MOPDA: MIXED-TRAJECTORY ON-POLICY DISTILLATION FOR LANGUAGE-GUIDED INDUSTRIAL ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Large vision-language models (LVLMs) have recently shown strong potential for industrial anomaly detection (IAD) by providing image-level anomaly judgments and interpretable defect reasoning. However, reliably translating generated language judgments into precise pixel-level anomaly localization remains challenging. To address this, we propose **M**ixed-Trajectory **O**n-**P**olicy **D**istillation for Language-Guided Industrial **A**nomaly Detection (**MOPDA**), the first framework to introduce on-policy self-distillation into LVLM-based IAD. Specifically, MOPDA introduces **Mixed-Trajectory Supervision**, which combines student-generated on-policy trajectories with teacher-generated evidence-conditioned trajectories under a shared token-level distillation objective for fine-grained judgment learning. To connect the learned judgment with dense anomaly perception, we further introduce **Language-guided Visual Anchoring**. The final judgment is used as a compact semantic condition to construct image-specific normal and abnormal anchors, which are contrasted with dense visual features to produce the anomaly map. In this way, language guides localization without directly determining the pixel-level response. Experiments on five IAD benchmarks show that MOPDA outperforms the evaluated LVLM-based baselines on most detection, localization, and judgment metrics, while remaining competitive with CLIP-based methods.

1 INTRODUCTION

Industrial anomaly detection (IAD) aims to identify defective products and highlight anomalous regions, which is essential for visual quality inspection. Existing IAD methods have established strong dense detection baselines through CLIP-style representations, prompt designs, and visual foundation features (Jeong et al., 2023; Zhou et al., 2024; Cao et al., 2024; Li et al., 2024; Hou et al., 2026). Though these methods have made substantial progress, they do not generate natural-language explanations of what defects are present, where they occur, or what visual evidence supports their predictions.

Recent LVLM-based IAD methods introduce generative vision-language models into IAD, enabling models to understand industrial images, answer defect-related questions, and generate image-specific anomaly judgments and explanations (Gu et al., 2024b; Li et al., 2026b; Xu et al., 2025; Chao et al., 2025; Li et al., 2026a). However, reliably converting these generated language judgments into precise pixel-level anomaly maps remains underexplored in LVLM-based IAD.

This limitation appears in two forms in current localization pathways, as illustrated in Figure 1. In decoupled designs (Figure 1a), language generation and anomaly localization follow separate pathways, so the generated language does not directly inform heatmap prediction (Gu et al., 2024b; Li et al., 2026b). In language-coupled designs (Figure 1b), localization is derived more directly from language-side grounding signals, such as segmentation or semantic-anchor tokens, mask text, box-like outputs, and reasoning-token attention (Qu et al., 2026; Deng et al., 2026; Jin et al., 2026; Zhao et al., 2025). Consequently, changes or errors in these language-side signals can propagate to the localization output, making the predicted heatmap sensitive to language-side variations. This motivates the design in Figure 1c, where the final judgment provides compact semantic guidance for localization, while pixel-level anomaly scores remain grounded primarily in dense visual features.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

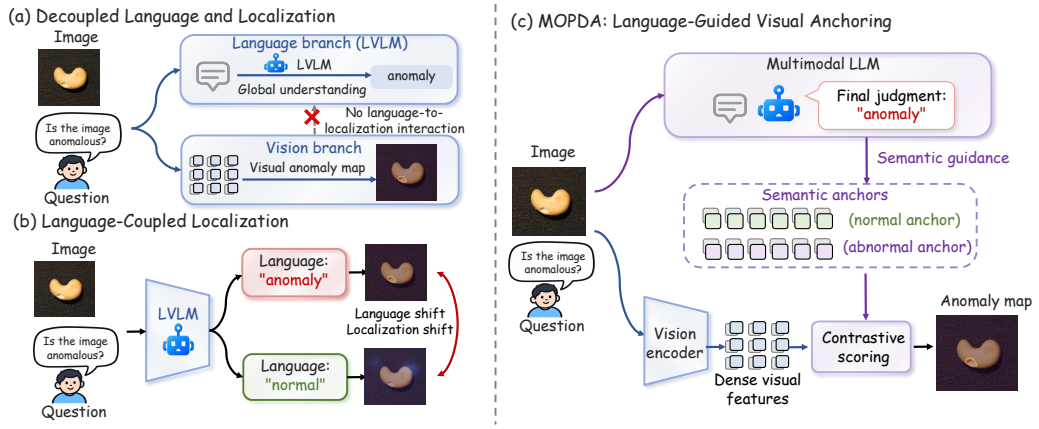

Figure 1: **Localization pathways in LVLM-based IAD.** (a) Decoupled methods separate language generation from anomaly localization. (b) Language-coupled methods derive localization from language-side signals, making localization sensitive to shifts in these signals. (c) MOPDA uses the final judgment as semantic guidance for normal/abnormal anchors, while dense visual features ground contrastive pixel-level anomaly scoring.

Furthermore, using generated judgments to guide localization introduces a distinct training challenge. Since the judgment is produced by the model itself and then used as a semantic condition for localization, its training should satisfy two requirements. First, supervision should target the model's own generation trajectory, which is the trajectory used to produce the judgment at deployment. Second, supervision should be fine-grained so that the judgment captures defect evidence rather than merely optimizing a final-answer reward.

Existing LVLM training paradigms each satisfy only one side of these requirements. Standard SFT provides dense token-level supervision, but learns from fixed offline responses and therefore does not supervise the states visited by the model at deployment. Reinforcement learning with verifiable rewards (RLVR), in contrast, samples on-policy rollouts but typically relies on sparse outcome-level signals, such as answer correctness, format, or coarse localization (Shao et al., 2024; Chao et al., 2025; Li et al., 2026a). Thus, SFT is fine-grained but off-policy, whereas RLVR is on-policy but sparse. On-policy distillation bridges this gap by applying token-level teacher supervision to student-generated on-policy rollouts, combining on-policy training with fine-grained feedback (Agarwal et al., 2024; Gu et al., 2024a; Zhao et al., 2026). However, in IAD, subtle defect evidence can lead the current student to incomplete or incorrect judgment trajectories, limiting student-only on-policy distillation to response paths sampled from the current policy.

Motivated by this limitation, we propose **M**ixed-Trajectory **O**n-**P**olicy **D**istillation for Language-Guided Industrial **A**nomaly Detection (**MOPDA**), which is the first framework to introduce on-policy self-distillation into LVLM-based IAD. For judgment learning, MOPDA introduces **Mixed-Trajectory Supervision**, which combines student on-policy rollouts with evidence-conditioned teacher rollouts. The student trajectories preserve supervision on deployment-relevant states, while the teacher trajectories complement them with evidence-conditioned judgment paths that the current student may not reliably generate.

Furthermore, we design **Language-guided Visual Anchoring** as a concrete mechanism for this guidance. Instead of using the full chain-of-thought as the localization context, MOPDA keeps only the final anomaly judgment and uses a judgment reforward to re-encode the image, question, and judgment into normal and abnormal anchor representations. A contrastive heatmap head then scores dense visual patch features against these anchors to produce the anomaly map. This design keeps pixel-level anomaly responses grounded in visual features rather than directly determined by the generated judgment. Figure 2 provides an output-level illustration of this behavior: in both examples, the generated reasoning identifies the visible defect, while the corresponding anomaly map highlights

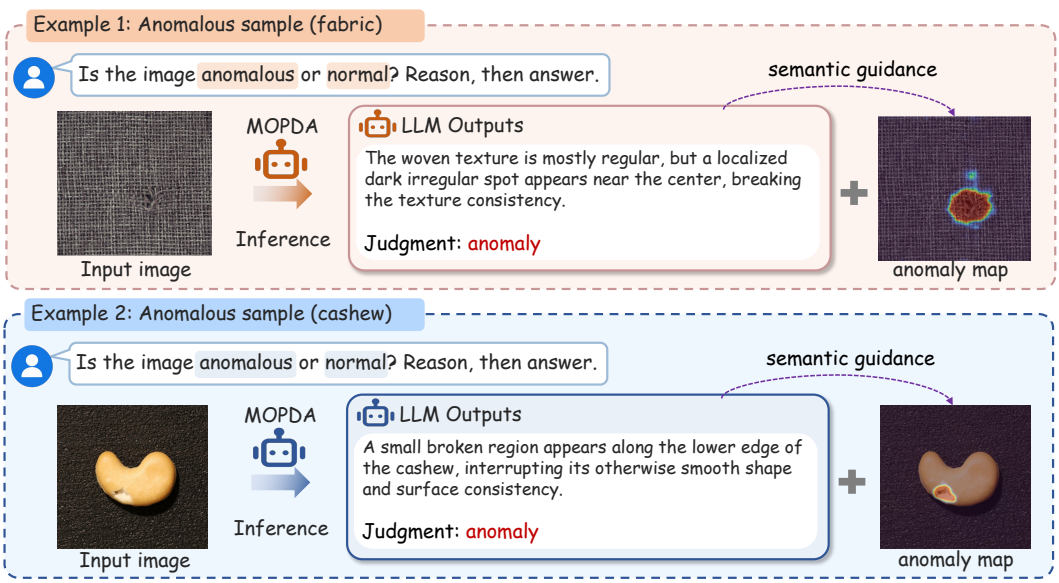

Figure 2: **Language-guided anomaly localization.** MOPDA correctly judges both the fabric (top) and cashew (bottom) as anomalous. Their anomaly maps highlight the visible defects, with the final judgment providing semantic guidance.

the same spatial region, showing qualitative consistency between language judgment and dense localization. Experiments on five IAD benchmarks validate the effectiveness of our framework.

Our contributions are summarized as follows.

1. We present **MOPDA**, the first framework to introduce on-policy self-distillation into LVLM-based IAD, and develop **Mixed-Trajectory Supervision** that combines student-generated on-policy trajectories with evidence-conditioned teacher-generated trajectories.
2. We design **Language-guided Visual Anchoring**, which conditions normal and abnormal anchors on the final judgment while grounding pixel-level anomaly scores in visual features.
3. Experiments on five IAD benchmarks demonstrate competitive visual anomaly detection against CLIP-based methods and improved detection, localization, and judgment performance over the evaluated LVLM baselines.

2 Related Work

2.1 Industrial Anomaly Detection

Zero-shot industrial anomaly detection commonly builds on pretrained vision-language or visual representations. WinCLIP (Jeong et al., 2023) derives image- and pixel-level anomaly scores from local CLIP features and semantic prompts, AnomalyCLIP (Zhou et al., 2024) learns object-agnostic prompts for cross-category generalization, and VisualAD (Hou et al., 2026) emphasizes dense visual matching with reduced dependence on language. These methods provide strong anomaly detection and localization, but do not generate image-specific language judgments.

Recent LVLM-based methods extend IAD toward language-based anomaly understanding. AnomalyGPT (Gu et al., 2024b) uses a separate localization pathway, Myriad (Li et al., 2026b) incorporates visual experts, ReAL (Jin et al., 2026) derives dense responses from reasoning attention, and AG-VAS (Qu et al., 2026) introduces semantic anchor tokens with anchor-guided mask decoding. In contrast, MOPDA uses the generated language judgment as semantic guidance while preserving a visual-feature-grounded localization pathway.

2.2 On-Policy Distillation

On-policy distillation provides token-level teacher supervision on student-generated trajectories, reducing training–inference mismatch (Agarwal et al., 2024; Gu et al., 2024a). Its self-distillation form obtains supervision from the same model under privileged contexts (Zhao et al., 2026), and recent work extends this paradigm to multimodal reasoning. In industrial anomaly detection, ADOPD (He et al., 2026) uses privileged references to improve query-only industrial reasoning.

MOPDA differs in two respects. First, it introduces **Mixed-Trajectory Supervision**, augmenting student-generated on-policy trajectories with evidence-conditioned teacher-generated trajectories under a shared token-level distillation objective. Second, the learned judgment is not only a reasoning output, but is further used to guide dense anomaly localization through Language-guided Visual Anchoring.

3 Method

3.1 Setting and Framework Overview

We study IAD under a cross-dataset setting, training on labeled source data and evaluating on unseen target domains. For each sample, the observable input is $x_i = (x_i^{\text{img}}, x_i^{\text{text}})$, where x_i^{img} is an industrial image and x_i^{text} is a defect-judgment question. The source training set is

$$\mathcal{D}_{\text{src}} = \{(x_i^{\text{img}}, x_i^{\text{text}}, a_i, M_i, r_i)\}_{i=1}^{N}, \tag{1}$$

where a_i is the image-level anomaly label, M_i is the pixel-level anomaly mask, and r_i is a defect-analysis reference containing a verified judgment and supporting evidence. These annotations and references are used only during source-domain training. At target-domain inference, the model receives only the image and question and produces a language response and a pixel-level anomaly map.

As shown in Figure 3, MOPDA uses the final language judgment to contextualize abnormal and normal anchor tokens through a judgment reforward. A contrastive heatmap head scores dense visual features against the resulting anchor representations to produce the anomaly map, whose spatial aggregation gives the image-level anomaly score. During training, Mixed-Trajectory Supervision combines student- and teacher-generated trajectories. Privileged defect evidence is available only to the teacher, while source labels and masks supervise anomaly prediction. Only the student is retained at inference.

3.2 Mixed-Trajectory On-Policy Distillation

The student receives only the image and question, while the teacher additionally observes the training-only defect-analysis reference r_i. Following privileged-context self-distillation (Zhao et al., 2026), their contexts are

$$c_i^S = \text{Prompt}_S(x_i^{\text{img}}, x_i^{\text{text}}), \qquad c_i^T = \text{Prompt}_T(x_i^{\text{img}}, x_i^{\text{text}}, r_i). \tag{2}$$

The teacher shares the student architecture and is maintained by EMA (Tarvainen & Valpola, 2017), as detailed in Appendix A.2.

To construct Mixed-Trajectory Supervision, we sample the trajectory source for each minibatch with $b \sim \text{Bernoulli}(\lambda)$:

$$\hat{y}_i \sim \begin{cases} \pi_\theta(\cdot \mid c_i^S), & b = 1, \\ \pi_{\bar{\theta}}(\cdot \mid c_i^T), & b = 0. \end{cases} \tag{3}$$

Student-generated trajectories retain supervision on response paths explored by the current policy and used at deployment. Teacher-generated trajectories, which form the Seq-KD branch, provide complementary evidence-conditioned judgment paths beyond those sampled from the student policy.

Both trajectory sources are optimized with the same token-level distillation objective. For a sampled response $\hat{y}_i$, the student and teacher evaluate the same response prefix under their respective contexts.

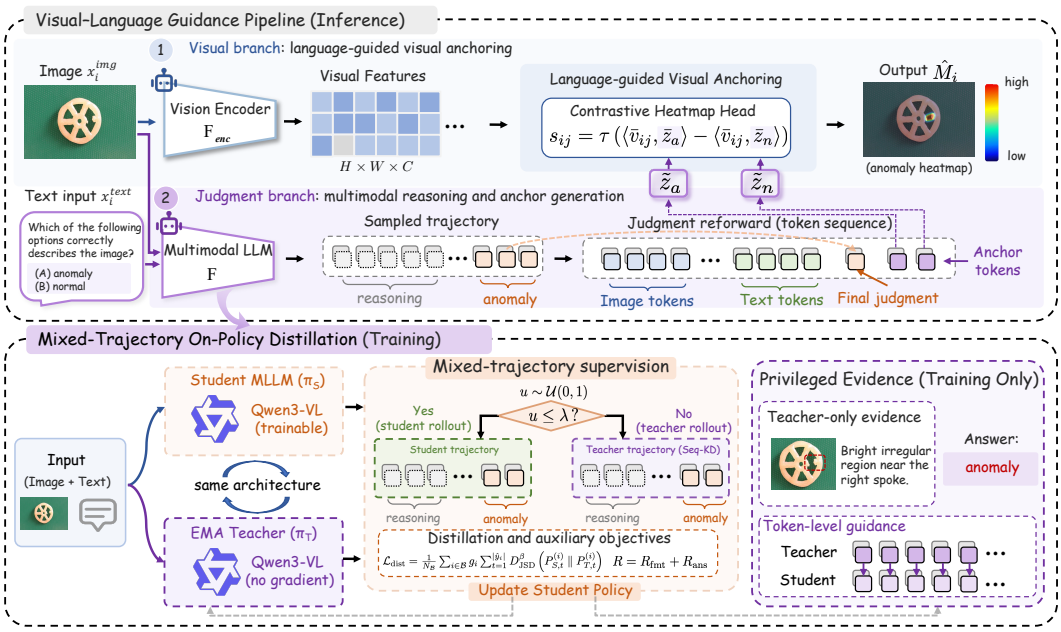

Figure 3: **Overview of MOPDA.** Mixed-Trajectory Supervision uses student-generated on-policy trajectories and evidence-conditioned teacher-generated trajectories (Seq-KD). The student learns from both through a shared token-level distillation objective, while only the teacher observes privileged evidence during training. Language-guided Visual Anchoring uses a judgment reforward to contextualize the base abnormal and normal anchor tokens z_a, z_n into image-specific anchor representations for dense anomaly localization.

Let $P_{S,t}^{(i)}$ and $P_{T,t}^{(i)}$ denote their corresponding next-token distributions at step t. We minimize

$$\mathcal{L}_{\text{dist}} = \frac{1}{N_{\mathcal{B}}} \sum_{i \in \mathcal{B}} g_i \sum_{t=1}^{|\hat{y}_i|} D_{\text{JSD}}^{\beta} \left(P_{S,t}^{(i)} \| P_{T,t}^{(i)} \right), \tag{4}$$

where $g_i = 1$ for student-generated trajectories, while for teacher-generated trajectories g_i acts as a validity gate that retains only responses with the required output format and a final judgment matching the source label. $N_{\mathcal{B}}$ denotes the total number of sampled response tokens in the minibatch. The exact next-token distributions, generalized Jensen–Shannon divergence, validity criterion, and normalization are provided in Appendix A.2.

3.3 Language-guided Visual Anchoring

Language-guided Visual Anchoring uses the final judgment $\hat{a}_i$, extracted from the `<answer></answer>` span of response $\hat{y}_i$, as a compact semantic condition for dense localization. The same judgment is returned as the language output and is also used to construct image-specific abnormal and normal anchor representations. Figure 2 illustrates this inference-time behavior: the model outputs the final judgment while the same judgment conditions the anchors used for dense anomaly localization. Table 4 evaluates this design against alternative localization interfaces.

Specifically, we append two base anchor tokens, denoted by z_a and z_n, representing abnormal and normal semantics, respectively. A judgment reforward with the same LVLM processes the image, question, final judgment $\hat{a}_i$, and the two anchor tokens:

$$(\tilde{z}_{i,a}, \tilde{z}_{i,n}) = f_\theta^{\text{ref}} \left(x_i^{\text{img}}, x_i^{\text{text}}, \hat{a}_i, z_a, z_n \right). \tag{5}$$

Here, $\tilde{z}_{i,a}$ and $\tilde{z}_{i,n}$ are the contextualized hidden states of the abnormal and normal anchor tokens after reforwarding. They therefore encode the image content, question context, and final-judgment semantics for sample i.

The contextualized anchor states are mapped into a shared localization space by two learnable projection heads g_a and g_n for the abnormal and normal anchors, respectively, and are then L2-normalized:

$$\bar{z}_{i,a} = \frac{g_a(\tilde{z}_{i,a})}{\|g_a(\tilde{z}_{i,a})\|_2}, \qquad \bar{z}_{i,n} = \frac{g_n(\tilde{z}_{i,n})}{\|g_n(\tilde{z}_{i,n})\|_2}. \tag{6}$$

Let $v_{i,p}$ denote the dense visual feature of patch p in sample i. A lightweight visual refinement module maps these features into the same localization space as the anchors, followed by L2 normalization to obtain $\bar{v}_{i,p}$. The anomaly logit for each patch is then computed by contrasting its similarity to the abnormal and normal anchors:

$$s_{i,p} = \tau \left(\langle \bar{v}_{i,p}, \bar{z}_{i,a} \rangle - \langle \bar{v}_{i,p}, \bar{z}_{i,n} \rangle \right), \tag{7}$$

where the learnable temperature τ multiplies the similarity difference. Upsampling the spatial logits and applying a sigmoid yields the pixel-level anomaly map $\hat{M}_i$. The detailed head architecture, projection dimensions, normalization, and image-score aggregation are provided in Appendix A.1.

3.4 Training Objective

Training jointly optimizes judgment learning and anomaly localization. We combine mixed-trajectory distillation with a lightweight outcome objective on student-generated trajectories and supervised anomaly localization:

$$\mathcal{L} = \mathcal{L}_{\text{dist}} + \lambda_R \mathcal{L}_{\text{reward}} + \lambda_H \mathcal{L}_{\text{map}}. \tag{8}$$

Here, $\mathcal{L}_{\text{dist}}$ provides token-level supervision for both student- and teacher-generated trajectories. On student rollouts, $\mathcal{L}_{\text{reward}}$ provides lightweight outcome supervision for response-format validity and final-judgment correctness, while it is disabled for teacher-generated Seq-KD trajectories. $\mathcal{L}_{\text{map}}$ combines image-level anomaly supervision, pixel-level localization supervision, and normal–abnormal anchor separation. Together, these objectives couple fine-grained judgment learning with dense anomaly localization. Exact loss definitions, component weights, and optimization details are provided in Appendix A.3.

The complete training procedure is summarized in Algorithm 1.

4 Experiments

4.1 Experimental setup

Datasets and protocol. We evaluate MOPDA under a strict cross-dataset zero-shot protocol on five benchmarks: MVTec-AD (Bergmann et al., 2019), VisA (Zou et al., 2022), SDD/KolektorSDD (Tabernik et al., 2019), WFDD (Chen et al., 2024), and KSDD2 (Božič et al., 2021). For MVTec-AD, we train on VisA and GoodsAD (Zhang et al., 2024). For VisA, we train on MVTec-AD and GoodsAD. For SDD, WFDD, and KSDD2, we train on MVTec-AD and GoodsAD. The complete assignments are tabulated in Appendix B.

Model and training. We use Qwen3-VL-8B-Instruct (Bai et al., 2025) as the backbone and fine-tune all model parameters, including the vision encoder. Models are trained for 30 epochs on 8 A100 GPUs using DeepSpeed ZeRO-2, with an effective batch size of 32, a learning rate of 1×10^{-6}, and a warmup ratio of 0.03. For MOPDA, we use an EMA decay of 0.999, set the generalized JSD parameter to $\beta = 0.5$, and sample student-generated trajectories with probability $\lambda = 0.8$. The auxiliary reward loss is weighted by 0.1. Additional optimization details are provided in Appendix A.3.

Metrics. We report image-level and pixel-level AUROC, AP, and F1, together with binary judgment accuracy (QA/Acc.), in percent. All MOPDA anomaly metrics are evaluated on the complete test set with the full-resolution protocol, while QA/Acc. is computed from the final binary language judgment.

4.2 Main results

Table 1: **Main comparison on five IAD benchmarks (%).** Image- and pixel-level cells report AUROC/AP/F1, and language judgment reports QA/Acc. Best results are shaded pink and second-best results blue. † marks our reproduction. "–" denotes an unavailable result.

Method	MVTec-AD	VisA	SDD	WFDD	KSDD2
Image-level *(AUROC / AP / F1)*					
CLIP-based					
WinCLIP†	70.70 / 84.91 / 86.46	65.22 / 70.94 / 74.62	69.72 / 29.04 / 36.67	72.52 / 74.42 / 75.01	83.28 / 85.90 / 76.40
AnomalyCLIP†	91.40 / 96.13 / 91.89	79.25 / 83.54 / 79.56	91.08 / 53.97 / 59.26	84.00 / 86.42 / 83.21	91.52 / 94.60 / 90.06
VisualAD†	92.22 / 96.69 / 92.34	84.58 / 87.33 / 82.45	83.10 / 31.39 / 47.25	84.23 / 85.98 / **83.82**	**98.42** / **98.65** / **94.57**
LVLM-based					
AnomalyGPT†	71.43 / 85.94 / 85.95	60.79 / 68.20 / 73.84	64.72 / 22.81 / 31.47	61.37 / 69.10 / 71.13	87.30 / 89.20 / 80.11
Myriad†	–	51.93 / 57.88 / 71.39	64.68 / 22.79 / 31.63	62.99 / 68.95 / 71.13	87.31 / 89.21 / 80.10
ReAL	79.80 / 87.10 / –	–	81.40 / 44.30 / –	79.90 / 78.80 / –	–
MOPDA	**93.76** / **96.99** / **92.58**	**89.01** / **91.91** / **85.98**	**96.15** / **77.77** / **72.22**	**88.53** / **90.43** / 82.85	95.48 / 96.30 / 88.70
Pixel-level *(AUROC / AP / F1)*					
CLIP-based					
WinCLIP†	80.93 / 16.82 / 23.20	70.54 / 6.32 / 9.79	72.18 / 1.14 / 6.19	67.25 / 9.05 / 13.54	96.79 / 22.69 / 31.05
AnomalyCLIP†	88.20 / 33.56 / 38.68	93.94 / 18.96 / 27.32	95.80 / 10.65 / 21.09	90.34 / **34.27** / **39.80**	91.45 / 45.33 / 56.26
VisualAD†	90.78 / 41.30 / 43.71	**95.74** / 28.47 / 35.09	94.55 / 5.67 / 12.51	82.33 / 18.77 / 25.04	98.21 / 60.42 / 59.69
LVLM-based					
AnomalyGPT†	77.61 / 21.41 / 28.76	81.68 / 6.64 / 11.48	62.33 / 0.18 / 1.86	75.62 / 7.25 / 13.24	90.70 / 49.31 / 51.44
Myriad†	–	82.06 / 4.12 / 8.57	62.34 / 0.18 / 1.87	75.87 / 8.28 / 11.53	90.70 / 49.31 / 51.45
ReAL	75.60 / 16.60 / –	–	82.30 / 2.60 / –	70.50 / 7.40 / –	–
MOPDA	**90.87** / **43.52** / **45.44**	95.49 / **30.58** / **35.61**	**99.40** / **21.25** / **32.52**	**91.68** / 25.26 / 31.47	**99.25** / **62.73** / **61.82**
Language judgment *(Accuracy)*					
AnomalyGPT†	44.75	27.29	11.93	58.72	15.73
Myriad†	–	55.50	24.01	60.00	28.93
MOPDA	**71.42**	**74.01**	**83.71**	**63.74**	**87.50**

Table 1 compares CLIP-based and LVLM-based methods on five industrial anomaly detection benchmarks. MVTec-AD and VisA cover diverse industrial objects and textures, while SDD, WFDD, and KSDD2 extend the evaluation to surface and fabric defects. MOPDA achieves competitive image-level discrimination and pixel-level localization against CLIP-based methods, while also producing explicit binary language judgments. Compared with the evaluated LVLM baselines, it improves most reported detection, localization, and judgment metrics. Figure 4 further compares representative localization results on six categories from the five benchmarks. AnomalyGPT often produces sparse or off-target responses, while VisualAD sometimes activates over broader or multiple regions. MOPDA generally concentrates its response on the annotated defects across object, texture, and surface categories. Together with Table 1, these examples illustrate how its competitive quantitative performance translates into spatially coherent localization.

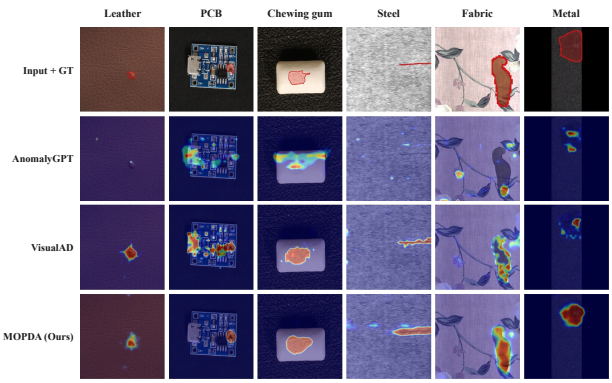

Figure 4: **Cross-benchmark qualitative comparison.** Anomaly maps for six categories from five benchmarks. Ground-truth masks are overlaid in red.

Figure 2 shows localized heatmaps for correctly judged fabric and cashew anomalies. Table 4 tests the role of the judgment directly: ground-truth judgments improve all six anomaly metrics, whereas incorrect judgments reduce them. Together, these results support language as semantic guidance rather than a hard gate for dense anomaly localization.

Table 2: **Training-paradigm comparison on VisA.** SFT, GRPO, OPSD, and MOPDA use the same data, localization branch, and trainable vision tower. Best results are shaded pink and second-best results blue.

Setting	Image-level			Pixel-level			Language
	AUROC	AP	F1	AUROC	AP	F1	QA/Acc.
SFT	87.53	91.19	85.46	84.76	21.66	29.66	71.88
GRPO	**89.68**	**92.13**	85.82	92.86	23.25	29.16	69.58
OPSD	88.83	91.82	**86.08**	95.38	28.67	34.26	71.65
MOPDA	89.01	91.91	85.98	**95.49**	**30.58**	**35.61**	**74.01**

4.3 Ablation studies

Training paradigm. We compare four training paradigms under the same data, localization branch, and trainable vision tower. SFT learns from fixed reference responses, GRPO optimizes sparse outcome rewards on the model's own samples, and OPSD applies privileged token-level distillation only to student-generated on-policy trajectories. MOPDA extends OPSD through **Mixed-Trajectory Supervision** by additionally using evidence-conditioned teacher-generated trajectories in the Seq-KD branch.

As shown in Table 2, MOPDA outperforms SFT on all six anomaly-detection metrics, indicating the benefit of supervising deployment-relevant trajectories rather than only imitating fixed references. Compared with GRPO, MOPDA sacrifices slightly on image-level AUROC/AP but achieves substantially stronger pixel-level localization, showing the advantage of dense token-level supervision over sparse outcome rewards for localization.

Effect of Mixed-Trajectory Supervision. Paired comparisons in Table 3 show that Mixed-Trajectory Supervision improves judgment accuracy on three of the four evaluated benchmarks, with substantially more errors corrected than correct predictions regressed overall. The class-wise analysis further shows that the correction patterns vary across target domains, with some datasets benefiting more from improved anomalous-sample predictions and others from improved normal-sample predictions. These results support the motivation for Mixed-Trajectory Supervision. Student-only on-policy distillation is restricted to response paths explored by the current policy, while evidence-conditioned teacher trajectories provide complementary judgment paths and additional supervision beyond student-generated trajectories.

Language-guided visual anchoring. Table 4 analyzes **Language-guided Visual Anchoring** by asking two questions: whether the localization branch should use the final judgment as its compact semantic condition, and how strongly the heatmap depends on the semantics of that judgment. The first block compares four MOPDA variants: full-response conditioning, learnable anchors alone, no judgment aggregation, and no language output. The no-aggregation variant still generates the language response but prevents the final judgment from being aggregated into the anchors. The no-language-output variant removes language output from the localization path. The second block keeps the inspected image and sampled response fixed, replacing only the judgment span fed to the judgment reforward with an incorrect judgment or, as an analysis-only oracle, a GT judgment.

The interface controls identify the final judgment as the strongest deployable condition: full responses introduce redundant noise, whereas learnable anchors alone lack sample-specific semantics. The no-aggregation and no-language-output controls further separate language generation from judgment aggregation: keeping language output without aggregating the final judgment remains close at the image level but drops on dense localization metrics, and removing language output weakens localization further. These results show that aggregating the final judgment into the visual anchors improves localization beyond generating a language response alone. An anchor-space visualization in Appendix E.3 illustrates the separation between normal and abnormal anchors. Judgment interventions support the same conclusion: GT judgments improve all six reported anomaly metrics, whereas incorrect judgments reduce them, confirming judgment semantics as an effective localization signal.

Table 3: **Paired judgment changes under Mixed-Trajectory Supervision across IAD benchmarks.** OPSD and MOPDA predictions are paired by sample. Corr./Regr. denote wrong-to-correct and correct-to-wrong transitions. Anom./Norm. denote class-wise contributions to ΔQA.

	QA (%)			Paired Changes (%)		Class-wise ΔQA (%)	
Dataset	OPSD	MOPDA	Δ ↑	Corr. ↑	Regr. ↓	Anom. ↑	Norm. ↑
VisA	71.65	74.01	+2.36	5.74%	3.38%	+5.32%	−2.96%
SDD	71.93	83.71	+11.78	13.78%	2.01%	−1.50%	+13.28%
WFDD	65.09	63.74	−1.35	0.90%	2.25%	−0.90%	−0.45%
KSDD2	86.94	87.50	+0.56	1.69%	1.12%	+0.28%	+0.28%

Table 4: **Language-guided Visual Anchoring ablation on MVTec-AD.** The two blocks vary the localization interface and reforwarded judgment. The MOPDA row matches Table 1. The gray GT-judgment row is an unranked oracle. Best non-oracle results are shaded pink and second-best blue.

Setting	Image-level			Pixel-level		
	AUROC	AP	F1	AUROC	AP	F1
Localization interface controls						
MOPDA w/ Full Response	91.49	95.75	91.84	87.63	33.21	38.30
MOPDA w/ Learnable Anchors Only	90.95	95.56	91.44	87.35	32.63	37.65
MOPDA w/o Judgment Aggregation	93.46	96.89	92.53	89.74	38.72	42.05
MOPDA w/o Language Output	90.73	95.06	91.22	85.95	27.34	33.41
Judgment intervention controls						
MOPDA w/ Incorrect Judgment	85.60	93.41	88.90	87.61	32.05	36.93
MOPDA	**93.76**	**96.99**	**92.58**	**90.87**	**43.52**	**45.44**
MOPDA w/ GT Judgment (oracle)	95.60	97.57	94.05	92.95	45.45	47.18

5 Conclusion

We presented MOPDA, a mixed-trajectory on-policy distillation framework for LVLM-based industrial anomaly detection. MOPDA combines **Mixed-Trajectory Supervision** for judgment learning with **Language-guided Visual Anchoring** for dense anomaly localization. Experiments on five IAD benchmarks demonstrate improved judgment and localization performance over the evaluated LVLM-based baselines. More importantly, our analysis shows that language judgments provide effective semantic guidance for dense anomaly localization, while dense visual features remain the primary evidence for pixel-level anomaly scoring. This provides a flexible connection between language understanding and dense visual perception without making localization directly determined by the language output.

AI USE STATEMENT

ChatGPT was used solely to polish the language of this manuscript.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-vl technical report, 2025.

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9584–9592. IEEE, June 2019.

Jakob Božič, Domen Tabernik, and Danijel Skočaj. Mixed supervision for surface-defect detection: From weakly to fully supervised learning. *Computers in Industry*, 129:103459, 2021.

Yunkang Cao, Jiangning Zhang, Luca Frittoli, Yuqi Cheng, Weiming Shen, and Giacomo Boracchi. Adaclip: Adapting CLIP with hybrid learnable prompts for zero-shot anomaly detection. In Ales Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part XXXV*, volume 15093 of *Lecture Notes in Computer Science*, pp. 55–72. Springer, 2024.

Yuhao Chao, Jie Liu, Jie Tang, and Gangshan Wu. Anomalyr1: A grpo-based end-to-end mllm for industrial anomaly detection, 2025.

Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization. In Ales Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part LXVII*, volume 15125 of *Lecture Notes in Computer Science*, pp. 37–54. Springer, 2024.

Jiaxi Cheng, Yuliang Xu, Shoupeng Wang, Tao Ma, Yuchen He, Jinghe Zhang, Sihang Cai, Jiawei Zhen, Jingyi Jia, Yao Wan, Yan Xia, and Zhou Zhao. AnomalyCoT: A multi-scenario chain-of-thought dataset for multimodal large language models. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-2332.

Huilin Deng, Hongchen Luo, Wei Zhai, Yanming Guo, Yang Cao, and Yu Kang. Vmad: Visual-enhanced multimodal large language model for zero-shot anomaly detection. *IEEE Transactions on Automation Science and Engineering*, 23:3607–3618, 2026.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Minillm: Knowledge distillation of large language models. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 32694–32717, 2024a.

Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. Anomalygpt: Detecting industrial anomalies using large vision-language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(3):1932–1940, Mar 2024b.

Jingtai He, Shiyuan Meng, Wenchao Meng, and Qinmin Yang. Adopd: Reference-privileged on-policy distillation for mllm-based industrial anomaly detection, 2026.

Yanning Hou, Peiyuan Li, Zirui Liu, Yitong Wang, Yanran Ruan, Jianfeng Qiu, and Ke Xu. Visualad: Language-free zero-shot anomaly detection via vision transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21346–21356, June 2026.

Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19606–19616, June 2023.

Xi Jiang, Jian Li, Hanqiu Deng, Yong Liu, Bin-Bin Gao, Yifeng Zhou, Jialin Li, Chengjie Wang, and Feng Zheng. MMAD: A comprehensive benchmark for multimodal large language models in industrial anomaly detection. In *ICLR*, 2025.

Yizhou Jin, Yuezhu Feng, Jinjin Zhang, Peng Wang, Qingjie Liu, and Yunhong Wang. Reasoning-driven anomaly detection and localization with image-level supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 43061–43071, June 2026.

Xurui Li, Ziming Huang, Feng Xue, and Yu Zhou. Musc: Zero-shot industrial anomaly classification and segmentation with mutual scoring of the unlabeled images. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 2172–2197, 2024.

Yanhui Li, Yunkang Cao, Chengliang Liu, Yuan Xiong, Xinghui Dong, and Chao Huang. Iad-r1: Reinforcing consistent reasoning in industrial anomaly detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(8):6583–6591, Mar 2026a.

Yuanze Li, Haolin Wang, Shihao Yuan, Ming Liu, Debin Zhao, Yiwen Guo, Chen Xu, Guangming Shi, and Wangmeng Zuo. Myriad: A large multimodal model applying vision experts for industrial anomaly detection. *Science China Information Sciences*, 69(9), Aug 2026b.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007. IEEE, Oct 2017.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571. IEEE, Oct 2016.

Zhen Qu, Xian Tao, Xiaoyi Bao, Dingrong Wang, ShiChen Qu, Zhengtao Zhang, and Xingang Wang. Ag-vas: Anchor-guided zero-shot visual anomaly segmentation with large multimodal models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14126–14136, June 2026.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024.

Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 31(3):759–776, May 2019.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pp. 1195–1204, 2017.

Jiacong Xu, Shao-Yuan Lo, Bardia Safaei, Vishal M. Patel, and Isht Dwivedi. Towards zero-shot anomaly detection and reasoning with multimodal large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20370–20382, June 2025.

Jian Zhang, Runwei Ding, Miaoju Ban, and Linhui Dai. Pku-goodsad: A supermarket goods dataset for unsupervised anomaly detection and segmentation. *IEEE Robotics and Automation Letters*, 9 (3):2008–2015, Mar 2024.

Shifang Zhao, Yiheng Lin, Lu Han, Yao Zhao, and Yunchao Wei. Omniad: Detect and understand industrial anomaly via multimodal reasoning, 2025.

Siyan Zhao, Zhihui Xie, Mengchen Liu, Jing Huang, Guan Pang, Feiyu Chen, and Aditya Grover. Self-distilled reasoner: On-policy self-distillation for large language models, 2026.

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 49705–49737, 2024.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. *SPot-the-Difference Self-supervised Pre-training for Anomaly Detection and Segmentation*, pp. 392–408. Springer Nature Switzerland, 2022.

MOPDA: MIXED-TRAJECTORY ON-POLICY DISTILLATION FOR LANGUAGE-GUIDED INDUSTRIAL ANOMALY DETECTION

SUPPLEMENTARY MATERIAL

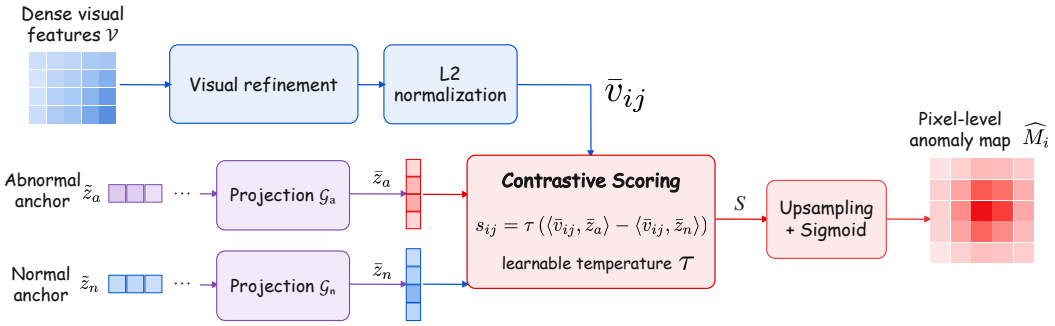

Figure 5: **Architecture of the contrastive heatmap head.** Dense visual features and judgment-conditioned abnormal and normal anchors are mapped to a shared 256-dimensional space. Their scaled cosine-similarity difference produces anomaly logits, which are upsampled and passed through a sigmoid to form the pixel-level anomaly map. Image-level aggregation is described below.

A IMPLEMENTATION AND TRAINING DETAILS

A.1 LANGUAGE-GUIDED VISUAL ANCHORING DETAILS

Judgment extraction. For a selected training response or the student's inference response $\hat{y}_i$, $\hat{a}_i = \mathrm{Trim}_{\mathrm{ans}}(\hat{y}_i)$ extracts the final span enclosed by `<answer></answer>`. This span is both the language-side judgment and the condition supplied to the judgment reforward.

Backbone and head inputs. Figure 5 shows the localization head used with Qwen3-VL-8B-Instruct. We reshape the query image's post-merger visual tokens into a single 4096-channel spatial grid V_i. The judgment reforward yields 4096-dimensional contextualized hidden states $\tilde{z}_a$ and $\tilde{z}_n$ at the abnormal and normal anchor tokens. These are the head's language inputs.

Visual refinement and anchor projection. The visual grid first passes through a learned residual channel adapter. It applies channel-wise LayerNorm and two 1×1 convolutions with a $4096 \to 1024 \to 4096$ bottleneck and GELU; its output is added to the input with a learnable scale initialized to zero. Two 3×3 Conv–GELU layers then map the grid to 256 channels. Separately, g_a and g_n project the two anchor states to 256 dimensions using Linear–GELU–Linear–LayerNorm. Denoting the refined visual grid by $h(V_i)$, the representations used for scoring are

$$\bar{v}_{ij} = \frac{h(V_i)_j}{\|h(V_i)_j\|_2}, \qquad \bar{z}_a = \frac{g_a(\tilde{z}_a)}{\|g_a(\tilde{z}_a)\|_2}, \qquad \bar{z}_n = \frac{g_n(\tilde{z}_n)}{\|g_n(\tilde{z}_n)\|_2}. \tag{9}$$

Each patch is scored by the abnormal-minus-normal cosine-similarity difference, multiplied by the learnable temperature $\tau = \min(\exp(\alpha), 100)$, with learnable α.

Inference outputs. At inference, the logits are bilinearly upsampled to the evaluation resolution and passed through Sigmoid to obtain the anomaly map. The image-level anomaly score is the mean of the largest 1% of values in this map. For Gaussian-smoothed evaluation, smoothing is applied to the probability map before computing the image score and localization metrics.

A.2 Distillation Implementation Details

EMA teacher. The teacher has the same architecture and vocabulary as the student. Its parameters are maintained by an exponential moving average (EMA) of the student parameters (Tarvainen & Valpola, 2017),

$$\bar{\theta}^{(k)} = \mu\bar{\theta}^{(k-1)} + (1-\mu)\theta^{(k)}, \qquad \mu \in (0,1). \tag{10}$$

We set $\mu = 0.999$ in the reported experiments.

Generalized Jensen–Shannon divergence. For either trajectory source, the student and teacher evaluate the same sampled response prefixes under their respective contexts. Their next-token distributions at position t are

$$P_{S,t}^{(i)} = p_\theta(\cdot \mid c_i^S, \hat{y}_{i,<t}), \qquad P_{T,t}^{(i)} = \mathrm{stopgrad}\big(p_{\bar{\theta}}(\cdot \mid c_i^T, \hat{y}_{i,<t})\big)\,. \tag{11}$$

We use generalized Jensen–Shannon divergence with $\beta \in (0,1)$,

$$\begin{aligned} m_{i,t}^{\beta} &= (1-\beta)P_{S,t}^{(i)} + \beta P_{T,t}^{(i)}, \\ D_{\mathrm{JSD}}^{\beta}(P_{S,t}^{(i)} \| P_{T,t}^{(i)}) &= \beta\,\mathrm{KL}(P_{T,t}^{(i)} \| m_{i,t}^{\beta}) + (1-\beta)\,\mathrm{KL}(P_{S,t}^{(i)} \| m_{i,t}^{\beta}). \end{aligned} \tag{12}$$

With $\beta = 0.5$, this is the symmetric JSD. Gradients flow only through the student distributions, not through the teacher distributions or sampled tokens.

Teacher-sequence validity and exact loss. Teacher-generated responses are checked for the required format and agreement of their final judgment with the source label a_i. We set $g_i = 1$ for student-generated trajectories and $g_i = \mathrm{Valid}(\hat{y}_i, a_i) \in \{0,1\}$ for teacher-generated trajectories. This gate applies only to distillation. The exact minibatch loss and token-count normalization are

$$\mathcal{L}_{\mathrm{dist}} = \frac{1}{N_{\mathcal{B}}} \sum_{i \in \mathcal{B}} g_i \sum_{t=1}^{|\hat{y}_i|} D_{\mathrm{JSD}}^{\beta}(P_{S,t}^{(i)} \| P_{T,t}^{(i)}), \qquad N_{\mathcal{B}} = \sum_{i \in \mathcal{B}} |\hat{y}_i|. \tag{13}$$

Invalid teacher responses contribute zero to the numerator. The denominator still counts all sampled tokens in the minibatch, as in the main-text objective.

A.3 Training Objective and Optimization Details

Localization supervision. During source-domain training, the mean of the largest 1% of low-resolution anomaly logits is aligned with the image label a_i by BCE, yielding $\mathcal{L}_{\mathrm{img}}$. This Top-1% pooling is omitted from Figure 5. Pixel-level supervision compares the anomaly response with mask M_i using BCE, focal loss (Lin et al., 2017), and Dice loss (Milletari et al., 2016), denoted by $\mathcal{L}_{\mathrm{pix}}$, $\mathcal{L}_{\mathrm{foc}}$, and $\mathcal{L}_{\mathrm{dice}}$. Following (Hou et al., 2026), the contextualized abnormal and normal anchors are separated with

$$\mathcal{L}_{\mathrm{con}} = \max(\langle \bar{z}_a, \bar{z}_n \rangle + m, 0). \tag{14}$$

The complete localization objective is

$$\mathcal{L}_{\mathrm{map}} = \lambda_{\mathrm{img}}\mathcal{L}_{\mathrm{img}} + \lambda_{\mathrm{pix}}\mathcal{L}_{\mathrm{pix}} + \lambda_{\mathrm{foc}}\mathcal{L}_{\mathrm{foc}} + \lambda_{\mathrm{dice}}\mathcal{L}_{\mathrm{dice}} + \lambda_{\mathrm{con}}\mathcal{L}_{\mathrm{con}}. \tag{15}$$

The reported Qwen3-VL-8B-Instruct main experiments use $\lambda_{\mathrm{img}} = 2.0$, $\lambda_{\mathrm{pix}} = 0.5$, and $\lambda_{\mathrm{foc}} = \lambda_{\mathrm{dice}} = \lambda_{\mathrm{con}} = 1.0$. The checkpoints at training steps 1200 and 2000 share these weights.

Outcome reward and branch updates. Student-generated trajectories receive a lightweight reward for response format and final-judgment correctness:

$$R_i = R_{\mathrm{fmt}}(\hat{y}_i) + R_{\mathrm{ans}}(\hat{a}_i, a_i). \tag{16}$$

The two binary components are

$$\begin{aligned} R_{\mathrm{fmt}}(\hat{y}_i) &= \mathbf{1}[\mathrm{Has}_{\mathrm{think}}(\hat{y}_i) \wedge \mathrm{Has}_{\mathrm{answer}}(\hat{y}_i)]\,, \\ R_{\mathrm{ans}}(\hat{a}_i, a_i) &= \mathbf{1}[\mathrm{Label}(\hat{a}_i) = a_i]\,. \end{aligned} \tag{17}$$

Each Has check requires paired opening and closing tags. Label maps a parsed `A. anomaly` or `B. normal` answer to its binary class and is invalid if parsing fails. The corresponding policy-gradient term $\mathcal{L}_{\mathrm{reward}}$ provides an auxiliary outcome signal. For the teacher-generated Seq-KD branch, $\mathcal{L}_{\mathrm{reward}} = 0$. The joint objective combines $\mathcal{L}_{\mathrm{dist}}$, $\lambda_R\mathcal{L}_{\mathrm{reward}}$, and $\lambda_H\mathcal{L}_{\mathrm{map}}$. The reported experiments use $\lambda_R = 0.1$ and $\lambda_H = 1.0$. The latter is `lambda_seg` in the training configuration and is separate from the five component weights in Eq. equation 15.

Algorithm 1 MOPDA Training via Mixed-Trajectory On-Policy Distillation

Require: Source data $\mathcal{D}_{\text{src}} = \{(x_i^{\text{img}}, x_i^{\text{text}}, a_i, M_i, r_i)\}$, student parameters θ
Hyperparameters: Student-rollout probability λ, JSD coefficient β, EMA decay μ, reward and localization weights λ_R, λ_H

Initialize EMA teacher $\bar{\theta} \leftarrow \theta$
for each minibatch $\mathcal{B} \subset \mathcal{D}_{\text{src}}$ **do**
  Construct student contexts c_i^S and evidence-privileged teacher contexts c_i^T for $i \in \mathcal{B}$
  Sample $u \sim \mathcal{U}(0, 1)$
  if $u \leq \lambda$ **then** ▷ student rollout
    Sample $\hat{y}_i \sim \pi_\theta(\cdot \mid c_i^S)$, set $g_i \leftarrow 1$, and compute reward loss $\mathcal{L}_{\text{reward}}$
  else ▷ teacher rollout (Seq-KD)
    Sample $\hat{y}_i \sim \pi_{\bar{\theta}}(\cdot \mid c_i^T)$, set $g_i \leftarrow \text{Valid}(\hat{y}_i, a_i)$ and $\mathcal{L}_{\text{reward}} \leftarrow 0$
  end if
  Compute $\mathcal{L}_{\text{dist}}$ using Eq. (13)
  Compute $\mathcal{L}_{\text{map}}$ from source labels, masks, and anchor separation
  Form $\mathcal{L} \leftarrow \mathcal{L}_{\text{dist}} + \lambda_R \mathcal{L}_{\text{reward}} + \lambda_H \mathcal{L}_{\text{map}}$
  Update student parameters θ by minimizing $\mathcal{L}$
  $\bar{\theta} \leftarrow \mu\bar{\theta} + (1 - \mu)\theta$
end for

A.4 Training Configuration and Diagnostics

Backbone and fine-tuning. We use Qwen3-VL-8B-Instruct as the vision-language backbone. The reported MOPDA model uses full-parameter fine-tuning, including the visual tower. Both language and dense visual representations adapt to source-domain training.

Controlled comparisons. All training-paradigm and localization-interface ablations use the same backbone, source data, and trainable visual tower. They differ only in the component being studied, such as replacing Mixed-Trajectory On-Policy Distillation with supervised fine-tuning, removing the generated judgment from the localization path, or changing the text used to condition the anchors.

Training dynamics and optimization diagnostics. Figure 6 shows the optimization dynamics of the language and localization objectives throughout training. The format reward rapidly saturates, indicating that the model quickly learns the required response structure, while the answer reward improves progressively as training proceeds. Meanwhile, both the localization and distillation losses decrease steadily. The image- and pixel-level BCE losses also rapidly converge to low values, whereas the Dice loss continues to decrease in later training, reflecting progressive refinement of spatial overlap. Finally, the cosine similarity between the normal and abnormal anchors becomes increasingly negative, indicating that the anchor-separation objective progressively encourages more discriminative normal–abnormal representations. Overall, these trends show stable joint optimization of language judgment, distillation, and dense anomaly localization.

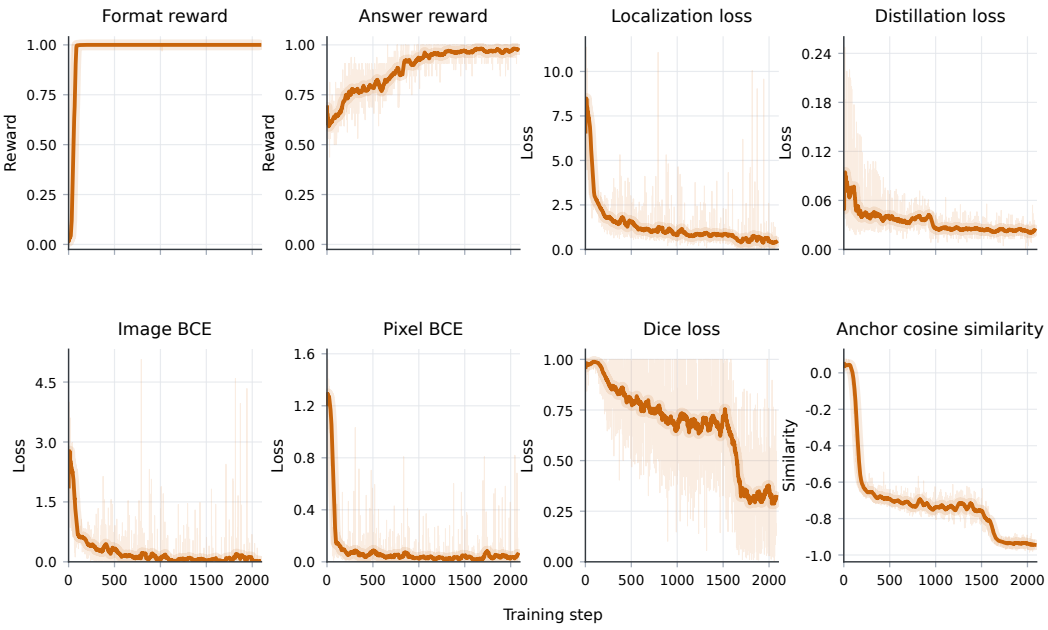

Figure 6: **Training dynamics of MOPDA.**

B Dataset Construction and Zero-Shot Protocol

Source-target separation. We follow a strict cross-dataset protocol: images from a target benchmark are not used to train the model evaluated on that benchmark. The target-domain characteristics and source-target assignments are summarized in Table 5. GoodsAD is used as an additional source dataset, while every reported target remains held out.

Table 5: **Cross-dataset source-target protocol.** The table summarizes the visual characteristics of each held-out target benchmark and the source-domain data used to train the corresponding model.

Target benchmark	Target characteristics	Source-domain training data
MVTec-AD	Objects and textures	VisA + GoodsAD
VisA	Industrial objects	MVTec-AD + GoodsAD
SDD/KolektorSDD	Steel-surface defects	MVTec-AD + GoodsAD
WFDD	Fabric defects	MVTec-AD + GoodsAD
KSDD2	Metal-surface defects	MVTec-AD + GoodsAD

Source-domain supervision. Source samples include normal and anomalous images. Pixel masks supervise dense localization, image-level labels supervise binary anomaly discrimination, and reference answers construct the source-domain teacher context.

Reference evidence. The training-only defect-analysis references r_i were sourced from the open AnomalyCoT (Cheng et al., 2025) and MMAD (Jiang et al., 2025) datasets. They were supplied only to the evidence-privileged teacher during source-domain training.

Target-domain evaluation. For every target benchmark, the model is evaluated directly without target-domain fine-tuning, target-specific prompt optimization, or access to target labels and masks. This protocol evaluates cross-domain transfer of the learned connection between language judgment and visual localization.

C PROMPT DESIGN

Prompt organization. Figures 7–9 make the information boundary between the student and teacher explicit. The shared system prompt defines the anomaly-detection role and structured output. The student prompt contains only the deployment-compatible image and question, whereas the teacher prompt preserves the same task but appends a verified reference answer during training. Thus, privileged evidence is confined to teacher supervision and never enters the deployed input.

System Prompt

You are an AI model for anomaly detection. Analyze the entire image and decide whether it is anomalous or normal.

Provide brief reasoning in <think>...</think> and finally output the answer letter and option text in <answer>...</answer>.

Figure 7: **System Prompt.** The shared anomaly-detection role and structured response format used by both branches.

Student Prompt

<image>
Is the image anomalous or normal? (A) anomaly (B) normal.
Please answer using exactly one of the following two formats.

If the image is anomalous:
<think>your reasoning here</think>
<answer>A. anomaly</answer>

If the image is normal:
<think>your reasoning here</think>
<answer>B. normal</answer>

Figure 8: **Student Prompt.** The deployment-compatible query contains only the image and the binary anomaly question.

Teacher Prompt training only

<image>
Is the image anomalous or normal? (A) anomaly (B) normal.
Please answer using exactly one of the following two formats.

If the image is anomalous:
<think>your reasoning here</think>
<answer>A. anomaly</answer>

If the image is normal:
<think>your reasoning here</think>
<answer>B. normal</answer>

Privileged reference
Reference answer:
{reference solution}

You also have access to the reference answer for this same sample.
Use the reference answer to answer the original task again. Preserve the exact output format shown in the reference answer, including its reasoning tags and answer tag.

Figure 9: **Teacher Prompt.** A verified source-domain reference answer is appended only for training-time teacher supervision.

Connection to localization. The student's final answer span supplies the compact semantic condition for visual anchoring. The complete response contributes to language learning, while localization uses the extracted judgment.

D Evaluation Details

Evaluation outputs. Each target image produces two complementary predictions: a binary language judgment and a dense anomaly map. The binary judgment is evaluated with accuracy. The anomaly map is evaluated at both image and pixel levels using AUROC, average precision, and F1.

Full-resolution localization. Pixel-level metrics are computed against the original target masks at their native evaluation resolution. Image-level anomaly scores are obtained by aggregating the most anomalous spatial responses in the predicted map.

No target supervision. Evaluation uses only the image and the common binary anomaly question. Target labels and masks are read only by the metric computation after prediction; they are never exposed to response generation, judgment conditioning, or heatmap prediction.

E More Case Studies

E.1 Qualitative Localization Comparisons

Heatmap localization comparison. Figures 10–14 use the same layout: columns are target-domain samples, while rows show the input with its ground-truth annotation followed by anomaly maps from WinCLIP, AnomalyCLIP, VisualAD, AnomalyGPT, Myriad, and MOPDA. The examples span object, texture, and surface categories with compact, elongated, and irregular defects.

Thin structural defects on SDD. In Figure 10, the baselines retain isolated fragments or activate off-defect regions around narrow discontinuities. MOPDA follows the annotated defect more continuously and preserves its elongated structure. Its heatmap extends beyond the binary annotation in some regions.

Fabric defects on WFDD. Figure 11 contains compact marks and extended defects on strongly textured cloth. The baselines often split their responses or follow competing texture patterns, whereas MOPDA concentrates on compact defects and tracks the long horizontal defect, indicating transfer beyond source-domain appearance.

Multi-instance objects on VisA. Figure 12 tests scenes containing several similar objects. MOPDA places its strongest responses on the damaged capsules, candle, and macaroni and follows their corrupted parts, whereas the baseline maps more often spread across intact instances or unrelated regions.

Object and texture categories on MVTec-AD. Figure 13 combines compact object defects with elongated texture anomalies. MOPDA remains localized on the capsule, carpet, leather, and screw while retaining the long spatial support of the grid and wood defects.

Surface defects on KSDD2. Figure 14 includes surface defects of different sizes and shapes. MOPDA covers a larger portion of several annotated regions, whereas the baselines often emphasize isolated fragments. Small and boundary-adjacent defects remain challenging, and some responses extend beyond the annotated regions.

Cross-dataset observation. Across Figures 10–14, the examples illustrate how localization behavior varies across target domains. MOPDA captures the spatial extent of many annotated defects, while missed defects and overly broad responses remain visible in some cases.

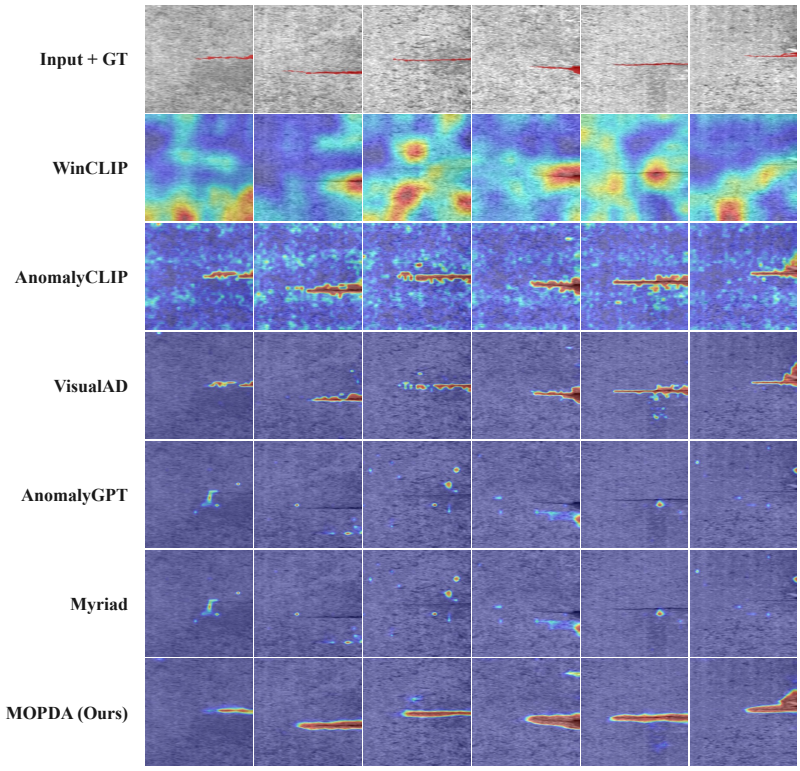

Figure 10: **Qualitative comparison on SDD (KolektorSDD).**

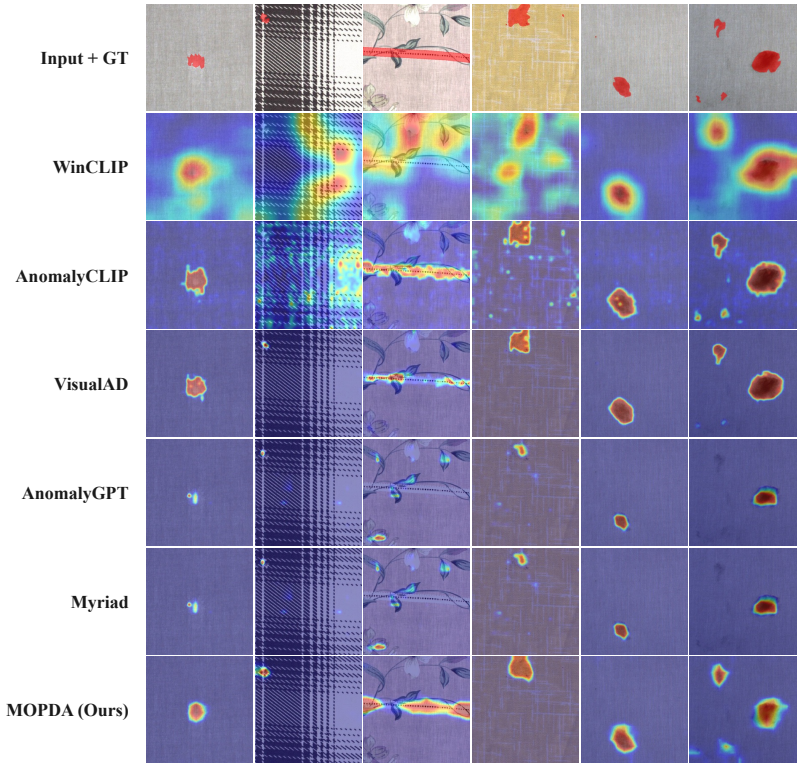

Figure 11: **Qualitative comparison on WFDD.**

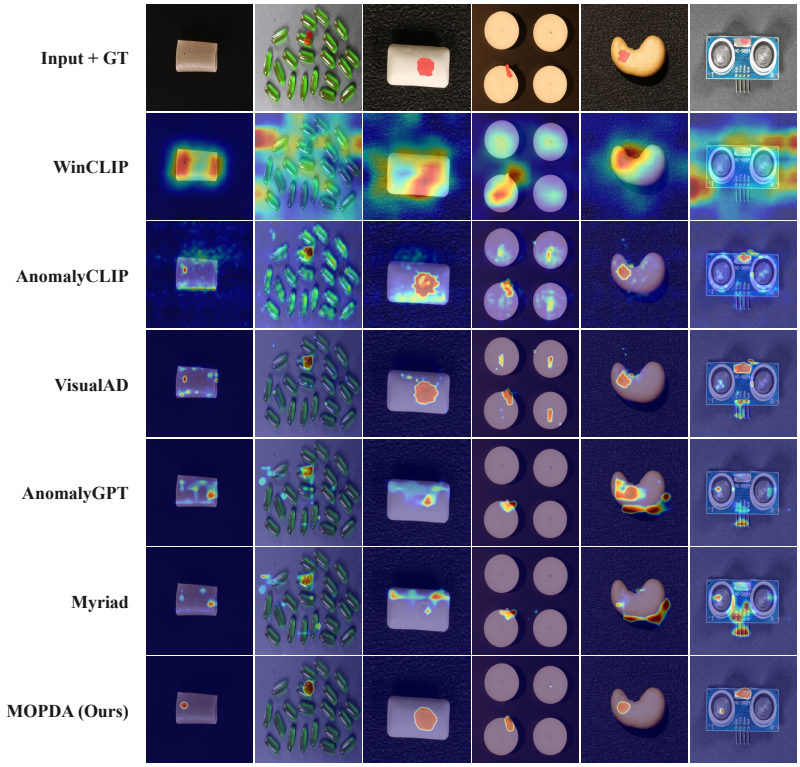

Figure 12: **Qualitative comparison on VisA.**

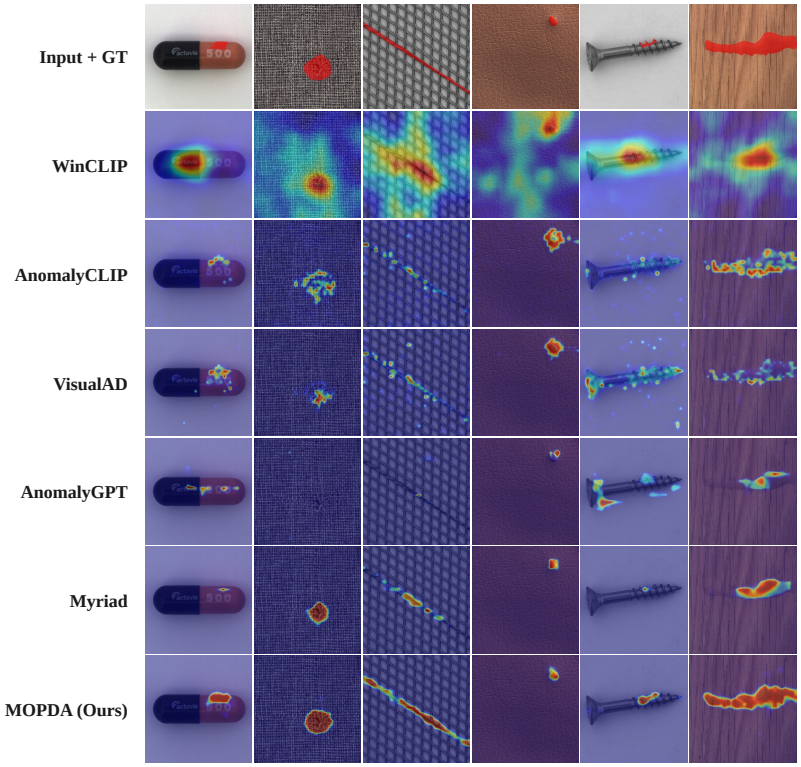

Figure 13: **Qualitative comparison on MVTec-AD.**

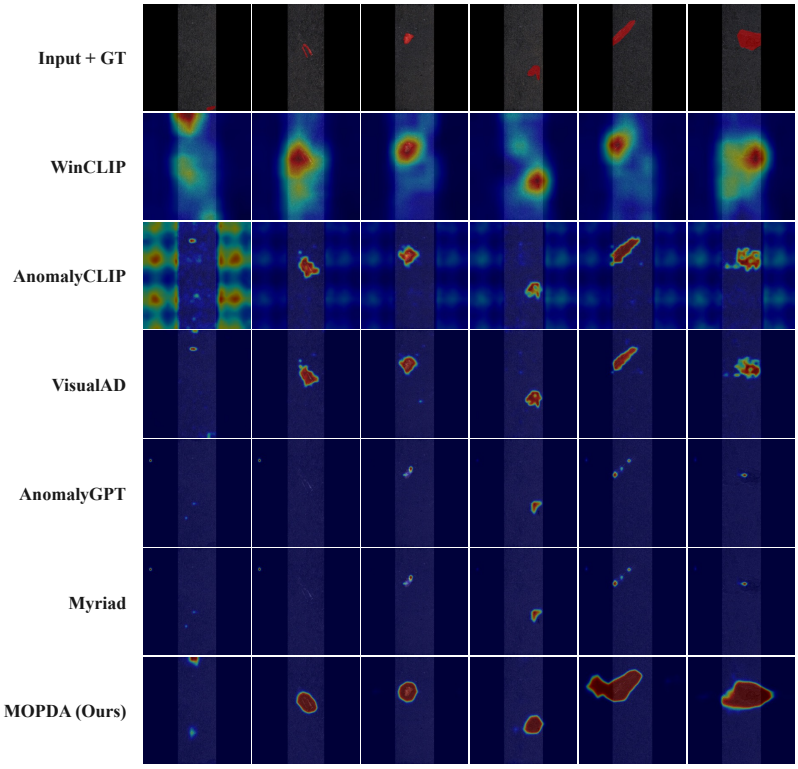

Figure 14: **Qualitative comparison on KSDD2.**

E.2 Judgment and Localization Case Studies

Figures 15– 19 show four selected cases per benchmark, with two normal and two anomalous samples in each. All selected image-level judgments are correct. On normal samples, the anomaly maps remain low-response overall and show no prominent spurious hotspots. On anomalous samples, the strongest responses concentrate around visible defect regions. These cases illustrate both low false-positive activation on normal surfaces and localized responses to defects across target domains.

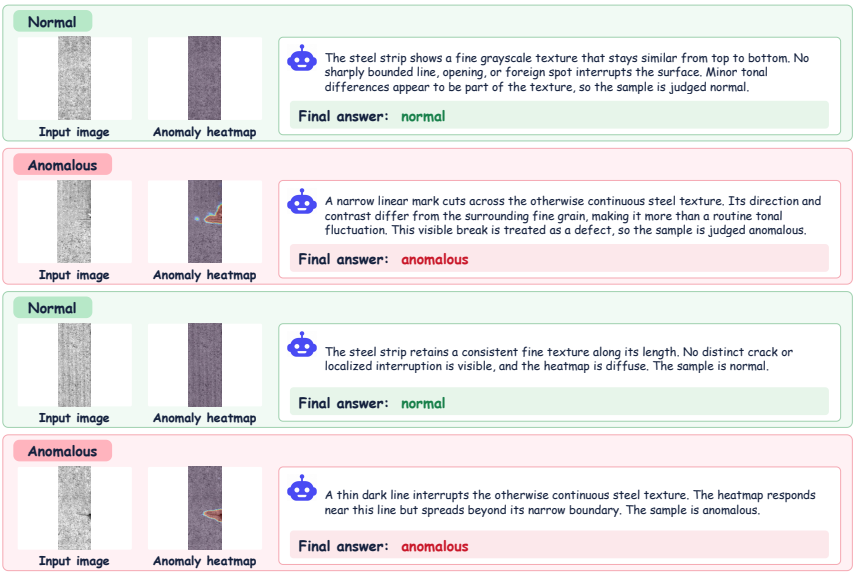

Figure 15: **Judgment and localization cases on SDD.**

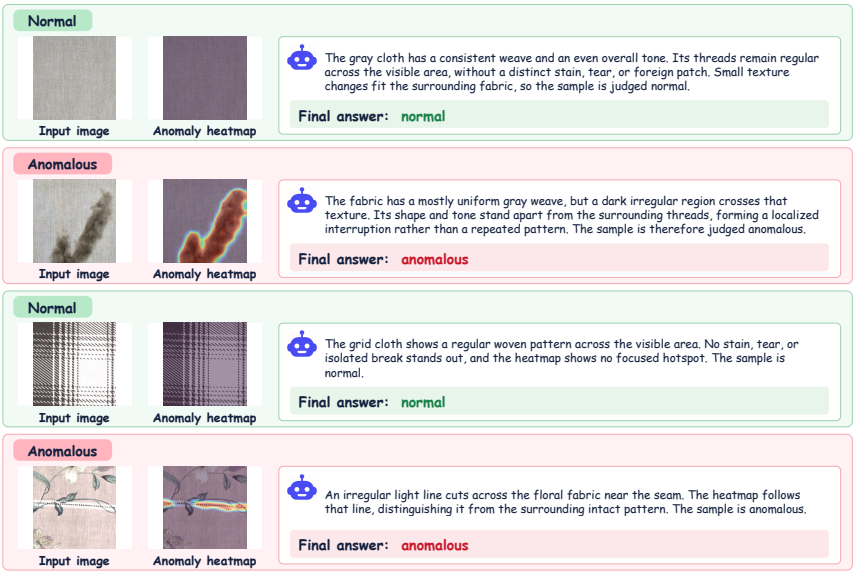

Figure 16: **Judgment and localization cases on WFDD.**

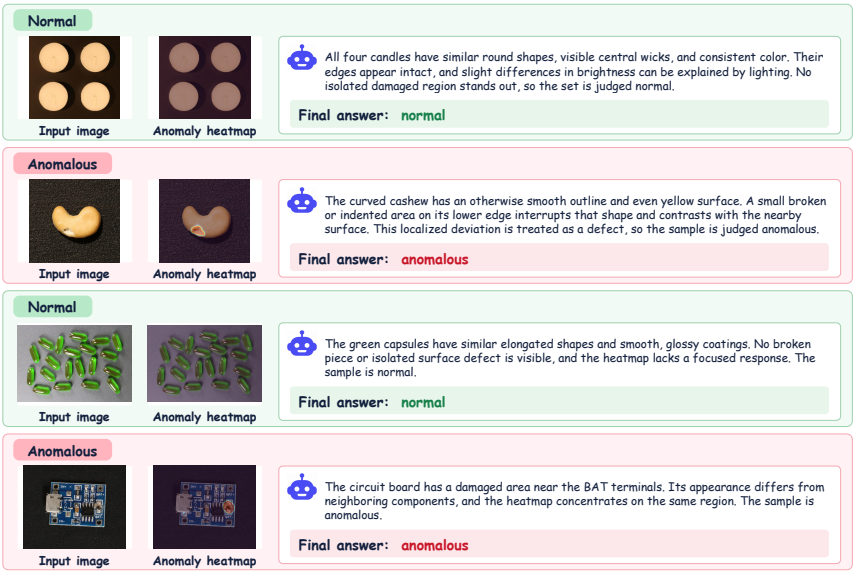

Figure 17: **Judgment and localization cases on VisA.**



Figure 18: **Judgment and localization cases on MVTec-AD.**

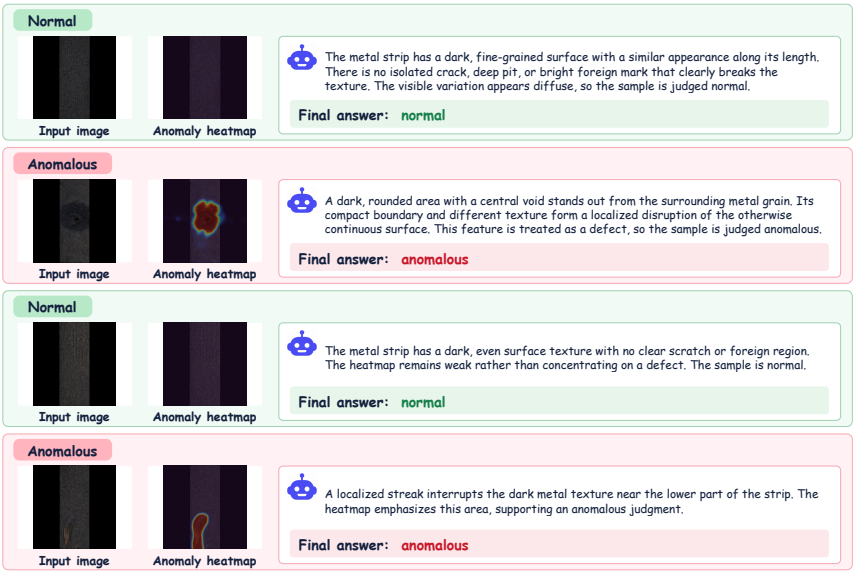

Figure 19: **Judgment and localization cases on KSDD2.**

E.3 Anchor-Space Visualization

Figure 20 visualizes the same effect in anchor space: compared with anchoring without language output, the final-judgment condition increases the separation between normal and abnormal anchors on both MVTec-AD and VisA.

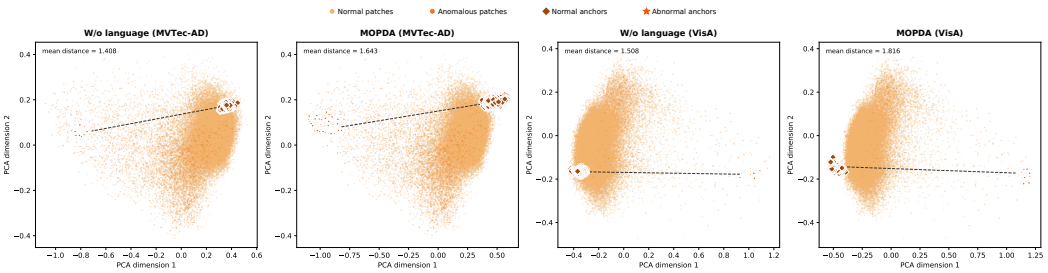

Figure 20: **Anchor separation under final judgment.** PCA visualization compares anchoring without language output and the final MOPDA condition on MVTec-AD and VisA. The final judgment increases the normal-abnormal anchor distance and provides a clearer semantic direction for visual-feature-grounded anomaly scoring.

F Failure Cases and Limitations

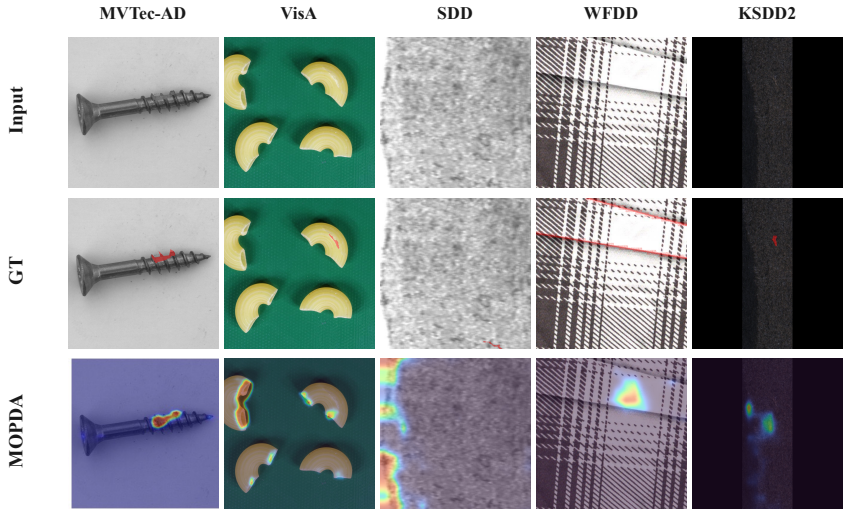

Figure 21: **Localization failure cases.** Input, GT, and MOPDA response for five benchmarks.

Localization failures. Figure 21 shows five selected cases in which the predicted heatmap misses part of the annotated defect or activates outside it.

Source-domain supervision. MOPDA uses source-domain image labels, pixel masks, and reference answers during training. Target-domain evaluation uses images without target annotations or reference answers.

Binary response schema. The two-option answer format produces a stable normal/anomalous judgment for visual anchoring. Richer response schemas could condition localization on defect categories or calibrated uncertainty.