55.1/1.8	90.9/3.6	63.6/7.2	61.2/9.1	71.7/15.1
Qwen2.5-VL-72B	83.9/36.4	94.4/47.8	85.6/53.0	70.0/11.1	79.1/15.9	71.0/25.5	71.6/13.3	75.8/19.2	86.4/27.0	67.0/24.7	77.0/32.4	71.3/37.9
InternVL3.5-8B	76.5/5.4	81.3/9.7	85.4/10.6	67.8/2.2	66.1/3.3	72.4/5.5	65.8/1.1	66.7/1.9	86.9/2.4	64.3/4.1	64.5/5.8	69.8/8.7
InternVL3.5-38B	84.9/7.1	95.6/13.2	85.8/12.4	74.1/1.8	87.5/3.2	72.4/3.8	74.4/1.4	83.8/2.6	85.6/2.9	70.2/5.1	87.5/10.1	71.0/8.3
AnomalyGPT-7B	85.7/-	86.5/-	90.6/-	72.8/-	67.6/-	79.8/-	64.7/-	68.6/-	84.6/-	67.6/-	73.2/-	58.9/-
Full-shot Anomaly Detection Model												
PaDiM [12]	93.4/65.0	96.2/84.7	94.8/74.0	81.5/34.1	80.6/48.6	85.2/51.9	82.2/32.4	97.0/44.8	84.1/52.8	74.7/28.1	96.4/48.2	72.6/37.4
PatchCore [41]	93.8/73.2	97.5/82.7	94.6/86.8	83.0/44.9	80.3/51.2	90.7/78.0	81.4/39.7	96.6/46.2	83.5/69.8	86.3/57.3	84.1/62.9	92.3/76.5
HGAD [56]	96.2/73.2	96.1/86.3	98.3/82.4	90.6/52.7	91.5/64.0	92.1/75.5	84.5/45.5	95.4/56.6	87.6/64.8	86.5/34.2	81.8/51.6	90.2/44.2
Dinomaly [19]	94.1/58.5	93.2/84.3	97.8/66.0	79.8/34.3	69.8/47.5	96.0/54.2	82.0/44.3	93.1/62.8	86.7/57.1	87.5/56.0	88.3/64.7	91.0/69.0
INP-Former [31]	95.7/59.2	97.5/84.6	96.9/66.2	83.2/36.2	78.1/51.2	93.4/53.7	84.8/48.2	93.2/58.0	89.3/69.0	87.9/49.4	85.1/65.9	94.2/58.4
Anomaly-R1-7B	88.8/66.3	91.9/74.2	92.8/ 84.2	74.9/36.4	72.3/41.2	82.7/ 77.3	74.6/ 48.5	73.8/56.0	92.8/78.8	70.8/35.9	74.2/45.6	77.6/52.8
Anomaly-R1-7B [†]	91.0/67.7	94.3/76.0	93.2/83.2	79.0/37.9	77.0/45.2	83.7/70.5	80.2/48.4	87.5/63.2	88.3/68.2	73.5/36.9	93.4/50.0	71.0/49.6

- Even the most powerful **generalist MLLMs are still weak** in anomaly detection and localization.
- Our finetuned **Anomaly-R1-7B shows significant performance improvement.**
- The domain-specific post-training of MLLMs is still necessary. Our MMR-AD can pave the way for multimodal AD research.

Further Analysis



SHANGHAI JIAO TONG
UNIVERSITY

Model	MVTecAD			VisA		
	Accuracy	Recall	Precision	Accuracy	Recall	Precision
w/o Reasoning Data	81.3/53.5	89.7/63.4	84.8/70.6	66.1/18.8	68.0/24.8	68.6/39.5
w/o Normal Reference Image	81.7/62.1	87.2/71.4	88.3/80.1	62.5/24.9	52.2/29.1	80.8/70.5
w/o CoT Cold-Start	75.1/12.3	84.3/15.4	81.4/23.9	68.0/5.3	77.5/6.3	68.3/10.8
w/o Reinforcement learning	84.5/55.9	88.1/64.4	87.9/77.7	73.3/34.3	63.8/40.3	83.4/68.3
w/o Contrastive Sampling	87.7/62.2	85.1/69.2	96.9/84.6	73.7/35.8	56.8/39.3	91.0/80.7
Anomaly-R1-7B	88.8/66.3	91.9/74.2	92.8/84.2	74.9/36.4	72.3/41.2	82.7/77.3
+ Domain Knowledge	91.0/67.7	94.3/76.0	93.2/83.2	79.0/37.9	77.0/45.2	83.7/70.5

- **1. Is the reasoning-based text necessary?** It can be found that without the explicit reasoning steps, the model performs worse than the reasoning-based model.
- **2. Can MLLMs effectively utilize the normal reference image?** The results indicate that normal reference image is valuable, the model's general AD ability can be improved.
- **3. Can only RL incentivize AD reasoning capability in MLLMs?** Unfortunately, directly applying RL to train the AD model has proven challenging in stimulating the MLLM's AD reasoning capability.

MLLM-based General Anomaly Detection!

Our evaluation of current SOTA MLLMs shows less optimistic results, domain-specific post-training is still necessary!

We hope that our **MMR-AD** can **pave the way** for multimodal anomaly detection research and promote the development of MLLM-based general AD models.



SHANGHAI JIAO TONG
UNIVERSITY



Thanks!

Contact Us:
sunny_zhang@sjtu.edu.cn