# AD-VRPD: ANOMALY DETECTION VIA VERIFIABLE ROLLOUT-GUIDED SELF-DISTILLATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Visual anomaly detection without reference images requires explicit inspection criteria, yet generic anomaly queries leave the expected normal state and relevant defects unspecified. We introduce diagnostic-text-guided, reference-free anomaly detection and adapt MMR-AD with reusable descriptions of normality and possible defects. These descriptions are constructed offline using examples and annotations from all source datasets, including test sources; anomalous queries receive descriptions selected by their ground-truth defect type. The detector is trained and tested on disjoint source datasets. Under this setting, accurate localization remains challenging, particularly for small defects. We propose AD-VRPD, a verifiable rollout-guided on-policy self-distillation method with a frozen teacher viewing GT-marked images and a student viewing unmarked images. A verifier triggers target reconstruction when teacher guidance conflicts with rollout outcomes. Experiments show that privileged-view self-distillation provides substantial localization gains, with AD-VRPD yielding further improvements over vanilla self-distillation. The resulting compact detector is competitive with larger and proprietary models under the same diagnostic-text interface.

## 1 INTRODUCTION

Reference-free visual anomaly detection removes visual exemplars from the inference input, but does not eliminate the need to specify what constitutes an anomaly. Language-guided methods such as WinCLIP and AnomalyCLIP establish the value of textual descriptions for inspection Jeong et al. (2023); Zhou et al. (2024). In generic-query-based formulations, however, models are asked whether an image is abnormal without explicit criteria for normality or relevant defects. The model must infer both the inspection requirements and whether the image violates them. This ambiguity matters when acceptable variations or required configurations depend on the inspection context, as illustrated by logical anomalies in MVTec LOCO AD Bergmann et al. (2022). We identify this under-specification as a limitation of generic anomaly queries and investigate the value of explicit diagnostic criteria.

To address this limitation, we introduce a *diagnostic-text-conditioned formulation*: an unmarked query image is paired with descriptions of the expected normal state and possible defects, and the model predicts an anomaly decision and bounding boxes (Figure 1). We instantiate this formulation by adapting MMR-AD Yao et al. (2026) with shared, defect-type-specific descriptions constructed from normal references, annotated anomalous examples, and associated reasoning records. We first construct the descriptions and then partition the benchmark by source dataset, using ten datasets for detector training and four for testing. Text construction uses examples and annotations from all fourteen datasets, including the four test sources. Anomalous queries receive their GT defect-type descriptions, while normal queries receive a randomly selected description from the same category. This evaluates reference-free inference on sources held out from detector training, with target-domain information supplied through offline diagnostic text.

Making the inspection criteria explicit addresses task ambiguity, but accurate localization remains challenging. In our evaluation set, 62.1% of the 4,380 annotated defect boxes occupy less than 1% of image area; medium (1–5%) and large ($> 5\%$) boxes account for 25.5% and 12.4%, respectively (Appendix F). This prevalence of small defects motivates supervision that directs attention to localized visual evidence. We therefore adopt on-policy self-distillation (OPSD) to transfer spatial



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.

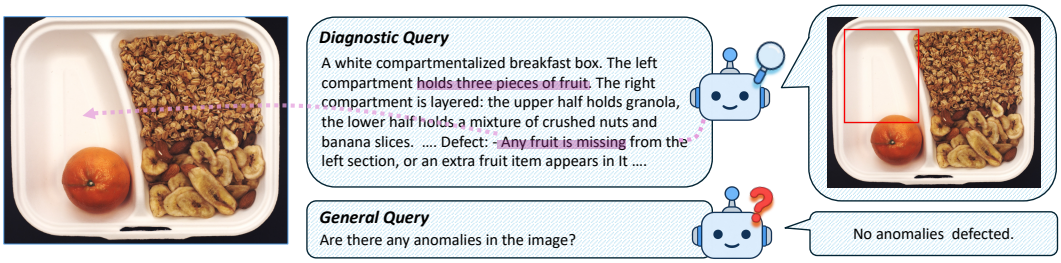



Figure 1: Diagnostic text specifies normality and relevant deviations that a generic query leaves implicit. Responses illustrate the task design, not a controlled performance comparison.

guidance from a frozen teacher initialized from the same model as the student. The teacher sees ground-truth-box-marked images and supervises student-generated trajectories, while the student sees unmarked images. This privileged view provides localization evidence but also introduces an input mismatch: in exploratory training, some student rollouts refer to a "red box" despite receiving unmarked images. This observation motivates checking teacher supervision rather than transferring it unconditionally.

Building on the verifiable-reward-guided distillation principle of OPDVR Lin et al. (2026), we propose **AD-VRPD** for diagnostic-text-guided anomaly detection and localization. A task verifier evaluates student predictions and identifies whether the teacher's probability change for a sampled token agrees with encouraging or discouraging the rollout. At conflicting positions, AD-VRPD reconstructs the distillation target by retaining the student's sampled-token probability and distributing the remaining mass according to the teacher. This retains supervision over alternative outputs without encoding the conflicting probability change in the target. Across four datasets held out from detector training, OPSD accounts for most of the localization gain. AD-VRPD further improves Box-F1 by 2.0 points at IoU 0.1 and 0.7 points at IoU 0.5 in the evaluated runs, enabling a compact model to compete with larger and proprietary models under the same diagnostic-text interface.

Our contributions are threefold:

- **Problem and formulation.** We identify under-specification in generic anomaly queries and introduce a diagnostic-text-guided, reference-free formulation that explicitly specifies normality and the defects to inspect.
- **Benchmark construction.** We adapt MMR-AD into a diagnostic-text-conditioned benchmark with reusable defect-type descriptions and a dataset-disjoint detector evaluation protocol, enabling analysis of how explicit inspection criteria affect detection and localization.
- **Method.** We propose AD-VRPD, a verifiable rollout-guided on-policy self-distillation method that regulates privileged teacher supervision through target reconstruction, improving localization with a compact multimodal model.

## 2 Related Work

### 2.1 Visual Anomaly Detection

Visual anomaly detection spans normal-feature distribution and memory modeling Defard et al. (2020); Roth et al. (2022), reconstruction and synthetic-anomaly learning Zavrtanik et al. (2021); Liu et al. (2023), and teacher–student feature discrepancy Deng and Li (2022); Batzner et al. (2024). These methods establish strong visual inspection baselines, typically using normal training examples to characterize deviations.

CLIP-based approaches transfer image–text representations to zero-/few-shot inspection Radford et al. (2021). WinCLIP, APRIL-GAN, AnomalyCLIP, and AdaCLIP explore local feature alignment, projection adaptation, and prompt learning for anomaly classification and segmentation Jeong et al. (2023); Chen et al. (2023); Zhou et al. (2024); Cao et al. (2024). Generative multimodal detectors extend inspection to instruction following and textual outputs, using supervised adaptation and

reinforcement learning Gu et al. (2024); Li et al. (2025); Chao et al. (2025); Yao et al. (2026). Our setting supplies explicit normal-state and defect specifications to a detector that directly generates decisions and bounding boxes.

### 2.2 Benchmarks and Task Formulations

MVTec LOCO AD distinguishes structural and logical anomalies Bergmann et al. (2022). MMAD evaluates MLLMs through seven inspection subtasks Jiang et al. (2025), while MMR-AD supports training and evaluation of general anomaly detectors Yao et al. (2026). We adapt MMR-AD with reusable normal-state and defect descriptions, evaluating decisions and boxes under explicit diagnostic queries. The description-assignment rules, offline use of annotations, and dataset-disjoint detector split are detailed in Section 3.

### 2.3 On-policy Distillation and Verifiable Feedback

On-policy distillation supervises the student's own generated trajectories with teacher predictions Agarwal et al. (2023). Closely related to our approach, On-policy Distillation with Verifiable Reward (OPDVR) uses trajectory correctness to gate the sampled-token teacher–student log-ratio reward Lin et al. (2026). Our verifier-guided design is inspired by this alignment of distillation direction with verifiable outcomes. We adapt this principle to diagnostic-text-guided anomaly detection with a frozen, same-initialization teacher receiving GT-marked images. At conflicting positions, AD-VRPD reconstructs a full distributional target, retaining the student's sampled-token mass while allocating the remaining mass according to the teacher, rather than masking only the sampled-token reward.

Within anomaly detection, ADOPD uses reference-privileged supervision and sequence-weight calibration, while OPD-IAD combines privileged self-distillation with language-guided visual anchoring He et al. (2026); Chen et al. (2026). AD-VRPD focuses on direct decision-and-box generation and uses a localization-aware verifier to regulate the distributional supervision.

## 3 Task and Data Construction

### 3.1 Task Formulation

Given an unmarked query image $x$ and diagnostic text $u = (u^{\text{normal}}, u^{\text{defect}})$, the detector predicts

$$F_\theta(x, u) = (\hat{a}, \widehat{\mathcal{B}}), \tag{1}$$

where $\hat{a} \in \{0, 1\}$ indicates whether an anomaly is present and $\widehat{\mathcal{B}}$ contains the predicted defect boxes. The text specifies the expected normal state and possible deviations, providing a reusable inspection criterion rather than a diagnosis of the query image. Normal images have a negative target label and no GT boxes. At inference, the detector receives only the query image and diagnostic text, without visual reference examples. Here, reference-free describes the inference input; it does not imply that diagnostic-text construction is independent of target-domain examples or annotations.

### 3.2 Benchmark Construction

We adapt all fourteen constituent datasets of MMR-AD Yao et al. (2026), retaining their images, anomaly labels, and bounding-box annotations. For each defect type, Qwen3.7-plus synthesizes a shared diagnostic description from normal references, annotated anomalous examples, and diagnostic observations summarized from the associated reasoning records. Each description contains a *Normal State* specification and a *Possible Defects* component (Figure 2). Automated consistency checks and revision are followed by manual review, with sampled review for realiad and manta.

Anomalous images are paired with the description of their GT defect type. Normal images are randomly paired with a defect-type description from the same category, testing whether the specified defect is correctly rejected when absent. Thus, the benchmark evaluates detection and localization under defect-type-conditioned queries.

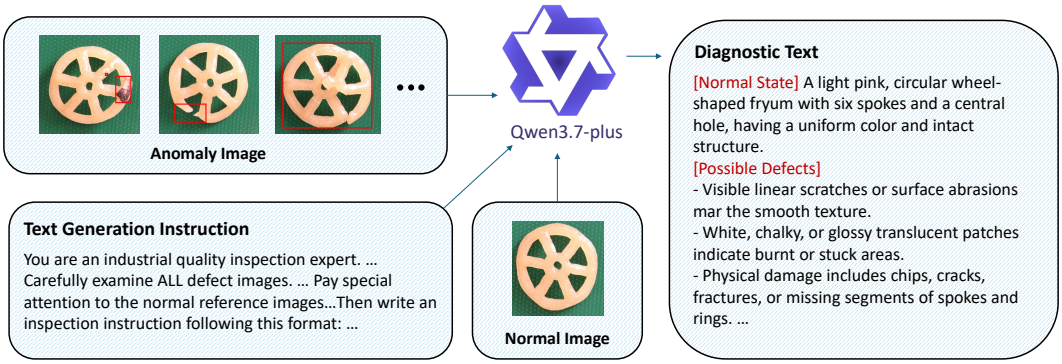



Figure 2: Offline construction of shared defect-type descriptions from normal references, annotated anomalies, and source reasoning records. Red boxes denote source annotations.

After text construction, we partition the benchmark by source dataset. Ten datasets supply detector training and separate validation samples; MVTec AD, VisA, MVTec 3D-AD, and MPDD are reserved for testing. The resulting splits contain 49,015 training, 1,000 validation, and 5,283 test image–text samples. No cases from these four datasets are used to train the detector. Offline text construction nevertheless covers all fourteen datasets, including the eventual test sources; the held-out designation therefore applies to detector training. Construction and review details are provided in Appendix C.

## 4 METHOD

Inspired by OPDVR Lin et al. (2026), AD-VRPD regulates on-policy self-distillation using verifiable trajectory feedback and GT-marked teacher views. Since privileged visual input does not guarantee useful supervision at every token, a verifier identifies conflicting distillation directions and adjusts their targets (Figure 3).

### 4.1 PRIVILEGED-VIEW ON-POLICY SELF-DISTILLATION

The student receives the unmarked query image $x$ and diagnostic text $u$. The teacher receives the same text and a privileged image $\tilde{x} = \mathcal{M}(x, \mathcal{B}^*)$ with GT box markings. Both originate from the same model initialization $\theta_0$, which remains fixed for the teacher throughout training. On student-generated trajectories, the two models predict the next token from the same prefix:

$$p_t(v) = \pi_\theta(v \mid x, u, y_{<t}), \qquad q_t(v) = \pi_{\theta_0}(v \mid \tilde{x}, u, y_{<t}). \tag{2}$$

This provides teacher supervision on prefixes encountered by the student while preserving the student's reference-free inference interface.

For efficient distillation, we retain the student's Top-$K$ token indices ($K = 100$), use the same indices for the teacher, and aggregate the remaining vocabulary into a tail bin. The sampled token $s = y_t$ is always represented separately, including when it falls outside Top-$K$. We denote these normalized, aligned distributions by $P_t$ and $Q_t$, with $P_{t,s} = p_t(s)$ and $Q_{t,s} = q_t(s)$. Appendix D gives the complete construction.

### 4.2 VERIFIER-GUIDED TARGET RECONSTRUCTION

**Trajectory-level feedback.** A verifier assigns each response a scalar reward $r$ using the GT anomaly label and boxes. For anomalous images predicted as anomalous with nonempty box predictions, it greedily matches boxes in descending IoU order, allowing only positive-overlap pairs. The reward averages localization-quality scores over GT boxes and penalizes unmatched predictions. A configurable threshold $\tau \in (0, 1)$ controls the required localization quality: the score is negative below $\tau$, zero at $\tau$, and positive above it. Correct normal decisions receive positive rewards; false

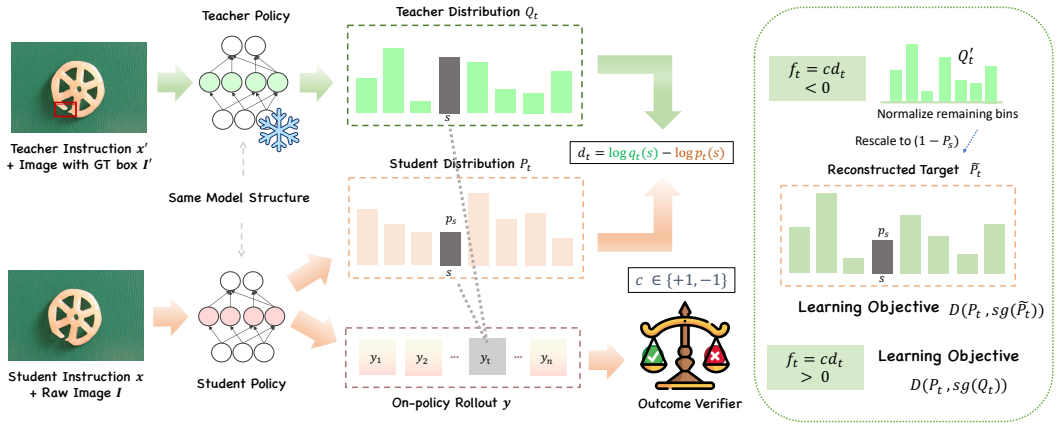



Figure 3: AD-VRPD uses a frozen GT-view teacher and an unmarked-image student. Verifier conflicts ($f_t < 0$) trigger target reconstruction; otherwise, ordinary distillation applies. Gray bars indicate the sampled token, illustrated within Top-$K$.

alarms, missed anomalies, and invalid responses receive negative rewards. The full scoring protocol appears in Appendix D.2.

We convert the reward into a direction shared by all tokens in the trajectory:

$$c = \begin{cases} +1, & r > 0, \\ -1, & r \le 0, \end{cases} \qquad d_t = \log q_t(s) - \log p_t(s), \qquad f_t = c\, d_t. \tag{3}$$

Here $d_t$ indicates whether the teacher favors increasing or decreasing the sampled-token probability. A negative $f_t$ identifies a conflict: the teacher would suppress a token from a positively rewarded response or reinforce one from a nonpositively rewarded response. This is a trajectory-level quality signal, not a token-level correctness annotation; the verifier and branch selection are not differentiated.

**Reconstructing the target.** When $f_t < 0$, we retain the student's current sampled-token mass and use the teacher to allocate the remaining mass. Let $\mathrm{sg}(\cdot)$ denote stop-gradient, and define

$$a_t = \mathrm{sg}(p_t(s)), \qquad Q'_{t,i} = \frac{Q_{t,i}}{1 - q_t(s)} \quad (i \neq s). \tag{4}$$

The reconstructed target is

$$\widetilde{P}_{t,i} = \begin{cases} a_t, & i = s, \\ (1 - a_t) Q'_{t,i}, & i \neq s. \end{cases} \tag{5}$$

Thus $\widetilde{P}_{t,s} = a_t$ and $\sum_{i \neq s} \widetilde{P}_{t,i} = 1 - a_t$, so the target remains normalized. Reconstruction removes the conflicting change in sampled-token mass from the target while retaining teacher supervision over the other bins, including the tail. This is a constraint on the distillation target, not a guarantee that the student's probability remains unchanged after optimization.

### 4.3 Training Objective

Let $D$ denote the distillation divergence, instantiated with equal-weight Jensen–Shannon divergence. The token loss compares the full student distribution with a detached target:

$$\ell_t = \begin{cases} D\left(P_t, \mathrm{sg}(\widetilde{P}_t)\right), & f_t < 0, \\ D(P_t, \mathrm{sg}(Q_t)), & f_t \ge 0. \end{cases} \tag{6}$$

All bins participate in the loss. For trajectory $n$, token position $t$, and valid-token indicator $h_{nt}$, the training objective is

$$\mathcal{L}_{\text{AD-VRPD}} = \frac{\sum_{n,t} h_{nt} \omega_{nt} \ell_{nt}}{\sum_{n,t} h_{nt}}, \tag{7}$$

where $\omega_{nt}$ combines the capped policy importance ratio and rollout-correction weight (Appendix D). We sample four responses per input and optimize the student only. At inference, neither the privileged teacher view nor the verifier is required.

# 5 EXPERIMENTS

## 5.1 SETUP AND EVALUATION

**Setup.** We train on ten source datasets and evaluate on the pooled 5,283-image test set from MVTec AD, VisA, MVTec 3D-AD, and MPDD, all held out from detector training (Section 3). We use Qwen3-VL-8B-Instruct (abbreviated as Qwen3-VL-8B in the tables) as the backbone. We compare the base model, vanilla OPSD, and AD-VRPD. OPSD and AD-VRPD share the same training configuration and a 700-step budget; for each method, we select the checkpoint with the highest validation Box-F1@0.1. Validation and evaluation use greedy decoding. Detailed training settings are provided in Appendix E.

**Evaluation.** We report precision, recall, and F1 for image-level anomaly classification and box-level localization at IoU thresholds of 0.1 and 0.5. Image-level metrics evaluate the predicted anomaly decision, while box-level metrics independently evaluate predicted boxes through one-to-one matching with GT boxes, treating all defect types as one anomaly class. All metrics are micro-averaged over the test set. Box-F1@0.5 is the primary localization metric, while Box-F1@0.1 measures localization under a more permissive overlap criterion. Formal definitions are provided in Appendix B.

## 5.2 MAIN RESULTS

We examine whether self-distillation can improve a compact detector sufficiently to compete with larger general-purpose models. Table 1 compares AD-VRPD with general multimodal models and specialized anomaly detectors, including its pretrained backbone to show the effect of training at a fixed model size. All methods receive unmarked images and diagnostic text under the pairing rules in Section 3. Test-source examples contribute to offline text construction, and anomalous queries use GT defect-type descriptions; the held-out condition applies to detector training. The open-weight comparisons include Qwen3-VL Qwen Team (2025), InternVL3 Zhu et al. (2025), and LLaVA-NeXT Li et al. (2024); specialized comparisons include IAD-R1 Li et al. (2025) and AnomalyR1 Chao et al. (2025).

**From conservative decisions to greater defect coverage.** Qwen3-VL-8B has high image precision (95.3%) but low recall (47.4%). AD-VRPD increases recall to 81.1% with precision of 88.9%, recovering more anomalies at the cost of some image-level false positives. At the box level, both precision and recall improve, with recall at IoU 0.5 rising from 27.1% to 43.0%. The localization gain therefore extends beyond a shift toward more positive image decisions.

**Localization under permissive and strict overlap requirements.** Box-F1 increases from 46.4% to 65.9% at IoU 0.1 and from 34.2% to 46.1% at IoU 0.5. Gains under the stricter criterion show improvement beyond partial-overlap detections, although the gap between thresholds indicates remaining difficulty in precise alignment. Classification and localization also differ across models: Gemini-3.8-Flash has higher Image-F1 than AD-VRPD (87.2% versus 84.8%) but lower Box-F1@0.5 (39.6% versus 46.1%).

**Competitive performance without increasing model size.** AD-VRPD ranks third on both box-level F1 metrics. At IoU 0.5, it exceeds the 27B Qwen baseline and several proprietary models, while remaining below Qwen3.8-Max and Kimi-K3. These gains use a frozen teacher initialized from the same 8B backbone, without increasing detector size. Appendix G illustrates localization improvements and a failure case.

Table 1: Image-level classification and box-level localization on the pooled test set of MVTec AD, VisA, MVTec 3D-AD, and MPDD (5,283 images; micro-averaged, %; higher is better). Bold and underlined scores denote the best and second-best results. Shading identifies AD-VRPD; "—" denotes unspecified parameter counts.

| | | Image-level | | | Box-level (IoU $\geq$ 0.1) | | | Box-level (IoU $\geq$ 0.5) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **Params** | Precision | Recall | F1-score | Precision | Recall | F1-score | Precision | Recall | F1-score |
| ***Closed-source models*** | | | | | | | | | | |
| Qwen3.8-Max | — | 87.5 | 85.9 | <u>86.7</u> | 69.9 | **66.6** | **68.2** | <u>53.5</u> | **50.9** | **52.2** |
| Kimi-K3 | — | 87.7 | 79.7 | 83.5 | 70.3 | 63.6 | <u>66.8</u> | 50.6 | <u>45.8</u> | <u>48.1</u> |
| Qwen3.7-Plus | — | 90.1 | 77.1 | 83.1 | <u>73.0</u> | 57.3 | 64.2 | 49.4 | 38.8 | 43.5 |
| GPT-5.6-Terra | — | 78.2 | 79.8 | 79.0 | 63.3 | 57.5 | 60.3 | 42.0 | 38.2 | 40.0 |
| Gemini-3.8-Flash | — | 92.1 | 82.8 | **87.2** | 63.7 | 50.5 | 56.4 | 44.8 | 35.5 | 39.6 |
| GPT-5.6-Luna | — | 77.8 | **88.9** | 83.0 | 57.0 | <u>65.5</u> | 61.0 | 36.8 | 42.3 | 39.3 |
| ***Open-weight models*** | | | | | | | | | | |
| Qwen3-VL-8B | 8B | <u>95.3</u> | 47.4 | 63.3 | 62.9 | 36.8 | 46.4 | 46.4 | 27.1 | 34.2 |
| Qwen3.8-27B | 27B | 82.1 | 82.4 | 82.3 | 65.7 | 59.6 | 62.5 | 48.0 | 43.6 | 45.7 |
| GLM-4.1V-9B | 9B | **96.8** | 50.5 | 66.4 | **80.6** | 33.5 | 47.4 | **60.2** | 25.0 | 35.3 |
| InternVL3-8B | 8B | 89.0 | 57.9 | 70.1 | 33.8 | 21.3 | 26.1 | 10.7 | 6.8 | 8.3 |
| LLaVA-NeXT-8B | 8B | 79.9 | 18.5 | 30.1 | 26.2 | 5.1 | 8.6 | 6.2 | 1.2 | 2.0 |
| ***Specialized anomaly detection methods*** | | | | | | | | | | |
| AnomalyR1 | 3B | 66.3 | <u>88.0</u> | 75.6 | 7.0 | 19.9 | 10.4 | 2.6 | 7.3 | 3.8 |
| IAD-R1 | — | 92.6 | 44.0 | 59.6 | 23.1 | 9.1 | 13.1 | 4.6 | 1.8 | 2.6 |
| **AD-VRPD (Ours)** | **8B** | 88.9 | 81.1 | 84.8 | 71.0 | 61.5 | 65.9 | 49.7 | 43.0 | 46.1 |

### 5.3 SELF-DISTILLATION

We compare Qwen3-VL-8B, OPSD, and AD-VRPD to separate the benefit of privileged-view self-distillation from verifier-guided target reconstruction. The trained variants share the backbone, data, training budget, and validation-based selection rule.

Table 2: Self-distillation on the pooled test set of MVTec AD, VisA, MVTec 3D-AD, and MPDD (5,283 images; micro-averaged, %). OPSD and AD-VRPD use identical training settings and checkpoint-selection rules. Bold denotes the best result.

| | Image-level | | | Box-level @ 0.1 | | | Box-level @ 0.5 | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | **P** | **R** | **F1** | **P** | **R** | **F1** | **P** | **R** | **F1** |
| Qwen3-VL-8B | **95.3** | 47.4 | 63.3 | 62.9 | 36.8 | 46.4 | 46.4 | 27.1 | 34.2 |
| + OPSD | 91.1 | 76.1 | 82.9 | 66.7 | 61.3 | 63.9 | 48.6 | 42.6 | 45.4 |
| + AD-VRPD | 88.9 | **81.1** | **84.8** | **71.0** | **61.5** | **65.9** | **49.7** | **43.0** | **46.1** |

**Contribution of privileged-view supervision.** OPSD accounts for most of the localization gain, raising Box-F1@0.5 from 34.2% to 45.4% (Table 2). Box recall increases by 15.5 points while precision also improves. With identical initial teacher and student parameters, this supports the value of privileged visual guidance and on-policy supervision without a larger teacher.

**Additional benefit of verifier-guided reconstruction.** AD-VRPD yields additional Box-F1 gains of 2.0 points at IoU 0.1 and 0.7 points at IoU 0.5. Relative to OPSD, box precision increases by 4.3 and 1.1 points, respectively, while recall changes by only 0.2 and 0.4 points. The additional benefit therefore primarily reflects improved precision at similar recall, rather than a large increase in defect coverage. These gains are modest, particularly under strict matching, and this comparison evaluates gating and reconstruction jointly.

### 5.4 EFFECT OF DIAGNOSTIC TEXT

We compare three text inputs across the full test set. *General* contains only the generic inspection instruction; *Normal* adds the normal-state description; and *Full* additionally includes possible defects. The output format is identical across conditions. We evaluate the base 8B model and AD-VRPD, reporting Image-F1 and Box-F1 at both IoU thresholds in Table 3.

Table 3: Effect of diagnostic text on the full test set. Full uses the diagnostic input of Tables 1 and 2 and reports the same results. Scores are percentages. Bold indicates the best text condition within each model and metric.

| Model | Text input | Image-F1 | Box-F1@0.1 | Box-F1@0.5 |
|---|---|---|---|---|
| Qwen3-VL-8B | General | 50.4 | 35.1 | 28.1 |
| | Normal | **78.4** | 39.8 | 28.9 |
| | Full | 63.3 | **46.4** | **34.2** |
| AD-VRPD | General | 64.7 | 48.6 | 35.5 |
| | Normal | 67.4 | 50.7 | 39.1 |
| | Full | **84.8** | **65.9** | **46.1** |

**Inspection criteria help, with different effects on decision and localization.** For the base model, adding the normal-state description raises Image-F1 from 50.4% to 78.4%. Adding possible defects further increases Box-F1 from 39.8% to 46.4% at IoU 0.1 and from 28.9% to 34.2% at IoU 0.5, but reduces Image-F1 to 63.3%. Thus, the full diagnostic input yields the best localization F1 for the base model, while normal-state text alone yields the best decision F1. More detailed text does not uniformly improve every metric.

**Training improves performance under the full diagnostic input.** AD-VRPD performs best with Full text, achieving 84.8%, 65.9%, and 46.1% in Image-F1, Box-F1@0.1, and Box-F1@0.5, respectively. These correspond to gains of 21.5, 19.5, and 11.9 percentage points over Qwen3-VL-8B under the same input condition. This supports explicit inspection criteria, although removing text also shifts the input away from AD-VRPD's training condition. The experiment evaluates text conditioning, not the gate in isolation. Logical-anomaly results show different trade-offs (Appendix A).

### 5.5 DEFECT SIZE ANALYSIS

**Evaluation by defect size.** We measure defect size by the relative area $a(b) = \text{area}(b)/(WH)$ of each ground-truth box in an image of width $W$ and height $H$. Motivated by the predominance of small defects discussed in the Introduction, we group GT boxes as small ($a < 1\%$), medium ($1\% \leq a \leq 5\%$), or large ($a > 5\%$). We evaluate whether the localization improvements extend to the small defects that dominate this benchmark.

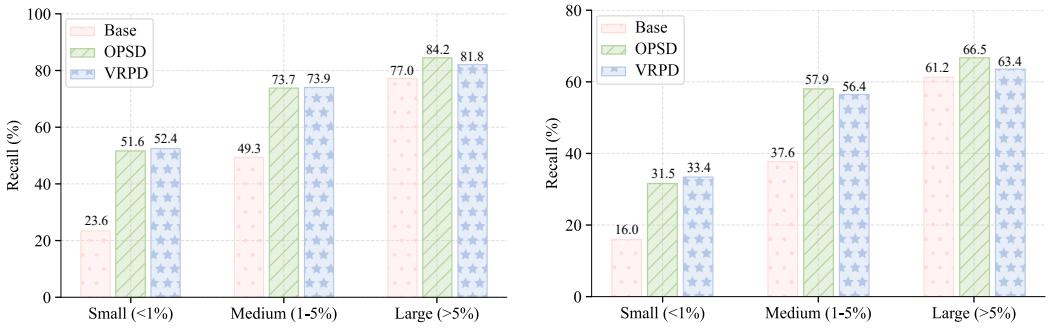
We match predictions to GT boxes before grouping by size, using the main evaluation protocol. Group-wise Box Recall is the fraction of GT boxes matched at the specified IoU threshold.

**Small-defect localization and trade-offs.** Small defects are hardest for the base model, with recall of 23.6% and 16.0% at IoU 0.1 and 0.5. OPSD increases these to 51.6% and 31.5%; AD-VRPD reaches 52.4% and 33.4% (Figure 4). Most improvement comes from self-distillation, with additional gains of 0.8 and 1.9 points from AD-VRPD.

This benefit is not uniform: at IoU 0.5, AD-VRPD has lower medium- and large-defect recall than OPSD (56.4%/63.4% versus 57.9%/66.5%). Overall recall remains close (Table 2), indicating a shift toward small defects rather than uniformly greater coverage.

## 6 LIMITATIONS

Our reference-free setting uses target-informed text: construction includes test-source examples and annotations, and anomalous queries use GT defect-type descriptions. Generalization to independently authored, category-wide instructions without these inputs remains untested.




(a) Box Recall at IoU $\geq 0.1$ (b) Box Recall at IoU $\geq 0.5$

Figure 4: Box Recall by relative GT box area for Qwen3-VL-8B, OPSD, and AD-VRPD (labeled Base and VRPD for Qwen3-VL-8B and AD-VRPD in the plots). AD-VRPD improves small-defect recall over both comparisons; gains over OPSD are not uniform across sizes. Recall does not measure false positives.

Training requires annotated boxes and a task-specific verifier. Trajectory feedback cannot identify individual erroneous tokens or boxes, and sensitivity to the configurable reward threshold is untested. Gains vary across defect sizes and anomaly types; precise box alignment remains challenging.

## 7 Conclusion

Reference-free anomaly detection still requires a specification of what constitutes normality. We make this requirement explicit through diagnostic text and adapt MMR-AD to evaluate joint anomaly decisions and box localization under this interface. Building on verifiable-reward-guided distillation, AD-VRPD transfers supervision from a GT-view teacher and reconstructs targets when the teacher's sampled-token direction conflicts with trajectory feedback. Privileged-view self-distillation provides the main localization improvement, and verifier-guided reconstruction yields an additional gain in the evaluated runs. The resulting compact detector is competitive with larger and proprietary models under the shared diagnostic-text protocol. Our findings support combining explicit inspection criteria with verified privileged supervision to improve compact anomaly detectors.

## References

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos, Matthieu Geist, and Olivier Bachem. On-Policy Distillation of Language Models: Learning from Self-Generated Mistakes. *arXiv:2306.13649*, 2023. https://arxiv.org/abs/2306.13649.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. *arXiv:2011.08785*, 2020. https://arxiv.org/abs/2011.08785.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards Total Recall in Industrial Anomaly Detection. In *CVPR*, 2022. https://openaccess.thecvf.com/content/CVPR2022/html/Roth_Towards_Total_Recall_in_Industrial_Anomaly_Detection_CVPR_2022_paper.html.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. DRAEM—A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection. In *ICCV*, 2021. https://arxiv.org/abs/2108.07610.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In *ICML*, 2021. https://arxiv.org/abs/2103.00020.

Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. AnomalyGPT: Detecting Industrial Anomalies Using Large Vision-Language Models. In *AAAI*, 2024. https://arxiv.org/abs/2308.15366.

Yanhui Li, Yunkang Cao, Chengliang Liu, Yuan Xiong, Xinghui Dong, and Chao Huang. IAD-R1: Reinforcing Consistent Reasoning in Industrial Anomaly Detection. *arXiv:2508.09178*, 2025. https://arxiv.org/abs/2508.09178.

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. https://openreview.net/forum?id=JDiER86r8v.

Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation. In *CVPR*, 2023. https://arxiv.org/abs/2303.14814.

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection. In *ICLR*, 2024. https://arxiv.org/abs/2310.18961.

Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond Dents and Scratches: Logical Constraints in Unsupervised Anomaly Detection and Localization. *International Journal of Computer Vision*, 2022. https://doi.org/10.1007/s11263-022-01578-9.

Xincheng Yao, Zefeng Qian, Chao Shi, Jiayang Song, and Chongyang Zhang. MMR-AD: A Large-Scale Multimodal Dataset for Benchmarking General Anomaly Detection with Multimodal Large Language Models. *arXiv:2604.10971*, 2026. https://arxiv.org/abs/2604.10971.

Jingtai He, Shiyuan Meng, Wenchao Meng, and Qinmin Yang. ADOPD: Reference-Privileged On-Policy Distillation for MLLM-Based Industrial Anomaly Detection. *arXiv:2608.09789*, 2026. https://arxiv.org/abs/2608.09789.

Shuimu Chen, Jing Jin, Nan Su, Hongbo Xu, Zebang Cheng, Wenming Yang, Fei Ma, and Guijin Wang. OPD-IAD: From Language Judgment to Industrial Anomaly Detection via On-Policy Self-Distillation. *arXiv:2607.18850*, 2026. https://arxiv.org/abs/2607.18850.

Wenze Lin, Jiale Zhao, Xitai Jiang, Songde Rao, Yining Li, Shenzhi Wang, Bingxiang He, and Gao Huang. On-policy Distillation with Verifiable Reward. *arXiv:2608.24696*, 2026. https://arxiv.org/abs/2608.24696.

Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. SimpleNet: A Simple Network for Image Anomaly Detection and Localization. In *CVPR*, 2023. https://arxiv.org/abs/2303.15140.

Hanqiu Deng and Xingyu Li. Anomaly Detection via Reverse Distillation From One-Class Embedding. In *CVPR*, 2022. https://openaccess.thecvf.com/content/CVPR2022/html/Deng_Anomaly_Detection_via_Reverse_Distillation_From_One-Class_Embedding_CVPR_2022_paper.html.

Kilian Batzner, Lars Heckler, and Rebecca König. EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies. In *WACV*, 2024. https://arxiv.org/abs/2303.14535.

Xuhai Chen, Yue Han, and Jiangning Zhang. APRIL-GAN: A Zero-/Few-Shot Anomaly Classification and Segmentation Method for CVPR 2023 VAND Workshop Challenge Tracks 1&2: 1st Place on Zero-shot AD and 4th Place on Few-shot AD. *arXiv:2305.17382*, 2023. https://arxiv.org/abs/2305.17382.

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. *arXiv:2407.15795*, 2024. https://arxiv.org/abs/2407.15795.

Qwen Team. Qwen3-VL Technical Report. *arXiv preprint arXiv:2511.21631*, 2025.

Jinguo Zhu, Weiyun Wang, Zhe Chen, et al. InternVL3: Exploring Advanced Training and Test-Time Recipes for Open-Source Multimodal Models. *arXiv preprint arXiv:2504.10479*, 2025.

Bo Li, Kaichen Zhang, Hao Zhang, et al. LLaVA-NeXT: Stronger LLMs Supercharge Multimodal Capabilities in the Wild. 2024. https://llava-vl.github.io/blog/2024-05-10-llava-next-stronger-llms/.

Yuhao Chao, Jie Liu, Jie Tang, and Gangshan Wu. AnomalyR1: A GRPO-based End-to-end MLLM for Industrial Anomaly Detection. *arXiv:2504.11914*, 2025. https://arxiv.org/abs/2504.11914.

## A Diagnostic Text: Logical-Anomaly Subgroup

Table 4 reports subgroup results from the diagnostic-text analysis. This subgroup does not reproduce the overall model ranking: with Full text, AD-VRPD has lower Image-F1 and slightly lower Box-F1 than Qwen3-VL-8B. Accordingly, the overall results do not establish uniform improvement across anomaly types. For Qwen3-VL-8B, Normal favors image-level recall, whereas Full increases precision with lower recall.

Table 4: Logical-anomaly subgroup results (%). Image precision and recall expose trade-offs hidden by F1 alone.

| Model | Text | Image-P | Image-R | Image-F1 | Box-F1@0.1 | Box-F1@0.5 |
|---|---|---|---|---|---|---|
| Qwen3-VL-8B | General | 28.3 | 11.0 | 15.9 | 11.4 | 4.4 |
| | Normal | 53.5 | 91.5 | 67.5 | 32.8 | 21.7 |
| | Full | 79.5 | 52.5 | 63.3 | 36.2 | 23.9 |
| AD-VRPD | General | 27.9 | 16.1 | 20.4 | 7.1 | 4.7 |
| | Normal | 38.8 | 28.0 | 32.5 | 15.0 | 12.0 |
| | Full | 51.5 | 58.5 | 54.8 | 31.0 | 23.1 |

## B Image and Box Evaluation Protocol

**Image-level metrics.** Let $y_i, \hat{y}_i \in \{0, 1\}$ be the GT and model anomaly decisions, with anomaly as the positive class. Define

$$T_I = \sum_i \mathbf{1}[y_i = 1, \hat{y}_i = 1], \tag{8}$$

$$F_I = \sum_i \mathbf{1}[y_i = 0, \hat{y}_i = 1], \qquad U_I = \sum_i \mathbf{1}[y_i = 1, \hat{y}_i = 0]. \tag{9}$$

Image precision, recall, and F1 are $T_I/(T_I + F_I)$, $T_I/(T_I + U_I)$, and $2T_I/(2T_I + F_I + U_I)$, respectively. Each image contributes at most once to these counts; box overlap does not affect classification correctness.

**Box-level metrics.** Let $G_i$ and $B_i$ be the GT and parsed predicted box sets. For each threshold $\tau \in \{0.1, 0.5\}$, let $M_i^\tau$ denote a one-to-one matching whose edges satisfy $\mathrm{IoU}(b, g) \geq \tau$, and write $m_i^\tau = |M_i^\tau|$. Matching is performed separately at each threshold, without defect-type constraints. Pooled counts are

$$T_B^\tau = \sum_i m_i^\tau, \tag{10}$$

$$F_B^\tau = \sum_i (|B_i| - m_i^\tau), \qquad U_B^\tau = \sum_i (|G_i| - m_i^\tau). \tag{11}$$

The reported metrics are

$$\text{Box-P@}\tau = \frac{T_B^\tau}{T_B^\tau + F_B^\tau}, \qquad \text{Box-R@}\tau = \frac{T_B^\tau}{T_B^\tau + U_B^\tau}, \tag{12}$$

$$\text{Box-F1@}\tau = \frac{2T_B^\tau}{2T_B^\tau + F_B^\tau + U_B^\tau}. \tag{13}$$

All test images participate. An unmatched prediction is an FP even on a normal image; duplicate predictions can match a GT box at most once. An unmatched GT is an FN. A normal image with no predicted boxes contributes no box TP, FP, or FN. A localization error can contribute both an FP and an FN. No confidence ranking or AP computation is implied by this matching protocol.

**Independent output channels and aggregation.** Image-level metrics use the predicted Yes/No decision, while box-level metrics use the parsed boxes independently. Box counts are pooled across all normal and anomalous images. A Yes response without boxes produces no box matches. The evaluation separates decision accuracy from localization quality, in contrast to the training verifier, which prioritizes the parsed decision.

## C Diagnostic-Text Construction Details

Table 5 summarizes the split sizes and normal/anomalous composition.

Table 5: Detector training and evaluation splits. Counts denote image–text samples.

| Split | Source datasets | Total | Normal | Anomalous |
|---|---|---|---|---|
| Training | 10 | 49,015 | 24,709 | 24,306 |
| Validation | 10 | 1,000 | 500 | 500 |
| Test | 4 | 5,283 | 1,854 | 3,429 |

**Source datasets and counting.** The ten detector-training sources are 3cad, cableinspect, goodsad, manta, miad, mvtecloco, realiad, realiad-d3, texturead, and wfdd. The test sources are mvtec, visa, mvtec3d, and mpdd; names here follow the dataset manifest. Validation uses separate samples from the ten training sources. The 49,015 training image–text samples correspond to 49,014 unique images, with one image occurring twice.

**Description generation and revision.** Each defect type is represented by one primary normal reference, optional additional normal references, and at most four anomalous examples. Up to four additional normal references are included to capture variation in normal appearance. Diagnostic observations summarized from the source reasoning records are supplied as auxiliary context, with instructions not to copy them verbatim. After generation, an automated check assesses coverage, missed observations, and contradictions. Incomplete coverage or detected inconsistencies trigger one rewrite and a second check. Descriptions undergo exhaustive manual review for twelve datasets and sampled review for realiad and manta.

**Pairing and information available during construction.** Descriptions are shared by defect type. Anomalous queries receive their GT-type description, whereas normal queries are randomly assigned a type from the same category. The intended descriptions specify inspection criteria without explicit query labels or box coordinates. However, GT-based query assignment and construction from target-source examples both supply information about the inspection task. Exclusion from detector training does not imply exclusion from text construction, and dataset-disjoint splitting does not by itself establish semantic category disjointness.

# D Distillation and Verifier Details

## D.1 Aligned Top-$K$ and tail distributions

Let $\mathcal{T}_t = \mathrm{TopK}(p_t)$, $K = 100$, and $\mathcal{A}_t = \mathcal{T}_t \setminus \{s\}$. Using the same ordered indices for student and teacher, we construct

$$P_t = \big[(p_t(v))_{v\in\mathcal{A}_t},\ p_t(s),\ p_{t,\perp}\big], \tag{14}$$

$$Q_t = \big[(q_t(v))_{v\in\mathcal{A}_t},\ q_t(s),\ q_{t,\perp}\big], \tag{15}$$

$$p_{t,\perp} = 1 - p_t(s) - \sum_{v\in\mathcal{A}_t} p_t(v), \qquad q_{t,\perp} = 1 - q_t(s) - \sum_{v\in\mathcal{A}_t} q_t(v). \tag{16}$$

The sampled-token bin is disjoint from the tail. When $s \in \mathcal{T}_t$, there are $K-1$ other token bins, one sampled-token bin, and one tail bin; otherwise there are $K$ other token bins and the two additional bins. These give $K+1$ and $K+2$ effective bins, respectively. A fixed-width representation can add a zero-mass placeholder in the former case. Support selection is held fixed during differentiation. Target reconstruction applies to all non-sampled bins, including the tail.

## D.2 Trajectory reward

Let $a^* \in \{0,1\}$ be the GT anomaly label and $\hat{a}$ the parsed decision. For an anomalous image with a positive decision and nonempty predictions, form all prediction–GT pairs with IoU strictly greater than zero. Sort these pairs by descending IoU and accept each pair if neither box has already been matched. Denote the matched IoU for GT box $j$ by $u_j$, assigning $u_j = 0$ to unmatched GT boxes. The localization score is

$$r_{\mathrm{box}}(u) = \begin{cases} -\beta\left(1 - \dfrac{u}{\tau}\right), & 0 \le u < \tau, \\ \dfrac{u-\tau}{1-\tau}, & \tau \le u \le 1, \end{cases} \qquad 0 < \tau < 1, \quad \beta > 0. \tag{17}$$

Here $\tau$ sets the localization requirement, and $\beta$ controls the penalty for low-overlap or unmatched GT boxes. The reward definition supports different choices of $\tau$ without changing the target-reconstruction rule. We use a fixed $\tau = 0.1$ and $\beta = 0.1$ in our experiments.

With $N_{\mathrm{gt}}$ GT boxes, $N_{\mathrm{pred}}$ predictions, and $N_{\mathrm{unmatched}}$ predictions remaining after matching, define

$$r_{\mathrm{loc}} = \frac{1}{N_{\mathrm{gt}}} \sum_{j=1}^{N_{\mathrm{gt}}} r_{\mathrm{box}}(u_j) - \lambda_{\mathrm{fp}} \frac{N_{\mathrm{unmatched}}}{\max(1, N_{\mathrm{pred}})}, \qquad \lambda_{\mathrm{fp}} = 0.1. \tag{18}$$

The complete reward cases are listed in Table 6.

Table 6: Trajectory rewards used by the training verifier. For a parsed No decision, accompanying boxes are ignored.

| GT status | Parsed response | Reward |
|---|---|---|
| Normal | No | $+1.0$ |
| Normal | Yes | $-1.0$ |
| Anomalous | No | $-1.0$ |
| Anomalous | Yes, no boxes | $-0.5$ |
| Anomalous | Yes, nonempty boxes | $r_{\mathrm{loc}}$ |
| Either | Invalid response format | $-2.0$ |

Matching eligibility and localization scoring use different thresholds. A positive-overlap match below $\tau$ receives a negative localization score but is not an unmatched prediction. At IoU exactly $\tau$, the localization score is zero; only larger IoUs receive positive scores. Predictions with no positive-overlap pair, or left over after competing for GT boxes, contribute to the unmatched-prediction penalty. Consequently, a matched box does not necessarily yield positive trajectory feedback. Equation 3 assigns $c = -1$ when the total reward is zero.

The verifier prioritizes the parsed anomaly decision: a No response is scored as a negative decision even if it includes boxes. Such boxes are ignored, without an additional consistency penalty. This training convention is distinct from the independent image- and box-level evaluation protocol in Appendix B. Each sampled response receives one scalar reward, shared across token gates; no per-box or per-token correctness labels are assigned.

### D.3 Loss weighting

Write $\pi_{\text{old}}$ for the old policy used in the importance ratio, distinct from the frozen teacher. The weighting in Equation 7 is

$$\omega_{nt} = \eta_{nt} w_{nt}, \qquad \eta_{nt} = \min\left\{ \frac{\pi_\theta(y_{nt} \mid x_n, u_n, y_{n,<t})}{\pi_{\text{old}}(y_{nt} \mid x_n, u_n, y_{n,<t})},\ 2.0 \right\}, \tag{19}$$

where $w_{nt} = \bar{\rho}^2$ is the rollout-correction weight. The weighted losses are averaged over valid tokens rather than normalized by the sum of weights.

## E Training Configuration

Table 7 summarizes the training configuration. The teacher uses the same initial checkpoint as the student and remains frozen. Validation and evaluation use greedy decoding.

Table 7: Training configuration for AD-VRPD.

| Setting | Value |
|---|---|
| Initial checkpoint | `Qwen/Qwen3-VL-8B-Instruct` |
| Teacher update | Frozen throughout training |
| Optimizer | AdamW |
| Learning rate | $10^{-5}$ |
| Weight decay | 0.01 |
| Learning-rate schedule | 10 warmup steps, then constant |
| Prompts per training batch | 64 |
| Responses per prompt | 4 |
| Sequences per training batch | 256 |
| Mini-batch setting | 32 |
| Micro-batch size per GPU | 2 |
| GPU configuration | 8× NVIDIA A100 |
| Training budget | 700 steps |
| Random seed | 42 |
| Checkpoint selection | Highest validation Box-F1@0.1 |
| Maximum prompt length | 12,800 tokens |
| Maximum response length | 512 tokens |
| Rollout temperature / top-$p$ | 1.0 / 1.0 |
| Validation and evaluation decoding | Greedy |
| Distillation support | Student Top-100, sampled token, and tail |
| Verifier threshold $\tau$ (experiments) | 0.1 |
| Importance-ratio cap | 2.0 |
| Loss aggregation | Token mean |

## F Defect-size Distribution

Table 8 gives the box counts and proportions for the relative-area groups used in Section 5.5.

Table 8: Defect size distribution over 4,380 GT boxes. Relative area is GT box area divided by image area; shares describe boxes, not images.

| Size group | Relative area | GT boxes | Share |
|---|---|---|---|
| Small | $< 1\%$ | 2,721 | 62.1% |
| Medium | 1–5% | 1,116 | 25.5% |
| Large | $> 5\%$ | 543 | 12.4% |

## G Qualitative Analysis

Figure 5 compares Qwen3-VL-8B and AD-VRPD on selected examples. The first four rows illustrate recovery of a missed defect, tighter localization of separate damaged regions, more complete coverage of multiple defects, and a shift from an object-level box to the local defect. The final row shows a failure: AD-VRPD predicts a region near the object boundary instead of the annotated interior defect, whereas Qwen3-VL-8B overlaps the GT region. These examples illustrate changes in localization behavior and a remaining failure mode; they do not estimate how frequently each behavior occurs.

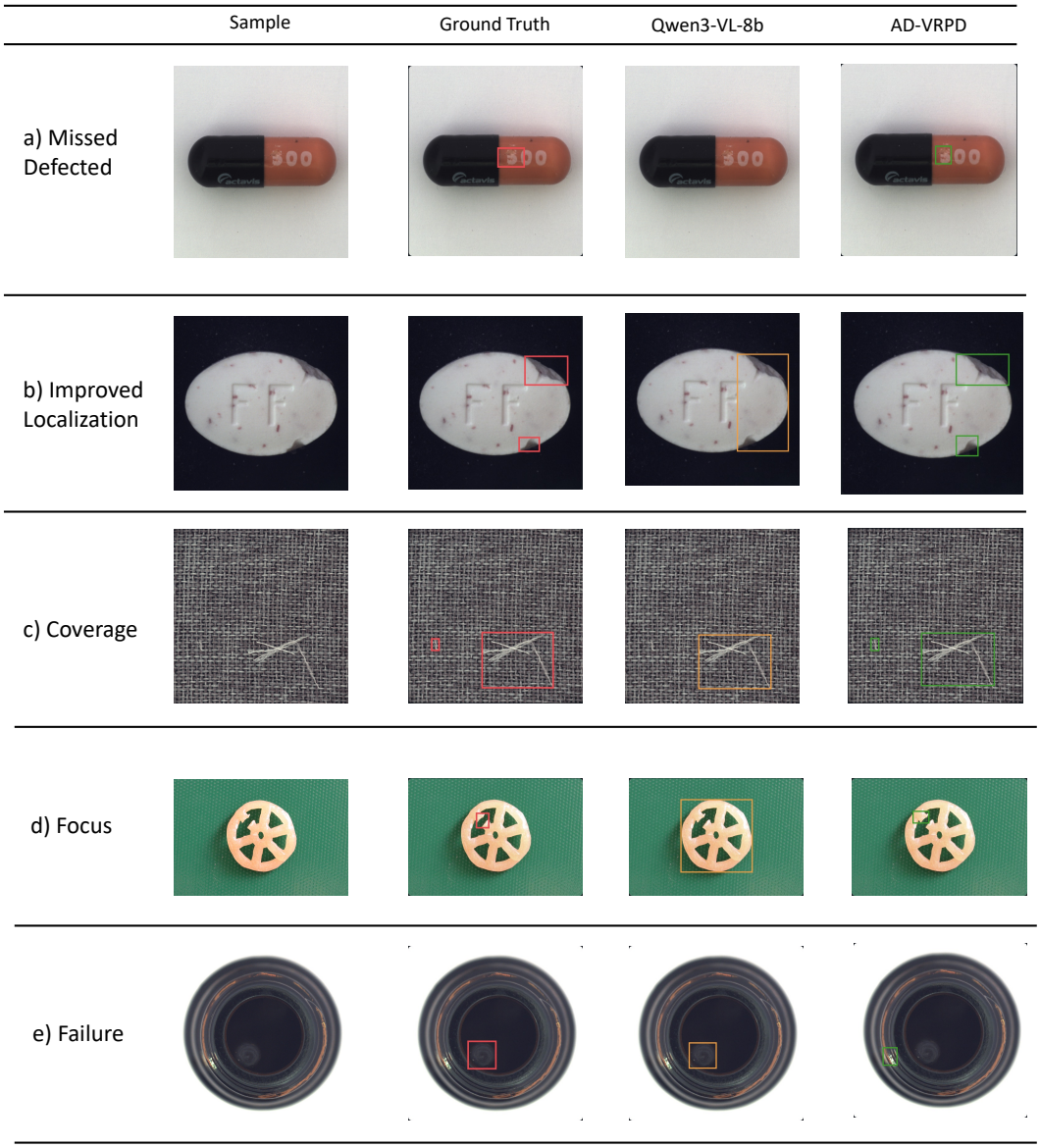



Figure 5: Qualitative comparison with the same backbone. Columns show the input image, ground truth, Qwen3-VL-8B, and AD-VRPD. Red, orange, and green boxes indicate GT, Qwen3-VL-8B predictions, and AD-VRPD predictions, respectively. Rows (a)–(d) illustrate recovered detection, tighter localization, broader defect coverage, and local defect focus; row (e) shows a failure.