

Research trends in Anomaly Detection and Efficient Learning



RILS LAB

Robot Intelligence Learning & System

유호현

2026. 07. 23

Papers

- **(CVPR 2026) ADSeeker: A Knowledge-Grounded Reasoning Framework for Industry Anomaly Detection and Reasoning**
- **(ICCV 2025) TR-PTS: Task-Relevant Parameter and Token Selection for Efficient Tuning**
- **(ICCV 2025) Triad: LMM-based Anomaly Detection with Expert-guided RoI Tokenizer**
- **(AAAI 2026) IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection**
- **(CVPR 2026) ReAL: Reasoning-Driven Anomaly Detection and Localization with Image-Level Supervision**

06

ADSeeker: A Knowledge-Grounded Reasoning Framework for Industry Anomaly Detection and Reasoning

Kai Zhang, Zekai Zhang, Xihe Sun, Anpeng Wang, Jingmeng Nie, Qinghui Chen,
Han Hao, Jianyuan Guo, Jinglin Zhang

CVPR 2026

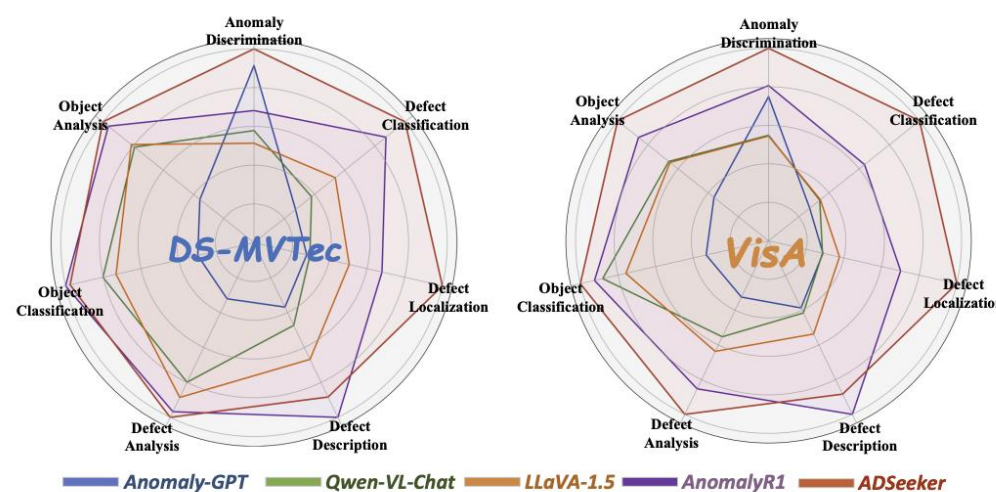
ADSeeker: A Knowledge-Grounded Reasoning Framework for Industry Anomaly Detection and Reasoning

Problem

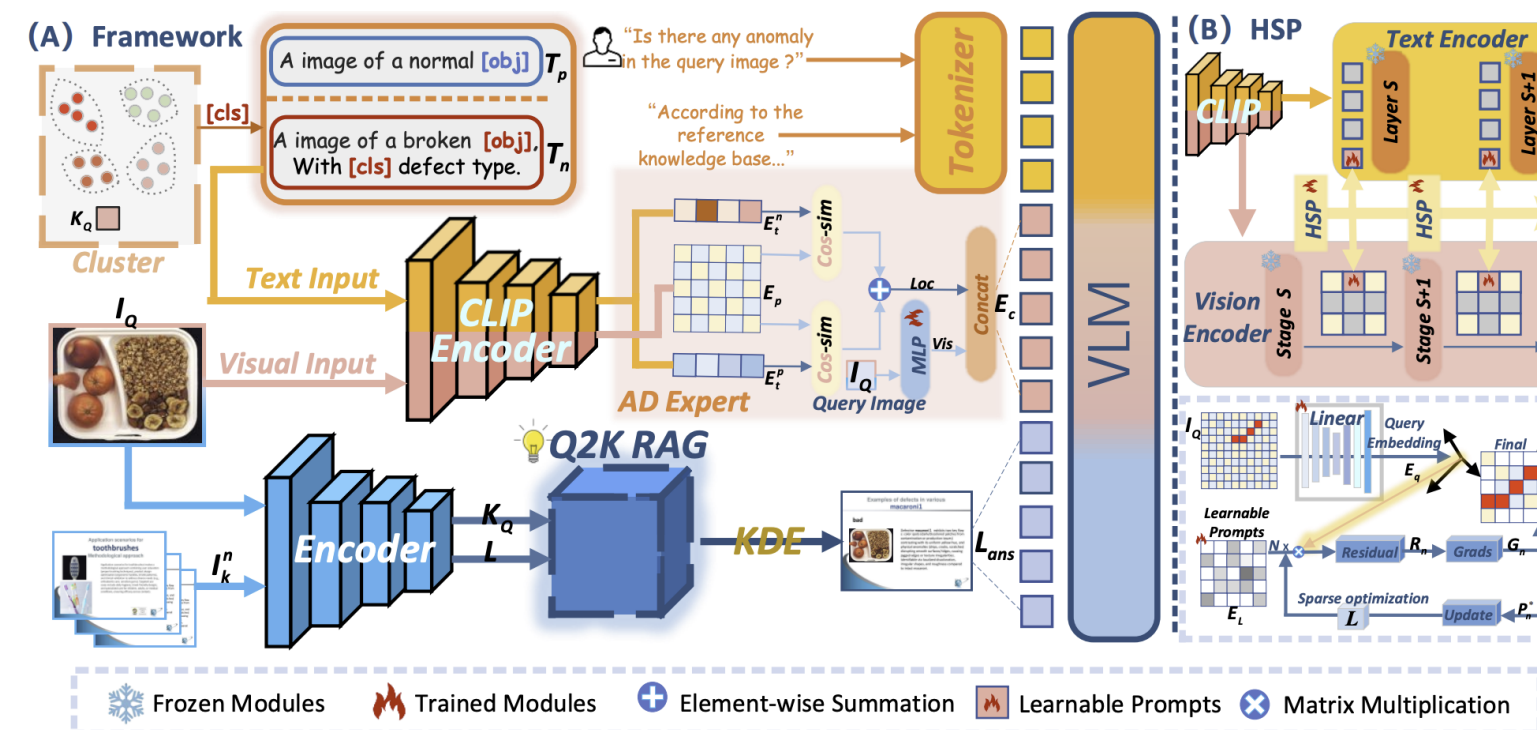
- MLLMs lag human experts in IAD due to insufficient AD-knowledge integration and imprecise anomaly-reasoning language; fine-tuning risks forgetting.

Method

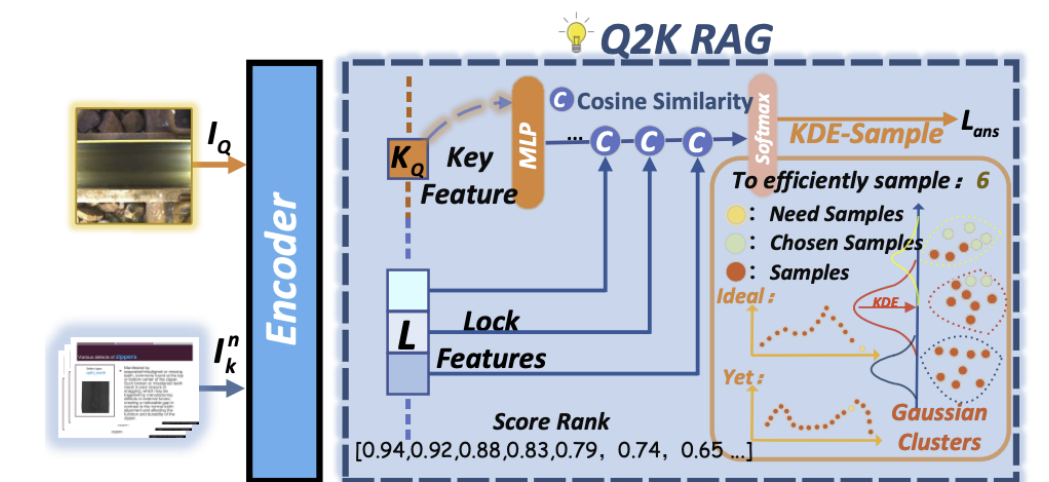
- Training-free ADSeeker retrieves relevant knowledge via Q2K RAG (visual-document KB, SEEK-M&V) and extracts type-level defect features via a Hierarchical Sparse Prompt (HSP) expert module, both fused into the backbone VLM.



<Visualization of the reasoning capability>



<The architecture of ADSeeker>



<Illustration of the core architecture of Q2K RAG>

ADSeeker: A Knowledge-Grounded Reasoning Framework for Industry Anomaly Detection and Reasoning

Results & Contribution

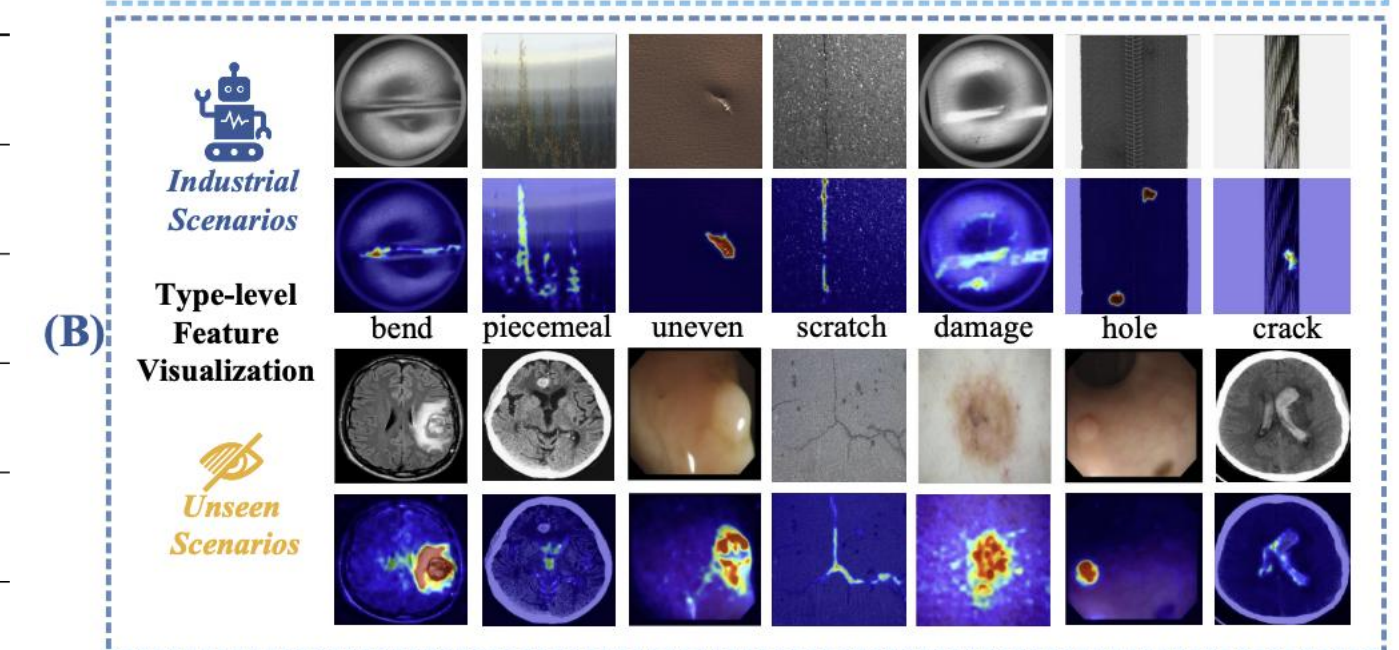
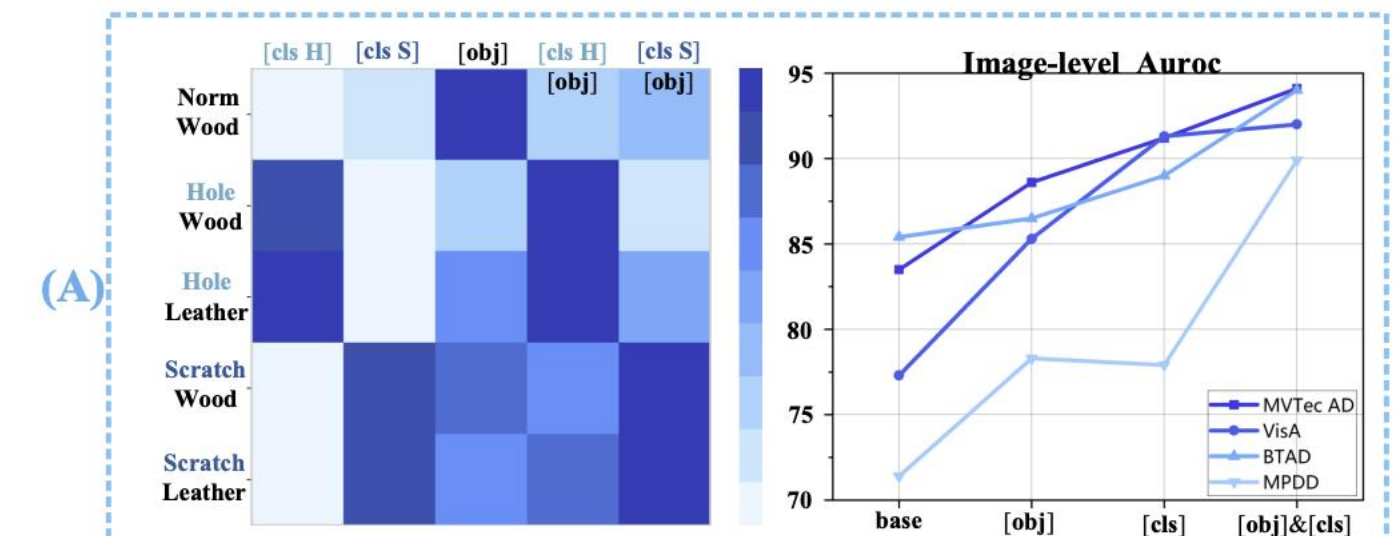
- SOTA ZSAD (94.0% avg. AUROC); boosts open-source MLLMs +3~6pts on MMAD reasoning without extra training, beating training-based SOTA at modest overhead.

Model	Industrial Defects				Medical Anomalies			Average
	MVTecAD	VisA	BTAD	MPDD	BrainMRI	HeadCT	Br35H	
CLIP[29]	74.1	66.4	34.5	54.3	73.9	56.5	78.4	62.6
WinCLIP[15]	91.8	78.8	68.2	63.6	92.6	90.0	80.5	80.8
AnomalyCLIP[40]	91.5	82.1	88.3	77.0	90.3	93.4	94.6	88.2
AdaCLIP[6]	89.2	85.8	88.6	76.0	94.8	91.4	97.7	89.1
AnomalyOV[34]	94.0	91.1	89.0	81.7	93.9	97.6	95.5	91.8
ADSeeker	94.3	91.5	94.0	85.9	97.5	96.6	97.9	94.0

<Image-Level AUROC performance of existing ZSAD approaches>

Model	Scale	Setting	Anomaly	Defect				Object		Average
			Discrimination	Classification	Localization	Description	Analysis	Classification	Analysis	
LLaVA-OV[21]	7B	Baseline	51.17	46.13	41.85	62.19	69.73	90.31	80.93	63.19
		Seek-Setting	55.35 (+4.18)	49.86 (+3.73)	54.34 (+12.49)	57.17 (-5.02)	73.20 (+3.47)	92.07 (+1.76)	84.31 (+3.38)	66.61(+3.42)
LLaVA-NeXT[25]	7B	Baseline	57.64	33.79	47.72	51.84	67.93	81.39	74.91	59.32
		Seek-Setting	64.10 (+6.46)	53.29 (+19.50)	58.13 (+10.41)	54.09 (+2.25)	71.20 (+3.27)	87.70 (+6.31)	73.20 (-1.71)	65.96(+6.64)
InternVL2[9]	8B	Baseline	59.97	43.85	47.91	57.60	78.10	74.18	80.37	63.14
		Seek-Setting	63.26 (+3.29)	58.71 (+14.86)	49.45 (+1.54)	56.45 (-1.15)	82.73 (+4.63)	84.79 (+10.61)	80.93 (+0.56)	68.05(+4.91)
Qwen2.5-VL[3]	3B	Baseline	51.10	41.07	44.87	61.43	75.99	86.54	79.58	62.94
		Seek-Setting	58.17 (+7.07)	48.71 (+7.64)	53.67 (+8.80)	69.79 (+8.36)	80.36 (+4.37)	86.78 (+0.24)	82.25 (+2.67)	68.53(+5.59)
Qwen2.5-VL[3]	7B	Baseline	55.35	49.86	54.34	57.18	73.20	92.07	84.31	66.62
		Seek-Setting	68.62 (+13.27)	53.82 (+3.96)	55.57 (+1.23)	60.51 (+3.33)	77.40 (+4.20)	89.91 (-2.16)	83.48 (-0.83)	69.90(+3.28)

<Anomaly Reasoning Performance>



<Validation of type-level feature representation>

07

TR-PTS: Task-Relevant Parameter and Token Selection for Efficient Tuning

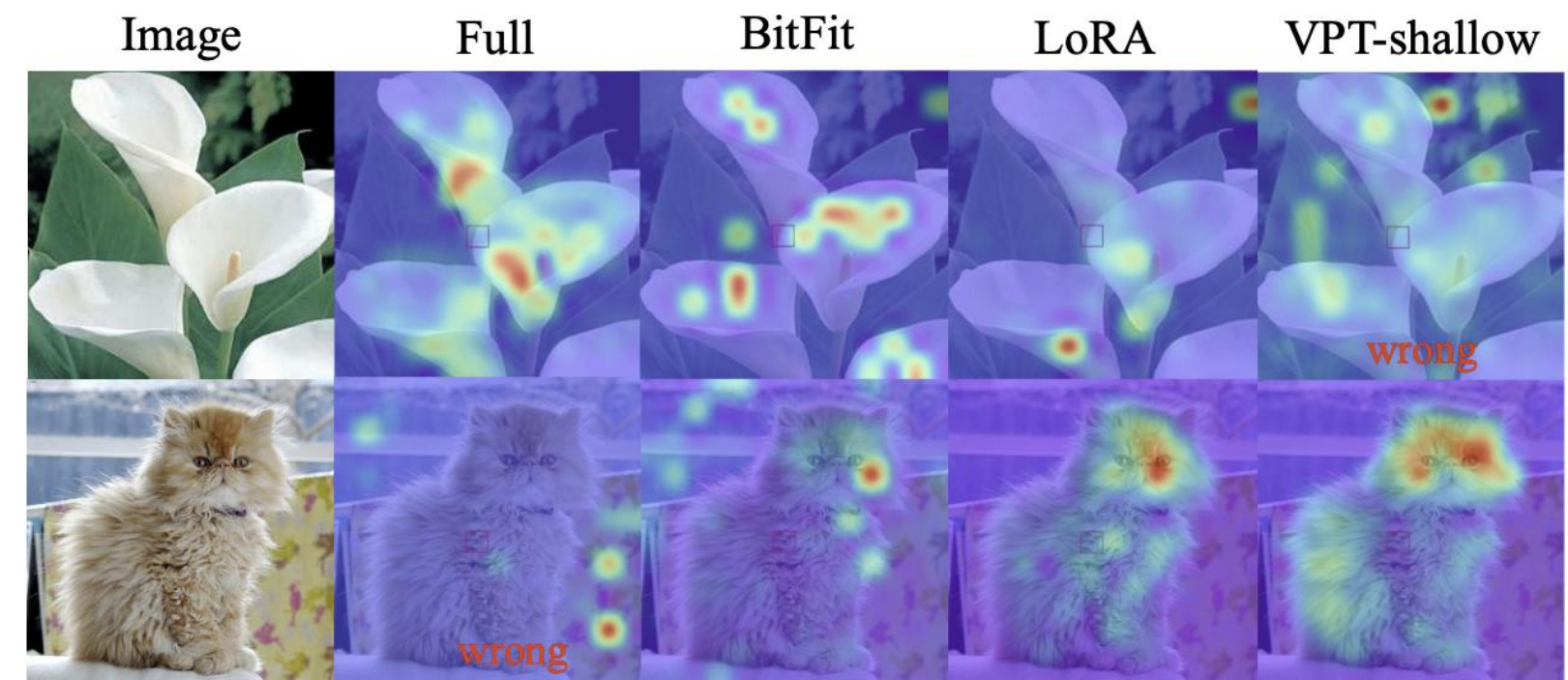
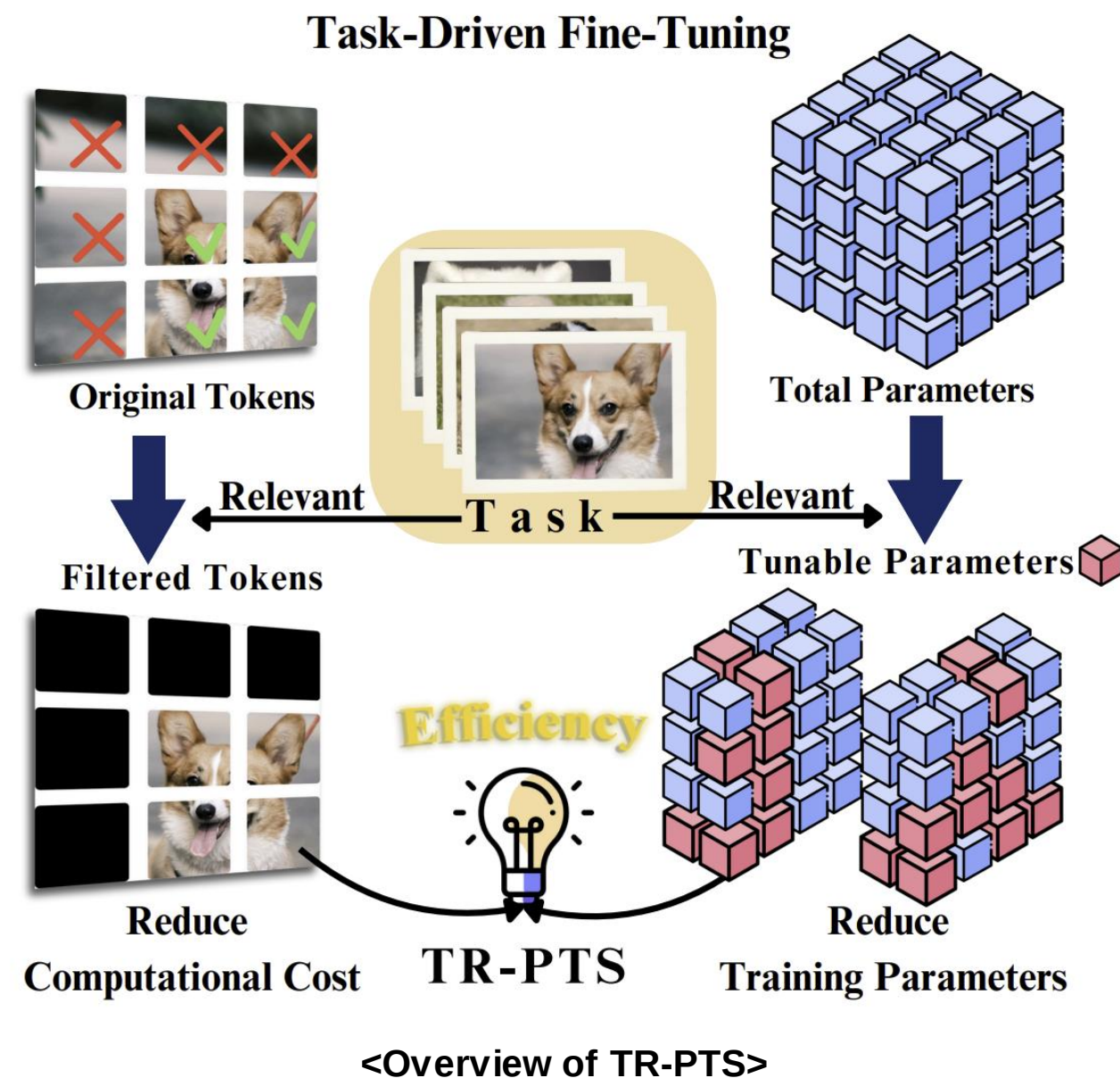
Siqi Luo, Haoran Yang, Yi Xin, Mingyang Yi, Guangyang Wu, Guangtao Zhai,
Xiaohong Liu

ICCV 2025

TR-PTS: Task-Relevant Parameter and Token Selection for Efficient Tuning

Problem

- Most PEFT methods are task-agnostic, failing to exploit task-specific adaptation and leaving both efficiency and accuracy suboptimal.

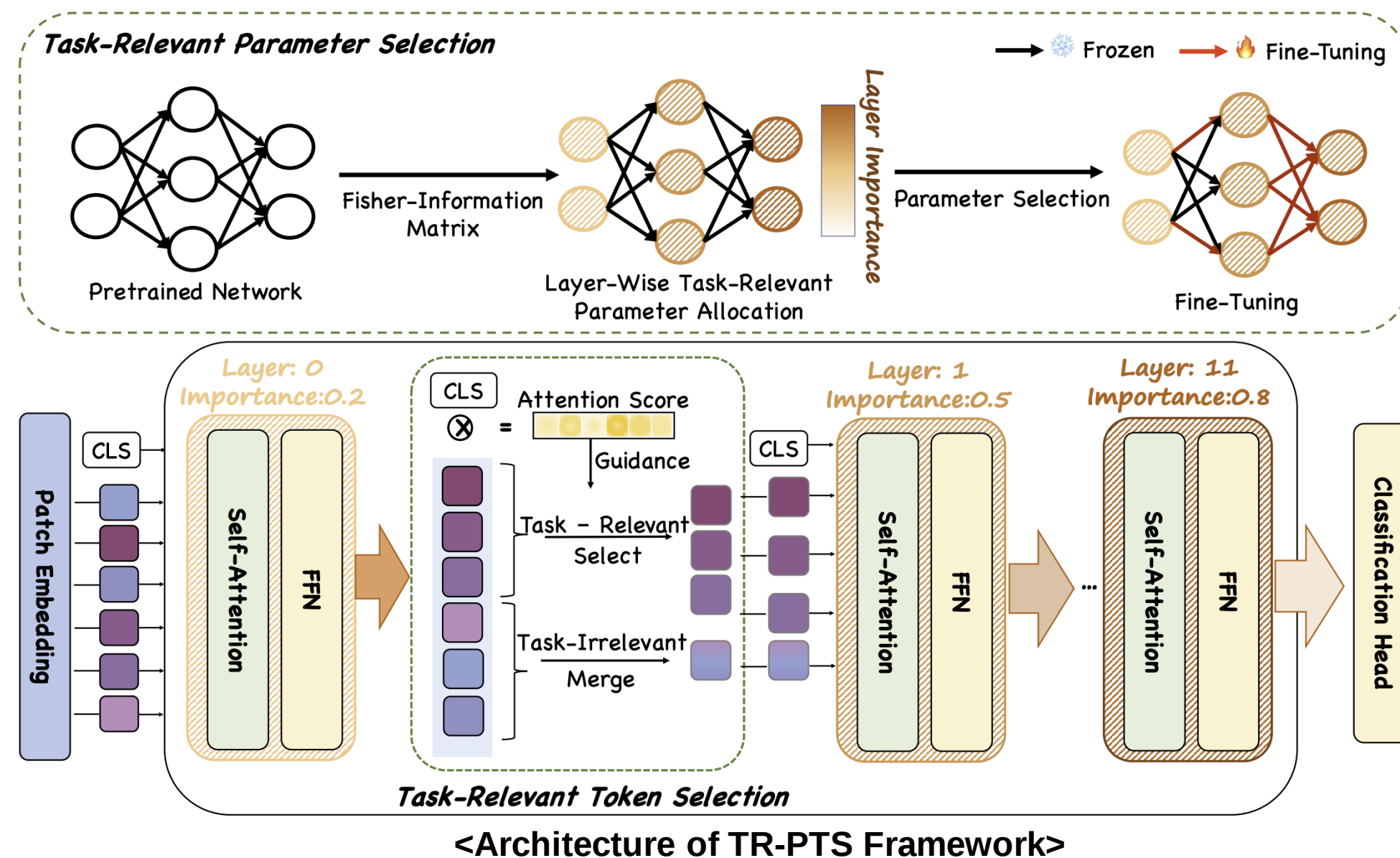


<Visualizing the Impact of Each Token on the Final Prediction>

TR-PTS: Task-Relevant Parameter and Token Selection for Efficient Tuning

Method

- Fisher-Information-guided selection of the most informative layer-wise parameters to tune.
- Task-relevant token selection — retain informative tokens, merge redundant ones — under a joint parameter-token optimization.

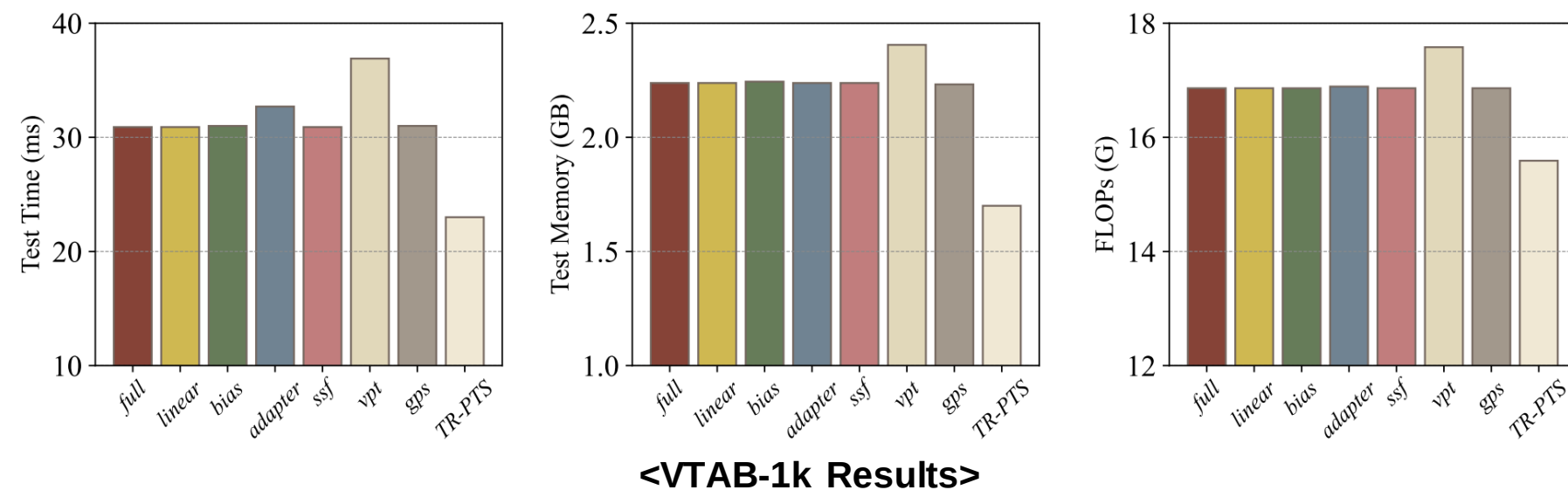


TR-PTS: Task-Relevant Parameter and Token Selection for Efficient Tuning

Results & Contribution

- SOTA on FGVC and VTAB-1k, surpassing full fine-tuning by +3.40% / +10.35% with as few as 0.60% trainable parameters.

Method \ Dataset	Natural							Specialized				Structured								Mean Acc.	Params. (%)
	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVHN	Sun397	Patch Camelyon	EuroSAT	Resisc45	Retinopathy	Clevr/count	Clevr/distance	DMLab	KITTI/distance	dSprites/loc	dSprites/ori	SmallNORB/azi	SmallNORB/ele		
Full [14]	68.9	87.7	64.3	97.2	86.9	87.4	38.8	79.7	95.7	84.2	73.9	56.3	58.6	41.7	65.5	57.5	46.7	25.7	29.1	65.57	100.00
Linear [14]	63.4	85.0	64.3	97.0	86.3	36.6	51.0	78.5	87.5	68.6	74.0	34.3	30.6	33.2	55.4	12.5	20.0	9.6	19.2	53.00	0.05
BitFit [36]	72.8	87.0	59.2	97.5	85.3	59.9	51.4	78.7	91.6	72.9	69.8	61.5	55.6	32.4	55.9	66.6	40.0	15.7	25.1	62.05	0.16
Adapter [11]	74.1	86.1	63.2	97.7	87.0	34.6	50.8	76.3	88.0	73.1	70.5	45.7	37.4	31.2	53.2	30.3	25.4	13.8	22.1	55.82	0.31
AdaptFormer [3]	70.8	91.2	70.5	99.1	90.9	86.6	54.8	83.0	95.8	84.4	76.3	81.9	64.3	49.3	80.3	76.3	45.7	31.7	41.1	72.32	0.24
LoRA [12]	68.1	91.4	69.8	99.0	90.5	86.4	53.1	85.1	95.8	84.7	74.2	83.0	66.9	50.4	81.4	80.2	46.6	32.2	41.1	72.63	0.90
VPT-Shallow [14]	77.7	86.9	62.6	97.5	87.3	74.5	51.2	78.2	92.0	75.6	72.9	50.5	58.6	40.5	67.1	68.7	36.1	20.2	34.1	64.85	0.13
VPT-Deep [14]	78.8	90.8	65.8	98.0	88.3	78.1	49.6	81.8	96.1	83.4	68.4	68.5	60.0	46.5	72.8	73.6	47.9	32.9	37.8	69.43	0.70
SSF [17]	69.0	92.6	75.1	99.4	91.8	90.2	52.9	87.4	95.9	87.4	75.5	75.9	62.3	53.3	80.6	77.3	54.9	29.5	37.9	73.10	0.28
GPS [39]	81.1	94.2	75.8	99.4	91.7	91.6	52.4	87.9	96.2	86.5	76.5	79.9	62.6	55.0	82.4	84.0	55.4	29.7	46.1	75.18	0.25
TR-PTS (Ours)	81.2	93.9	75.1	99.5	91.9	91.0	54.5	88.1	95.7	87.8	76.6	83.5	63.2	54.8	82.8	87.7	56.9	31.8	46.1	75.92	0.34



Method \ Dataset	CUB-200-2011	NABirds	Oxford Flowers	Stanford Dogs	Stanford Cars	Mean	Params.(%)
Full [14]	87.3	82.7	98.8	89.4	84.5	88.54	100.00
Linear [14]	85.3	75.9	97.9	86.2	51.3	79.32	0.21
BitFit [36]	88.4	84.2	98.8	91.2	79.4	88.40	0.33
Adapter [11]	87.1	84.3	98.5	89.8	68.6	85.66	0.48
AdaptFormer[3]	88.4	84.7	99.2	88.2	81.9	88.48	0.75
LoRA [12]	85.6	79.8	98.9	87.6	72.0	84.78	0.90
VPT-Shallow [14]	86.7	78.8	98.4	90.7	68.7	84.62	0.29
VPT-Deep [14]	88.5	84.2	99.0	90.2	83.6	89.11	0.99
SSF [17]	89.5	85.7	99.6	89.6	89.2	90.72	0.45
GPS [39]	89.9	86.7	99.7	92.2	90.4	91.78	0.77
TR-PTS (Ours)	90.0	87.1	99.6	92.4	90.6	91.94	0.60

<FGVC Results>

08

Triad: LMM-based Anomaly Detection with Expert-guided RoI Tokenizer

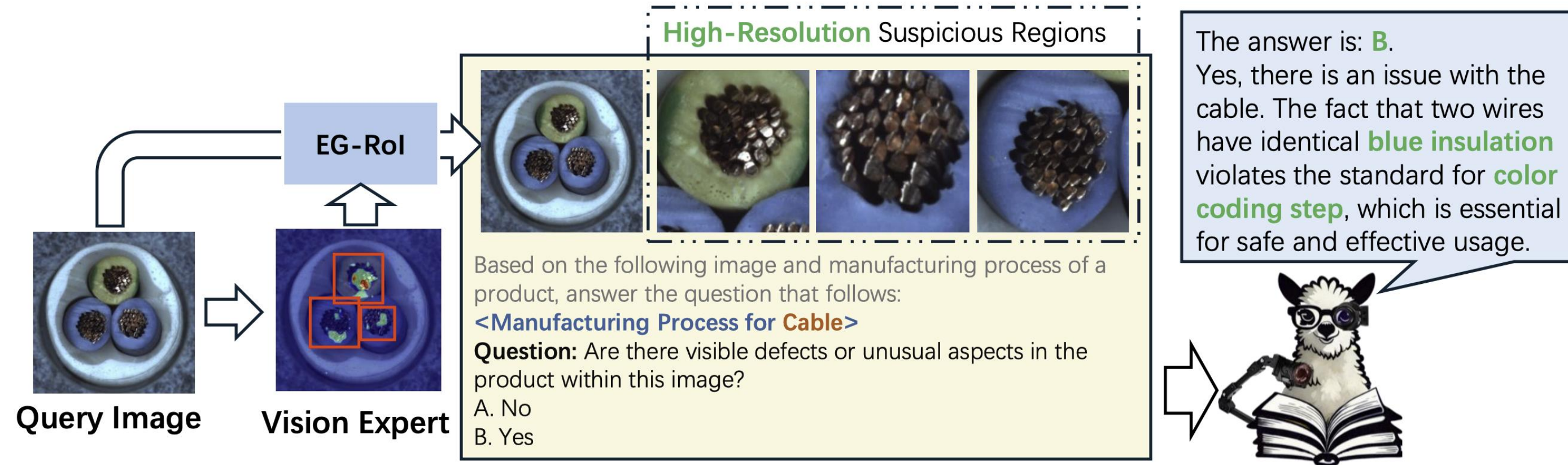
Yuanze Li, Shihao Yuan, Haolin Wang, Qizhang Li, Ming Liu, Chen Xu, Guangming Shi, Wangmeng Zuo

ICCV 2025

Triad: LMM-based Anomaly Detection with Expert-guided RoI Tokenizer

Problem

- Generalist LMMs lack defect cognition and fail to attend to defect regions.
- Existing methods identify defect patterns but cannot reason about their underlying manufacturing cause.

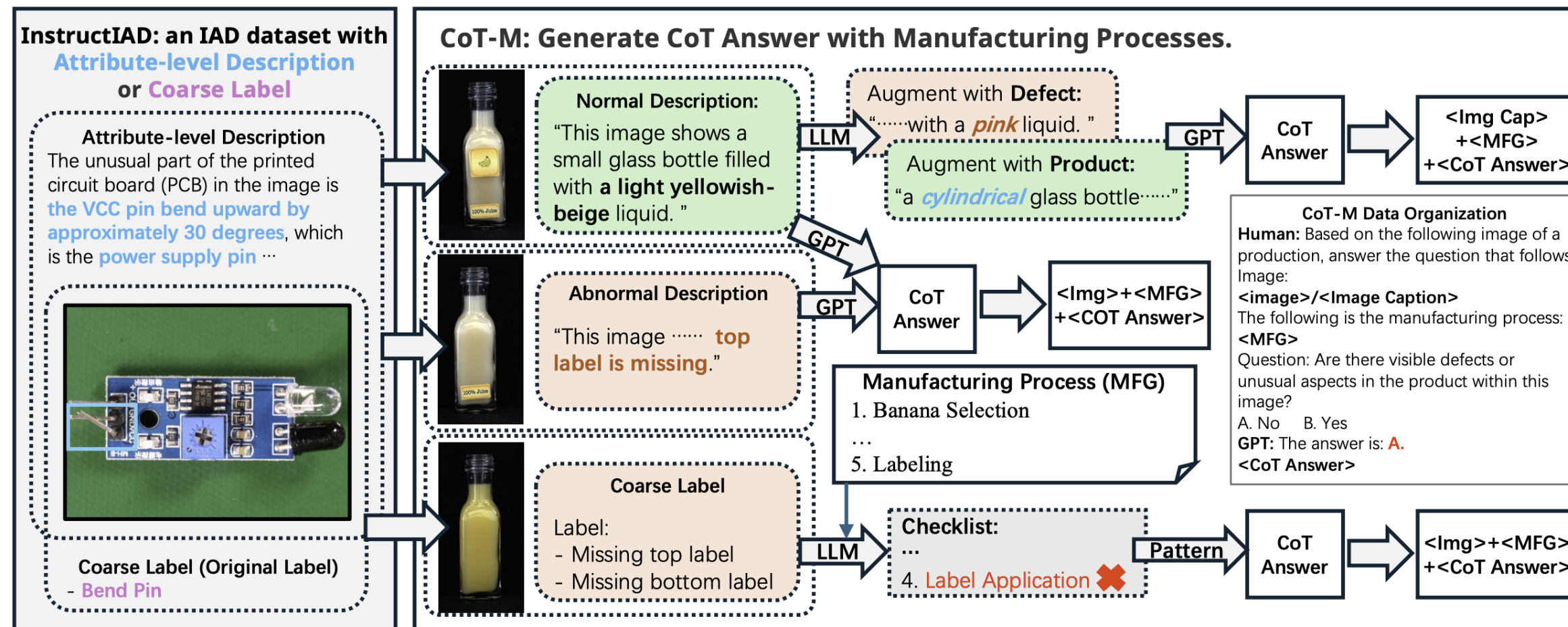


<Main workflow of Triad>

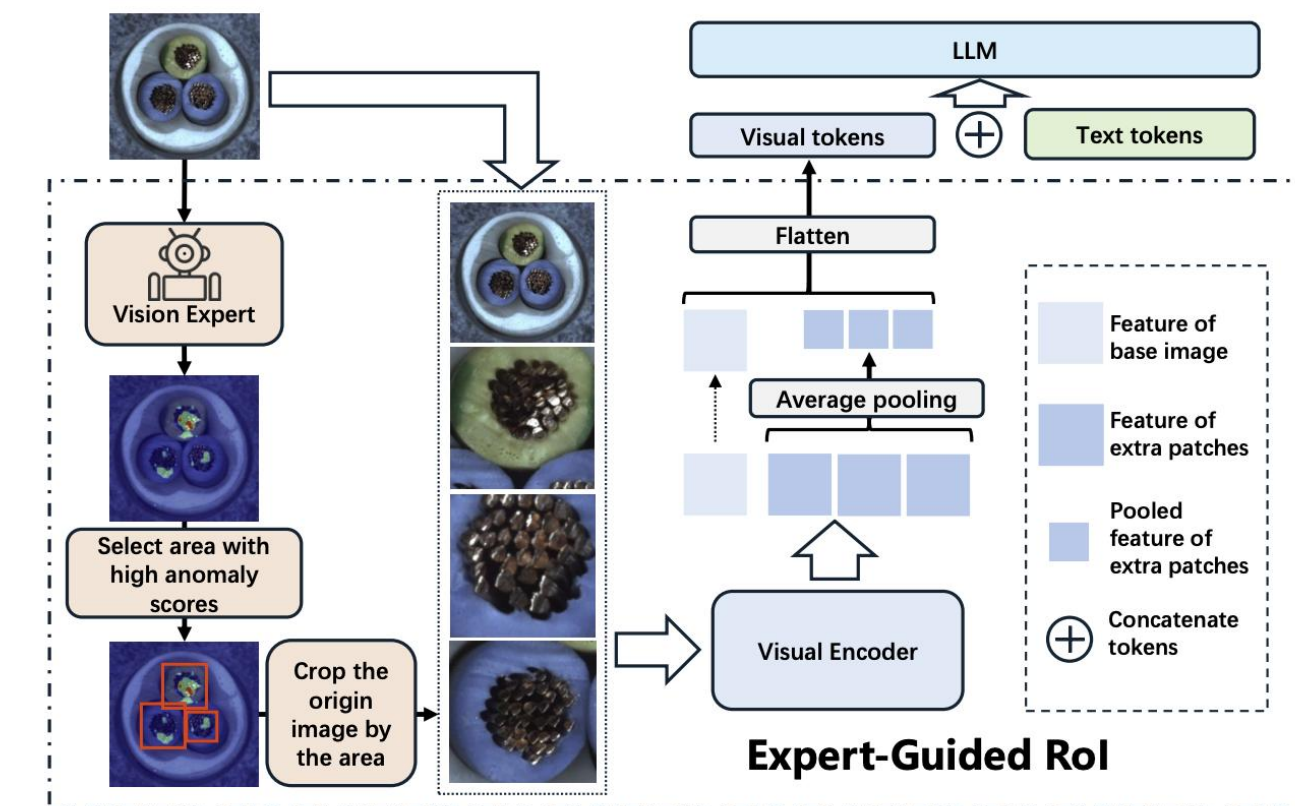
Triad: LMM-based Anomaly Detection with Expert-guided RoI Tokenizer

Method

- EG-RoI tokenizer — a modified AnyRes that injects expert-detected suspicious regions as additional visual tokens.
- InstructIAD + CoT-M — a manufacturing-grounded chain-of-thought data organization.



<Overview of the proposed CoT-M data organization pipeline>

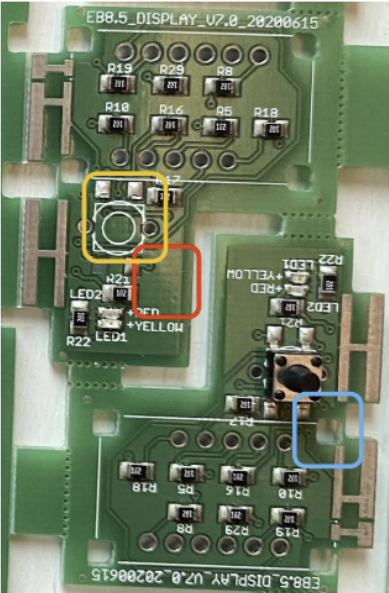


<The whole workflow of EG-RoI>

Triad: LMM-based Anomaly Detection with Expert-guided RoI Tokenizer

Results & Contribution

- EG-RoI attains competitive accuracy against current LMMs; CoT-M yields further gains when the manufacturing process is provided.
- Introduces the domain-knowledge axis, at the cost of an external expert dependency.



Human: Based on the following image of a product, answer the question that follows:
Image:
<image>
Question: Can you confirm if the product has any defects?

Manufacturing Process for PCB:

- Raw PCB Manufacturing:**
 - Substrate Preparation: The PCB starts with a raw fiberglass substrate (FR4 or similar), coated with copper on both sides.
 - Copper Etching: Using a subtractive process, unwanted copper is removed to create the copper traces that will connect components. This is done through chemical etching based on the design pattern.
 - Drilling and Plating: Holes are drilled to accommodate through-hole components (if any), and these holes are plated with copper to create electrical connections between different layers (if it is a multilayer PCB).
- Solder Mask Application:**
 - A solder mask, typically blue in this case, is applied to the board. This mask protects the copper traces and defines the areas where solder will be applied for mounting components.
- Silkscreen Printing:**
 - The silkscreen layer is printed onto the PCB, marking component locations, labels, and orientations (e.g., 'B+', 'B-', 'M+', etc.).
- Solder Paste Application:**
 - Solder paste is applied to the appropriate pads on the PCB where surface-mounted components will be placed, using a stencil aligned with the board.
- Pick and Place:**
 - An automated pick-and-place machine picks the surface-mounted components (like the IC, resistors, capacitors, LEDs, and two black tactile button) and places them onto the solder-pasted pads based on the design.
- Reflow Soldering:**
 - The entire PCB is passed through a reflow oven. Here, the solder paste melts and creates strong electrical and mechanical connections between the surface-mounted components and the board.
- Through-Hole Component Insertion (if applicable):**
 - Components like the micro USB port may be through-hole and manually or automatically inserted into the pre-drilled holes in the PCB.
- Wave Soldering (if applicable):**
 - If through-hole components are used, the board is passed through a wave soldering machine, where a wave of molten solder coats the exposed component leads to establish the electrical connections.
- Component Soldering (Manual/Additional):**
 - Some specialized components that cannot be attached using automated processes, such as certain connectors, may be soldered manually after the reflow or wave soldering.

Industry Anomaly Detection (IAD)

The answer is: **A.**

Based on the image provided, the unusual part of the printed circuit board (PCB) is the **missing component** in the **lower right corner**. This missing component is a **rectangle-shaped component**, which appears to be a capacitor...

LLaVA-OneVision-ov:

... it is **not possible to definitively confirm** ... a thorough inspection would be required ...

GPT4:

...**Answer: A. Yes** ... Defects observed include: ... **areas with inconsistent soldering** ... Some components appear to be slightly **misaligned** ...

Qwen2-VL-72B:

There are **no visible defects** on the product.

Myriad: The answer is: **No** ...

+ Manufacturing Process (MDIAD)

The answer is: **A. Yes**, there are two defects on the PCB. **Missing component:** The black button ... located in the **upper right corner** ... because there is a **white square mark** ... **Missing capacitor:** ...as indicated by the white square mark and **two empty solder joints** ...

LLaVA-OneVision-ov:

Based on the image provided, there are **no visible defects** in the PCB...

GPT4:

Based on the production process ..., I **do not observe** any visible defects in the product.

Qwen2-VL-72B:

Based on the provided production process and ..., there are **no apparent defects** in the product.

Myriad: The answer is: **No** ...

<Comparison with and without manufacturing processes on a PCB7 example from PCB-Bank>

09

IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection

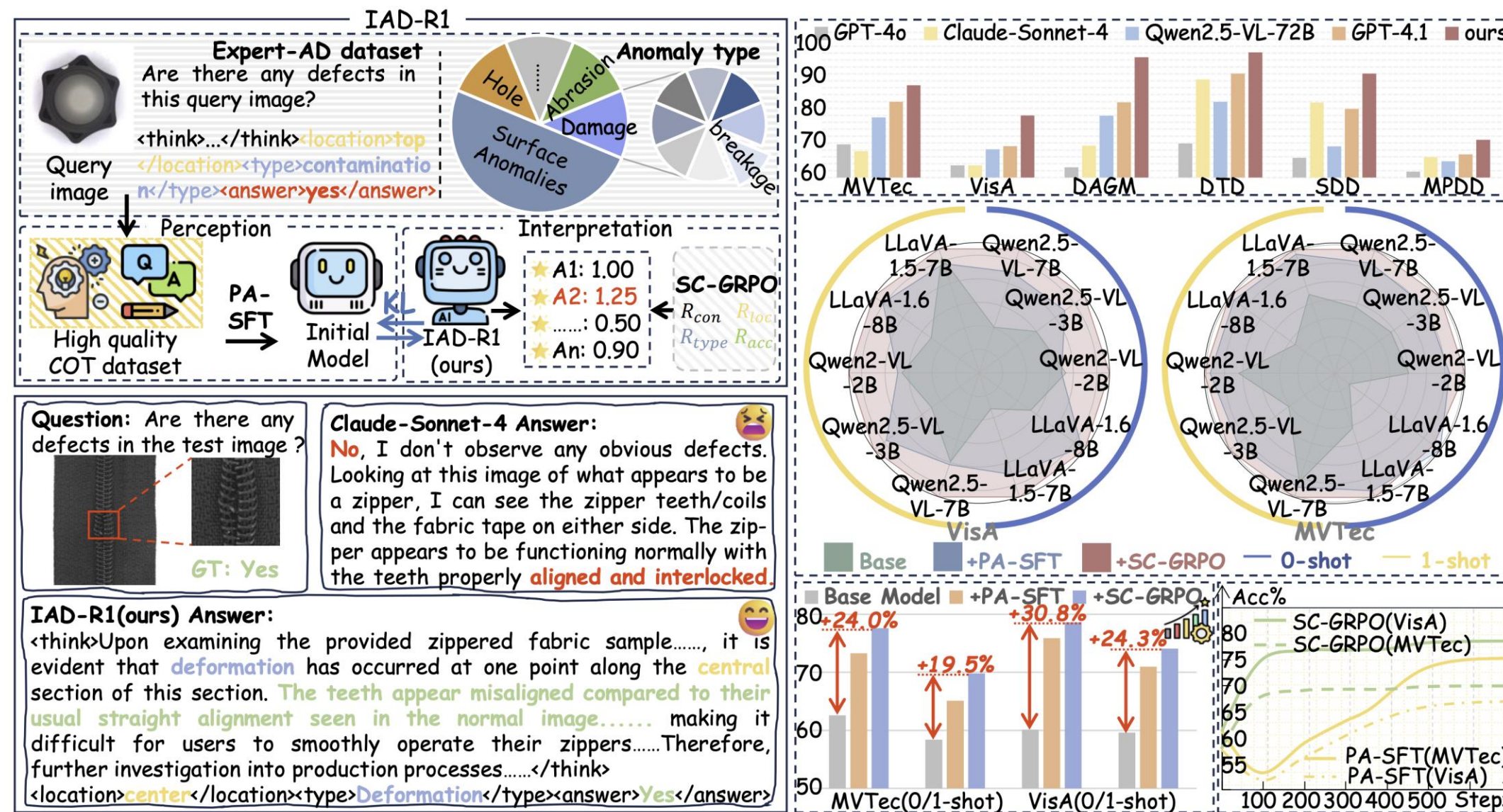
Yanhui Li, Yunkang Cao, Chengliang Liu, Yuan Xiong, Xinghui Dong, Chao Huang

AAAI 2026

IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection

Problem

- Vision-expert-assisted approaches are upper-bounded by the expert's performance.
- Naive end-to-end fine-tuning suffers from reasoning-answer inconsistency.

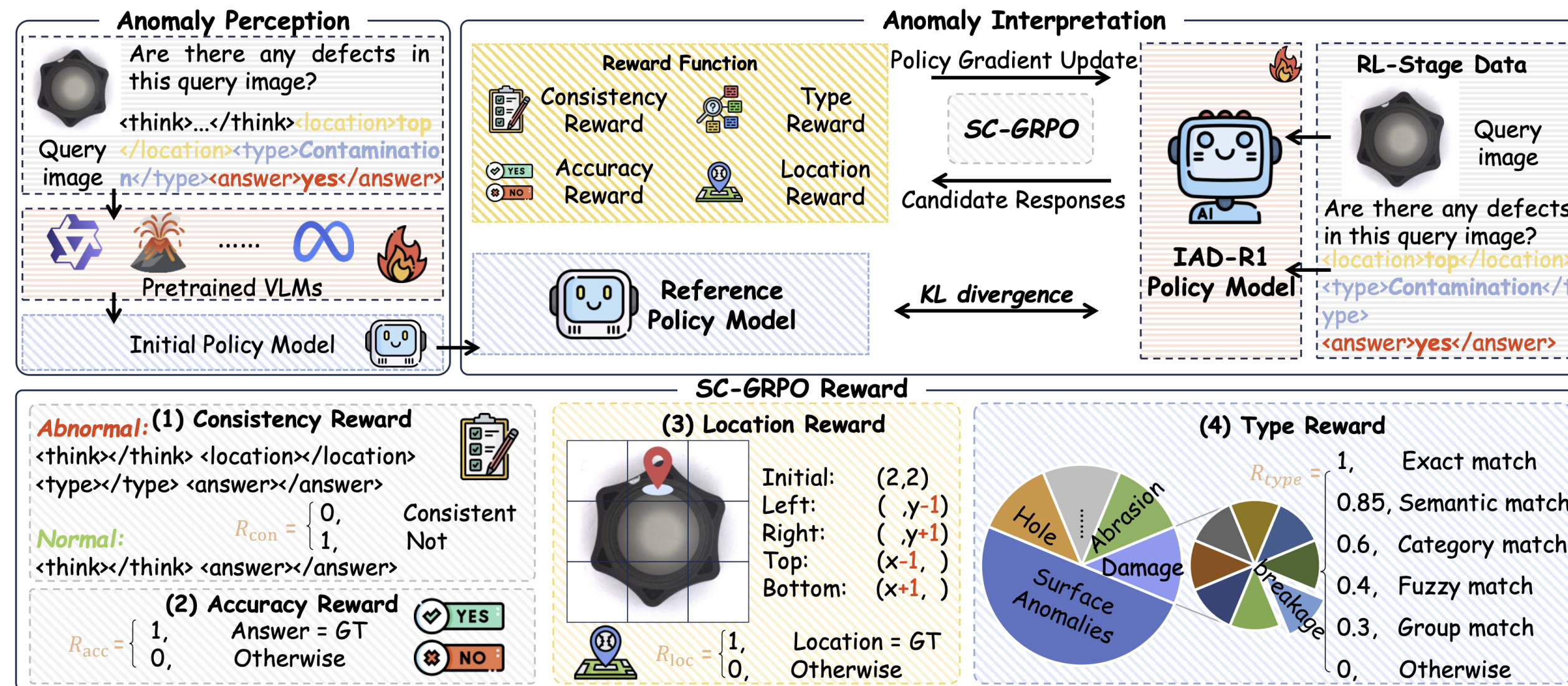


<Overview of the IAD-R1>

IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection

Method

- Architecture-agnostic post-training: PA-SFT activates anomaly perception and establishes the reasoning–answer correlation via the Expert-AD CoT dataset.
- SC-GRPO then applies structured rewards to drive a “perception → interpretation” leap.



<Method Overview>

IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection

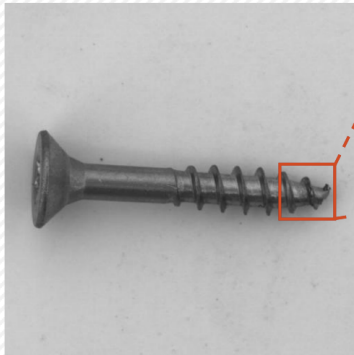
Results & Contribution

- Consistent gains across 7 VLMs (avg. accuracy +43.3% on DAGM at 0.5B)
- SFT+RL emerges as the IAD post-training standard, removing expert dependency through RL-based reasoning.

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	82.0	57.0	59.6	75.4	76.8	55.1	67.7
	Anomaly-OV[CVPR 2025]	7B	74.3	70.3	74.3	77.5	90.7	88.7	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	93.3	95.5	88.6	83.8
	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	75.4	85.2	90.8	83.4	80.4
	IAD-R1(LLaVA-OneVision-SI)	7B	86.7	70.9	78.0	94.8	96.2	90.1	86.1

<Performance comparison of different models on industrial workpieces and surface texture benchmarks>

Question: Are there any defects in the test image ?



GT: YES

The image shows a screw, and upon visual inspection, there are **no obvious defects** such as cracks, bends, or missing threads. However, the tip of the screw appears slightly damaged or worn, which could be considered a defect depending on the intended use and quality standards. For a more thorough analysis, especially in terms of material integrity and precise measurements, specialized equipment and tests would be necessary.

Qwen3

Looking at this screw in the image, I can see what appears to be some surface irregularities or potential defects. There seem to be **some spots or marks on the screw's surface that could indicate corrosion, staining, or manufacturing defects**. The head of the screw and parts of the threaded shaft show some discoloration or surface imperfections. Yes, there appear to be defects in this test image.

Claude-Sonnet-4

<think>During my thorough inspection of this screw, focusing particularly on **its right side** due to prior guidance regarding expected locations..... Upon closer observation near the upper-right section, it appears there's evidence suggesting some **form of damage** has occurred—possibly indicating **wear or fracture** at those specific points along the thread pattern. This **irregularity** stands out against our established standards which require smooth continuity across all components without visible anomalies like cracks or deformities. Therefore requiring further investigation into potential causes leading up until now observed state. </think> <location>**right**</location> <type>**Damage**</type><answer>**Yes**</answer>

IAD-R1(ours)

<Comparison of model output>

10

ReAL: Reasoning-Driven Anomaly Detection and Localization with Image-Level Supervision

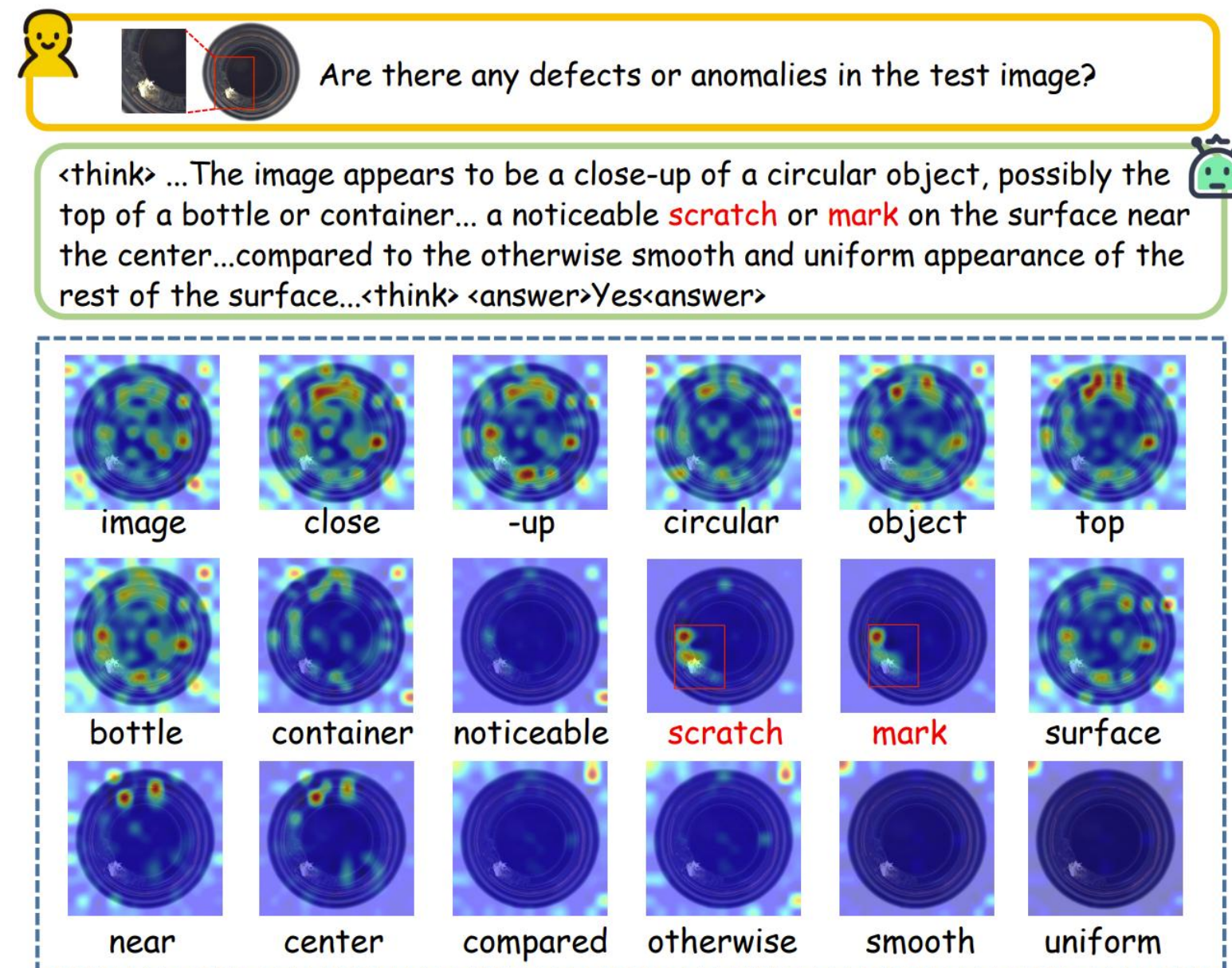
Yizhou Jin, Yuezhu Feng, Jinjin Zhang, Peng Wang, Qingjie Liu, Yunhong Wang

CVPR 2026

ReAL: Reasoning-Driven Anomaly Detection and Localization with Image-Level Supervision

Problem

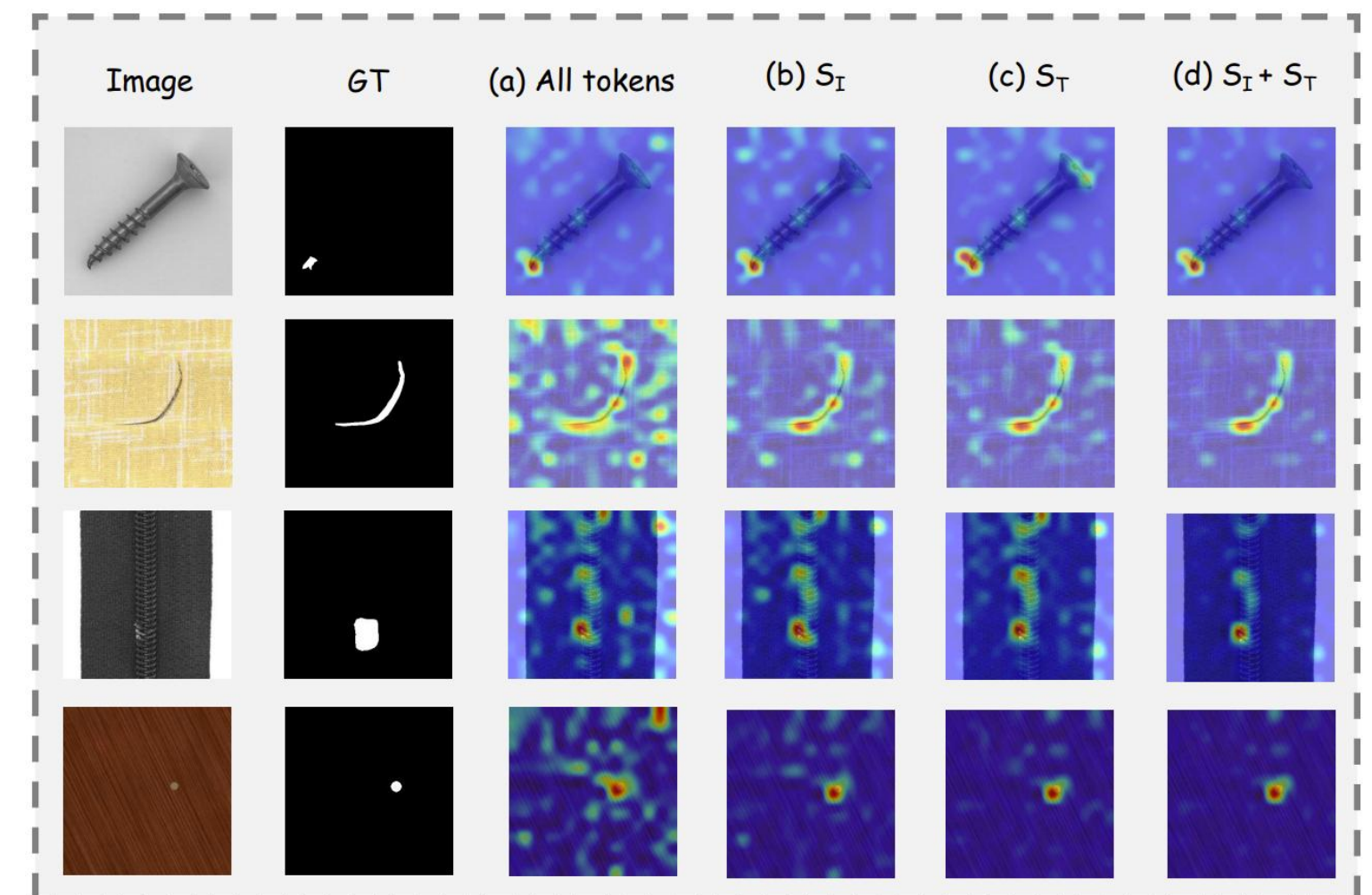
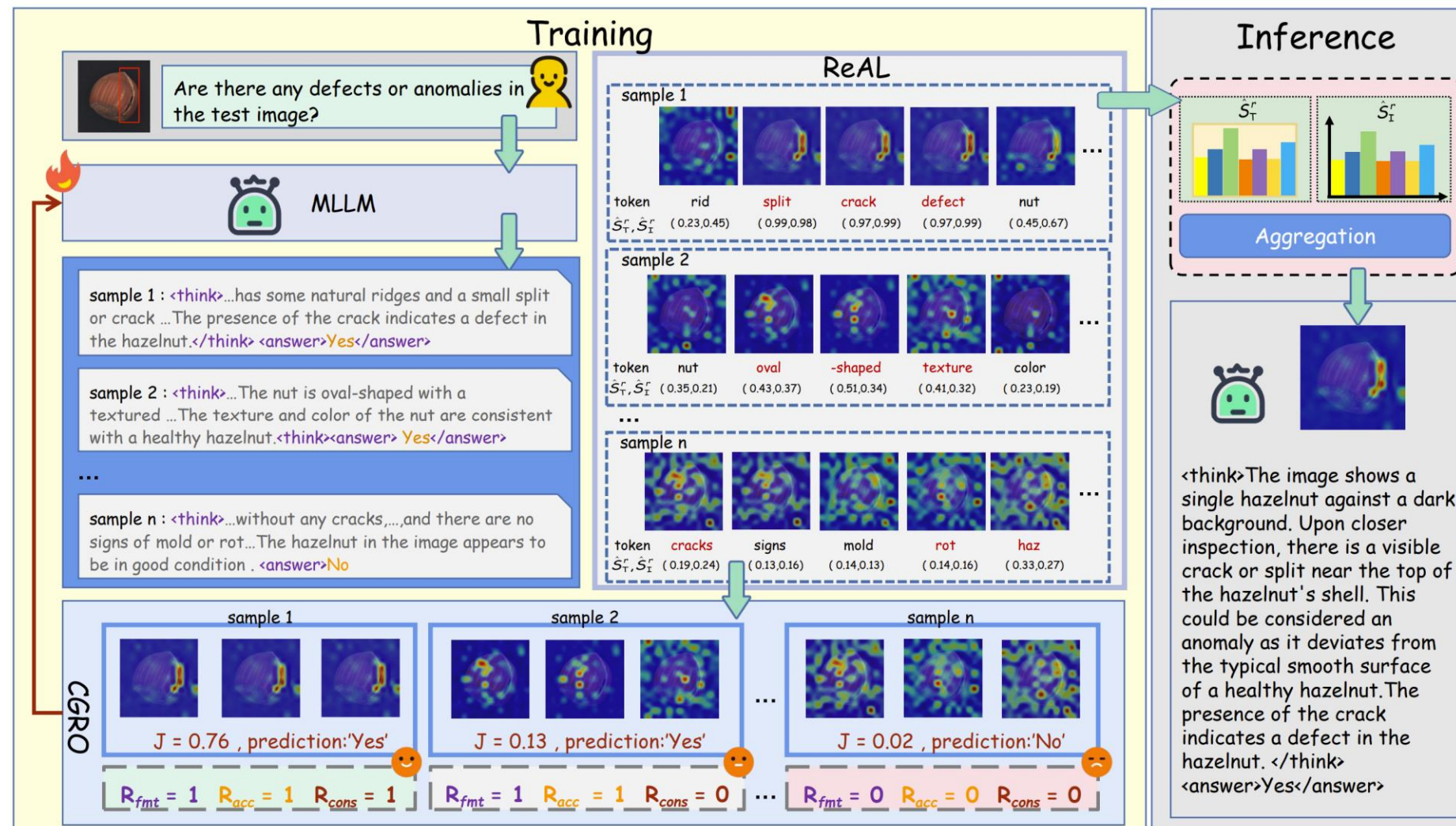
- Pixel-level localization in MLLM IAD relies on external modules (SAM, vision experts) and dense masks.
- Observation: only a few reasoning tokens (e.g., "scratch") already attend precisely to the true anomaly regions.



ReAL: Reasoning-Driven Anomaly Detection and Localization with Image-Level Supervision

Method

- ReAL selects tokens by semantic relevance and spatial entropy, then aggregates weighted attention into a pixel anomaly map.
- CGRO & GRPO with a Jaccard-consensus consistency reward (plus accuracy/format)



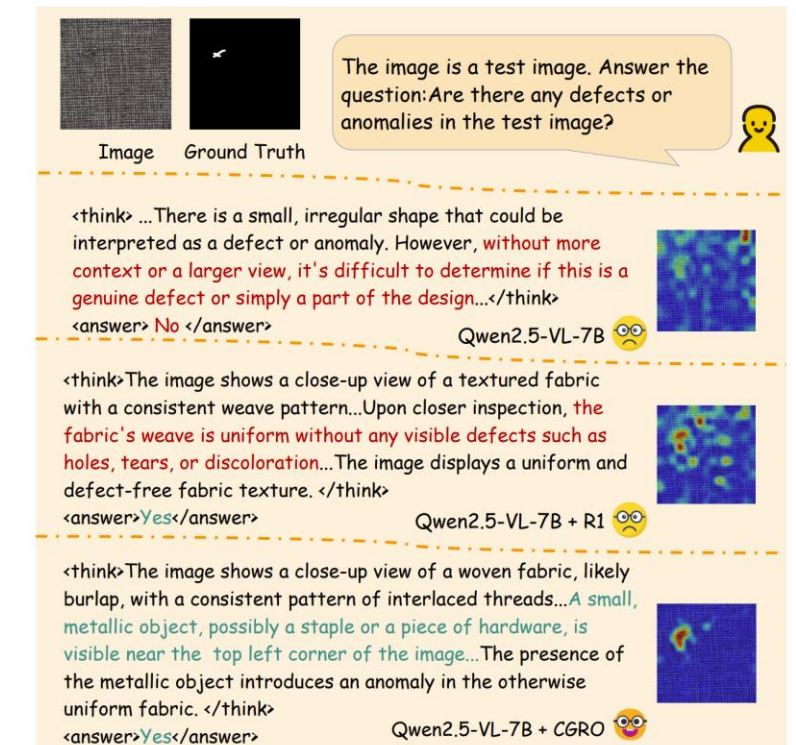
ReAL: Reasoning-Driven Anomaly Detection and Localization with Image-Level Supervision

Results & Contribution

- With image-level supervision only, competitive with dense-supervision methods on 4 benchmarks (ablation: pixel AUROC 72.7→80.7, image AUROC 63.4→83.9).
- Unifies detection, localization, and reasoning within a single MLLM — the endpoint of the 08→09→10 arc.

Model	Param.	Anno.	Image-level					Pixel-level					Reasoning
			AVG	SDD	DTD	WFDD	MVTec	AVG	SDD	DTD	WFDD	MVTec	MVTec-COT
GPT-4o-mini [15]	/	/	70.9, 71.4	66.2, 92.8	79.4, 70.3	66.4, 63.5	71.5, 59.0	N/A	N/A	N/A	N/A	N/A	19.0, 65.2
GPT-4o [15]	/	/	79.4, 85.3	77.5, 90.7	89.2, 90.2	78.2, 78.6	72.7, 81.7	N/A	N/A	N/A	N/A	N/A	19.9, 60.4
GPT-4.1	/	/	87.2, 88.4	84.4, 92.1	94.3, 94.3	86.8, 86.0	83.3, 81.2	N/A	N/A	N/A	N/A	N/A	20.8, 69.9
GPT-5-nano	/	/	80.3, 81.6	82.6, 88.6	90.4, 87.2	80.2, 79.7	68.0, 70.7	N/A	N/A	N/A	N/A	N/A	14.0, 50.0
Gemini-2.5-Flash [9]	/	/	79.9, 76.3	71.8, 49.5	88.6, 91.3	82.6, 82.2	76.5, 82.3	N/A	N/A	N/A	N/A	N/A	17.6, 68.1
LLaVA-OneVision-SI [23]	7B	/	78.3, 75.7	65.5, 67.8	91.5, 89.9	77.0, 75.0	79.4, 70.1	N/A	N/A	N/A	N/A	N/A	10.5, 39.1
LLaVA-OneVision-OV [23]	7B	/	78.5, 81.4	68.1, 94.7	92.4, 89.1	75.5, 73.4	77.8, 68.5	N/A	N/A	N/A	N/A	N/A	23.7, 70.5
LLaVA-1.6 [30]	7B	/	62.9, 63.5	67.0, 50.4	70.2, 74.2	59.5, 60.6	54.8, 68.9	N/A	N/A	N/A	N/A	N/A	18.6, 54.2
Qwen2.5-VL [3]	3B	/	52.9, 56.6	46.3, 69.9	64.1, 51.3	50.0, 52.3	51.0, 52.8	N/A	N/A	N/A	N/A	N/A	16.8, 49.0
Qwen2.5-VL [3]	7B	/	63.4, 63.3	65.0, 93.8	68.3, 56.2	60.5, 57.2	59.6, 45.8	N/A	N/A	N/A	N/A	N/A	20.8, 65.5
Qwen2.5-VL [3]	32B	/	74.6, 71.9	82.4, 90.1	81.0, 73.0	68.4, 70.5	66.5, 54.0	N/A	N/A	N/A	N/A	N/A	20.0, 66.1
Qwen2.5-VL [3]	72B	/	81.1, 80.4	84.5, 95.6	91.2, 87.6	74.5, 72.8	74.1, 65.5	N/A	N/A	N/A	N/A	N/A	21.1, 67.5
Qwen3VL-thinking [43]	8B	/	78.8, 73.0	70.5, 52.1	94.6, 92.3	75.7, 74.3	74.4, 73.5	N/A	N/A	N/A	N/A	N/A	19.3, 61.9
Qwen3VL-thinking-A3B [43]	30B	/	82.6, 78.7	69.5, 55.3	96.4, 94.9	85.5, 84.5	79.0, 80.2	N/A	N/A	N/A	N/A	N/A	16.9, 58.9
InternVL-2.5 [8]	8B	/	69.8, 72.9	63.7, 96.3	81.0, 72.8	65.9, 63.1	68.7, 59.3	N/A	N/A	N/A	N/A	N/A	19.7, 57.5
InternVL-3 [55]	78B	/	81.0, 83.4	82.5, 88.6	91.4, 92.0	75.2, 75.9	74.7, 77.2	N/A	N/A	N/A	N/A	N/A	22.2, 70.1
MiniCPM-V-2.6 [48]	8B	/	70.3, 70.6	70.2, 97.1	76.8, 66.3	67.0, 64.2	67.0, 54.8	N/A	N/A	N/A	N/A	N/A	21.7, 68.3
Triad [26]	7B	T+I	<u>85.5</u> , 83.8	79.8, 68.1	95.2, 95.4	82.5, 82.4	84.7, 89.4	N/A	N/A	N/A	N/A	N/A	8.6, 35.9
IAD-R1(Qwen2.5-VL) [27]	3B	T+I	/, /	/, 83.4	/, 89.1	/, /	/, 77.6	N/A	N/A	N/A	N/A	N/A	/, /
IAD-R1(Qwen2.5-VL) [27]	7B	T+I	/, /	/, 85.4	/, 90.8	/, /	/, 81.9	N/A	N/A	N/A	N/A	N/A	/, /
AnomalyGPT [11]	7B	T+I+P	71.1, 53.9	72.7, 13.0	75.7, 74.2	71.0, 56.2	65.1, 72.2	77.8, 98.4	75.1, 99.9	83.5, 98.9	72.4, 98.6	80.2, 96.4	11.9, 36.7
EIAD [52]	7B	T+I	68.4, 73.3	52.7, 87.2	83.4, 74.8	59.6, 56.1	77.7, 75.2	57.6, 91.4	52.1, 99.7	61.2, 85.9	53.7, 96.7	63.4, 83.5	21.1, 65.7
Qwen2.5-VL* [3]	3B	/	52.9, 56.6	46.3, 69.9	64.1, 51.3	50.0, 52.3	51.0, 52.8	55.6, 57.2	45.4, 10.6	71.4, 91.8	52.3, 79.9	53.4, 46.5	16.8, 49.0
Qwen2.5-VL+R1* [12]	3B	I	58.1, 60.5	54.1, 94.8	57.9, 38.0	55.0, 51.2	65.3, 57.8	72.6, 88.0	74.4, 99.9	85.6, 98.5	63.2, 97.0	67.1, 56.7	24.2, 71.7
Qwen2.5-VL+CGRO*	3B	I	67.6, 72.4	67.9, 82.5	78.3, 74.4	61.4, 59.9	62.7, 72.6	73.5, 94.3	75.2, 99.8	86.1, 98.4	62.1, 97.5	70.6, 81.5	25.6, 72.4
Qwen2.5-VL* [3]	7B	/	63.4, 63.3	65.0, 93.8	68.3, 56.2	60.5, 57.2	59.6, 45.8	61.7, 85.6	56.6, 99.8	76.0, 98.5	56.5, 70.3	57.5, 73.8	20.8, 65.5
Qwen2.5-VL+R1* [12]	7B	I	80.0, 82.0	77.3, 97.2	90.7, 87.2	74.9, 74.2	77.1, 69.3	78.5, 96.7	80.8, 99.8	92.7, 98.6	67.0, 97.6	73.6, 90.9	26.3, 73.8
Qwen2.5-VL+CGRO*	7B	I	83.9, <u>86.9</u>	81.4, 97.0	94.5, 93.6	79.9, 79.8	79.8, 77.3	80.7, <u>97.1</u>	82.3, 99.8	94.2, 98.6	70.5, 97.8	75.6, 92.2	27.1, 74.7

Model	Image-level			Pixel-level		
	AUROC	AUPR	ACC	AUROC	AUPR	ACC
Vanilla	63.4	59.4	63.3	64.7	2.6	73.0
Vanilla + ReAL	63.4	59.4	63.3	61.7	5.1	85.6
Vanilla + CGRO	83.9	76.7	86.9	<u>72.7</u>	<u>5.9</u>	<u>92.6</u>
Full	83.9	76.7	86.9	80.7	13.3	97.1



Thank you

Question & Discussion



RILS LAB