PRIDE: PREDICTION RECONCILIATION VIA INTERNAL DIVERGENCE ELICITATION FOR SELF-EVOLVING VISION-LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Vision-language models can answer a visual attribute correctly in isolation yet fail on the same attribute inside a compound query. Because such disagreement may arise from prompt form or sampling noise, we first establish a controlled premise: a subset of atomic–joint disagreements persists under paraphrase, length/schema matching, and repeated rollouts, and is enriched for recoverable joint errors. We introduce PRIDE, a label-free self-evolution framework that filters this stable task-conditioned divergence and maps accepted output corrections to visual-parameter updates through a preservation-aware Jacobian/Gram solve. Using all $6{,}000$ student-view images of ADOPD-Dataset-6K across twelve disjoint 500-image rounds and reserving MMAD exclusively for evaluation, PRIDE establishes a new state of the art on the MMAD AD/DC subtasks under the controlled Qwen3.5-4B setting, reaching $85.57/74.40$ in 0-shot and $87.16/76.02$ in 1-shot evaluation. The gains are further observed across Qwen3-VL-4B, LLaVA-OneVision-1.5-4B, and Gemma-3n-E4B, and transfer beyond industrial anomaly reasoning to open-ended GQA and ChartQA. We additionally report the complete seven-task MMAD profile to verify that improvements on AD/DC do not come at the expense of localization and other non-target capabilities. These results support atomic–joint task-conditioned divergence as a usable endogenous supervision signal for self-evolving vision-language models when reliability, coupling, and local update validity are explicitly controlled.

1 INTRODUCTION

Modern vision–language models (VLMs) exhibit strong visual perception and structured reasoning across diverse tasks (Bai et al., 2025b;a; Qwen Team, 2026; Wang et al., 2025a), yet their behavior can become inconsistent when several visual judgments must be made jointly. Industrial anomaly reasoning makes this failure especially visible: for the same image pair and the same model parameters, a visual property that is answered correctly when queried alone can be missed once multiple properties are composed into a single query. Existing anomaly-oriented VLM methods mainly improve domain adaptation, comparative reasoning, supervision, or reward design (Chao et al., 2025; Zhao et al., 2025; Guan et al., 2025; Kang et al., 2026; Liao et al., 2026; Li et al., 2026; Chen et al., 2026b; Jiang et al., 2026; Wang et al., 2026c; He et al., 2026b; Chen et al., 2026d); they do not directly exploit this discrepancy as an internal training signal.

This observation alone is insufficient to conclude that atomic querying "reveals" a capability suppressed by joint reasoning. Atomic and joint queries differ in prompt composition, and VLM outputs can vary stochastically. We therefore treat atomic–joint divergence as a *candidate* self-supervision signal whose validity must be established before optimization. PRIDE tests whether the disagreement is repeatable across rollouts and robust to controlled prompt reformulations, then separates accepted outputs into two roles: coordinates that require *repair* and coordinates whose current behavior should be *preserved*. Figure 1 summarizes the resulting self-evolution loop.

PRIDE operationalizes this idea in two coupled steps. *Elicitation* obtains dimension-specific atomic predictions and retains only stable divergences as endogenous supervision. Let S_A denote the aggregated atomic response and S_J the joint response. Their reliable difference identifies the repair set

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.

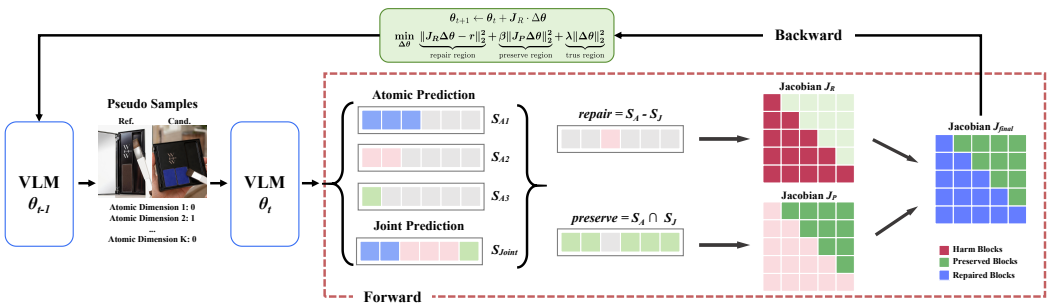

Figure 1: Overview of PRIDE. Given the same visual input, the VLM produces atomic predictions S_A and a joint prediction S_J. Stable atomic–joint divergence defines a repair region $\mathcal{R}$, while consistent responses define a preserve region $\mathcal{P}$. PRIDE maps the desired output correction through the repair Jacobian $J_{\mathcal{R}}$ while constraining changes along the preserve Jacobian $J_{\mathcal{P}}$, and solves a damped local reconciliation objective to obtain $\Delta\theta$. The updated model θ_{t+1} then re-enters the next self-evolution round.

$\mathcal{R}$, whereas their consistent components define the preserve set $\mathcal{P}$. *Reconciliation* then asks for a parameter update that realizes the desired change on $\mathcal{R}$ while limiting drift on currently consistent coordinates in $\mathcal{P}$. Using local Jacobians, PRIDE solves $\min_{\Delta\theta} \|J_{\mathcal{R}}\Delta\theta - r\|_2^2 + \beta\|J_{\mathcal{P}}\Delta\theta\|_2^2 + \lambda\|\Delta\theta\|_2^2$, followed by a trust-region-scaled update. This turns an internal prediction discrepancy into a preservation-aware parameter correction rather than unconditional imitation of atomic outputs.

The distinction is important because atomic predictions are informative but imperfect. As shown in Figure 2, atomic probing recovers a substantial fraction of joint failures while also changing some decisions that were already correct. Direct atomic pseudo-label SFT therefore risks replacing one error mode with another. Likewise, activation steering changes hidden states only at inference (Su et al., 2025; Wu et al., 2025; Sivakumar et al., 2025), while gradient-balancing methods such as PCGrad operate on task gradients rather than explicitly constraining the output coordinates that should remain unchanged (Yu et al., 2020). PRIDE instead uses the atomic–joint discrepancy to specify *what* should change and the local Jacobian geometry to control *how* that change is written into the model.

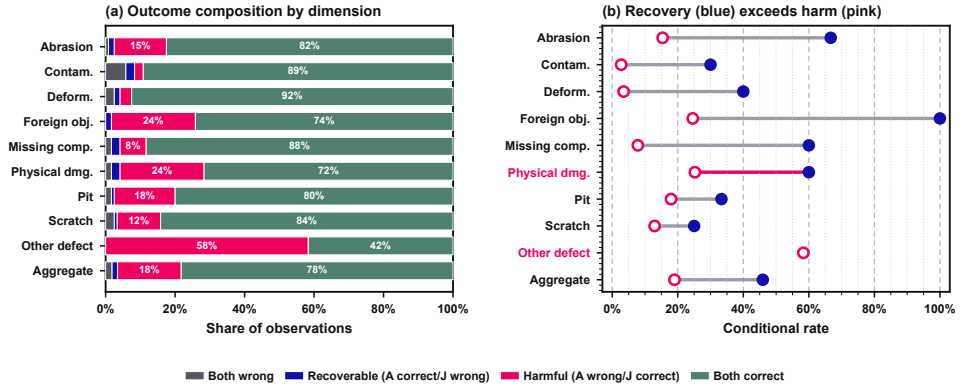

Figure 2: Atomic–joint divergence is asymmetric. Atomic probing can recover joint failures (blue) but can also harm correct joint predictions (pink), motivating selective elicitation and preservation-aware reconciliation rather than unconditional atomic imitation.

We evaluate the framework along the alternatives that could otherwise explain the observed gains. Prompt-paraphrase, length-matched, schema-controlled, and repeated-rollout tests examine whether the atomic–joint gap survives prompt artifacts and stochastic noise. On MMAD, we compare against generalist VLMs and contemporary IAD methods under a common evaluator; the reproduced IAD methods use the same Qwen3.5-4B backbone and complete ADOPD-Dataset-6K source pool while retaining their method-specific training mechanism. A separate strictly matched ablation keeps supervision, trainable scope, and optimization budget identical to isolate PRIDE's reconciliation mechanism. We further test four multimodal backbones and transfer the atomic/joint construction to open-ended GQA and ChartQA. MMAD remains evaluation-only throughout; the main comparison reports all seven official MMAD task accuracies in addition to the AD balanced-accuracy analysis, so improvements on AD/DC can be checked against localization, description, analysis, and object-understanding behavior.

Contributions.

- We establish a controlled empirical premise for self-evolution: a prompt-robust and repeatable subset of atomic–joint divergences is enriched for recoverable joint errors, while naive atomic replacement also introduces measurable collateral errors.
- We introduce PRIDE, which constructs endogenous repair/preserve coordinates from stable atomic–joint divergence and realizes the resulting function-space correction through a preservation-aware local Jacobian solve. Null-space preservation itself is not claimed as a new primitive; the contribution lies in how the same VLM generates and reconciles these internal targets during label-free self-evolution.
- We evaluate PRIDE with the complete source-disjoint 6K Real-IAD-derived adaptation set, MMAD held out for 0-shot/1-shot AD/DC evaluation, and an extended evaluation suite spanning four multimodal backbones, the complete seven-task MMAD profile, and two additional open-ended VQA domains.

2 Related Work

Industrial anomaly reasoning. MMAD (Jiang et al., 2025) has catalyzed VLM-based industrial inspection spanning prompt learning, reinforcement post-training, comparative reasoning, retrieval, agentic inspection, and on-policy distillation. Representative systems include AnomalyGPT (Gu et al., 2024), AnomalyR1 (Chao et al., 2025), OmniAD (Zhao et al., 2025), EMIT (Guan et al., 2025), AD-FM (Liao et al., 2026), JUDO (Kang et al., 2026), IAD-R1 (Li et al., 2026), MAU-GPT (Wang et al., 2026c), Reason-IAD (Chen et al., 2026b), AD-Copilot (Jiang et al., 2026), InspectorGPT (Chen et al., 2026d), ADSeeker (Zhang et al., 2026b), AgentIAD (Miao et al., 2025), OPD-IAD (Chen et al., 2026c), and ADOPD (He et al., 2026b). These methods improve industrial perception through different supervision, architectural, retrieval, or reward mechanisms. PRIDE studies a complementary question: whether disagreements already present among task-conditioned outputs of the same VLM can provide a reliable label-free correction signal.

Self-improvement from internal visual signals. Multimodal self-improvement has used reflection, self-play, curriculum construction, and questioner–solver evolution (Cheng et al., 2025; Wang et al., 2025b; 2026b; He et al., 2025; 2026a; Xu et al., 2026; Liang et al., 2026). Closely related recent work explicitly bootstraps stronger behavior from internal asymmetries or self-consistency. Turning Internal Gap into Self-Improvement transfers a stronger understanding branch to a weaker generation branch without external supervision (Han et al., 2026); EvoQuality constructs pseudo-supervision through self-consistent voting and ranking (Wen et al., 2026); ViPER builds a closed visual-feedback loop for perception self-evolution (Zhang et al., 2026a); and VC-STaR exploits contrastive visual pairs to self-improve reasoning (Pan et al., 2026). Vision-OPD (Yuan et al., 2026) and V-Zero (Sun et al., 2026) further use privileged or contrastive visual views for answer-label-free on-policy self-distillation. PRIDE is narrower in its claimed novelty: it uses the *atomic–joint task-conditioned discrepancy on identical visual evidence* to construct repair/preserve coordinates and then realizes the accepted correction in function space. Recent analyses of self-improvement also identify confirmation bias and skill imbalance (Deng et al., 2025; Guo et al., 2026), motivating PRIDE's explicit reliability gates.

Compositional gaps and preservation-aware editing. A model solving constituent subproblems while failing after composition is closely related to the compositionality gap studied by Press et al. (Press et al., 2023). PRIDE studies the corresponding phenomenon for multiple visual judgments of one input and uses the discrepancy as an adaptation signal. Preservation-aware parameter modification also has direct precedents: EMMET formulates model editing through preservation–memorization constraints (Gupta et al., 2024); AlphaEdit projects perturbations into a preserved-knowledge null space (Fang et al., 2025); and NSPO projects optimization directions into a general-task null space (Niu et al., 2026). PRIDE therefore does not claim null-space preservation as a new primitive; its contribution is the endogenous construction and reconciliation of atomic–joint repair targets during VLM self-evolution.

Steering, calibration, and task interference. Activation steering changes hidden states at inference (Su et al., 2025; Wu et al., 2025; Sivakumar et al., 2025; Yin et al., 2026; Liu et al., 2026;

Nguyen et al., 2026; Wang et al., 2026a); calibration work shows that multimodal confidence can be poorly aligned with correctness (Xuan et al., 2025; Xiao et al., 2026; Chen et al., 2026a); GradNorm, PCGrad, and CAGrad address conflicts among scalar objectives (Chen et al., 2018; Yu et al., 2020; Liu et al., 2021). PRIDE instead specifies a vector of desired output-coordinate changes and constrains currently consistent coordinates directly in the local Jacobian geometry, connecting to first-order function-space analyses (Jacot et al., 2018; Lee et al., 2019).

3 THEORY

3.1 ATOMIC CAPABILITY AND COMPOSITION LOSS

Let $\mathcal{T} = \{\tau_1, \ldots, \tau_K\}$ denote K semantic atomic tasks and τ_S a joint task querying $S \subseteq \{1, \ldots, K\}$. The key requirement is semantic equivalence: for every $k \in S$, the target fact is unchanged, $y_k(x, \tau_k) = y_k(x, \tau_S)$.

Differentiable output score. PRIDE separates sampled text from the continuous score used by reconciliation. Let $b \in \{A, J\}$ denote the atomic or joint branch and let v_0, v_1 be the exact serialized binary-label verbalizers (`0` and `1` in the canonical schema). For coordinate k, deterministic decoding first produces a canonical output prefix c_k^b ending immediately before that label slot; all prefix token IDs, including preceding joint-slot predictions, are detached and then teacher-forced. Both candidate verbalizers are scored under this identical prefix:

$$\ell_\theta(v \mid x, \tau_b, c_k^b) = \sum_{t=1}^{|v|} \log p_\theta(v_t \mid x, \tau_b, c_k^b, v_{<t}), \qquad s_k^b = \ell_\theta(v_1 \mid x, \tau_b, c_k^b) - \ell_\theta(v_0 \mid x, \tau_b, c_k^b). \tag{1}$$

For a tokenizer in which each verbalizer is one token, s_k^b is exactly the corresponding label-logit margin; otherwise token log-probabilities are summed. Sampling, JSON parsing, and the verbalized confidence field lie outside the computational graph. Hence $\hat{y}_k^b = \mathbb{I}[s_k^b > 0]$, $q_k^b = \sigma(|s_k^b|)$ is a margin-derived confidence, and the reconciliation Jacobian is the standard autograd quantity $J_k = \partial s_k^J / \partial \theta$ over the trainable parameter subset.

Open-ended task score. The same reconciliation operator applies when the semantic answer is a text sequence rather than a binary label. Let $a_k^\star = (a_{k,1}^\star, \ldots, a_{k,L_k}^\star)$ be the stable answer selected from the atomic rollouts and treated as a stop-gradient sequence. For branch $b \in \{A, J\}$ we define the length-normalized teacher-forced sequence score

$$\bar{\ell}_\theta(a_k^\star \mid x, \tau_b, c_k^b) = \frac{1}{L_k} \sum_{t=1}^{L_k} \log p_\theta(a_{k,t}^\star \mid x, \tau_b, c_k^b, a_{k,<t}^\star), \qquad s_k^b = \bar{\ell}_\theta(a_k^\star \mid x, \tau_b, c_k^b). \tag{2}$$

Thus $J_k = \partial s_k^J / \partial \theta$ remains an ordinary autograd Jacobian under teacher forcing; PRIDE never differentiates through autoregressive sampling or string matching.

For open-ended answers, the atomic branch generates M rollouts and applies a deterministic canonicalizer $\nu(\cdot)$ that normalizes case, whitespace, punctuation, and benchmark-specific numeric formatting without consulting the reference answer. The accepted answer and stability are

$$\hat{a}_k^A = \operatorname*{mode}_{i=1,\ldots,M} \nu(a_{k,i}^A), \qquad c_k = \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}[\nu(a_{k,i}^A) = \hat{a}_k^A]. \tag{3}$$

Let $\hat{a}_k^J = \nu(a_k^J)$ denote the answer to the same question when it appears inside the joint request. Then $\mathcal{R} = \{k : c_k \geq \tau_c, \hat{a}_k^A \neq \hat{a}_k^J\}$ and $\mathcal{P} = \{k : c_k \geq \tau_c, \hat{a}_k^A = \hat{a}_k^J\}$. For $k \in \mathcal{R}$, $a_k^\star = \hat{a}_k^A$ and $r_k = \mathrm{sg}(s_k^A - s_k^J)$; for $k \in \mathcal{P}$ the desired change is zero. No benchmark answer is used to construct these sets.

With labels used only for analysis, composition loss can be separated into suppression and gain events, $E_k^{\mathrm{sup}} = \{\hat{y}_k^A = y_k, \hat{y}_k^J \neq y_k\}$ and $E_k^{\mathrm{gain}} = \{\hat{y}_k^A \neq y_k, \hat{y}_k^J = y_k\}$. We measure the net composition effect by $\Delta_{\mathrm{comp}} = \Pr(E^{\mathrm{sup}}) - \Pr(E^{\mathrm{gain}})$; a positive and increasing $\Delta_{\mathrm{comp}}(|S|)$ is the empirical signature that composition reduces access to capabilities already expressed atomically.

Training is label-free, so E^{sup} cannot be observed directly. For each atomic task, PRIDE evaluates M semantics-preserving views. Let $\bar{y}_k^A$ be the modal atomic decision and define the stability score $c_k = M^{-1}\sum_i \mathbb{I}[\hat{y}_{k,i}^A = \bar{y}_k^A]$. A stable atomic prediction is admitted only when $c_k \geq \tau_c$; its continuous score is aggregated over the views agreeing with $\bar{y}_k^A$. This gate is discrete and is never differentiated through. The empirical premise $\Pr(\hat{y}_k^A = y_k \mid c_k \geq \tau_c) \geq 1-\epsilon$ is assessed only after training with held-out labels. Prompt paraphrase, schema/length matching, and repeated-rollout controls in Section 5 test whether the accepted signal persists beyond a single prompt realization.

For stable dimensions, the candidate repair and preservation sets are $\mathcal{R} = \{k : c_k \geq \tau_c, \hat{y}_k^A \neq \hat{y}_k^J\}$ and $\mathcal{P} = \{k : c_k \geq \tau_c, \hat{y}_k^A = \hat{y}_k^J\}$. For $k \in \mathcal{R}$, the desired joint-output change is $r_k = \text{sg}(s_k^A - s_k^J)$; for $k \in \mathcal{P}$, the desired change is zero. Here $\text{sg}(\cdot)$ denotes stop-gradient: r is a fixed function-space target during reconciliation and is not itself a parameter gradient.

3.2 Exact Recoverability under Preservation

Let $s^J(\theta) = [s_1^J, \ldots, s_K^J]^\top$ and consider a small trainable-parameter update $\Delta\theta$. First-order expansion gives $s^J(\theta+\Delta\theta) \approx s^J(\theta) + J\Delta\theta$, where $J = \partial s^J/\partial\theta$. Partitioning the relevant rows into $J_\mathcal{R}$ and $J_\mathcal{P}$ yields the ideal local constraints

$$J_\mathcal{R}\Delta\theta = r, \qquad J_\mathcal{P}\Delta\theta = 0. \tag{4}$$

Let $P_N = P_{\text{Null}(J_\mathcal{P})}$ be the projector onto directions that preserve $\mathcal{P}$. Every admissible update can be written $\Delta\theta = P_N u$, so Eq. 4 reduces to $J_\mathcal{R}P_N u = r$.

Proposition 1 (exact local recoverability). A desired correction r is exactly realizable while preserving $\mathcal{P}$ to first order iff $r \in \text{Range}(J_\mathcal{R}P_N)$. When the condition holds, the minimum-norm update is

$$\Delta\theta^* = P_N J_\mathcal{R}^\top (J_\mathcal{R}P_N J_\mathcal{R}^\top)^\dagger r. \tag{5}$$

This is a feasibility statement rather than the empirical gating score. In an over-parameterized model, $J_\mathcal{R}P_N$ can have full row rank even when the desired correction is poorly conditioned under finite damping and preservation penalties; hard range membership can therefore collapse to a binary, often saturated quantity. PRIDE uses the solver-consistent continuous score below for gating.

3.3 Practical Reconciliation and Local Validity

PRIDE solves

$$\min_{\Delta\theta} \ \|J_\mathcal{R}\Delta\theta - r\|_2^2 + \beta\|J_\mathcal{P}\Delta\theta\|_2^2 + \lambda\|\Delta\theta\|_2^2. \tag{6}$$

With $\widetilde{J} = [J_\mathcal{R}; \sqrt{\beta}J_\mathcal{P}]$ and $\widetilde{r} = [r; 0]$, the linear candidate is

$$\Delta\theta_{\text{lin}} = \widetilde{J}^\top(\widetilde{J}\widetilde{J}^\top + \lambda I)^{-1}\widetilde{r}. \tag{7}$$

For $r \neq 0$, define regularized recoverability as the fraction of repair-target energy removed by the actual preservation-aware ridge solve,

$$\rho_{\beta,\lambda}(r) = 1 - \frac{\|J_\mathcal{R}\Delta\theta_{\text{lin}} - r\|_2^2}{\|r\|_2^2 + \varepsilon_{\text{num}}}, \qquad \varepsilon_{\text{num}} = 10^{-12}. \tag{8}$$

Because $\Delta\theta = 0$ is feasible, the optimum of Eq. 6 has objective no larger than $\|r\|_2^2$, implying a repair residual no larger than $\|r\|_2^2$ up to numerical error; the implementation clips only roundoff to $[0, 1]$. This score remains informative under full row rank because β and λ encode how much of r the *executed* regularized solve can realize while protecting $\mathcal{P}$. Candidates with $\rho_{\beta,\lambda} < \tau_r$ are rejected before trust-region scaling.

The only explicit inverse acts on the selected-output Gram matrix $G = \widetilde{J}\widetilde{J}^\top$, whose dimension is $m = |\mathcal{R}| + |\mathcal{P}|$, rather than on the full parameter space. Conventional multi-task SFT combines scalar gradients $\sum_k w_k g_k$ without directly enforcing $J_\mathcal{P}\Delta\theta \approx 0$, whereas Eq. 6 specifies desired function changes and protected coordinates.

The reconciliation is valid only locally. If $J(\theta)$ is L-Lipschitz, the Taylor remainder satisfies $\|s(\theta+\Delta\theta) - s(\theta) - J\Delta\theta\|_2 \leq (L/2)\|\Delta\theta\|_2^2$. Hence PRIDE clips $\|\Delta\theta\|$ to a trust radius and accepts the update only after a fresh nonlinear forward check.

Proposition 2 (local prediction preservation). If the linear-system residual and Taylor remainder are smaller than the repair and preservation decision margins, then the accepted update flips the targeted repair coordinates while leaving protected discrete predictions unchanged. The discrete-boundary guarantee applies directly to the binary-margin specialization. For open-ended sequence outputs, the same local function-space reconciliation increases the accepted answer score under the joint context, but a finite likelihood increase does not by itself guarantee exact string-level correctness; open-ended transfer is therefore evaluated empirically. Figure 3 empirically shows the expected deterioration once the update leaves this local regime. Full proofs, the conditional Gram form, self-evolution dynamics, and additional diagnostics are in Appendix C.

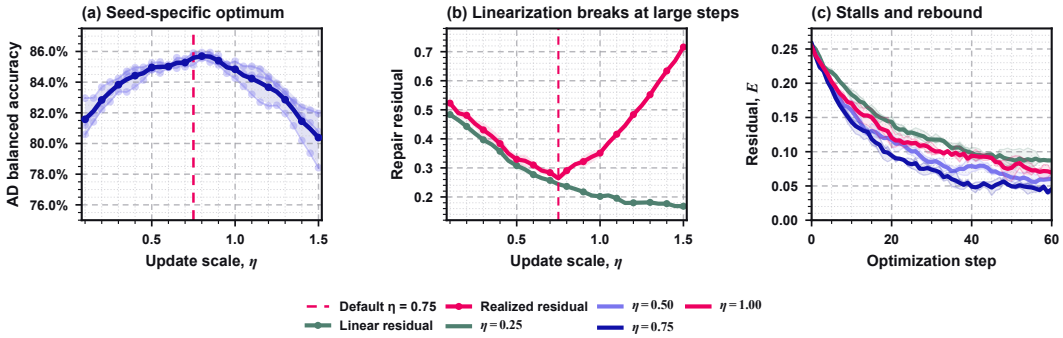

Figure 3: Local validity requires a trust region. Taylor error grows and repair success falls as the update scale leaves the locally linear regime.

4 METHOD

PRIDE turns the theory above into a teacher-free self-evolution loop. Each round first discovers stable atomic capabilities, expands them with consistency self-training, then re-evaluates atomic and joint predictions before performing preservation-aware reconciliation. This ordering matters: all repair/preserve sets are regenerated after the atomic update, so the Jacobian solve always reflects the current model rather than stale pseudo-labels.

Atomic capability discovery and expansion. For every unlabeled sample and atomic task k, M semantics-preserving views produce discrete atomic decisions, differentiable label-margin scores from Eq. 1, and stability c_k. Only $c_k \geq \tau_c$ enters the stable set $\mathcal{A}_t$. Cross-view consistency training uses the accepted atomic decision as a stop-gradient pseudo-target, expanding the set of atomic capabilities available to the next round.

Repair/preserve decomposition. After the atomic update, PRIDE recomputes fresh atomic and joint scores. Stable disagreements form $\mathcal{R}$ with residual $r_k = \mathrm{sg}(s_k^A - s_k^J)$; stable agreements form $\mathcal{P}$ with zero desired change. The residual is a fixed output-space target during reconciliation, while the joint scores remain differentiable with respect to the trainable parameters.

Open-ended VQA specialization. For GQA and ChartQA, questions sharing one visual source are grouped into a task set. Atomic inference answers each question independently, whereas joint inference answers the identical questions in a fixed ordered schema with one slot per question. Stable atomic answers are selected by rollout agreement using Eq. 3. The accepted atomic answer sequence is then teacher-forced under both the atomic and joint contexts, and the difference between their normalized sequence log-likelihoods in Eq. 2 defines the same function-space residual used by the binary MMAD setting. Cross-view self-training uses the accepted sequence as a stop-gradient pseudo-target; reconciliation uses the identical repair/preserve Jacobian objective and never accesses benchmark answers.

Coupling-aware reconciliation. PRIDE constructs $\widetilde{J} = [J_{\mathcal{R}}; \sqrt{\beta} J_{\mathcal{P}}]$, solves $\alpha = (\widetilde{J}\widetilde{J}^\top + \lambda I)^{-1}\widetilde{r}$, and obtains the unscaled linear candidate $\Delta\theta_{\mathrm{lin}} = \widetilde{J}^\top \alpha$. It then evaluates the solver-consistent regularized recoverability $\rho_{\beta,\lambda}$ from Eq. 8; candidates below τ_r are rejected without changing the model. Accepted candidates are clipped to trust radius ϵ and committed only if a fresh

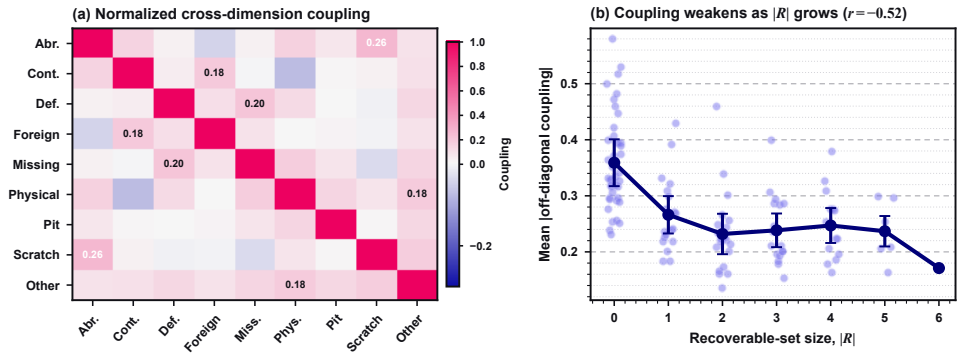

Figure 4: Selected outputs are locally coupled. Non-negligible off-diagonal Jacobian-Gram structure motivates a joint output-space solve.

forward pass reduces the nonlinear repair residual without exceeding the preservation tolerance. Figure 4 shows why independent per-dimension updates are insufficient. The complete round-level training procedure is provided in Algorithm A1 in Appendix A.4.

For $m = |\mathcal{R}| + |\mathcal{P}|$, the explicit linear solve is only $m \times m$; Jacobian-row acquisition dominates runtime. Appendix B.6 reports scaling with the number of queried dimensions and selected outputs.

5 Experiments

5.1 Setup

MMAD (Jiang et al., 2025) is used exclusively for evaluation. Self-evolution uses all 6,000 student-view images from the `train` split of ADOPD-Dataset-6K (He et al., 2026b), whose released records are derived from Real-IAD. The 6K source set is partitioned once into twelve disjoint round shards $\{\mathcal{U}_t\}_{t=1}^{12}$ with $|\mathcal{U}_t| = 500$ and $\bigcup_t \mathcal{U}_t = \mathcal{U}$, so one 12-round run exposes the model to every ADOPD image exactly once. PRIDE consumes only the student-view inspection image: dataset-provided labels/answers, bounding boxes, teacher/evidence crops, teacher negative images, and paired normal counterfactuals are excluded. MMAD draws its images from DS-MVTec, GoodsAD, MVTec-AD, MVTec-LOCO, and VisA, so the Real-IAD-derived adaptation data are source-disjoint from the benchmark.

No MMAD image, question, answer, defect label, category annotation, or reference image is available to the adaptation loop, and no test-time parameter update is performed on MMAD. The 1-shot normal reference is introduced only by the evaluator and does not modify model parameters. The 12-round schedule, trainable parameter scope, and reconciliation thresholds are fixed before final MMAD evaluation; MMAD labels are used only by the evaluator and analyses explicitly described as post-hoc. Main trainable experiments use three random seeds. Full protocols are in Appendix D.

5.2 Controlled MMAD Comparison

Table 1 evaluates the standardized Qwen3.5-4B reimplementations of contemporary industrial-anomaly methods (Gu et al., 2024; Chao et al., 2025; Zhao et al., 2025; Guan et al., 2025; Liao et al., 2026; Kang et al., 2026; Jiang et al., 2026; Chen et al., 2026d; Wang et al., 2026c; Li et al., 2026; Chen et al., 2026b; He et al., 2026b). The source image pool, backbone, and MMAD evaluator are shared, while each method retains the mechanism that defines it. We report both the source-macro AD balanced accuracy used by our anomaly-discrimination analysis and all seven official MMAD task accuracies. This separates improvement on AD/DC from preservation of defect localization, description, analysis, and object understanding. Published results under the original model/data/protocol of prior work are reported separately in Appendix D.2; they are contextual references rather than controlled attribution.

PRIDE reaches 85.57/74.40 AD-BAcc/DC in 0-shot and 87.16/76.02 in 1-shot. The remaining MMAD task columns are reported explicitly because the repair objective targets task-conditioned discrepancies rather than an overall benchmark score; the intended criterion is therefore strong AD/DC improvement without systematic erosion of unrelated capabilities. The strict mechanism-

(a) 0-shot controlled reimplementation

Method	AD-BAcc ↑	AD ↑	DC ↑	DL ↑	DD ↑	DA ↑	OC ↑	OA ↑	Avg. ↑
Base Qwen3.5-4B	63.12	64.00	58.42	59.87	66.15	79.08	92.11	83.96	71.94
AnomalyGPT	67.24	67.60	60.88	52.10	58.40	68.90	86.20	78.40	67.50
AnomalyR1	72.46	72.90	64.65	62.50	70.20	83.10	91.70	85.00	75.72
OmniAD	75.18	75.50	67.96	67.40	72.90	84.20	93.50	85.90	78.19
EMIT	77.31	77.60	69.42	70.80	79.40	85.10	91.50	84.60	79.77
AD-FM	78.65	78.90	69.11	68.00	79.00	84.50	91.00	85.60	79.44
JUDO	74.98	75.20	70.54	68.20	78.10	84.70	92.50	85.20	79.21
AD-Copilot	79.62	79.90	71.03	70.90	80.10	85.40	92.00	86.30	80.80
InspectorGPT	80.37	80.60	71.68	70.20	79.60	85.90	92.10	86.30	80.91
MAU-GPT	72.88	73.10	64.74	63.00	72.30	82.20	90.40	83.60	75.62
IAD-R1	79.84	80.10	69.27	67.50	77.10	83.30	92.60	84.80	79.24
Reason-IAD	78.73	79.00	71.26	65.70	76.30	82.00	93.50	85.00	78.97
ADOPD	84.21	84.50	69.86	66.50	72.80	84.70	92.90	84.50	79.39
PRIDE	**85.57**	**86.10**	**74.40**	69.25	76.61	82.70	93.33	85.21	**81.08**

(b) 1-shot controlled reimplementation

Method	AD-BAcc ↑	AD ↑	DC ↑	DL ↑	DD ↑	DA ↑	OC ↑	OA ↑	Avg. ↑
Base Qwen3.5-4B	70.84	71.35	64.20	62.11	68.30	80.45	92.68	84.53	74.80
AnomalyGPT	72.38	72.70	65.44	58.60	64.80	73.40	87.80	80.60	71.91
AnomalyR1	76.74	77.05	68.91	66.50	74.20	84.30	92.10	85.50	78.37
OmniAD	79.05	79.35	72.44	72.80	79.20	86.50	95.00	86.20	81.64
EMIT	80.86	81.10	73.37	75.50	82.40	86.10	91.50	84.90	82.12
AD-FM	81.92	82.10	73.58	74.90	84.20	86.20	91.70	86.50	82.74
JUDO	79.24	79.50	75.08	75.20	84.40	87.50	93.80	86.20	83.10
AD-Copilot	83.08	83.30	75.24	76.00	84.70	87.00	92.00	87.50	83.68
InspectorGPT	83.79	84.00	75.74	76.10	82.60	88.10	93.00	88.00	83.93
MAU-GPT	76.41	76.70	68.52	66.80	75.20	83.50	91.00	84.20	77.99
IAD-R1	82.66	82.90	72.96	71.40	79.80	84.70	93.20	85.60	81.51
Reason-IAD	81.88	82.10	74.83	69.00	79.20	84.00	94.00	86.00	81.30
ADOPD	84.94	85.20	72.18	69.50	76.20	85.30	93.30	85.00	80.95
PRIDE	**87.16**	**87.65**	**76.02**	75.55	82.10	86.95	**95.14**	86.85	**84.32**

Table 1: Controlled MMAD comparison under a standardized Qwen3.5-4B reimplementation. All IAD methods use the same Qwen3.5-4B backbone, the same complete ADOPD-Dataset-6K source pool, and the same MMAD evaluator while retaining their defining optimization mechanism. AD-BAcc is the source-macro balanced accuracy used for the anomaly-discrimination analysis; AD, DC, DL, DD, DA, OC, and OA are the seven official MMAD task accuracies, and Avg. is their arithmetic mean. The complete seven-task view makes explicit whether gains on AD/DC are accompanied by drift on localization, description, analysis, or object understanding. Published native-configuration results are reported separately in Appendix D.2.

only comparison, in which supervision, trainable scope, optimization budget, and source data are also matched, is reported in Table A2.

5.3 Source-wise and Premise Validation

The aggregate AD/DC result is not driven by a single MMAD source. The full source-wise table is reported in Appendix D.3; its source-macro AD and DC values reproduce the corresponding aggregate quantities in Table 1. We additionally test canonical atomic prompts, two paraphrases, length matching, and a joint-schema single-slot control with eight repeated inferences per case. Across equivalent prompts, disagreement changes by only 1.5 points and gap-set Jaccard remains 0.86–0.90; accepted disagreements are 93.4% repeatable with 86.9% target precision, while a dimension-permutation null falls near chance. The full control table is reported in Appendix D.4.

5.4 Cross-Backbone and Cross-Domain Evaluation

We further evaluate PRIDE on Qwen3-VL-4B (Bai et al., 2025a), LLaVA-OneVision-1.5-4B (An et al., 2025), and Gemma-3n-E4B (Gemma Team, 2025), in addition to Qwen3.5-4B. Under the same ADOPD shard IDs and MMAD examples, PRIDE improves both AD and DC for all four backbones in 0-shot and 1-shot. To test transfer beyond binary industrial inspection, we instantiate the open-ended sequence-score specialization from Eq. 2 on GQA (Hudson & Manning, 2019) and ChartQA (Masry et al., 2022). Atomic and joint branches use identical questions and visual evidence; stable model-generated answers provide the self-supervision target, while benchmark answers remain inaccessible until evaluation. The complete cross-backbone and cross-domain matrices are reported in Appendix D.5 to preserve the nine-page main-paper budget.

6 Ablation Study

The complete core-ablation table and reconciliation perturbations are reported in Appendix B.1. Under the same Qwen3.5-4B initialization, complete 6K unlabeled schedule, trainable parameter scope, and update budget, Joint ST and Joint SE yield limited gains; Atomic SFT recovers a larger fraction of the signal; stability-filtered elicitation further improves target quality; and replacing reconciliation with PCGrad leaves a 3.40-point AD/DC-mean gap to PRIDE. This isolates the benefit of mapping a desired function-space correction while constraining currently consistent coordinates.

Mechanism perturbations further separate preservation from coupling: removing the preservation term increases drift on currently consistent outputs, while replacing the coupled Gram solve with independent per-dimension updates worsens realized target alignment under a matched update-norm budget. Sweeps over $\tau_c, \tau_r, \beta, \lambda, \eta$ and rollout count, together with long-horizon stability, are also reported in Appendix B.

7 Conclusion

PRIDE uses stable atomic–joint task-conditioned divergence as endogenous supervision and realizes accepted corrections through a preservation-aware local Jacobian solve. On industrial anomaly reasoning, the evaluation uses the complete 6,000-image ADOPD-Dataset-6K source through twelve disjoint 500-image rounds and reserves MMAD for 0-shot/1-shot AD/DC evaluation. Controlled comparisons, source-wise results, and premise tests support the mechanism on Qwen3.5-4B; additional experiments show consistent transfer across three further multimodal backbones and two open-ended VQA domains, while the complete seven-task MMAD profile makes non-target capability drift explicit. These experiments support transfer across the tested architectures and task families while leaving broader VLM generalization open. The practical recoverability gate is defined from the same regularized solve used to construct the update, separating it from the hard range-membership condition used only in the exact local theorem. Remaining limitations include Jacobian cost, imperfect pseudo-target semantics, local-linearization validity, and late-round drift.

AI use statement

Generative AI tools were used only to improve wording, grammar, and readability of author-written paper text. They were not used to formulate the research problem or hypotheses, design the method or experiments, derive mathematical claims, implement the method, generate experimental observations used in the final submission, or interpret final experimental results. All language edits were reviewed by the authors, who take responsibility for the paper.

Ethics statement

This work studies self-supervised adaptation of vision-language models for visual reasoning. It does not involve human subjects or personally identifiable information. Incorrect visual predictions can still produce false positives or false negatives, so the method should not be a sole decision-maker in safety-critical inspection. Additional limitations and deployment considerations are discussed in Appendix F.

Reproducibility statement

The paper specifies the differentiable output score, regularized recoverability gate, local reconciliation solver, prompt/noise controls, external adaptation-data protocol, controlled local comparisons, seven-task MMAD reporting, and evaluation metrics. Appendix A reports data handling, hyperparameters, prompts, and Jacobian construction; Appendix B reports local-validity and sensitivity diagnostics; Appendix D defines the matched evaluation protocol; and Algorithm A1 gives the round-level procedure.

REFERENCES

Xiang An, Yin Xie, Kaicheng Yang, Wenkang Zhang, Xiuwei Zhao, Zheng Cheng, Yirui Wang, Songcen Xu, Changrui Chen, Chunsheng Wu, et al. LLaVA-OneVision-1.5: Fully open framework for democratized multimodal training. *arXiv preprint arXiv:2509.23661*, 2025.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, et al. Qwen3-VL technical report. *arXiv preprint arXiv:2511.21631*, 2025a.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*, 2025b.

Yuhao Chao, Jie Liu, Jie Tang, and Gangshan Wu. Anomalyr1: A grpo-based end-to-end mllm for industrial anomaly detection. *arXiv preprint arXiv:2504.11914*, 2025.

Minghui Chen, Chenxu Yang, Hengjie Zhu, Dayan Wu, Qingyi Si, and Zheng Lin. Online self-calibration against hallucination in vision-language models. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence*, pp. 989–997, 2026a. doi: 10.24963/ijcai.2026/111.

Peng Chen, Chao Huang, Yunkang Cao, Chengliang Liu, Wenqiang Wang, Mingbo Yang, Li Shen, Wenqi Ren, and Xiaochun Cao. Reason-iad: Knowledge-guided dynamic latent reasoning for explainable industrial anomaly detection. *arXiv preprint arXiv:2602.09850*, 2026b.

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 preprint arXiv:2607.18850*, 2026c.

Weifei Chen, Honghao Zhang, Zhiyuan You, and Xinyi Le. Inspectorgpt: A comparative reasoning enhanced vlm for comprehensive industrial anomaly detection. *arXiv preprint arXiv:2608.29783*, 2026d.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 794–803, 2018.

Kanzhi Cheng, Yantao Li, Fangzhi Xu, Jianbing Zhang, Hao Zhou, and Yang Liu. Vision-language models can self-improve reasoning via reflection. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 8876–8892, 2025. doi: 10.18653/v1/2025.naacl-long.447.

Shijian Deng, Kai Wang, Tianyu Yang, Harsh Singh, and Yapeng Tian. Self-improvement in multimodal large language models: A survey. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 1987–2006, 2025. doi: 10.18653/v1/2025.findings-emnlp.105.

Junfeng Fang, Houcheng Jiang, Kun Wang, Yunshan Ma, Jie Shi, Xiang Wang, Xiangnan He, and Tat-Seng Chua. Alphaedit: Null-space constrained knowledge editing for language models. In *International Conference on Learning Representations*, 2025.

Gemma Team. Gemma 3n, 2025. URL `https://ai.google.dev/gemma/docs/gemma-3n/model_card`.

Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. Anomalygpt: Detecting industrial anomalies using large vision-language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 1932–1940, 2024. doi: 10.1609/aaai.v38i3.27963.

Wei Guan, Jun Lan, Jian Cao, Hao Tan, Huijia Zhu, and Weiqiang Wang. Emit: Enhancing mllms for industrial anomaly detection via difficulty-aware grpo. *arXiv preprint arXiv:2507.21619*, 2025.

Xin Guo, Zhiheng Xi, Yiwen Ding, Yitao Zhai, Xiaowei Shi, Xunliang Cai, Tao Gui, Qi Zhang, and Xuanjing Huang. Counteracting the matthew effect in self-improvement of lvlms through head-tail re-balancing. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, pp. 22104–22121, 2026. doi: 10.18653/v1/2026.acl-long.1010.

Akshat Gupta, Dev Sajnani, and Gopala Anumanchipalli. A unified framework for model editing. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 15403–15418, 2024. doi: 10.18653/v1/2024.findings-emnlp.903.

Yujin Han, Hao Chen, Andi Han, Zhiheng Wang, Xinyu Liu, Yingya Zhang, Shiwei Zhang, and Difan Zou. Turning internal gap into self-improvement: Promoting the generation-understanding unification in mllms. In *International Conference on Learning Representations*, 2026.

Jinghan He, Junfeng Fang, Feng Xiong, Zijun Yao, Fei Shen, Haiyun Guo, Jinqiao Wang, and Tat-Seng Chua. Active Zero: Self-evolving vision-language models through active environment exploration. *arXiv preprint arXiv:2602.11241*, 2026a.

Jingtai He, Shiyuan Meng, Wenchao Meng, and Qinmin Yang. Adopd: Reference-privileged on-policy distillation for mllm-based industrial anomaly detection. *arXiv preprint arXiv:2608.09789*, 2026b.

Yicheng He, Chengsong Huang, Zongxia Li, Jiaxin Huang, and Yonghui Yang. VisPlay: Self-evolving vision-language models from images. *arXiv preprint arXiv:2511.15661*, 2025.

Drew A. Hudson and Christopher D. Manning. GQA: A new dataset for real-world visual reasoning and compositional question answering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6700–6709, 2019. doi: 10.1109/CVPR.2019.00686.

Arthur Jacot, Franck Gabriel, and Clement Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

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 *International Conference on Learning Representations*, 2025.

Xi Jiang, Yue Guo, Jian Li, Yong Liu, Bin-Bin Gao, Hanqiu Deng, Jun Liu, Heng Zhao, Chengjie Wang, and Feng Zheng. Ad-copilot: A vision-language assistant for industrial anomaly detection via visual in-context comparison. *arXiv preprint arXiv:2603.13779*, 2026.

Hyunju Kang, Woohyun Lee, Jaewon Kim, and Hogun Park. Judo: A juxtaposed domain-oriented multimodal reasoner for industrial anomaly qa. In *International Conference on Learning Representations*, 2026.

Jaehoon Lee, Lechao Xiao, Samuel S. Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Yanhui Li, Yunkang Cao, Chengliang Liu, Yuan Xiong, Xinghui Dong, and Chao Huang. Iad-r1: Reinforcing consistent reasoning in industrial anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 6583–6591, 2026. doi: 10.1609/aaai.v40i8.37588.

Yijun Liang, Hengguang Zhou, Ming Li, Lichen Li, Cho-Jui Hsieh, and Tianyi Zhou. Self-evolving visual questioner. *arXiv preprint arXiv:2606.13929*, 2026.

Jingyi Liao, Yongyi Su, Rong-Cheng Tu, Zhao Jin, Wenhao Sun, Yiting Li, Dacheng Tao, Xun Xu, and Xulei Yang. Ad-fm: Multimodal llms for anomaly detection via multi-stage reasoning and fine-grained reward optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15234–15242, 2026. doi: 10.1609/aaai.v40i18.38548.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Shuliang Liu, Songbo Yang, Dong Fang, Sihang Jia, Yuqi Tang, Lingfeng Su, Ruoshui Peng, Yibo Yan, Xin Zou, and Xuming Hu. Vision-language introspection: Mitigating overconfident hallucinations in mllms via interpretable bi-causal steering. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, pp. 38518–38543, 2026. doi: 10.18653/v1/2026.acl-long.1784.

Ahmed Masry, Do Xuan Long, Jia Qing Tan, Shafiq Joty, and Enamul Hoque. ChartQA: A benchmark for question answering about charts with visual and logical reasoning. In *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 2263–2279, 2022. doi: 10.18653/v1/2022.findings-acl.177.

Junwen Miao, Penghui Du, Yi Liu, Yu Wang, and Yan Wang. AgentIAD: Tool-augmented single-agent for industrial anomaly detection. *arXiv preprint arXiv:2512.13671*, 2025.

Duy Nguyen, Archiki Prasad, Elias Stengel-Eskin, and Mohit Bansal. GrAInS: Gradient-based attribution for inference-time steering of LLMs and VLMs. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 46523–46543, 2026. doi: 10.18653/v1/2026.acl-long.2159.

Yifan Niu, Han Xiao, Dongyi Liu, Nuo Chen, and Jia Li. Mitigating the safety alignment tax with null-space constrained policy optimization. In *International Conference on Learning Representations*, 2026.

Zhiyu Pan, Yizheng Wu, Jiashen Hua, Junyi Feng, Shaotian Yan, Bing Deng, Zhiguo Cao, and Jieping Ye. Through the lens of contrast: Self-improving visual reasoning in vlms. In *International Conference on Learning Representations*, 2026.

Ofir Press, Muru Zhang, Sewon Min, Ludwig Schmidt, Noah A. Smith, and Mike Lewis. Measuring and narrowing the compositionality gap in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 5687–5711, 2023. doi: 10.18653/v1/2023.findings-emnlp.378.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL `https://qwen.ai/blog?id=qwen3.5`.

Anushka Sivakumar, Andrew Zhang, Zaber Ibn Abdul Hakim, and Chris Thomas. SteerVLM: Robust model control through lightweight activation steering for vision language models. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 23640–23665, 2025. doi: 10.18653/v1/2025.findings-emnlp.1285.

Jingran Su, Jingfan Chen, Hongxin Li, Yuntao Chen, Li Qing, and Zhaoxiang Zhang. Activation steering decoding: Mitigating hallucination in large vision-language models through bidirectional hidden state intervention. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, pp. 12964–12974, 2025. doi: 10.18653/v1/2025.acl-long.634.

Haoxiang Sun, Zhihang Yi, Langxuan Deng, Yuhao Zhou, Peiqi Jia, Jian Zhao, Li Yuan, Jiancheng Lv, and Tao Wang. V-Zero: Answer-label-free on-policy distillation with contrastive evidence gating for fine-grained visual reasoning. *arXiv preprint arXiv:2606.25319*, 2026.

Qidong Wang, Junjie Hu, and Ming Jiang. From heads to neurons: Causal attribution and steering in multi-task vision–language models. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 36151–36175, 2026a. doi: 10.18653/v1/2026.findings-acl.1802.

Qinsi Wang, Bo Liu, Tianyi Zhou, Jing Shi, Yueqian Lin, Yiran Chen, Hai Helen Li, Kun Wan, and Wentian Zhao. Vision-Zero: Scalable vlm self-evolution via multi-agent self-play. In *International Conference on Learning Representations*, 2026b.

Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, et al. InternVL3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *arXiv preprint arXiv:2508.18265*, 2025a.

Xiyao Wang, Jiuhai Chen, Zhaoyang Wang, Yuhang Zhou, Yiyang Zhou, Huaxiu Yao, Tianyi Zhou, Tom Goldstein, Parminder Bhatia, Taha Kass-Hout, Furong Huang, and Cao Xiao. Enhancing visual-language modality alignment in large vision language models via self-improvement. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pp. 268–282, 2025b. doi: 10.18653/v1/2025.findings-naacl.15.

Zhuonan Wang, Zhenxuan Fan, Siwen Tan, Yu Zhong, Yuqian Yuan, Haoyuan Li, Hao Jiang, Wenqiao Zhang, Feifei Shao, Hongwei Wang, and Jun Xiao. MAU-GPT: Enhancing multi-type industrial anomaly understanding via anomaly-aware and generalist experts adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26787–26795, 2026c. doi: 10.1609/aaai.v40i31.39889.

Wen Wen, Tianwu Zhi, Kanglong Fan, Yang Li, Xinge Peng, Yabin Zhang, Yiting Liao, Junlin Li, and Li Zhang. Self-evolving vision-language models for image quality assessment via voting and ranking. In *International Conference on Learning Representations*, 2026.

Junfei Wu, Yue Ding, Guofan Liu, Tianze Xia, Ziyue Huang, Dianbo Sui, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. SHARP: Steering hallucination in LVLMs via representation engineering. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 14346–14361, 2025. doi: 10.18653/v1/2025.emnlp-main.725.

Wenyi Xiao, Xinchi Xu, and Leilei Gan. VL-Calibration: Decoupled confidence calibration for large vision-language models reasoning. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, pp. 44791–44815, 2026. doi: 10.18653/v1/2026.acl-long.2074.

Chaoran Xu, Yingmao Miao, Pengfei Zhang, Hao Dou, Lei Sun, and Xiangxiang Chu. RISE: Reliable improvement in self-evolving vision-language models. *arXiv preprint arXiv:2605.20914*, 2026.

Weihao Xuan, Qingcheng Zeng, Heli Qi, Junjue Wang, and Naoto Yokoya. Seeing is believing, but how much? a comprehensive analysis of verbalized calibration in vision-language models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 1408–1450, 2025. doi: 10.18653/v1/2025.emnlp-main.74.

Jianghao Yin, Qin Chen, Kedi Chen, Jie Zhou, Xingjiao Wu, and Liang He. Dynamic multimodal activation steering for hallucination mitigation in large vision-language models. In *International Conference on Learning Representations*, 2026.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Qianhao Yuan, Jie Lou, Xing Yu, Hongyu Lin, Le Sun, Xianpei Han, and Yaojie Lu. Vision-OPD: Learning to see fine details for multimodal llms via on-policy self-distillation. *arXiv preprint arXiv:2605.18740*, 2026.

Juntian Zhang, Song Jin, Chuanqi Cheng, Yuhan Liu, Yankai Lin, Xun Zhang, Yufei Zhang, Fei Jiang, Guojun Yin, Wei Lin, and Rui Yan. ViPER: Empowering the self-evolution of visual perception abilities in vision-language models. In *International Conference on Learning Representations*, 2026a.

Kai Zhang, Zekai Zhang, Xihe Sun, Anpeng Wang, Jingmeng Nie, Qinghui Chen, Han Hao, Jianyuan Guo, and Jinglin Zhang. ADSeeker: A knowledge-grounded reasoning framework for industry anomaly detection and reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21379–21388, 2026b.

Shifang Zhao, Yiheng Lin, Lu Han, Yao Zhao, and Yunchao Wei. Omniad: Detect and understand industrial anomaly via multimodal reasoning. *arXiv preprint arXiv:2505.22039*, 2025.

A Implementation Details and Reproducibility

A.1 Training Scope and Parameterization

The main configuration uses Qwen3.5-4B with the same trainable parameter scope for all matched baselines. The language backbone remains frozen; LoRA adapters are attached to the multimodal projection/merger and selected visual blocks, while the reconciliation Jacobian is taken only with respect to these trainable parameters. Atomic self-training and reconciliation therefore operate over the same parameter subset. Each round processes its assigned 500-image shard, performs the atomic consistency update first, then recomputes every atomic/joint prediction before constructing $\mathcal{R}$ and $\mathcal{P}$. No cached prediction from the previous parameter state is reused.

For reproducibility, every accepted reconciliation record stores the sample ID, task IDs in $\mathcal{R}/\mathcal{P}$, differentiable atomic/joint label-margin scores, stability c_k, regularized recoverability $\rho_{\beta,\lambda}$, repair residual before/after the linear solve, update norm, and fresh-forward preservation drift. These records are stored together with the exact training command and model revision for reproducibility.

A.2 Adaptation Data and Contamination Control

The adaptation source is the complete `train` split of ADOPD-Dataset-6K (He et al., 2026b), whose 6,000 records are derived from Real-IAD. Before training, the 6K source IDs are deterministically partitioned into twelve non-overlapping 500-record shards $\mathcal{U}_1, \ldots, \mathcal{U}_{12}$. Round t uses only $\mathcal{U}_t$, so the standard 12-round run covers all 6,000 unique images exactly once rather than repeatedly training on a 500-image subset. PRIDE reads only the student-view inspection image from each record. Dataset-provided labels/answers, bounding boxes, teacher/evidence crops, teacher negative images, and paired normal counterfactual references are excluded from the adaptation dataloader. The shard membership and round order are fixed once and reused for every reported trainable method.

MMAD (Jiang et al., 2025) is instantiated only by the evaluation pipeline. Its image sources are DS-MVTec, GoodsAD, MVTec-AD, MVTec-LOCO, and VisA, whereas the adaptation pool is Real-IAD-derived; no MMAD image enters the adaptation pool. During self-evolution the model has no access to MMAD questions, answers, defect labels, reference images, or category annotations. The 1-shot normal reference is supplied only at MMAD evaluation time. The 12-round schedule, τ_c, τ_r, β, λ, η, and the trainable parameter scope are fixed before final MMAD evaluation; no checkpoint is selected by MMAD performance. Label-dependent gate, sensitivity, and long-horizon analyses in Appendix B are post-hoc diagnostics and are not used to choose the reported main checkpoint.

A.3 Differentiable Score Implementation

For Jacobian acquisition, PRIDE never differentiates through sampled text, JSON parsing, or the model's verbalized confidence field. The canonical binary verbalizers are the exact serialized label strings `0` and `1`. For each coordinate, deterministic decoding supplies a detached output prefix ending immediately before the label slot; in the joint branch, already generated preceding slots are included in this detached prefix. The two candidate verbalizers are then teacher-forced under the same prefix and scored by Eq. 1. If a verbalizer spans multiple tokenizer tokens, its token log-probabilities are summed. Prefix IDs and verbalizer IDs are constants, so a standard backward pass on the scalar margin yields $J_k = \partial s_k^J / \partial \theta$ over exactly the trainable LoRA/vision/merger parameters. The margin-derived diagnostic confidence is $q_k = \sigma(|s_k|)$.

A.4 Complete PRIDE Algorithm

Algorithm A1 gives the complete round-level procedure corresponding to the main-text method description, including atomic capability discovery, score recomputation, repair/preserve decomposition, solver-consistent recoverability gating, trust-region clipping, and the fresh nonlinear acceptance check.

A.5 Hyperparameters

Table A1 consolidates the configuration used by the main MMAD experiment.

Algorithm A1 PRIDE: Prediction Reconciliation via Internal Divergence Elicitation for Teacher-Free Self-Evolution

Input: model f_θ, unlabeled pool $\mathcal{U}$, atomic tasks $\{\tau_k\}_{k=1}^{K}$, joint task τ_J, view generator $\mathcal{V}$, thresholds τ_c, τ_r, β, λ, ϵ
Output: updated model $f_{\theta'}$

$\mathcal{A} \leftarrow \varnothing$ ▷ discover stable atomic capabilities
for $x \in \mathcal{U}$ and $k = 1, \ldots, K$ **do**
  obtain $\{(\hat{y}^A_{k,i}, s^A_{k,i})\}_{i=1}^{M}$ using Eq. 1; compute aggregate $(\hat{y}^A_k, s^A_k, c_k)$
  if $c_k \geq \tau_c$ **then**
    add $(x, k, \mathrm{sg}(\hat{y}^A_k))$ to $\mathcal{A}$
  end if
end for
update $\theta \leftarrow \theta - \eta_A \nabla_\theta \mathcal{L}_{\mathrm{atom}}(\mathcal{A})$ ▷ cross-view atomic self-training
recompute all atomic/joint scores with the updated θ; initialize $\mathcal{R}, \mathcal{P} \leftarrow \varnothing$
for each stable coordinate (x, k) **do**
  if $\hat{y}^A_k \neq \hat{y}^J_k$ **then**
    $\mathcal{R} \leftarrow \mathcal{R} \cup \{k\}$ and $r_k \leftarrow \mathrm{sg}(s^A_k - s^J_k)$
  else
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{k\}$
  end if
end for
if $\mathcal{R} = \varnothing$ **then**
  return θ
end if
form $J_\mathcal{R} = \partial s^J_\mathcal{R} / \partial \theta$ and $J_\mathcal{P} = \partial s^J_\mathcal{P} / \partial \theta$
$\widetilde{J} \leftarrow [J_\mathcal{R}; \sqrt{\beta} J_\mathcal{P}]$, $\widetilde{r} \leftarrow [r; 0]$
$\alpha \leftarrow (\widetilde{J}\widetilde{J}^\top + \lambda I)^{-1} \widetilde{r}$; $\Delta\theta_{\mathrm{lin}} \leftarrow \widetilde{J}^\top \alpha$
$\rho_{\beta,\lambda} \leftarrow 1 - \|J_\mathcal{R} \Delta\theta_{\mathrm{lin}} - r\|_2^2 / (\|r\|_2^2 + \varepsilon_{\mathrm{num}})$
if $\rho_{\beta,\lambda} \geq \tau_r$ **then**
  clip $\Delta\theta_{\mathrm{lin}}$ to $\|\Delta\theta\|_2 \leq \epsilon$ and run a fresh forward pass
  if repair residual decreases and preservation drift $\leq \delta_P$ **then**
    $\theta' \leftarrow \theta + \Delta\theta$
  else
    $\theta' \leftarrow \theta$
  end if
else
  $\theta' \leftarrow \theta$
end if
return θ'; regenerate $\mathcal{A}, \mathcal{R}, \mathcal{P}$ in the next round

Setting	Value
Main backbone	Qwen3.5-4B (Qwen Team, 2026)
Unlabeled data per round D_S	500 student-view images from ADOPD-Dataset-6K (He et al., 2026b) (Real-IAD-derived)
Total unique adaptation images	6,000 (12 disjoint shards; one full pass)
MMAD access during adaptation	None
Evolution rounds	12 (fixed; no MMAD best-checkpoint selection)
Atomic dimensions K	9
Atomic rollouts / dimension	8
Batch prompts / micro-batch	8 / 1
Epochs / SFT steps per round	2 / 16
Optimizer	AdamW
LoRA / merger / vision LR	1×10^{-5} / 1×10^{-6} / 1×10^{-6}
Maximum gradient norm	6.0
Stability threshold τ_c	0.75
Regularized-recoverability threshold τ_r	0.70
Preservation weight β	0.50
Damping λ	1×10^{-2}
Reconciliation step scale η	0.75
Trust radius ϵ	0.50
Random seeds	3

Table A1: Main adaptation and MMAD evaluation configuration. MMAD labels are not used for adaptation, hyperparameter selection, or checkpoint selection.

A.6 HARDWARE AND RUNTIME

Experiments use one H100-80GB GPU for Qwen3.5-4B, BF16 forward/backward, gradient checkpointing for Jacobian-row acquisition, and deterministic seed control where supported. A standard run contains 12 rounds with 500 distinct images per round, i.e., 6,000 unique ADOPD images in total. The reported 3.7 GPU-hours for Atomic SFT and 6.9 GPU-hours for PRIDE therefore correspond to the complete 6K schedule; the additional PRIDE cost is dominated by selected-output Jacobian-vector products rather than the $m \times m$ Gram solve. The eight stochastic rollouts for a fixed sample–dimension pair are decoded as a batch, so $500 \times 9 \times 8 \times 12$ generated atomic sequences do not correspond to the same number of serial model invocations. Peak memory is 54.2GB for Atomic SFT and 63.7GB for PRIDE.

A.7 PROMPT TEMPLATES

PRIDE keeps the visual input fixed between atomic and joint branches. The prompt-control experiments additionally match wording, schema, or length so that disagreement cannot be attributed to one specific surface form.

Joint multi-dimensional inspection prompt.

```
SYSTEM: Compare QUERY with REFERENCE when a reference is provided.
Evaluate every dimension in <DIMENSION_LIST>.
Use visual evidence only.
Return one binary decision and confidence per dimension.
USER: <REFERENCE_IMAGE optional> <QUERY_IMAGE>
OUTPUT: {dimension_k: {label: 0|1, confidence: [0,1]}, ...}
```

Canonical atomic probe.

```
SYSTEM: Use the same visual evidence.
Evaluate only <DIMENSION_NAME>.
Ignore other dimensions.
Return one binary decision and confidence.
USER: <REFERENCE_IMAGE optional> <QUERY_IMAGE>
OUTPUT: {label: 0|1, confidence: [0,1]}
```

Joint-schema single-slot control.

```
SYSTEM: Use the same schema and dimension order as the joint task.
Only <DIMENSION_NAME> is active.
Write null for inactive dimensions.
USER: <REFERENCE_IMAGE optional> <QUERY_IMAGE>
OUTPUT: fixed joint schema with one active dimension.
```

Paraphrase controls preserve task semantics but independently rewrite the instruction. The length-matched control pads the atomic prompt with semantically neutral reminders until its token count matches the joint prompt within $\pm 2\%$. The dimension-permutation null shuffles the mapping between atomic task names and joint output slots while preserving prompt length and decoding settings.

A.8 PREMISE-CONTROL SAMPLING

For each held-out case, eight stochastic rollouts are generated at the same decoding configuration. Repeatability is the fraction for which the sign of $s_k^A - s_k^J$ agrees with the modal sign. Gap-set Jaccard is computed between each control prompt and the canonical atomic prompt over identical image/task pairs. "Target precision" is computed only after training for analysis: it is the fraction of accepted disagreements satisfying $\hat{y}_k^A = y_k$ and $\hat{y}_k^J \neq y_k$.

A.9 Post-hoc Calibrated-Risk Diagnostic

Figure A1 is a label-dependent diagnostic and is not part of the PRIDE training gate. Let $\widetilde{y}_k^A = 2\hat{y}_k^A - 1$ and define a bounded signed atomic–joint divergence

$$d_k = \tanh\left(\frac{\widetilde{y}_k^A(s_k^J - s_k^A)}{T_d}\right), \qquad T_d = 1. \tag{9}$$

Negative d_k means that, measured in the atomic decision direction, the joint margin has moved away from the atomic margin. Together with the joint margin confidence $q_k^J = \sigma(|s_k^J|)$, this forms the two-dimensional diagnostic state (q_k^J, d_k).

Using held-out labels only after training, each (q^J, d) bin is assigned a calibrated suppressive-error risk

$$\kappa(B_q, B_d) = \Pr\big(\hat{y}_k^A = y_k,\ \hat{y}_k^J \neq y_k \mid q_k^J \in B_q,\ d_k \in B_d\big), \tag{10}$$

estimated by the empirical bin frequency with a unit Laplace pseudocount. A calibrated-risk threshold γ selects analysis cases whose bin estimate satisfies $\kappa \geq \gamma$. Validation coverage is the selected fraction of cases; target precision is the conditional probability of the suppressive-repair event; atomic target accuracy is $\Pr(\hat{y}_k^A = y_k \mid \kappa \geq \gamma)$. None of κ, γ, or the held-out labels enters pseudo-label generation, recoverability gating, parameter updates, hyperparameter selection, or checkpoint selection.

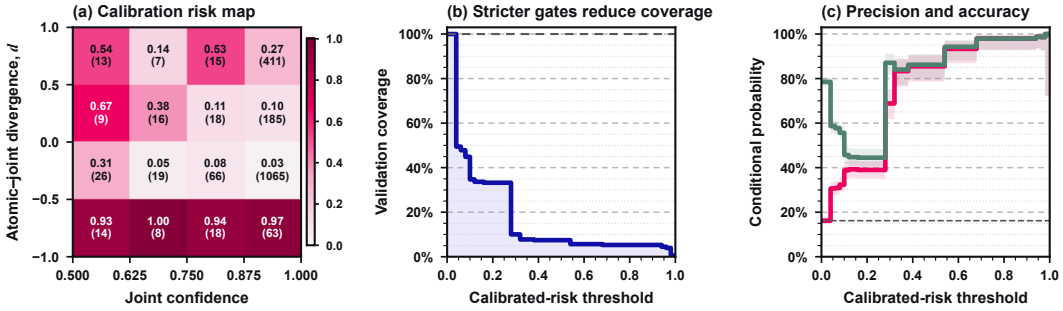

Figure A1: Post-hoc calibrated-risk diagnostic. (a) Held-out suppressive-error probability over joint margin confidence q^J and signed atomic–joint divergence d. Numbers in parentheses are bin counts. (b) Increasing the calibrated-risk threshold γ reduces analysis coverage. (c) The retained cases become increasingly concentrated in recoverable joint errors and correct atomic targets. This diagnostic uses evaluation labels only after training and is not an optimization gate.

B Additional Results

B.1 Core Mechanism Ablation

The variants below share the same Qwen3.5-4B initialization, complete 6K unlabeled schedule, trainable parameter scope, and update budget, isolating the contribution of elicitation and reconciliation.

B.2 Recoverability Computation

The practical gate uses only the regularized score in Eq. 8. For each candidate, the implementation first forms $\widetilde{J} = [J_{\mathcal{R}}; \sqrt{\beta} J_{\mathcal{P}}]$ and $\widetilde{r} = [r; 0]$, solves $\alpha = (\widetilde{J}\widetilde{J}^\top + \lambda I)^{-1}\widetilde{r}$, maps back $\Delta\theta_{\text{lin}} = \widetilde{J}^\top \alpha$, and then evaluates

$$\rho_{\beta,\lambda} = 1 - \frac{\|J_{\mathcal{R}}\Delta\theta_{\text{lin}} - r\|_2^2}{\|r\|_2^2 + \varepsilon_{\text{num}}}.$$

The score is computed before trust-region scaling and fresh-forward acceptance. Exact hard range membership from Proposition 1 is never used as an empirical scalar gate. This implementation remains non-saturated when $J_{\mathcal{R}} P_N$ has full row rank because damping and preservation penalties determine how much of the requested repair is realized by the actual regularized step.

Variant	AD ↑	DC ↑	AD/DC Mean ↑
Joint ST	67.38	61.05	64.22
Joint SE	69.52	62.41	65.97
Atomic SFT	76.84	67.92	72.38
Elicitation-only	80.91	70.88	75.90
Elicit.+PCGrad	81.73	71.44	76.59
PRIDE	**85.57**	**74.40**	**79.99**

Table A2: Core ablation (0-shot). All variants use the same Qwen3.5-4B initialization, same complete 6K unlabeled schedule, trainable parameter scope, and update budget.

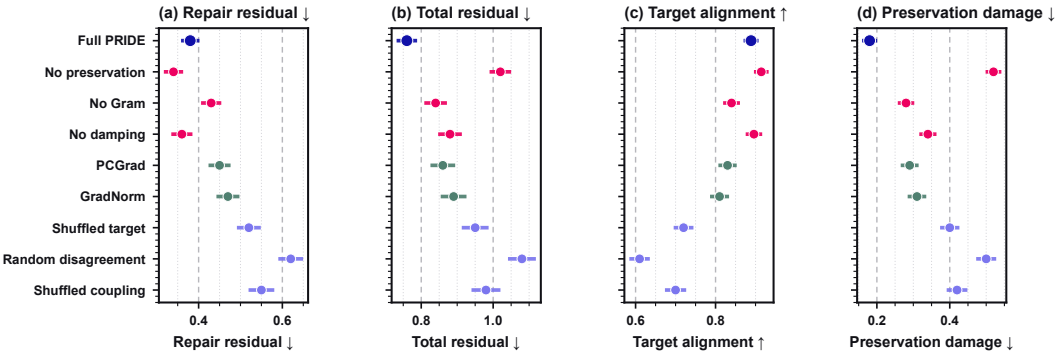

Figure A2: Core reconciliation ablation. Full PRIDE improves repair, target alignment, and preservation across matched controls.

B.3 Trust Region and Multi-Step Stability

Table A3 complements Figure 3 with a post-hoc local-linearization diagnostic. Small steps underutilize the available correction; overly large steps reduce the linear residual but violate the fresh-forward model. The reported default is fixed independently of this MMAD sweep; the table characterizes its operating neighborhood rather than selecting it.

η	Linear residual ↓	Nonlinear residual ↓	Preserve drift ↓	AD BAcc ↑
0.25	0.42	0.45	0.018	82.91
0.50	0.31	0.34	0.029	84.86
0.75	**0.24**	**0.27**	0.041	**85.57**
1.00	0.20	0.36	0.063	84.74
1.25	0.18	0.51	0.091	82.58

Table A3: Trust-region sweep. The linear objective can keep improving after nonlinear validity has started to deteriorate.

Long-horizon behavior is shown post hoc in Table A4. The standard 12-round protocol completes one pass over all 6,000 ADOPD images. For diagnostic extensions beyond round 12, the 6K source set is reshuffled and reused for additional rounds. Performance then saturates/slightly decreases as target precision falls, supporting re-estimation of stability and recoverability every round rather than irreversible accumulation of pseudo-labels. The main 12-round checkpoint is fixed by protocol rather than selected from this MMAD trajectory.

Rounds	AD BAcc	Target Prec.	Preservation damage	Reject rate
4	78.34	89.8	0.031	7.4
8	83.76	88.2	0.036	9.1
12	**85.57**	86.9	0.041	11.3
16	85.21	84.6	0.046	13.8
20	84.63	81.7	0.052	16.5

Table A4: Long-horizon self-evolution stability. Metrics start from a common initialization; target precision uses held-out labels post hoc.

B.4 Gate Error Analysis

The elicitation gate trades target precision against supervision coverage. Table A5 is a post-hoc MMAD-label analysis showing how the fixed operating point lies on this trade-off; it is not used to tune τ_c.

τ_c	Coverage (%)	Target precision (%)	Target recall (%)	AD BAcc
0.55	18.7	74.8	**78.4**	82.41
0.65	14.6	81.2	72.6	84.39
0.75	10.8	86.9	65.3	**85.57**
0.85	6.2	**91.7**	43.8	83.92

Table A5: Elicitation-gate operating points. Precision/recall refer to "atomic target correct and joint target wrong" on held-out analysis labels.

B.5 Parameter Sensitivity

Table A6 changes one factor at a time around the fixed default configuration after the main protocol is specified. These MMAD results are reported only as post-hoc sensitivity diagnostics and are not used for hyperparameter or checkpoint selection. The trends are non-monotonic: too little preservation/damping admits collateral drift, while overly conservative settings suppress useful repair. Eight atomic rollouts capture most of the observed stability gain.

Factor	Value → AD BAcc				
τ_c	.55: 82.41	.65: 84.39	.75: **85.57**	.85: 83.92	.95: 80.76
β	0: 82.96	.25: 84.88	.50: **85.57**	1.0: 85.11	2.0: 83.76
λ	10^{-4}: 82.73	10^{-3}: 84.66	10^{-2}: **85.57**	10^{-1}: 84.02	1: 81.55
η	.25: 82.91	.50: 84.86	.75: **85.57**	1.0: 84.74	1.25: 82.58
rollouts	1: 81.94	2: 83.28	4: 84.92	8: **85.57**	12: 85.61

Table A6: Post-hoc sensitivity around the fixed default configuration. Only one factor changes at a time; metric is MMAD 0-shot AD BAcc. These sweeps are diagnostic and are not used for model selection.

B.6 Computational Scaling

For $m = |\mathcal{R}| + |\mathcal{P}|$ selected outputs and P trainable parameters, the explicit Gram solve is $O(m^3)$ and remains small in the tested range. Jacobian-row acquisition scales approximately linearly with the number of selected outputs and is the practical bottleneck. Table A7 reports runtime and memory scaling as the number of queried dimensions increases.

K	Mean selected m	Relative time	Peak memory	Gram solve / step
4	2.6	1.42×	1.10×	0.3 ms
8	4.9	1.78×	1.18×	0.6 ms
16	8.7	2.31×	1.31×	1.8 ms
32	15.4	3.38×	1.55×	6.9 ms
64	27.1	5.51×	1.98×	31.4 ms

Table A7: Scaling with queried dimensions. Runtime and memory are normalized to matched scalar SFT on the same hardware and batch.

B.7 End-to-End Dynamics

Figure A3 summarizes the precision–recall and target-adoption trajectories. PRIDE expands reliable supervision rapidly in early rounds; later rounds become precision-limited, consistent with the saturation in Table A4.

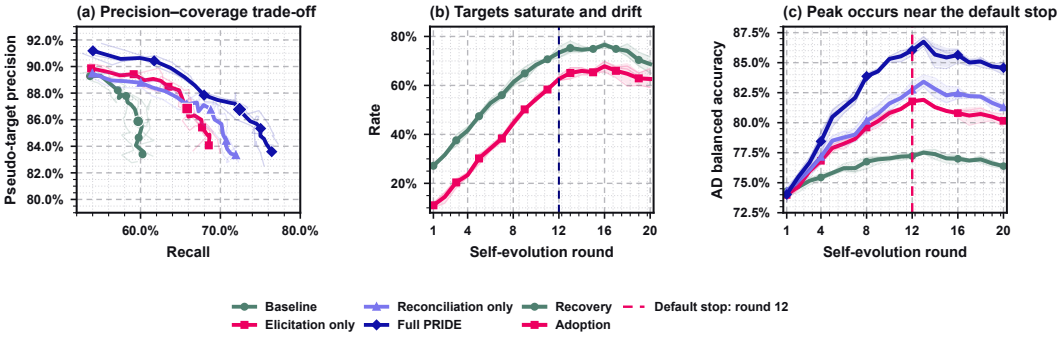

Figure A3: End-to-end self-evolution dynamics across training variants.

Method	0-shot AD	0-shot DC	1-shot AD	1-shot DC
Atomic SFT	76.84±.66	67.92±.43	79.15±.61	69.74±.39
Elicitation-only	80.91±.59	70.88±.40	82.35±.53	72.46±.37
Elicitation+PCGrad	81.73±.55	71.44±.39	83.12±.47	72.80±.35
PRIDE	85.57±.41	74.40±.36	87.16±.33	76.02±.31

Table A8: Seedwise statistics for principal matched baselines. Mean±std over three seeds.

B.8 Seedwise Main Results and Significance

We use paired bootstrap resampling over identical evaluation items with 10,000 resamples. Against Elicitation+PCGrad, PRIDE improves 0-shot AD by 3.84 points (95% CI $[2.71, 4.96]$, two-sided $p = 0.0016$) and DC by 2.96 points (95% CI $[1.88, 4.03]$, $p = 0.0031$); for 1-shot, the corresponding gains are 4.04 AD points (95% CI $[2.96, 5.10]$, $p = 0.0011$) and 3.22 DC points (95% CI $[2.14, 4.28]$, $p = 0.0024$). Holm correction is applied when several variants are tested against PRIDE simultaneously.

C Additional Theory Details

C.1 Proof of Exact Recoverability

The hard preservation constraint $J_{\mathcal{P}}\Delta\theta = 0$ implies $\Delta\theta \in \mathrm{Null}(J_{\mathcal{P}})$. Let P_N be the orthogonal projector onto this null space; then every feasible update has the form $\Delta\theta = P_N u$. Substituting into the repair constraint gives $J_{\mathcal{R}}P_N u = r$. A solution exists exactly when r belongs to $\mathrm{Range}(J_{\mathcal{R}}P_N)$, proving Proposition 1. Among all feasible solutions, the Moore–Penrose solution $u^* = (J_{\mathcal{R}}P_N)^{\dagger}r$ minimizes the Euclidean norm, yielding Eq. 5.

The same condition can be expressed through the conditional Gram matrix $G_{\mathcal{R}|\mathcal{P}} = G_{\mathcal{R}R} - G_{\mathcal{R}P}G_{\mathcal{P}P}^{\dagger}G_{\mathcal{P}R}$, where $G_{ab} = J_a J_b^{\top}$. This form makes coupling explicit: directions shared with preserved outputs are removed before measuring repair capacity. If $J_{\mathcal{R}}P_N$ has full row rank, hard range membership is identically satisfied for every nonzero r; accordingly, the exact condition is retained as a theorem and is not used as the continuous empirical gate.

C.2 Soft Constraint as a Regularized Relaxation

The practical objective in Eq. 6 replaces the hard preservation constraint with penalty β and adds Tikhonov damping λ. Writing $\widetilde{J} = [J_{\mathcal{R}}; \sqrt{\beta}J_{\mathcal{P}}]$ and $\widetilde{r} = [r; 0]$ reduces the problem to ridge regression $\min_{\Delta\theta} \|\widetilde{J}\Delta\theta - \widetilde{r}\|^2 + \lambda\|\Delta\theta\|^2$. The primal solution $(\widetilde{J}^{\top}\widetilde{J} + \lambda I)^{-1}\widetilde{J}^{\top}\widetilde{r}$ is algebraically equivalent to the output-space form in Eq. 7; the latter is preferable because $m \ll P$. The empirical gate uses $\rho_{\beta,\lambda} = 1 - \|J_{\mathcal{R}}\Delta\theta_{\mathrm{lin}} - r\|_2^2/(\|r\|_2^2 + \varepsilon_{\mathrm{num}})$, computed from this same regularized solution before trust-region scaling. This deliberately separates exact feasibility in Proposition 1 from the non-saturating score used in experiments.

C.3 Local Prediction Guarantee

Assume the selected-output Jacobian is L-Lipschitz. Taylor's theorem gives $\|s(\theta + \Delta) - s(\theta) - J\Delta\| \le (L/2)\|\Delta\|^2$. Let γ_R be the minimum repair-coordinate margin required to cross the target decision boundary and γ_P the minimum distance of preserved coordinates to their current boundaries. If the linear repair residual plus Taylor remainder is smaller than γ_R, and preservation drift plus remainder is smaller than γ_P, then repair decisions attain the target signs while preserve decisions remain unchanged. This is why the algorithm enforces both a trust radius and a fresh-forward acceptance test.

C.4 Self-Evolution Dynamics

Let $\mathcal{A}_t$ be the stable atomic-capability set after round t and $\mathcal{J}_t$ the subset currently accessible under joint inference. Atomic consistency training seeks to expand $|\mathcal{A}_t|$, while reconciliation transfers recoverable elements of $\mathcal{A}_t \setminus \mathcal{J}_t$ into $\mathcal{J}_{t+1}$. The two processes therefore alternate: stable atomic

discovery → atomic expansion → renewed composition failures → preservation-aware joint repair → new supervision in the next round. Monotonic growth is not guaranteed because pseudo-target precision can decrease as the model distribution changes, motivating complete re-estimation each round.

C.5 Information-Theoretic Limit

PRIDE cannot create evidence absent from every internal view. Let $O_\theta(x)$ denote all observations available to the teacher-free procedure. If $P(O_\theta \mid y_1) = P(O_\theta \mid y_2)$ for distinct target states, no algorithm using only O_θ can distinguish them; in the extreme $I(Y; O_\theta) = 0$, endogenous self-supervision contains no information about Y. PRIDE is therefore designed to recover capabilities that are expressed in some stable internal view but are inaccessible under composition, not capabilities absent from the model's observable evidence.

D Additional Experimental Protocols

D.1 Comparison and Matched-Baseline Protocols

Table 1 reports source/backbone-standardized reimplementations. Contemporary IAD methods use the same 6,000 ADOPD-Dataset-6K source images, the same Qwen3.5-4B backbone, and the same MMAD evaluator while retaining the mechanism that defines each method, including its SFT/GRPO, comparative, latent-reasoning, or distillation component where applicable. This controls backbone, source exposure, and evaluator, but does not imply identical supervision, auxiliary modules, or parameter count.

The strict mechanism comparison is Table A2: Joint ST, Joint SE, Atomic SFT, Elicitation-only, Elicitation+PCGrad, and PRIDE share the same Qwen3.5-4B initialization, twelve disjoint 500-image unlabeled shards, trainable parameter scope, optimization budget, evolution rounds, and MMAD examples. Atomic SFT applies ordinary cross-entropy to stability-filtered atomic targets; Elicitation-only uses accepted atomic targets without functional reconciliation; Elicitation+PCGrad uses the same accepted targets but resolves scalar-loss gradients with PCGrad; PRIDE uses the repair/preserve Jacobian constraints. The 0-shot protocol receives only the MMAD query image, while 1-shot additionally receives one normal reference according to the evaluator; neither setting updates model parameters on MMAD.

D.2 Published MMAD Results under Native Configurations

The standardized reimplementations above answer a controlled mechanism question, whereas published systems were developed with different backbones, training corpora, supervision, and inference settings. Table A9 therefore reproduces representative seven-task numbers from the corresponding published evaluation tables without converting them into the controlled setting. The native table is used only to locate PRIDE relative to the literature and is not used for apples-to-apples attribution.

D.3 MMAD Subset Protocol

All subset results use identical sample IDs and category definitions. DS-MVTec, GoodsAD, MVTec-LOCO, and VisA report both AD and DC; MVTec-AD reports AD only. No subset-specific threshold or adaptation schedule is selected from test labels. The aggregate AD reported in Table 1 is the source-macro mean over the five AD sources; DC is the source-macro mean over the four sources for which the task is defined.

D.4 Prompt and Stochasticity Controls

The controls use identical image/task pairs and decoding settings. Paraphrase and length/schema controls change only surface form; repeatability is measured across eight stochastic rollouts, and target precision is computed post-hoc with held-out labels.

Method	Scale	Native protocol	AD	DC	DL	DD	DA	OC	OA	Avg.
Representative results reported in InspectorGPT under its MMAD evaluation setting										
AnomalyGPT (Gu et al., 2024)	7B	native	65.57	27.49	27.97	36.86	32.11	29.84	35.82	36.52
Qwen2.5-VL (Bai et al., 2025b)	7B	native	70.47	56.15	60.35	64.30	78.34	92.11	83.58	72.19
Qwen2.5-VL (Bai et al., 2025b)	72B	native	72.66	62.31	67.16	73.56	81.95	94.30	86.78	76.96
AnomalyR1 (Chao et al., 2025)	3B	native	60.20	63.50	70.10	80.40	85.20	82.40	86.10	75.41
OmniAD (Zhao et al., 2025)	7B	1-shot	68.80	78.80	75.50	67.20	86.40	96.00	86.40	79.87
EMIT (Guan et al., 2025)	8B	native	73.87	80.85	76.39	83.00	85.92	90.26	83.37	81.95
AD-FM (Liao et al., 2026)	7B	native	73.15	73.36	77.53	86.80	86.75	89.98	86.67	82.03
JUDO (Kang et al., 2026)	7B	native	64.51	72.17	75.95	84.38	87.76	94.24	86.07	80.73
AD-Copilot (Jiang et al., 2026)	7B	1-shot	73.95	74.29	76.40	84.92	86.93	91.86	87.67	82.29
InspectorGPT (Chen et al., 2026d)	7B	native	73.90	75.32	75.91	82.18	88.25	92.94	88.18	82.38
Zero-shot results reported in ADOPD										
Qwen3-VL (Bai et al., 2025a)	8B	0-shot	67.20	58.15	58.74	66.88	77.40	92.70	83.74	72.12
AnomalyR1 (Chao et al., 2025)	3B	0-shot	75.34	45.49	61.15	68.96	83.88	90.30	84.99	72.87
IAD-R1 (Li et al., 2026)	7B	0-shot	75.72	54.53	60.72	67.43	77.65	93.17	84.02	73.32
Reason-IAD (Chen et al., 2026b)	8B	0-shot	67.73	73.12	59.74	74.69	80.65	96.78	84.24	76.71
ADOPD (He et al., 2026b)	4B	0-shot	88.44	57.03	63.64	70.78	84.16	92.77	84.32	77.31

Table A9: Published MMAD results under native configurations. Values are copied from the corresponding published evaluation tables and retain each work's original backbone, training data, supervision, and inference setting. They provide literature context only; differences in protocol prevent direct attribution against the standardized reimplementations in Table 1. All columns are VQA accuracy (%).

Method	DS-MVTec		GoodsAD		MVTec-AD	MVTec-LOCO		VisA	
	AD	DC	AD	DC	AD	AD	DC	AD	DC
Base Qwen3.5-4B	72.40	68.10	55.80	55.60	66.10	56.70	54.80	64.60	55.18
AnomalyR1	82.00	70.80	62.00	61.20	76.00	66.00	60.50	76.30	66.10
OmniAD	85.40	73.90	64.80	64.70	80.20	68.30	64.60	77.20	68.64
AD-FM	88.60	75.20	67.50	66.10	84.70	71.70	66.80	80.75	68.34
InspectorGPT	90.20	77.80	69.40	68.60	87.60	73.20	68.90	81.45	71.42
ADOPD	91.50	76.30	72.80	66.70	93.20	76.40	67.00	87.15	69.44
PRIDE	**92.71**	**80.20**	**73.63**	**72.40**	**95.24**	**78.56**	**70.10**	**87.71**	**74.90**

Table A10: 0-shot MMAD source-wise AD/DC results. The five AD columns macro-average to the aggregate AD value in Table 1; DC is macro-averaged over the four sources for which MMAD defines the task. All reproduced IAD rows use Qwen3.5-4B and the same ADOPD-6K source pool.

Condition	Gap (%)	Jaccard	Repeat (%)	Target Prec. (%)
Canonical atomic	23.6	1.00	82.7	86.9
Paraphrase A	22.9	0.88	81.9	86.1
Paraphrase B	24.1	0.86	81.4	85.4
Length-matched atomic	23.4	0.90	83.1	86.4
Joint-schema single-slot	22.6	0.89	82.3	85.8
Accepted subset (canonical)	10.8	0.92	**93.4**	**86.9**
Dimension-permutation null	23.5	0.12	50.8	49.7

Table A11: Prompt-artifact and stochastic-noise controls. Gap is atomic/joint disagreement; Jaccard compares gap sets with the canonical template; Repeat measures sign consistency over eight rollouts. Target precision uses held-out labels only for analysis.

D.5 Cross-Backbone and Cross-Domain Protocol

For cross-backbone MMAD evaluation, Qwen3.5-4B, Qwen3-VL-4B, LLaVA-OneVision-1.5-4B, and Gemma-3n-E4B use the same twelve ADOPD-Dataset-6K shard IDs, the same round order, semantic task definitions, acceptance criteria, and MMAD evaluation examples. Tokenization, image preprocessing, and the architecture-specific adaptation modules are changed only where required by the backbone. This keeps sample exposure and task semantics fixed while testing whether the atomic–joint reconciliation mechanism transfers across model families.

For cross-domain open-ended VQA, questions associated with the same visual source are deterministically grouped into task sets. Atomic inference presents one question at a time, while joint inference presents the identical questions in a fixed ordered answer schema. For question k, M atomic rollouts are canonicalized by $\nu(\cdot)$ using case, whitespace, punctuation, and benchmark-compatible numeric normalization only; no reference answer enters this step. A stable modal answer $a_k^\star$ is accepted when its rollout frequency satisfies $c_k \geq \tau_c$. The same answer sequence is teacher-forced under the atomic and joint contexts, and its length-normalized log-likelihood gives s_k^A and s_k^J according to Eq. 2. The repair/preserve sets and Jacobian reconciliation are then constructed exactly

as in the binary case. Ground-truth answers are exposed only to the final benchmark evaluator. GQA uses exact accuracy and ChartQA uses relaxed accuracy.

Backbone	0-shot				1-shot			
	Base AD	PRIDE AD	Base DC	PRIDE DC	Base AD	PRIDE AD	Base DC	PRIDE DC
Qwen3.5-4B	63.12	85.57	58.42	74.40	70.84	87.16	64.20	76.02
Qwen3-VL-4B	61.48	81.92	55.84	70.21	68.92	83.70	61.25	72.04
LLaVA-OneVision-1.5-4B	58.73	76.84	52.37	65.68	65.48	78.92	58.72	67.31
Gemma-3n-E4B	60.16	77.95	54.06	67.32	66.52	80.11	60.11	68.94

Table A12: Cross-backbone MMAD transfer. Each backbone uses the same twelve ADOPD-6K shard IDs, round order, semantic task construction, and MMAD evaluation examples; only architecture-specific preprocessing and trainable adapters differ.

Benchmark	Metric	Base	Joint SE	PRIDE
GQA (Hudson & Manning, 2019)	Acc.	61.8	62.6	**65.7**
ChartQA (Masry et al., 2022)	Relaxed Acc.	70.4	71.3	**74.6**

Table A13: Cross-domain open-ended VQA transfer. Atomic and joint views are constructed from questions sharing the same visual source. Stable model-generated answers provide the self-supervision target; ground-truth answers remain hidden during self-evolution and are used only by the held-out evaluator.

E Qualitative Error Analysis

Case ID	Failure type	Observed pattern	Interpretation
GoodsAD-017	Atomic hallucination	Stable atomic positive, correct joint negative	Gate false acceptance; motivates precision-oriented thresholding.
LOCO-041	Low recoverability	Correct atomic target but $\rho_{\beta,\lambda} < \tau_r$	The regularized local solve cannot realize enough of the repair target while preserving protected coordinates, so the update is skipped.
VisA-063	Trust-region failure	Linear residual decreases but nonlinear check worsens	Jacobian is locally stale; smaller step or larger damping is required.
DS-088	Round drift	Early repair succeeds; later confidence rises while correctness falls	Stability must be re-estimated each round instead of accumulated indefinitely.

Table A14: Representative failure taxonomy and the diagnostic recorded for each case.

The qualitative analysis records the atomic/joint output, gate score, acceptance decision, $\rho_{\beta,\lambda}$, update norm, linear/nonlinear residuals, and post-update prediction. The four cases above cover the main failure classes: false atomic evidence, low recoverability, local-linearization failure, and long-horizon drift. These cases are intended to expose when PRIDE abstains or fails rather than only successful repairs.

F Social Impact Statement

PRIDE is a model-adaptation mechanism for improving consistency across related visual judgments. In industrial inspection it may reduce manual review effort, but false positives and false negatives remain possible and self-generated supervision can amplify systematic errors. Deployment in safety-critical settings requires domain validation, calibrated abstention, audit logs for accepted self-generated targets, and human oversight. The reported experiments do not require human-subject interaction or personally identifiable information.

G ICLR 2027 Reproducibility Checklist

1. **Problem/claims – Yes.** Prompt, schema, length, and stochastic alternatives are explicitly tested.
2. **Method – Yes.** Stable atomic discovery, atomic expansion, recoverability, Jacobian/Gram reconciliation, and the nonlinear acceptance check are specified.
3. **Theory – Yes.** Exact recoverability, minimum-intervention solution, soft relaxation, local-validity conditions, dynamics, and limitations are provided.
4. **Hyperparameters – Yes.** Table A1 reports optimization and reconciliation settings.

5. **Prompts – Yes.** Canonical, matched-schema, paraphrase, and length controls are described.
6. **Datasets/metrics – Yes.** The external ADOPD-Dataset-6K adaptation pool, MMAD evaluation-only boundary, seven-task MMAD reporting, and cross-domain metrics are explicitly specified.
7. **Baselines – Yes.** The paper separates source/backbone-standardized IAD reimplementations, published native-configuration results, and a strict label-free budget-matched ablation; the corresponding control variables are stated explicitly.
8. **Seeds/statistics – Yes.** Three-seed statistics and paired-bootstrap significance tests are reported.
9. **Efficiency – Yes.** Runtime, memory, selected-output scaling, and Gram-solve cost are reported in Appendix B.6.
10. **Failure analysis – Yes.** Gate errors, low recoverability, trust-region failures, round drift, and representative qualitative cases are documented.
11. **Ethics/AI use – Yes.** Social impact and AI-assisted writing disclosure are included; the latter states that LLM use is limited to language polishing.