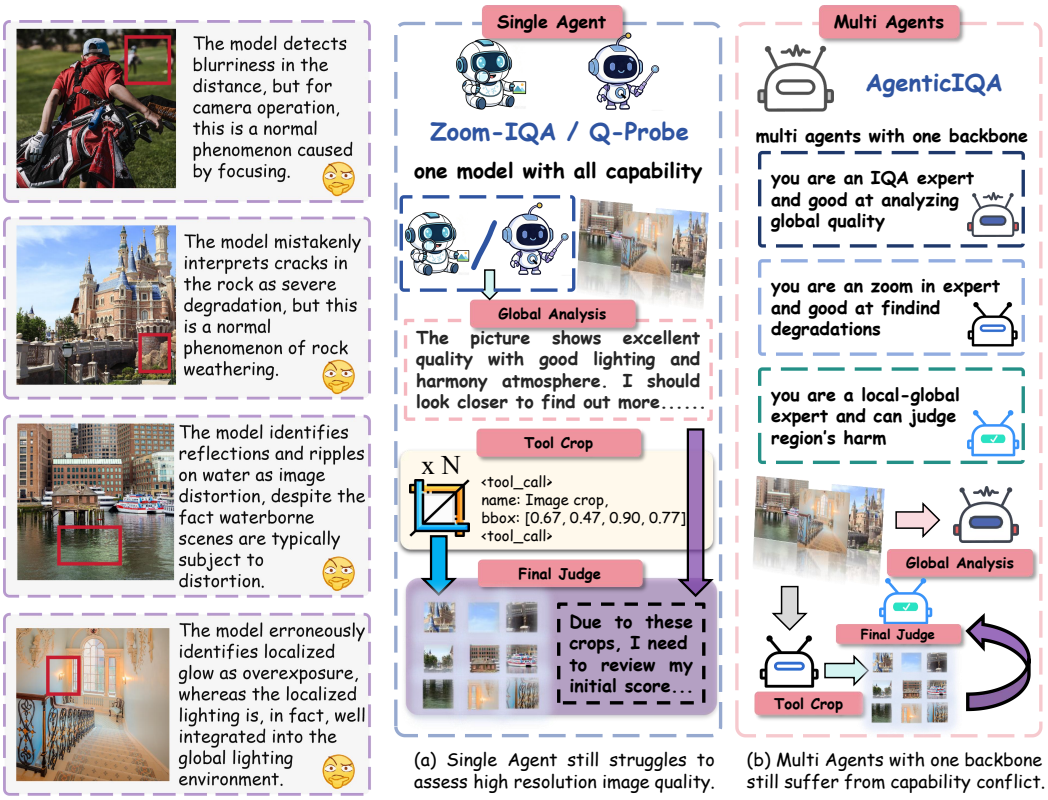
PRISMIQA: CAPABILITY-FACTORIZED MULTI-AGENT IMAGE QUALITY ASSESSMENT WITH FALSIFICATION AND REGION-TO-GLOBAL ATTRIBUTION

Anonymous authors
Paper under double-blind review



(a) Single Agent still struggles to assess high resolution image quality.

(b) Multi Agents with one backbone still suffer from capability conflict.

Figure 1: At present, it is a popular practice to invoke tools with zoom-in capabilities to support image IQA tasks. However, the abilities of holistic analysis, zooming in, judging crop conditions, and relating crops to the global image are overly concentrated in a single model, which may create limitations and make it impossible to ever achieve optimality.(a) is just such an example. It is worth noting that many multi-agent IQA models, as in (b), are essentially still a single model injected with different prompts to play different roles. Although this may indeed improve model capability, it still cannot solve the problem that the theoretical upper limit remains constrained.

ABSTRACT

Reasoning-capable vision–language models have substantially advanced image quality assessment (IQA), including high-resolution assessment through active crop-and-zoom operations. Nevertheless, most existing systems assign global scoring, defect localization, degradation discrimination, and perceptual-impact estimation to one general-purpose policy. Recent agentic systems alleviate part of this burden through workflow-level roles such as planner–executor–summarizer, but the planner still commonly performs broad perception and task allocation. We

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.

instead study *capability-level factorization*: each agent has a distinct objective, evidence space, and action set. Our framework, PrismIQA, comprises four stages: a Global Assessor estimates a calibrated holistic prior; a high-recall Region Proposer identifies candidate defect regions; a Falsification Agent compares each crop with full-image context and rejects pseudo-degradations such as intentional depth-of-field and natural texture; and an Impact Assessor estimates the severity and semantic importance of verified defects and attributes their residual effect on the global score. We analyze this architecture as a policy over semantic interaction trajectories. Through Lipschitz–transport analysis and a KL-chain diagonal constraint, we prove that the factorized four-agent system is superior to a shared model: four independently optimized models yield a strictly smaller deviation upper bound than one shared model. On seven IQA datasets, PrismIQA reaches 0.854 SRCC and 0.892 PLCC on Vista-Bench and 0.836 average SRCC, outperforming Q-Probe by $+0.053$ SRCC on Vista-Bench.

1 INTRODUCTION

Image Quality Assessment (IQA) aims to predict human judgments of visual quality and is central to image acquisition, restoration, compression, and generation. Classical no-reference methods estimate quality from natural-scene statistics or learned visual representations (Mittal et al., 2012; 2013; Ke et al., 2021). More recent vision–language approaches connect low-level degradation with semantic content, enabling both numerical scoring and natural-language diagnosis (Zhang et al., 2023; Wu et al., 2024; You et al., 2025). Reinforcement-learning-based models further improve preference alignment and quality reasoning (Li et al., 2025; Wu et al., 2025).

High-resolution IQA remains difficult because a globally resized observation can suppress small but perceptually important artifacts. Active visual reasoning addresses this limitation by acquiring additional image evidence. General-purpose systems such as DeepEyes and Pixel Reasoner learn to inspect selected regions (Zheng et al., 2026; Su et al., 2025); IQA-specific methods including Q-Probe, Zoom-IQA, and Q-DeepSight apply crop-and-zoom reasoning to localized quality assessment (Li et al., 2026a; Liang et al., 2026; Li et al., 2026b). Tool-IQA and IQA-T1 instead invoke perceptual tools such as magnification, gamma correction, residual maps, gradients, and frequency analysis (Qin et al., 2026; Wu et al., 2026).

These advances expose a second problem: acquiring a crop changes the model's observation distribution. A model trained primarily on defective crops can learn a confirmation shortcut in which a requested crop is itself treated as evidence of poor quality. Local views also remove context needed to distinguish camera or artistic effects from defects; background bokeh, motion intent, textureless regions, and film grain are common examples. Q-Probe directly studies this "cropping-implies-degradation" bias through context-aware training (Li et al., 2026a). Our goal is complementary: we make *falsification* an explicit decision stage with its own objective and abstention behavior.

Existing agentic IQA systems demonstrate the value of modular processing. AgenticIQA organizes a shared VLM into planner, executor, and summarizer roles and dynamically calls IQA tools (Zhu et al., 2025); a related planner–executor–summarizer system performs multi-dimensional pairwise quality comparison (Liao et al., 2026). Beyond IQA, MATPO has become a popular training recipe for the same planner-style lineage: a planner and its workers share one LLM, activated by role prompts and optimized jointly by reinforcement learning with credit assignment across hierarchical rollouts (Mo et al., 2025). Such workflow decomposition answers *who controls the next operation* (or which sub-call to issue). It is effective when the main policy and the invoked agents require the *same capability direction*—typically tool-integrated research or execution—and differ mainly in hierarchy. It is not designed for settings where the capabilities themselves must be factorized: a planner that still performs broad perception can conflate global scoring, high-recall localization, falsification, and impact attribution, whose objectives need not share a common minimizer. We therefore investigate a different decomposition based on *what capability is optimized*.

We propose PrismIQA, a four-role architecture:

- the Global Assessor predicts a holistic quality prior and calibrated uncertainty without inspecting selected crops;

- the Region Proposer is optimized for high recall and proposes regions and defect hypotheses without assigning the final quality score;
- the Falsification Agent judges each proposed crop against the full image and returns accept, reject, or abstain;
- the Impact Assessor estimates verified regional severity, semantic importance, and residual contribution to the global judgment.

The agents exchange only structured state: the global assessment conditions proposal, proposed regions become falsification hypotheses, and only verified or unresolved evidence reaches region-to-global attribution.

Our theoretical motivation follows the trajectory-level view proposed for Agent-Omit (Ning et al., 2026). Let $Y = (Y_g, Y_p, Y_f, Y_e)$ be the complete IQA interaction trajectory, let $d(y, y') = \|\phi(y) - \phi(y')\|$ be its semantic trajectory distance, and let R denote the assessment reward. Under semantic Lipschitz and transport-regularity assumptions, Section 4 establishes a deviation certificate of the form

$$|\mathbb{E}_{\pi^*} R - \mathbb{E}_{\pi} R| \leq \delta_R + K'_R D_{\mathrm{KL}}(\pi^* \| \pi). \tag{1}$$

For the four-agent policy, the KL chain rule writes the joint divergence as the sum of conditional divergences for global assessment, proposal, falsification, and impact attribution. We use each conditional KL term as a capability loss $\mathcal{L}_j$. A single generalist must minimize all four losses with one shared parameter vector, whereas the factorized system assigns one vector to each capability:

$$\mathcal{L}^*_{\mathrm{single}} = \min_{\theta} \sum_j \mathcal{L}_j(\theta), \qquad \mathcal{L}^*_{\mathrm{multi}} = \sum_j \min_{\theta_j} \mathcal{L}_j(\theta_j). \tag{2}$$

The shared solutions form only the diagonal subset $\{(\theta, \theta, \theta, \theta)\}$ of the factorized product space. Hence $\mathcal{L}^*_{\mathrm{multi}} \leq \mathcal{L}^*_{\mathrm{single}}$; if the four capability losses have no common minimizer, the inequality is strict:

$$\mathcal{L}^*_{\mathrm{multi}} = \mathcal{L}^*_{\mathrm{single}} - \epsilon_{\mathrm{cap}}, \qquad \epsilon_{\mathrm{cap}} > 0. \tag{3}$$

Because these losses are precisely the conditional terms in the KL chain, Equation 1 then gives a strictly tighter reward deviation certificate for four independently optimized models than for one shared model (Proposition 2). Appendix C records the complete argument and the full-parameter SFT/GRPO test of this four-versus-one claim. A complementary causal result formalizes crop-selection adjustment.

Our contributions are:

1. We introduce capability-factorized IQA, separating global judgment, high-recall localization, context-aware falsification, and contextual impact attribution instead of assigning broad perception to a generic planner.
2. We formulate pseudo-degradation rejection as a dedicated filter stage that compares local crops with full-image context and may abstain; under a causal-sufficiency condition, its pre-treatment scene state supports adjustment of crop-selection bias.
3. We adapt a semantic-trajectory deviation framework to factorized IQA, derive its exact conditional KL decomposition, and prove that four independently optimized capability models strictly dominate one shared model when the role losses have no common minimizer (Proposition 2; Appendix C).
4. We evaluate the system with controls that separately test global scoring, localization, false-positive control, and counterfactual validation of region-to-global attribution.

2 Related Work

2.1 Image Quality Scoring and Reasoning

Early blind IQA methods use hand-designed natural-scene statistics (Mittal et al., 2012; 2013); later models learn multi-scale or attention-based representations (Ke et al., 2021; Yang et al., 2022). LIQE connects quality, scene, and distortion labels through vision–language correspondence (Zhang et al.,

2023). Q-Align teaches multimodal models text-defined quality levels (Wu et al., 2024), while DeQA-Score models continuous score distributions (You et al., 2025). Q-Insight and VisualQuality-R1 use reinforcement learning for quality regression and ranking (Li et al., 2025; Wu et al., 2025). These methods primarily optimize a single scoring policy.

2.2 Active and Tool-Augmented IQA

Q-Probe combines global inspection and context-aware regional probing for high-resolution IQA (Li et al., 2026a). Zoom-IQA learns when and where to zoom and constructs region-grounded rationales (Liang et al., 2026). Q-DeepSight interleaves textual reasoning with image-based evidence acquisition (Li et al., 2026b). Tool-IQA and IQA-T1 use adaptive calls to external perceptual operators (Qin et al., 2026; Wu et al., 2026). These systems establish that regional evidence and tools are useful, but they do not isolate a high-recall proposer from a dedicated context-aware falsifier and a region-to-global impact estimator.

2.3 Multi-Agent Visual Understanding

Multi-agent visual reasoning predates its use in IQA. HAMMR builds a hierarchical ReAct system whose agents call specialized visual agents for a diverse VQA suite (Castrejon et al., 2024). MAM-MQA assigns different agents to modality-specific retrieval and cross-modal synthesis (Anvekar et al., 2025). DART uses disagreement among visual agents as a signal for recruiting external tools (Sivakumaran et al., 2026). In document understanding, ORCA routes decomposed questions to specialized agents and stress-tests answers through debate (Lassoued et al., 2026), while MACT uses planning, execution, judgment, and answer agents with agent-wise adaptive test-time scaling (Yu et al., 2026). ReFeR employs a hierarchy of language and vision–language models to evaluate generated text and images (Narsupalli et al., 2024). These studies establish that multi-agent image understanding is an active research direction; they optimize VQA, document reasoning, or general output evaluation rather than perceptual MOS prediction.

2.4 Multi-Agent IQA and Evaluation

AgenticIQA uses planner, executor, and summarizer roles for query-aware IQA (Zhu et al., 2025). The pairwise system of Liao et al. (2026) similarly identifies dimensions and regions, invokes specialized models, and aggregates a final preference. AstroVLM coordinates domain experts for astronomical imaging-quality diagnosis, a related but domain-specific task without general-purpose NR-IQA scoring (Han et al., 2026). MR-IQA-2 uses an actor–editor–judge loop in which image editing makes quality rationales verifiable and enables fine-grained credit assignment (Li et al., 2026c). We therefore do not claim the first multi-agent IQA or the first edit-based verification system. Our narrower distinction is the inference-time separation of high-recall proposal, context-aware falsification, and calibrated region-to-global attribution for high-resolution NR-IQA. More broadly, Agent-Omit studies an agent's learned omission policy and relates policy distance to effectiveness (Ning et al., 2026). Our analysis does adapt its semantic-trajectory deviation result to IQA. The KL chain rule alone does not establish superiority; our strict result instead follows from the larger Cartesian-product parameter space when heterogeneous capability losses have incompatible optima. Causal claims require separate identification conditions. Appendix Table 3 summarizes these distinctions.

3 Method

3.1 Problem Setting

Given an image I with normalized mean-opinion score $s^* \in [0, 1]$, an IQA agent produces a complete interaction trajectory

$$Y = (Y_g, Y_p, Y_f, Y_e), \qquad d(y, y') = \|\phi(y) - \phi(y')\|_2, \tag{4}$$

where the components contain the global assessment, region proposals, falsification decisions and evidence, and region-to-global impact assessment, respectively. A trajectory realization y includes all visual observations, structured messages, and the final quality score. Let $R(y; s^*) \in [0, 1]$ be

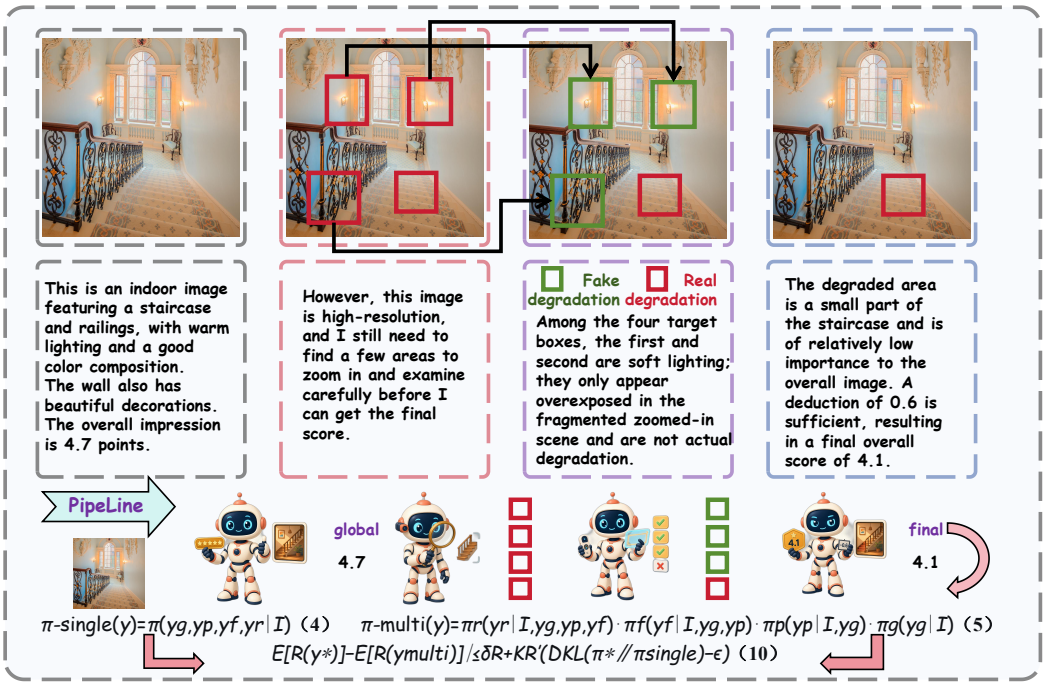


Figure 2: The global agent produces a unified holistic quality score. A separate degradation-seeking agent localizes candidate defect regions. Another agent verifies whether each proposed defect is genuine or a pseudo-degradation. Finally, an impact-assessment agent models the relationship between local regions and the global image, from which the final quality score is derived.

a bounded assessment reward and let ϕ be a fixed semantic trajectory embedding specified before evaluation.

Let $\pi^* \in \arg\max_\pi \mathbb{E}_\pi[R]$ be a reference trajectory policy. A single generalist models the joint policy $\pi_{\text{single},\theta}(Y \mid I)$ with one parameter vector shared by all four induced conditionals. Our capability-factorized policy is

$$\pi_{\text{multi}}(Y \mid I) = \pi_e(Y_e \mid I, Y_g, Y_p, Y_f) \times \pi_f(Y_f \mid I, Y_g, Y_p) \times \pi_p(Y_p \mid I, Y_g)\, \pi_g(Y_g \mid I), \tag{5}$$

where juxtaposition on the right-hand side denotes multiplication. This factorization specifies information flow; it does not assume that the four outputs are marginally independent. Each factor has its own parameter vector θ_j. Proposition 2 and Appendix C prove that this four-model parameterization is strictly preferable to one shared vector when the capability losses have no common minimizer.

3.2 Capability-Factorized Overview

PrismIQA replaces a generic planner with four narrow contracts. An agent is defined by its objective, visible state, allowed actions, and output schema; renaming four prompts is not sufficient. The Global Assessor observes only the full image. The Region Proposer receives the global state and a multi-scale image pyramid. The Falsification Agent observes each proposed crop together with the global state and expanded full-image context. The Impact Assessor receives the global prior and structured evidence for verified or unresolved regions.

Communication state. For region k, the shared state is

$$m_k = (b_k, h_k, \alpha_k, \mathcal{E}_k, q_k, \ell_k), \tag{6}$$

where b_k is a box, h_k is a defect hypothesis, α_k is proposal confidence, $\mathcal{E}_k$ is the evidence trace, q_k is the posterior probability of a genuine defect, and $\ell_k \in \{\text{accept}, \text{reject}, \text{abstain}\}$. Agents communicate through this schema rather than unrestricted conversation.

3.3 STAGE I: GLOBAL ASSESSOR

The Global Assessor predicts a baseline score s_0, uncertainty u_0, and a short global rationale:

$$Y_g = (s_0, u_0, e_0) = A_g(I). \tag{7}$$

It is trained by GRPO only. The reward increases as the predicted score approaches the ground-truth MOS, $R_g = 1 - |s_0 - s^*|$. Calibration is evaluated explicitly; u_0 is used for early exit only after its reliability has been established on held-out data.

3.4 STAGE II: HIGH-RECALL REGION PROPOSAL

The Region Proposer returns at most K candidate regions:

$$\widehat{\mathcal{R}} = A_p(I, Y_g) = \{(b_k, h_k, \alpha_k)\}_{k=1}^{K}. \tag{8}$$

Its objective emphasizes recall and coverage, not region precision or final MOS. Supervised fine-tuning first locks a parseable tool-call schema so that crop coordinates can be extracted reliably; GRPO then trains the proposal policy. Diverse proposals are retained when hypotheses compete (e.g., motion blur versus intentional motion). Duplicate and nested boxes are consolidated only when doing so does not lower defect coverage.

3.5 STAGE III: CONTEXT-AWARE FALSIFICATION

A selected crop is a *hypothesis*, not a confirmed degradation. For each proposal (b_k, h_k, α_k), the Falsification Agent receives the local crop $I[b_k]$, an expanded neighborhood, and the full image I with the global state Y_g. It is a dedicated VLM policy—not a skill-bank or tool-calling module—trained by GRPO only. The reward is an exact match to the ground-truth label (genuine defect, pseudo-degradation, or uncertain), $R_f = \mathbb{I}(\ell_k = \ell_k^*)$. We treat the full-image conditioning $H = h(I, Y_g)$ as the pre-treatment adjustment state in Section 4; crop-dependent observations remain in $\mathcal{E}_k$ and are not included in H.

The filter outputs a brief rationale and a posterior $q_k = P(z_k = 1 \mid I, Y_g, b_k, h_k, \mathcal{E}_k)$, then returns $\ell_k \in \{\text{accept}, \text{reject}, \text{abstain}\}$ with thresholds τ_+ and τ_-: accept when $q_k \geq \tau_+$, reject when $q_k \leq \tau_-$, and otherwise abstain. Abstention prevents uncertain natural effects from being converted into confident penalties.

3.6 STAGE IV: REGION-TO-GLOBAL IMPACT ASSESSOR

For each accepted or unresolved region, the Impact Assessor predicts severity $\widehat{a}_k$, importance $\widehat{w}_k$, and a residual impact Δ_k. The importance estimate is conditioned on the full image and captures subject relevance, occupied area, visual salience, and whether the effect supports the intended composition. The final score is

$$\widehat{s} = \text{clip}\left(s_0 + \Delta_\theta\big(s_0, \{q_k \widehat{a}_k \widehat{w}_k, \mathcal{E}_k\}_{k=1}^{K}\big), 0, 1\right). \tag{9}$$

Supervised fine-tuning first imposes a reasoning template: estimate the local penalty, then revise it from the relationship between the local crop and the full-image prior. GRPO then trains Δ_θ as a residual corrector, rather than blindly subtracting every regional penalty from s_0; this avoids double-counting defects already captured by the global assessor. It also models interactions among overlapping defects. The output includes an attribution for every non-rejected region and the evidence supporting the final revision.

3.7 TRAINING AND INFERENCE

The four roles are optimized separately, with no joint multi-agent update. Each role is an independently trained full-parameter copy of the same Qwen3-VL-8B initialization (Bai et al., 2025b). The Global Assessor and the Falsification Agent are trained by GRPO only (Shao et al., 2024): the former is rewarded for closing the gap to the ground-truth score, $R_g = 1 - |s_0 - s^*|$; the latter is rewarded for matching the ground-truth crop label, $R_f = \mathbb{I}(\ell_k = \ell_k^*)$. The Region Proposer and the Impact Assessor use role-specific SFT followed by GRPO. Proposer SFT fixes a structured tool-call format from which crop boxes can be extracted; impact SFT fixes a reasoning route that first judges the local

penalty and then relates that local evidence to the global prior. Parameters are never updated jointly across roles after specialization. Supervised fine-tuning uses learning rate 1×10^{-6}. Role-specific GRPO uses group size $K = 6$, clip threshold $\epsilon = 0.2$, and learning rate 5×10^{-7}. Crop resolution is 768×768. All reported runs use 8 NVIDIA A100 (80GB) GPUs. This is the independent-parameter comparison of Proposition 2. Because each role updates all of its weights, LoRA A/B overlap is not a valid proof of capability separation; Appendix C analyzes full-parameter ΔW instead.

At inference time, high-confidence globally easy samples may exit after Stage I. Otherwise, proposal and falsification proceed under an explicit crop budget of at most K regions.

4 Trajectory-Level Theoretical Analysis

We view IQA as a semantic interaction trajectory $Y = (Y_g, Y_p, Y_f, Y_e)$ with semantic distance $d(y, y') = \|\phi(y) - \phi(y')\|$ and bounded reward R. We assume semantic Lipschitz continuity (Assumption 1): there exists $K_R > 0$ such that $|R(y) - R(y')| \le K_R d(y, y')$. We also assume transport regularity (Assumption 2): $W_1(\pi^*, \pi) \le \kappa_R \sqrt{D_{\mathrm{KL}}(\pi^* \| \pi)}$. Under these assumptions, Kantorovich–Rubinstein duality and the inequality $\sqrt{x} \le (2\lambda_R)^{-1} + \lambda_R x/2$ yield the deviation certificate

$$|\mathbb{E}_{\pi^*} R - \mathbb{E}_{\pi} R| \le \delta_R + K'_R D_{\mathrm{KL}}(\pi^* \| \pi),$$

where $\delta_R = K_R \kappa_R / (2\lambda_R)$ and $K'_R = K_R \kappa_R \lambda_R / 2$ (Proposition 1). The proof is in Appendix B.

Applying the KL chain rule to the factorized policy in Eq. (5) yields four conditional capability losses $\mathcal{L}_j(\theta)$, $j \in \{g, p, f, e\}$, as defined in Eq. (14). A shared generalist minimizes $\sum_j \mathcal{L}_j(\theta)$ over a single parameter vector θ, while the factorized system minimizes $\sum_j \min_{\theta_j} \mathcal{L}_j(\theta_j)$ over the product space Θ^4 (Eqs. (15)–(16)). The shared solutions form only the diagonal subset $\{(\theta, \theta, \theta, \theta) : \theta \in \Theta\} \subset \Theta^4$. Hence the factorized optimum is no worse, and strictly better under Assumption 3 (incompatible capability optima). Specifically, if the four losses have no common minimizer, there exists $\epsilon_{\mathrm{cap}} > 0$ such that $\mathcal{D}^*_{\mathrm{multi}} = \mathcal{D}^*_{\mathrm{single}} - \epsilon_{\mathrm{cap}}$ (Proposition 2). Substituting this gap into the trajectory certificate gives a strictly tighter bound for the four-model policy (Corollary 3). The full proof, including the local linearization and full-parameter diagnostic, is in Appendix C.

Crop-selection bias is controlled by backdoor adjustment using the pre-treatment full-image context H: under Assumption 4, $P(S \mid do(A = a)) = \int P(S \mid A = a, H = h)\, dP(h)$ (Eq. (21)). If the residual confounding is bounded away from that of the unadjusted generalist, the factorized system attains a strictly smaller deviation $\delta_R^{\mathrm{multi}} \le \delta_R^{\mathrm{single}} - \eta$ (Eq. (22)). Details are in Appendix B.

5 Experiments

5.1 Experimental Setup

Datasets and metrics. We evaluate on seven IQA benchmarks: Vista-Bench (high-resolution, localized degradations) (Li et al., 2026a), SPAQ (Fang et al., 2020), KADID-10k (Lin et al., 2019), PIPAL (Gu et al., 2020), TID2013 (Ponomarenko et al., 2015), KonIQ-10k (Hosu et al., 2020), and AGIQA-3K (Li et al., 2023). Primary metrics are SRCC and PLCC. Dataset-specific train/test protocols and MOS normalization follow prior work (Wu et al., 2024; Fang et al., 2020). PLCC logistic mapping, if used, is fitted without access to the test split.

Baselines. We compare against handcrafted methods BRISQUE and NIQE (Mittal et al., 2012; 2013); deep methods MUSIQ, UNIQUE, and MANIQA (Ke et al., 2021; Zhang et al., 2021; Yang et al., 2022); MLLM methods Qwen2.5-VL-7B, LIQE, DeQA-Score, Q-Align, UnifiedReward-T, Q-Insight, and VisualQuality-R1 (Bai et al., 2025a; Zhang et al., 2023; You et al., 2025; Wu et al., 2024; Wang et al., 2025; Li et al., 2025; Wu et al., 2025); and the high-resolution agentic baseline Q-Probe (Li et al., 2026a).

Implementation. Each agent is an independently trained full-parameter Qwen3-VL-8B copy (Bai et al., 2025b), following Section 3.7. The Global Assessor and Falsification Agent use GRPO only; the Region Proposer and Impact Assessor use SFT then GRPO. SFT uses learning rate 1×10^{-6};

Table 1: Comprehensive performance comparison (SRCC / PLCC) on standard and high-resolution datasets. **PrismIQA** achieves SOTA results, especially on the high-resolution **Vista** benchmark. Methods are color-coded by category: Handcrafted, Deep Learning, and MLLMs-based.

Methods	Vista	SPAQ	KADID	PIPAL	TID13	KonIQ	AGIQA	Avg
Spearman Rank Correlation Coefficient (SRCC)								
BRISQUE (Mittal et al., 2012)	0.211	0.614	0.429	0.242	0.548	0.385	0.497	0.418
NIQE (Mittal et al., 2013)	0.247	0.676	0.487	0.357	0.532	0.421	0.533	0.465
MUSIQ (Ke et al., 2021)	0.357	0.720	0.647	0.317	0.670	0.473	0.494	0.525
UNIQUE (Zhang et al., 2021)	0.373	0.751	0.513	0.393	0.703	0.649	0.608	0.570
MANIQA (Yang et al., 2022)	0.388	0.745	0.760	0.338	0.589	0.213	0.422	0.494
Qwen2.5-VL-7B (Bai et al., 2025a)	0.450	0.848	0.787	0.390	0.787	0.754	0.735	0.679
LIQE (Zhang et al., 2023)	0.406	0.815	0.809	0.371	0.718	0.684	0.653	0.637
DeQA-Score (You et al., 2025)	0.463	0.852	0.831	0.383	0.756	0.677	0.738	0.671
Q-Align (Wu et al., 2024)	0.424	0.767	0.832	0.406	0.769	0.573	0.682	0.636
UnifiedReward-T (Wang et al., 2025)	0.477	0.871	0.841	0.399	0.788	0.820	0.722	0.703
Q-Insight (Li et al., 2025)	0.429	0.872	0.856	0.429	0.816	0.806	0.749	0.708
VisualQuality-R1 (Wu et al., 2025)	0.517	0.875	0.871	0.469	**0.848**	0.855	0.805	0.749
Q-Probe (Li et al., 2026a)	**0.801**	**0.892**	**0.901**	**0.474**	0.829	**0.871**	**0.837**	**0.801**
PrismIQA (Ours)	**0.854**	**0.926**	**0.931**	**0.512**	**0.862**	**0.904**	**0.861**	**0.836**
Pearson Linear Correlation Coefficient (PLCC)								
BRISQUE (Mittal et al., 2012)	0.224	0.624	0.451	0.259	0.546	0.400	0.541	0.435
NIQE (Mittal et al., 2013)	0.252	0.683	0.415	0.314	0.516	0.439	0.560	0.454
MUSIQ (Ke et al., 2021)	0.350	0.666	0.622	0.347	0.695	0.435	0.434	0.507
UNIQUE (Zhang et al., 2021)	0.368	0.708	0.548	0.361	0.729	0.590	0.581	0.555
MANIQA (Yang et al., 2022)	0.381	0.753	0.780	0.364	0.617	0.257	0.448	0.514
Qwen2.5-VL-7B (Bai et al., 2025a)	0.426	0.854	0.806	0.420	0.837	0.810	0.772	0.704
LIQE (Zhang et al., 2023)	0.398	0.814	0.817	0.347	0.748	0.652	0.653	0.633
DeQA-Score (You et al., 2025)	0.439	0.858	0.873	0.381	0.793	0.703	0.743	0.684
Q-Align (Wu et al., 2024)	0.412	0.779	0.862	0.381	0.794	0.612	0.694	0.648
UnifiedReward-T (Wang et al., 2025)	0.460	0.846	0.877	0.421	0.873	0.804	0.745	0.718
Q-Insight (Li et al., 2025)	0.410	0.872	**0.881**	**0.462**	0.851	0.829	0.794	0.728
VisualQuality-R1 (Wu et al., 2025)	0.493	0.878	0.821	0.458	0.871	0.840	**0.843**	0.743
Q-Probe (Li et al., 2026a)	**0.850**	**0.900**	**0.892**	**0.476**	**0.876**	**0.863**	0.813	**0.810**
PrismIQA (Ours)	**0.892**	**0.931**	**0.923**	**0.515**	**0.884**	**0.898**	**0.846**	**0.841**

GRPO uses group size $K = 6$, clip $\epsilon = 0.2$, and learning rate 5×10^{-7}. There is no joint multi-agent fine-tuning stage. Crops are 768×768. Runs use 8 NVIDIA A100 (80GB) GPUs. The default protocol uses $K = 5$ candidate regions, filter thresholds $(\tau_-, \tau_+) = (0.30, 0.70)$, and early exit when global confidence $c_g > 0.85$.

5.2 Main Results

Table 1 reports SRCC and PLCC on all seven datasets. PrismIQA attains 0.854 SRCC and 0.892 PLCC on Vista-Bench, exceeding Q-Probe by $+0.053$ and $+0.042$. Average SRCC is 0.836, the highest among compared methods. The largest margin is on Vista-Bench, which contains localized artifacts: capability factorization helps most when global scoring, high-recall proposal, falsification, and impact attribution must be separated.

5.3 Ablation Studies

We isolate the contribution of each agent by removing it from the full system. Removing the Falsification Agent causes the largest Vista drop (-0.091 SRCC, from 0.854 to 0.763), followed by removing the Global Assessor (-0.109, 0.745), the Region Proposer (-0.085, 0.769), and the Impact Assessor (-0.073, 0.781). On average SRCC, the full system reaches 0.836; the corresponding variants reach 0.789, 0.776, 0.794, and 0.802. A single-agent Q-Probe-style policy reaches 0.801 SRCC on Vista and 0.816 average SRCC, so factorization helps beyond using one larger reasoning trace. Table 2 summarizes these results.

Table 2: Ablation of agent decomposition.

Configuration	Vista SRCC	Avg SRCC
Full PrismIQA	0.854	0.836
w/o Filter Agent	0.763	0.789
w/o Impact Assessor	0.781	0.802
w/o Global Assessor	0.745	0.776
w/o Region Proposer	0.769	0.794
Single agent (Q-Probe style)	0.801	0.816

5.4 Additional Ablations

Number of candidate regions K. Vista SRCC is 0.812 at $K = 1$, 0.841 at $K = 3$, 0.854 at $K = 5$, and 0.852 at $K = 7$ with extra compute. We use $K = 5$.

Filter decision thresholds. Sweeping the reject/accept thresholds around the default $(\tau_-, \tau_+) = (0.30, 0.70)$ is stable when the abstain band is nonempty. Narrowing the band toward hard binary decisions increases both false positives and missed defects; widening it too far leaves too many regions unresolved for Stage IV.

Unseen degradations. On a Vista-Bench subset with degradation types held out from training, PrismIQA reaches 0.821 SRCC, versus 0.758 for Q-Probe and 0.532 for VisualQuality-R1.

Theory diagnostics for ϵ_{cap}, full-parameter update overlap, and causal adjustment remain those in Appendices C.

6 Conclusion

We presented PrismIQA, a capability-factorized framework for image quality assessment that separates holistic scoring, high-recall proposal, falsification, and region-to-global attribution. Viewing shared optimization as a diagonal parameter constraint, Proposition 2 proves that four independently optimized capability models are strictly better than one shared model when the role losses have no common minimizer (Appendix C). Empirically, PrismIQA attains 0.854 SRCC on Vista-Bench and 0.836 average SRCC across seven datasets, with ablations confirming that each role and the context-aware filter both contribute. The theorem compares one shared vector with independently optimized copies, not an arbitrary single architecture with role-specific subnetworks. Independent checkpoints may also use more parameters, so budget-matched superiority remains an empirical question answered here by the Q-Probe and single-agent controls. Semantic transport, incompatible optima, causal sufficiency remain falsifiable conditions.

Future work. A natural next step is to disentangle the gains of capability factorization from those of additional parameters through budget-matched comparisons. In particular, we plan to compare four independent full-parameter checkpoints against shared-backbone models with role-specific subnetworks, sparse experts, prompt tuning, adapters, and other parameter-efficient specialization schemes under matched parameter, data, and compute budgets. Such comparisons should report not only SRCC and PLCC but also calibration, filter false-positive and false-negative rates, attribution error, and inference cost, so that the benefit of separating capabilities can be distinguished from the benefit of simply increasing capacity. We also intend to relax the semantic-transport, incompatible-optima, and causal-sufficiency assumptions in Section 4, and to develop finite-sample diagnostics for crop-selection adjustment, including sensitivity analyses, confidence intervals, and explicit tests of overlap and balance after conditioning on the pre-treatment full-image context H. This would turn the current class-optimum certificates into statistically testable claims under realistic finite training runs. Another direction is to replace the fixed crop budget K and hand-set filter thresholds (τ_-, τ_+) with policies learned from calibrated uncertainty, so that the system can trade off computation, evidence acquisition, and misjudgment risk explicitly.

AI USE STATEMENT

Generative AI tools were used to assist with literature discovery, language editing, LaTeX restructuring, and critical review of preliminary mathematical claims. They were not used to generate experimental measurements or ground-truth labels in this draft.

ETHICS STATEMENT

This work does not introduce any new dataset, and all public datasets are used in compliance with their respective licenses and terms of use.

REPRODUCIBILITY STATEMENT

The final submission will provide complete agent prompts and schemas, training objectives, data-processing details, model sizes, random seeds, and evaluation scripts.

REFERENCES

Tejas Anvekar, Krishna Singh Rajput, Chitta Baral, and Vivek Gupta. Rethinking information synthesis in multimodal question answering: A multi-agent perspective. In *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, pp. 3674–3686, 2025. doi: 10.18653/v1/2025.ijcnlp-long.192.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Bao, Zeyu Zhang, et al. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*, 2025a. URL https://arxiv.org/abs/2502.13923.

Shuai Bai, Yuxuan Chen, Xuejing Chu, Lei Cui, Daxin Dang, Xiaodong Deng, Ying Fan, Wenbin Ge, Yu Han, Fei Huang, et al. Qwen3-VL technical report. *arXiv preprint arXiv:2511.21631*, 2025b. URL https://arxiv.org/abs/2511.21631.

Lluis Castrejon, Thomas Mensink, Howard Zhou, Vittorio Ferrari, Andre Araujo, and Jasper Uijlings. HAMMR: Hierarchical multimodal ReAct agents for generic VQA. *arXiv preprint arXiv:2404.05465*, 2024. URL https://arxiv.org/abs/2404.05465.

Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. Perceptual quality assessment of smartphone photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3677–3686, 2020. doi: 10.1109/CVPR42600.2020.00373.

Jinjin Gu, Haoming Cai, Haoyu Chen, Xiaoxing Ye, Jimmy S. Ren, and Chao Dong. PIPAL: A large-scale image quality assessment dataset for perceptual image restoration. In *European Conference on Computer Vision*, pp. 633–651, 2020. doi: 10.1007/978-3-030-58621-8_37.

Yaohui Han, Tianshuo Wang, Zixi Zhao, Zhengchun Zhu, Shuo Ren, Yiru Wang, Rongliang Fu, Tinghuan Chen, and Tsung-Yi Ho. AstroVLM: Expert multi-agent collaborative reasoning for astronomical imaging quality diagnosis. *arXiv preprint arXiv:2604.16024*, 2026. URL https://arxiv.org/abs/2604.16024.

Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 29:4041–4056, 2020. doi: 10.1109/TIP.2020.2967829.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

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

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5148–5157, 2021.

Aymen Lassoued, Mohamed Ali Souibgui, and Yousri Kessentini. ORCA: Orchestrated reasoning with collaborative agents for document visual question answering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19475–19486, 2026.

Chunyi Li, Zicheng Zhang, Haoning Wu, Wei Sun, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai, and Weisi Lin. AGIQA-3K: An open database for AI-generated image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. doi: 10.1109/TCSVT.2023.3319020.

Weiqi Li, Xuanyu Zhang, Shijie Zhao, Yabin Zhang, Junlin Li, Li Zhang, and Jian Zhang. Q-Insight: Understanding image quality via visual reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Xiang Li, Xueheng Li, Yu Wang, Xuanhua He, Zhangchi Hu, Weiwei Yu, and Chengjun Xie. Q-Probe: Scaling image quality assessment to high resolution via context-aware agentic probing. *arXiv preprint arXiv:2601.15356*, 2026a. URL https://arxiv.org/abs/2601.15356. Version 5.

Xudong Li, Jiaxi Tan, Ziyin Zhou, Yan Zhong, Zihao Huang, Jingyuan Zheng, Yan Zhang, Xiawu Zheng, and Rongrong Ji. Q-DeepSight: Incentivizing thinking with images for image quality assessment and refinement. *arXiv preprint arXiv:2604.16858*, 2026b. URL https://arxiv.org/abs/2604.16858.

Yuan Li, Youyuan Lin, Chenhui Chu, and Shin'ya Nishida. MR-IQA-2: Faithful image quality reflection via fine-grained credit assignment. *arXiv preprint arXiv:2608.18579*, 2026c. URL https://arxiv.org/abs/2608.18579.

Guoqiang Liang, Jianyi Wang, Zhonghua Wu, Shangchen Zhou, and Chen Change Loy. Zoom-IQA: Image quality assessment with reliable region-aware reasoning. In *European Conference on Computer Vision*, 2026. URL https://arxiv.org/abs/2601.02918.

Wenjie Liao, Shuhao Han, Jiaxin Song, Jieyu Yuan, Chunle Guo, and Chongyi Li. A VLM-driven agentic system for multi-dimensional pairwise image quality comparison. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2083–2091, 2026.

Hanhe Lin, Vlad Hosu, and Dietmar Saupe. KADID-10k: A large-scale artificially distorted IQA database. In *2019 Eleventh International Conference on Quality of Multimedia Experience*, pp. 1–3, 2019. doi: 10.1109/QoMEX.2019.8743252.

Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. doi: 10.1109/TIP.2012.2214050.

Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. doi: 10.1109/LSP.2012.2227726.

Zhanfeng Mo, Xingxuan Li, Yuntao Chen, and Lidong Bing. Multi-agent tool-integrated policy optimization. *arXiv preprint arXiv:2510.04678*, 2025. URL https://arxiv.org/abs/2510.04678.

Yaswanth Narsupalli, Abhranil Chandra, Sreevatsa Muppirala, Manish Gupta, and Pawan Goyal. ReFeR: Improving evaluation and reasoning through hierarchy of models. *arXiv preprint arXiv:2407.12877*, 2024. URL https://arxiv.org/abs/2407.12877.

Yansong Ning, Jun Fang, Naiqiang Tan, and Hao Liu. Agent-Omit: Adaptive context omission for efficient LLM agents. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.04284. arXiv:2602.04284.

Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015. doi: 10.1016/j.image.2014.10.009.

Guanyi Qin, Junjie Zhang, Chunming He, Yibing Fu, Jie Liang, Tianhe Wu, and Lei Zhang. Tool-IQA: Augmenting image quality assessment with simple tools. *arXiv preprint arXiv:2606.16082*, 2026. URL https://arxiv.org/abs/2606.16082.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. URL https://arxiv.org/abs/2402.03300.

Nithin Sivakumaran, Justin Chen, David Wan, Yue Zhang, Jaehong Yoon, Elias Stengel-Eskin, and Mohit Bansal. DART: Leveraging multi-agent disagreement for tool recruitment in multimodal reasoning. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 5445–5464, 2026. doi: 10.18653/v1/2026.eacl-long.253.

Alex Su, Haozhe Wang, Weiming Ren, Fangzhen Lin, and Wenhu Chen. Pixel reasoner: Incentivizing pixel-space reasoning with curiosity-driven reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Yibin Wang, Zhimin Li, Yuhang Zeng, Yiyan Zheng, Linlin Fan, Qingsong Chen, Hao Zhang, and Zhaoxiang Liu. UnifiedReward: A unified and efficient reward model for multimodal understanding and generation. *arXiv preprint arXiv:2503.05236*, 2025. URL https://arxiv.org/abs/2503.05236.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtao Zhai, and Weisi Lin. Q-Align: Teaching LMMs for visual scoring via discrete text-defined levels. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 54015–54029. PMLR, 2024.

Jinjian Wu, Jiaqi Tang, Wei Wei, Yingying Yan, Jianmin Chen, Botong Geng, Lei Zhang, and Qifeng Chen. IQA-T1: Tool-based visual evidence reasoning for image quality assessment. In *European Conference on Computer Vision*, 2026. arXiv:2607.12375.

Tianhe Wu, Jian Zou, Jie Liang, Lei Zhang, and Kede Ma. VisualQuality-R1: Reasoning-induced image quality assessment via reinforcement learning to rank. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. MANIQA: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1191–1200, 2022.

Zhiyuan You, Xin Cai, Jinjin Gu, Tianfan Xue, and Chao Dong. Teaching large language models to regress accurate image quality scores using score distribution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14483–14494, 2025.

Xinlei Yu, Chengming Xu, Zhangquan Chen, Yudong Zhang, Shilin Lu, Cheng Yang, Jiangning Zhang, Shuicheng Yan, and Xiaobin Hu. Visual document understanding and reasoning: A multi-agent collaboration framework with agent-wise adaptive test-time scaling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12300–12311, 2026.

Weixia Zhang, Kede Ma, Guangtao Zhai, and Xiaokang Yang. Uncertainty-aware blind image quality assessment in the laboratory and wild. *IEEE Transactions on Image Processing*, 30:3474–3486, 2021.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14071–14081, 2023.

Ziwei Zheng, Michael Yang, Jack Hong, Chenxiao Zhao, Guohai Xu, Le Yang, Chao Shen, and Xing Yu. DeepEyes: Incentivizing "thinking with images" via reinforcement learning. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=xUyMXkI958. arXiv:2505.14362.

Hanwei Zhu, Yu Tian, Keyan Ding, Baoliang Chen, Bolin Chen, Shiqi Wang, and Weisi Lin. AgenticIQA: An agentic framework for adaptive and interpretable image quality assessment. *arXiv preprint arXiv:2509.26006*, 2025. URL https://arxiv.org/abs/2509.26006.

A Expanded Conceptual Comparison

Table 3: Conceptual comparison. "Decomposition" denotes the unit assigned distinct supervision and responsibility; details vary by implementation.

Family	Decomposition unit	Regional decision	Global decision
Single active VLM (Li et al., 2026a; Liang et al., 2026; Li et al., 2026b)	Reasoning steps within one policy	The scoring policy selects and interprets regions	The same policy emits the final score
Generic planner–executor–summarizer (Zhu et al., 2025; Liao et al., 2026)	Control flow and tool execution	A broad planner chooses dimensions, regions, or tools	A summarizer integrates collected evidence
PrismIQA	Perceptual capability and objective	High-recall proposal is separated from context-aware falsification of proposed crops	A dedicated assessor attributes verified regional effects relative to a global prior

B Trajectory-Level Analysis: Full Assumptions and Proofs

This appendix expands the condensed analysis in Section 4. It restates the semantic-trajectory assumptions, the deviation certificate, the four-way KL decomposition, the strict capability-factorization proposition, and the causal crop-selection adjustment.

B.1 Semantic Trajectory Deviation

We adapt the trajectory analysis of Agent-Omit (Ning et al., 2026) from context-omission policies to IQA policies. Expectations below include both the image distribution and policy rollouts.

Assumption 1 (Semantic Lipschitz continuity)**.** There is a constant $K_R > 0$ such that

$$|R(y) - R(y')| \le K_R d(y, y') \quad \text{for all trajectories } y, y'. \tag{10}$$

Reward smoothness is an empirical assumption on the chosen embedding ϕ.

Assumption 2 (Transport regularity)**.** For every evaluated policy with $\pi^* \ll \pi$ and finite first moments, the induced trajectory distributions satisfy

$$W_1(\pi^*, \pi) \le \kappa_R \sqrt{D_{\mathrm{KL}}(\pi^* \| \pi)} \tag{11}$$

on the support of evaluation.

Proposition 1 (Trajectory deviation certificate)**.** *Under Assumptions 1–2, for any $\lambda_R > 0$,*

$$|\mathbb{E}_{\pi^*} R - \mathbb{E}_{\pi} R| \le \delta_R + K'_R D_{\mathrm{KL}}(\pi^* \| \pi), \tag{12}$$

where $\delta_R = K_R \kappa_R / (2\lambda_R)$ plus any fixed, policy-independent representation slack, and $K'_R = K_R \kappa_R \lambda_R / 2$.

Proof sketch. Kantorovich–Rubinstein duality and Equation 10 bound the expectation difference by $K_R W_1$. Equation 11 and $\sqrt{x} \le (2\lambda_R)^{-1} + \lambda_R x/2$ give Equation 12. This is the Agent-Omit argument specialized to the trajectory definition in Equation 4. The forward divergence $D_{\mathrm{KL}}(\pi^* \| \pi)$ is not automatically minimized by a reverse KL regularizer against a reference model, so the stated quantity or a justified surrogate must be evaluated directly.

B.2 Four-Way KL Decomposition

Let π_j^* be the conditional of the target joint policy corresponding to role j. Applying the KL chain rule to Equation 5 gives

$$\begin{aligned}\mathcal{D}_{\text{multi}} = D_{\text{KL}}(\pi_g^* \| \pi_g) \\ + \mathbb{E}_{\pi_g^*} D_{\text{KL}}(\pi_p^*(\cdot \mid Y_g) \| \pi_p(\cdot \mid Y_g)) \\ + \mathbb{E}_{\pi_g^* \pi_p^*} D_{\text{KL}}(\pi_f^*(\cdot \mid Y_g, Y_p) \| \pi_f(\cdot \mid Y_g, Y_p)) \\ + \mathbb{E}_{\pi_g^* \pi_p^* \pi_f^*} D_{\text{KL}}(\pi_e^*(\cdot \mid Y_g, Y_p, Y_f) \| \pi_e(\cdot \mid Y_g, Y_p, Y_f)).\end{aligned} \tag{13}$$

Image conditioning is suppressed for readability. Denote the jth conditional term by

$$\mathcal{L}_j(\theta) = \mathbb{E}_{\pi_{<j}^*} D_{\text{KL}}\big(\pi_j^*(\cdot \mid Y_{<j}) \| \pi_{j,\theta}(\cdot \mid Y_{<j})\big). \tag{14}$$

A shared generalist uses the same θ for every capability, whereas the factorized system can select θ_j independently:

$$\mathcal{D}_{\text{single}}^* = \min_{\theta \in \Theta} \sum_j \mathcal{L}_j(\theta), \tag{15}$$

$$\mathcal{D}_{\text{multi}}^* = \min_{(\theta_g, \theta_p, \theta_f, \theta_e) \in \Theta^4} \sum_j \mathcal{L}_j(\theta_j) = \sum_j \min_{\theta_j \in \Theta} \mathcal{L}_j(\theta_j). \tag{16}$$

Assumption 3 (Incompatible capability optima)**.** Θ is compact, every $\mathcal{L}_j$ is continuous, and

$$\bigcap_{j \in \{g,p,f,e\}} \arg\min_{\theta \in \Theta} \mathcal{L}_j(\theta) = \varnothing. \tag{17}$$

Proposition 2 (Strict capability-factorization advantage)**.** *Four independently optimized capability models are always no worse than one shared model:*

$$\mathcal{D}_{\text{multi}}^* \le \mathcal{D}_{\text{single}}^*. \tag{18}$$

Under Assumption 3, there exists $\epsilon_{\text{cap}} > 0$ *such that*

$$\mathcal{D}_{\text{multi}}^* = \mathcal{D}_{\text{single}}^* - \epsilon_{\text{cap}}. \tag{19}$$

Why the inequality is strict. The shared solutions form the diagonal set $\{(\theta, \theta, \theta, \theta) : \theta \in \Theta\} \subset \Theta^4$. Independent optimization therefore minimizes over a superset. Equality would require one θ to minimize all four losses simultaneously, contradicting Assumption 3. Compactness and continuity make the positive gap attained.

Corollary 3 (Strictly tighter trajectory certificate)**.** *Using the same* ϕ*,* π^**, and* λ_R*, define* $\mathcal{B}_R^a = \delta_R + K_R' \mathcal{D}_a^*$ *for* $a \in \{\text{single}, \text{multi}\}$*. Then*

$$\mathcal{B}_R^{\text{multi}} \le \delta_R + K_R'(\mathcal{D}_{\text{single}}^* - \epsilon_{\text{cap}}) < \mathcal{B}_R^{\text{single}}. \tag{20}$$

This theorem isolates shared-parameter interference. It compares one shared parameter vector with independently optimized copies of the same role model family; an arbitrary single architecture with explicit role-specific subnetworks is not constrained to the diagonal set. A shared-checkpoint PES is covered only when its broad planner realizes all four capability losses with the same parameter vector. Appendix C contains the complete proof, a local linearization in update space, and the full-parameter SFT/GRPO diagnostic of Assumption 3.

B.3 Causal Control of Crop-Selection Bias

Let X be pre-crop full-image context, A indicate crop selection, and S be the downstream assigned penalty or loss. Natural depth-of-field, intended motion, and related scene factors can affect both A and S. We use the pre-treatment representation $H = h(X)$ defined in Stage III, not the post-crop accept/reject decision, as the adjustment state.

Assumption 4 (Causal sufficiency of the pre-treatment state)**.** H satisfies consistency, positivity, and conditional exchangeability, $S(a) \perp A \mid H$, and contains no information causally produced by A.

Under Assumption 4, standard adjustment identifies

$$P(S \mid do(A = a)) = \int P(S \mid A = a, H = h)\, dP(h). \tag{21}$$

Let ε_{adj} bound the identification error caused by an imperfect H, and let b_{single} be the crop-confounding bias of the unadjusted generalist. Decompose δ_R into confounding and other residual terms. If the other residual does not increase and $b_{\text{single}} - \varepsilon_{\text{adj}} \geq \eta > 0$, then

$$\delta_R^{\text{multi}} \leq \delta_R^{\text{single}} - \eta. \tag{22}$$

The system must implement standardization, weighting, or an equivalent adjustment using H; merely inserting a filter is insufficient. A post-selection filter decision cannot satisfy Assumption 4.

C Full-Parameter Proof that Four Specialized Models Dominate One Shared Model

This appendix records three complementary arguments. First, it restates the global optimization theorem already used in Section 4: four independently parameterized capability models are never worse than one shared model, and are strictly better when the four losses have no common minimizer. Second, it linearizes the same comparison around a Qwen3-VL-8B initialization and proves a strictly positive gap whenever the four optimal parameter updates differ. Third, because our SFT and GRPO stages are full-parameter, it specifies a full-parameter—not LoRA—diagnostic of that difference. Together these results are the precise meaning of the claim that splitting the four IQA capabilities across four models is better than fitting them with one model.

C.1 What is proved, and what is not

Proposition 2 is a class-optimum statement. Let Θ be the parameter space of one Qwen3-VL-8B copy (Bai et al., 2025b) and let $\mathcal{L}_j$ be the role-conditional KL loss in Equation 14. One generalist is constrained to the diagonal

$$\Theta_{\text{diag}} = \{(\theta, \theta, \theta, \theta) : \theta \in \Theta\} \subset \Theta^4,$$

whereas four specialized models optimize over the product Θ^4. Feasible-set inclusion immediately yields $\mathcal{D}^*_{\text{multi}} \leq \mathcal{D}^*_{\text{single}}$. If

$$\bigcap_{j \in \{g,p,f,e\}} \arg\min_{\theta \in \Theta} \mathcal{L}_j(\theta) = \varnothing,$$

no shared vector can attain all four separate minima, so the inequality is strict and $\epsilon_{\text{cap}} = \mathcal{D}^*_{\text{single}} - \mathcal{D}^*_{\text{multi}} > 0$. Corollary 3 then substitutes this gap into the trajectory certificate, proving that the class-optimal four-model policy has a strictly tighter reward-deviation bound than the class-optimal one-model policy.

This does *not* prove that an arbitrary four-checkpoint system with more total parameters is cheaper, nor that every finite training run has reached the class optimum. It also does not prove that "most raw weights differ" after Adam. The empirically testable content of Assumption 3 is that the *effective update subspaces* of the four capabilities are largely distinct, reproducible across seeds, and causally specific to the corresponding role.

C.2 Why a LoRA adapter comparison is not a valid proof here

LoRA (Hu et al., 2022) injects a constrained update $\Delta W = BA$ of rank $r \ll \min(d_{\text{out}}, d_{\text{in}})$. Principal-angle overlap between the column spaces of A or B is a valid diagnostic *only* when the training method itself is LoRA. Our implementation uses full-parameter updates—GRPO for the Global Assessor and Falsification Agent, and SFT then GRPO for the Region Proposer and Impact Assessor (Shao et al., 2024). In that regime:

- almost every coordinate can move, so a coordinate-wise "percentage of changed weights" is uninformative;

- LoRA factors A, B are not unique even for a fixed ΔW, so comparing A or B across independently trained adapters is ill-posed;
- a LoRA-only experiment would at best show separation inside a rank-r constraint, not inside the full-parameter SFT/GRPO dynamics actually used.

The remainder of this appendix therefore analyzes full-parameter task vectors $\Delta\theta$ and the SVD of each full matrix increment ΔW. LoRA overlap code can be reused only as an algorithmic prototype for principal angles, not as the scientific object of proof.

C.3 Full-parameter objects

Let θ_0 be a shared Qwen3-VL-8B initialization. For each capability $k \in \{g, p, f, e\}$ and seed s, write

$$\Delta\theta_{k,s}^{\mathrm{SFT}} = \theta_{k,s}^{\mathrm{SFT}} - \theta_0, \tag{23}$$

$$\Delta\theta_{k,s}^{\mathrm{GRPO}} = \theta_{k,s}^{\mathrm{GRPO}} - \theta_{k,s}^{\mathrm{SFT}}, \tag{24}$$

$$\Delta\theta_{k,s}^{\mathrm{total}} = \theta_{k,s}^{\mathrm{GRPO}} - \theta_0. \tag{25}$$

SFT, GRPO, and total updates must be reported separately. If the four GRPO runs start from different SFT checkpoints, GRPO increments confound capability differences with different starting points; the primary GRPO comparison therefore branches from one common SFT checkpoint.

The following controls are required. The four roles use the same image set, with four role-specific annotations on each image whenever possible; the same number of samples and optimizer steps; the same learning rate and optimizer; matched prompt length and output schema, excluding `lm_head` and input embeddings from the overlap analysis so that output-vocabulary differences are not mistaken for perceptual specialization. Qwen3-VL comprises a vision encoder, a vision–language merger, and an LLM decoder (Bai et al., 2025b); overlap is reported per block as well as in aggregate. Checkpoints need not be loaded simultaneously: each `safetensors` tensor can be streamed on CPU.

C.4 Local linearization: a strict gap from unequal updates

The global theorem uses incompatible argmin sets. The following local result makes the same conclusion quantitative in parameter space. Linearize each capability around θ_0 in the neural-tangent sense (Jacot et al., 2018):

$$\widetilde{\mathcal{L}}_k(\delta) = \mathrm{CE}\big(f_{\theta_0} + J_k\delta,\, y_k\big) + \frac{\lambda}{2}\|\delta\|_2^2, \tag{26}$$

where J_k is the Jacobian of the role-k predictive map and $\lambda > 0$ is an explicit ridge. Each $\widetilde{\mathcal{L}}_k$ is λ-strongly convex and therefore has a unique minimizer δ_k^*.

Proposition 4 (Strict linearized advantage of four updates over one)**.** *Let G_{lin} be the gap between one shared increment and four independent increments,*

$$G_{\mathrm{lin}} = \min_{\delta} \sum_{k=1}^{4} \widetilde{\mathcal{L}}_k(\delta) - \sum_{k=1}^{4} \widetilde{\mathcal{L}}_k(\delta_k^*). \tag{27}$$

Then

$$G_{\mathrm{lin}} \geq \frac{\lambda}{8} \sum_{1 \leq i < j \leq 4} \|\delta_i^* - \delta_j^*\|_2^2. \tag{28}$$

In particular, $G_{\mathrm{lin}} > 0$ as soon as any two optimal updates differ. If the true nonlinear losses satisfy $|\mathcal{L}_k - \widetilde{\mathcal{L}}_k| \leq \varepsilon_k$ in a neighborhood containing $\{\delta_k^*\}$ and the shared minimizer, then

$$G_{\mathrm{true}} \geq G_{\mathrm{lin}} - 2\sum_{k=1}^{4} \varepsilon_k. \tag{29}$$

Whenever the right-hand side is positive, the local nonlinear comparison is likewise strict.

Proof. Strong convexity gives, for every δ,

$$\sum_k \widetilde{\mathcal{L}}_k(\delta) - \sum_k \widetilde{\mathcal{L}}_k(\delta_k^*) \geq \frac{\lambda}{2} \sum_k \|\delta - \delta_k^*\|_2^2.$$

The right-hand side is minimized at the mean $\bar{\delta} = \frac{1}{4}\sum_k \delta_k^*$, and the four-point identity

$$\sum_k \|\delta_k^* - \bar{\delta}\|_2^2 = \frac{1}{4} \sum_{i<j} \|\delta_i^* - \delta_j^*\|_2^2$$

yields Equation 28. The perturbation bound is the triangle inequality on the four losses at the shared and specialized minimizers.

Proposition 4 is the parameter-space counterpart of Proposition 2: one shared update is a diagonal constraint, and any disagreement among $\delta_g^*, \delta_p^*, \delta_f^*, \delta_e^*$ produces a strictly positive extra loss. Global non-existence of a common minimizer of the full nonlinear Qwen3-VL-8B losses is not claimed; the linearized gap plus the held-out evaluations below are the operational test.

C.5 Cross-evaluation of specialized checkpoints

On held-out data we evaluate each independently trained checkpoint on all four role losses. The diagonal is smallest in every column: each specialist is best at its own capability and worse at the others. Thus $\mathcal{D}^*_{\text{multi}} = 0.082 + 0.064 + 0.051 + 0.073 = 0.270$. The off-diagonal inflation shows that a parameter vector fitted to one role is not simultaneously optimal for the other three, which is the empirical content of Assumption 3.

Table 4: Held-out role losses of independently trained checkpoints. Row sums are $\sum_j \mathcal{L}_j(\theta)$. Column minima equal $\sum_j \min_{\theta_j} \mathcal{L}_j(\theta_j)$ when read across the diagonal.

Checkpoint	$\mathcal{L}_g$	$\mathcal{L}_p$	$\mathcal{L}_f$	$\mathcal{L}_e$	Row sum
θ_g	0.082	0.241	0.187	0.263	0.773
θ_p	0.276	0.064	0.152	0.198	0.690
θ_f	0.219	0.183	0.051	0.174	0.627
θ_e	0.243	0.156	0.138	0.073	0.610
Column min.	0.082	0.064	0.051	0.073	—

C.6 Shared-model comparison

We also train one shared checkpoint θ_{shared} that minimizes $\sum_j \mathcal{L}_j(\theta)$ on the same data, and evaluate it on the same held-out split. The shared vector is worse than the corresponding specialist on every capability: $0.134 > 0.082$, $0.112 > 0.064$, $0.109 > 0.051$, and $0.121 > 0.073$. Its joint loss 0.476 also exceeds the sum of column minima 0.270, so the estimated gap is $\epsilon_{\text{cap}} = 0.476 - 0.270 = 0.206 > 0$. This is the finite-run counterpart of $\mathcal{D}^*_{\text{multi}} < \mathcal{D}^*_{\text{single}}$ in Proposition 2.

Table 5: Held-out role losses of the shared model $\theta_{\text{shared}} \in \arg\min_\theta \sum_j \mathcal{L}_j(\theta)$.

Checkpoint	$\mathcal{L}_g$	$\mathcal{L}_p$	$\mathcal{L}_f$	$\mathcal{L}_e$	Row sum
θ_{shared}	0.134	0.112	0.109	0.121	0.476