PCPD: Partial Complementary Policy Distillation for Multi-Dimensional AIGC Image Quality Assessment via Self-Evolving Skills

Anonymous ACL submission

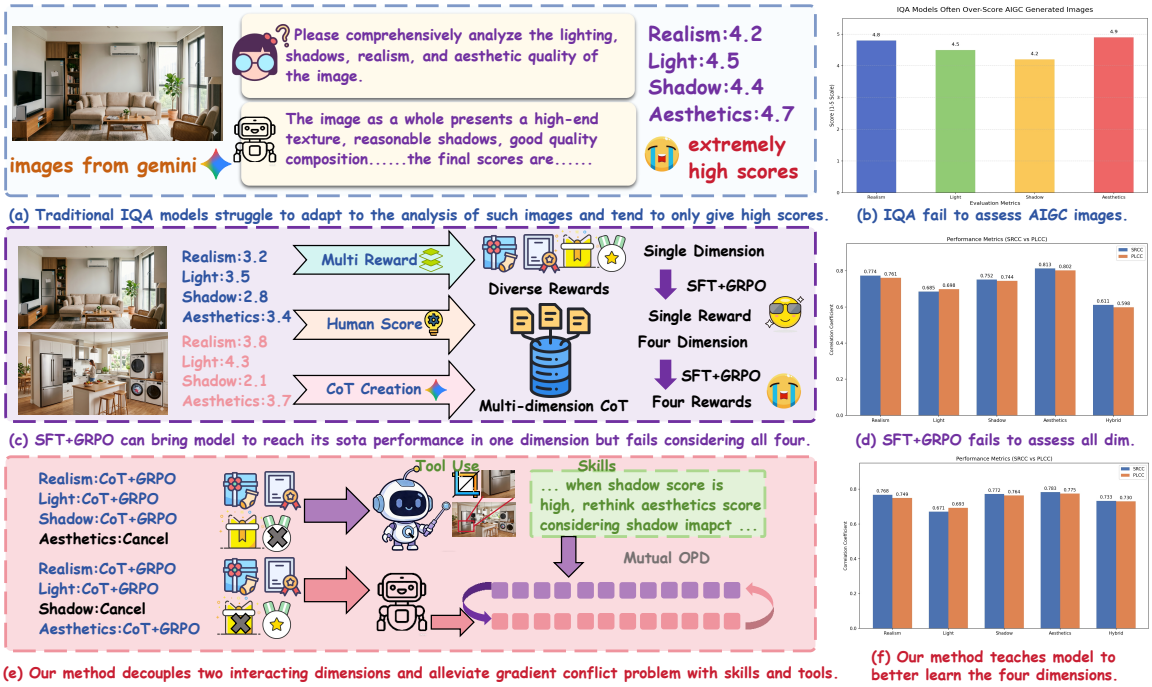

Figure 1: (a) and (b): Traditional IQA models struggle to differentiate the quality of high-fidelity AIGC images, typically outputting uniformly inflated scores. (c) and (d): Using human-annotated scores for AIGC images under the SFT + GRPO learning paradigm leads to gradient conflicts across multiple training dimensions. (e) and (f): Partial Complementary Policy Distillation: Our proposed method, PCPD, decouples these conflicting dimensions into two complementary groups. It utilizes skills and tools to resolve dimensional conflicts and performs mutual distillation to achieve optimal performance.

Abstract

Currently, AI generated content (AIGC) for image generation has achieved impressive results. Since the training process of AIGC models often requires evaluating the quality of the generated images, a high performance Image Quality Assessment (IQA) model is urgently needed. However, due to the exceptionally high quality of current AIGC images, IQA models trained based on degradation injection levels tend to yield spuriously high scores, making it difficult to accurately assess AIGC images. To address this issue, the common practice is to construct dedicated datasets of AIGC images with human-annotated quality scores and train IQA models on them. Nevertheless, the evaluation metrics for AIGC generated images are inherently multi-dimensional. Relying solely on Supervised Fine Tuning and Reinforcement Learning for model training often leads to conflicting training gradients across different dimensions, preventing the model from achieving optimal performance. To address this, we propose Partial Complementary Policy Distillation (**PCPD**). This approach partitions the multiple dimensions into two mutually overlapping subsets. By leveraging tool calling and skill self-evolution, we guide the model to induce solutions for resolving these dimensional conflicts, ultimately internalizing these solutions

into the model parameters. Extensive experiments demonstrate that our model achieves state-of-the-art (SOTA) performance on AIGC image evaluation tasks compared to other IQA models, effectively resolving the issue of dimensional conflicts.

1 Introduction

As image generation technology evolves from traditional painting and photography to Artificial Intelligence Generated Content (AIGC), the underlying logic of Image Quality Assessment (IQA) (Fang et al., 2020; Saha et al., 2023) has undergone a fundamental transformation. The evaluation standards for modern AIGC images have expanded from merely identifying low-level degradations (such as noise and blur) to analyzing the multi-dimensional comprehensive impressions conveyed by the images, encompassing high-level semantic attributes including aesthetics, lighting, shadow, and physical realism. In this process, high-level IQA models are particularly crucial. IQA plays a dual core role in AIGC: on the one hand, it is used to filter out low-quality generated content; on the other hand, it serves as a reward signal for human preference alignment in the post-training of generative models.

Driven by this demand, IQA methods have experienced rapid development from traditional handcrafted features (Mittal et al., 2012a,b) to deep learning (Ke et al., 2021; Yang et al., 2022), and subsequently to Multimodal Large Language Models (MLLMs/VLMs) (Zhang et al., 2023; Wu et al., 2023). MLLMs, in particular, effectively integrate low-level distortions with high-level semantic attributes through the Chain of Thought (CoT) mechanism, significantly enhancing generalization capabilities (Wu et al., 2024; You et al., 2025). Current research is further shifting towards utilizing Reinforcement Learning (RL) to align with human perception (Cai et al., 2025; Zhao et al., 2025), overcoming the bottlenecks of traditional Supervised Fine-Tuning (SFT) limited by high annotation costs and insufficient reasoning depth. Emerging RL paradigms based on regression (e.g., Q-Insight (Li et al., 2025), pursuing absolute precision alignment) and comparison (e.g., VisualQuality-R1 (Wu et al., 2025), pursuing statistical alignment of preference ranking) attempt to achieve strong generalization and deep reasoning within an interpretable framework. However, due to the exceptionally high quality of current AIGC images, existing IQA models trained on degradation injection tend to assign uniformly high scores, facing the problem of "score saturation." They struggle to perform fine-grained differentiation among high-quality images with similar perceptual quality. More critically, while AIGC image quality assessment is inherently multi-dimensional, traditional models fail to analyze these dimensions independently, often resulting in score convergence across different dimensions.

To address these challenges, the prevailing approach in the industry is to construct high-quality AIGC datasets annotated with multi-dimensional scores and adopt a two-stage training paradigm of SFT $\rightarrow$ GRPO. This involves mixing multi-dimensional query data and performing reinforcement learning using multi-dimensional rewards. However, this implicit mixed learning leads to severe Dimensionality Conflict, where different capabilities point towards different optimization directions, making it difficult for a single training process to simultaneously advance performance across all dimensions. Through in-depth analysis during training, we find that while lighting and realism dimensions can harmoniously coexist with other dimensions, an irreconcilable learning contradiction exists between Aesthetics and Shadow: the more accurate the aesthetic assessment is, the more inaccurate the physical assessment of shadows tends to be, and vice versa. Details can be found in Fig1.

In-depth empirical and bad case analyses reveal that models fall into two typical dilemmas of "Attention Hijacking" during multi-dimensional joint evaluation:

- **Local-to-Global Masking:** When shadows in an image exhibit excellent low-level features (such as edge sharpness and texture), the model develops local feature dependence. Even if these high-quality shadows are incorrectly cast onto indoor sofas or refrigerators, severely disrupting the composition and semantic rationality of the image, the model is deceived by the excellent texture of the shadows themselves. Trapped in a mindset that "the shadow quality is excellent," it consequently ignores the disruption to the global structure during aesthetic analysis.

- **Global-to-Local Halo Effect:** Conversely, when an image presents a highly appealing global aesthetic atmosphere, the strong stylistic prior inversely hijacks the local reasoning

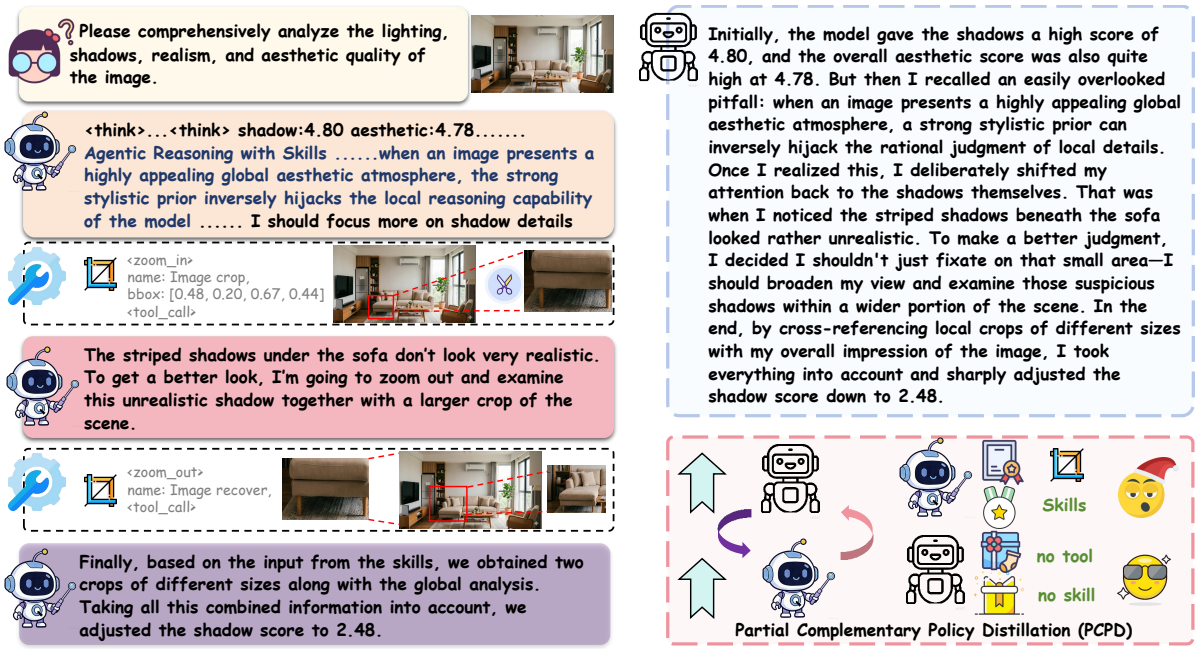

Figure 2: The model on the left side of the figure underwent SFT (Supervised Fine-Tuning) and reward-based GRPO (Group Relative Policy Optimization) using only three corresponding types of data: realism, lighting, and shadows. The model on the right, however, underwent SFT and reward-based GRPO using corresponding data for realism, lighting, and aesthetics. During the Partial Complementary Policy Distillation process, the models learn from each other. Furthermore, the left model utilizes skills and tools to explicitly resolve the conflict between shadows and aesthetics. Ultimately, the right model internalizes these capabilities into its parameters, saving resources and improving efficiency.

capability of the model. Under this "Halo Effect," as long as the shadows match the overall tone of the image in color and atmosphere, the model abandons rigorous analysis of their physical realism and geometric optical consistency, resulting in high scores even for erroneous shadows that severely violate physical laws.

Such receptive field competition at the spatial scale, where aesthetics require a broad global prior while shadows demand precise local high-frequency coordination, leads to severe mutual gradient interference. To this end, we propose the Partial Complementary Policy Distillation (PCPD) framework and introduce an Agentic reasoning mechanism based on Self-evolving Visual Skills. We split the conflicting dimensions into two subsets with overlapping items ("Aesthetics, Lighting, Realism" and "Shadow, Lighting, Realism") and train two expert models independently.

On the side of the expert model inclined towards shadow analysis, we abandon the practice of allowing the model to implicitly fuse conflicting features in a single forward pass, and instead adopt the "Thinking with Images" paradigm (OpenAI, 2025). We equip the model with a `Zoom-in` tool and encapsulate it as a conditionally triggered self-evolving skill. This effectively prevents the halo effect, blocking aesthetics from interfering with shadow evaluation.

Meanwhile, the expert model inclined towards aesthetic analysis focuses on global aesthetics to prevent the Masking Effect. Even if the shadow structure and physical properties are exceptionally good, it will still highlight their negative impact on the overall aesthetics.

By decoupling the originally mixed evaluation tasks into two complementary sequential logical chains, the two models learns to break the aforementioned "Masking Effect" and "Halo Effect." Ultimately, through On-policy Distillation, the Unified Policy Model successfully internalizes the structured solutions and parameters of this "global-local" dynamic switching.

The main contributions of this paper are as follows:

- **Proposing Partial Complementary Policy Distillation (PCPD):** This is the first distillation method specifically designed for the decoupling and fusion of multi-dimensional

capabilities in IQA. By orthogonally splitting mutually exclusive dimensions for training and eventually merging them, it fundamentally resolves the problem of gradient conflict in multi-dimensional joint optimization.

- **Introducing Agentic Self-evolving Skills Targeting Cognitive Biases:** We reveal the "Halo Effect" and "Masking Effect" brought about by different receptive fields between aesthetics and shadow in IQA tasks, and pioneeringly introduce self-evolving skills such as local `Zoom-in` during the distillation process, transforming spatial feature collisions into sequential logical reasoning.

- **Achieving SOTA Performance in Multi-dimensional AIGC-IQA:** Extensive experiments verify that our unified model not only effectively solves the problem of dimensional repulsion, but also outperforms existing IQA baseline models across multiple fine-grained evaluation dimensions, demonstrating exceptional fine-grained assessment capabilities.

2 Related Work

2.1 Image Quality Assessment

Image Quality Assessment (IQA) has evolved from handcrafted features (Mittal et al., 2012a,b) to deep learning models, where recent Multimodal Large Language Models (MLLMs) shift the paradigm via Supervised Fine-Tuning (SFT) methods like Q-Align (Wu et al., 2023) and DeQA-Score (You et al., 2025). To overcome SFT's reliance on costly annotations and shallow reasoning, Reinforcement Learning (RL) has emerged to align MLLMs with human preferences, as seen in Q-Insight's (Li et al., 2025) GRPO application and Visual Quality-R1's (Wu et al., 2025) RL-to-Rank. Concurrently, domain-specific methods like SA-IQA (Gao and Song, 2025) assess complex spatial aesthetics by fusing multi-dimensional ratings (e.g., layout, harmony, lighting, distortion). However, entangling these heterogeneous capabilities within a unified training process forces a single model to balance global semantics with fine-grained local anomaly detection. This causes severe feature interference and capacity bottlenecks, restricting the model to a coarse-grained view that fails to perform the deep localized scrutiny necessary for high-resolution image evaluation.

2.2 Thinking with Images

To overcome superficial perception limitations, the Thinking with Images paradigm (OpenAI, 2025) has inspired agentic visual reasoning MLLMs. While DeepEyes (Zheng et al., 2025) demonstrates that end-to-end RL can incentivize fine-grained visual analysis, pure RL struggles with complex interactions; thus, Pixel Reasoner (Wang et al., 2025a) utilizes a two-phase approach to overcome tool-use learning traps. Recently, Q-Probe (Li et al., 2026) adapted this agentic probing for high-resolution IQA, using context-aware cropping to mitigate the spurious correlation that zooming implies degradation. However, Q-Probe remains constrained to low-level technical degradations and suffers from "Attention Hijacking" during multi-dimensional AIGC evaluation, as a shared parameter space struggles to balance conflicting objectives like broad aesthetic appreciation and physical anomaly detection. In contrast, our PCPD framework physically decouples these conflicting dimensions into specialized expert models via Partial Complementary Policy Distillation, structurally isolating spatial feature collisions to avoid the capacity bottlenecks and optimization pitfalls inherent in single-policy agentic models.

3 Method

To address the inherent dimensional conflicts in AIGC image quality assessment—specifically the gradient interference between global aesthetics and local physical constraints (e.g., shadows)—we propose the Partial Complementary Policy Distillation (PCPD) framework. As illustrated in this section, PCPD decouples the evaluation process into two specialized branches before fusing them into a unified, high-performance policy.

3.1 Agentic Visual Reasoning via Zoom-in and Zoom-out

Human experts evaluate complex image quality by dynamically adjusting their receptive fields: observing the global composition first, scrutinizing local details, and then returning to the global context to verify physical consistency. To simulate this, we introduce an agentic "Thinking with Images" mechanism for the evaluation of physical realism.

We equip the model with parameterized visual tools: `<zoom_in(bbox)>` and `<zoom_out()>`. When the model processes the multi-dimensional

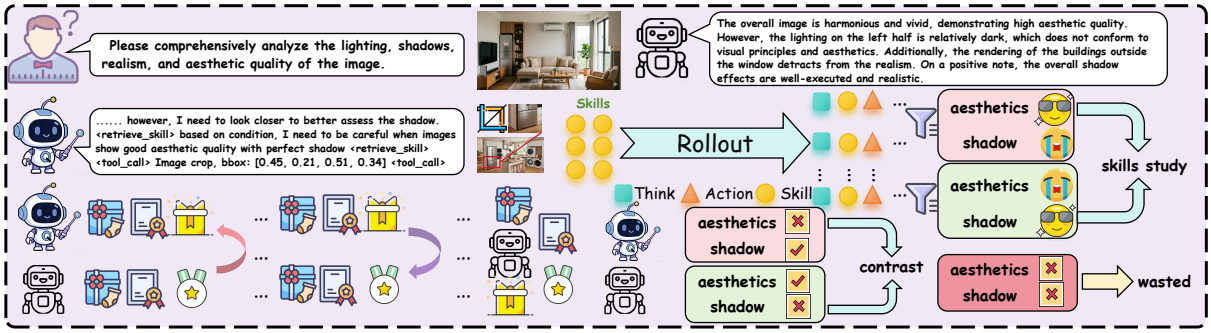

Figure 3: The model responsible for realism, lighting, and shadows extracts skills during the GRPO training process. Specifically, when the aesthetic analysis is accurate but the shadow analysis is inaccurate—or vice versa—it conducts an in-depth analysis of the specific trajectories to identify conflicts, subsequently summarizing these insights into skills.

evaluation prompt, it can autonomously output the <zoom_in> token along with bounding box coordinates to crop and magnify suspicious shadow regions. Crucially, evaluating a shadow in isolation is insufficient for determining its physical rationality. Therefore, after local scrutiny, the model is guided to utilize the <zoom_out()> tool. This action restores the global visual context, allowing the model to locate the physical entity casting the shadow and the corresponding light source. By explicitly chaining these actions, the model establishes a rigorous causal reasoning path: *Global Overview* $\rightarrow$ *Local Shadow Scrutiny* $\rightarrow$ *Global Entity/Light Source Verification*. This effectively breaks the "Global-to-Local Halo Effect," where a globally aesthetic image might otherwise mask physically impossible local shadows.

3.2 Self-Evolving Skills via GRPO Reflection

During the Group Relative Policy Optimization (GRPO) phase, we implement a self-reflection mechanism that enables the model to autonomously summarize and evolve evaluation "skills", which is shown in Fig 3.

For a given AIGC image, the model generates a group of reasoning trajectories and corresponding quality scores across four dimensions: Aesthetics (S_A), Lighting (S_L), Shadow (S_S), and Realism (S_R). We design an automated discrepancy detector that monitors the variance between specific dimensional scores within the generated group. Mathematically, for the conflicting dimensions of Aesthetics and Shadow, the discrepancy detector calculates the cross-dimensional penalty $\mathcal{D}$:

$$\mathcal{D} = |S_A - S_S| \cdot \mathbb{I}\big(\max(S_A, S_S) > \tau_{high}\big) \quad (1)$$

where $\mathbb{I}(\cdot)$ is the indicator function and τ_{high} is a predefined confidence threshold. When $\mathcal{D} > \delta$ (indicating a severe conflict, e.g., accurate shadow evaluation but massive discrepancy with the aesthetic score), the training pipeline pauses the standard parameter update and triggers a *Reflection Prompt*. The model is fed its own conflicting trajectories and asked to explicitly articulate why its aesthetic judgment blinded its physical judgment (or vice versa). The model then outputs a summarized rule (e.g., *"Skill 04: Do not let vibrant neon color palettes distract from checking the angular consistency of ground shadows"*).

These summarized rules are accumulated into a dynamic "Skill Bank." In subsequent training iterations, relevant skills are retrieved and prepended to the model's prompt. During the GRPO phase, the policy π_θ is optimized by maximizing the relative advantage across the sampled group of G reasoning trajectories.

$$\mathcal{J}_{GRPO}(\theta) = \mathbb{E}_{q,\{o_i\}}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\min\Big(\rho_i(\theta)\hat{A}_i, \mathrm{clip}\big(\rho_i(\theta), 1-\epsilon, 1+\epsilon\big)\hat{A}_i\Big) - \beta\mathbb{D}_{KL}\Bigg] \quad (2)$$

where $\rho_i(\theta) = \frac{\pi_\theta(o_i|q)}{\pi_{ref}(o_i|q)}$ is the probability ratio, and $\hat{A}_i$ is the group-relative advantage calculated by normalizing the rewards within the group. This

allows the model's evaluation capability to self-evolve and structurally mitigate cognitive biases.

3.3 Mutual On-Policy Complementary Distillation

To fundamentally resolve the gradient conflicts during training, we initialize two identical MLLMs but diverge their optimization paths into distinct specializations.

Model α (Agentic Expert): This model is granted full access to the <zoom_in>/<zoom_out> tools and the self-evolved Skill Bank. During its GRPO training, its reward function is explicitly formulated based on objective physical metrics, specializing in *Shadow*, *Realism*, and *Lighting* verification. **Model β (Global Expert):** This model operates as a standard end-to-end evaluator, relying entirely on a single-pass global image view without explicit tools. Its GRPO reward is driven by human preference alignment, specializing in *Aesthetics*, *Realism*, and *Lighting*.

Crucially, the input prompts for *both* models strictly require them to generate comprehensive reasoning and scores for *all four dimensions* (Aesthetics, Lighting, Shadow, Realism). Consequently, although Model α excels at localized physical verification such as shadows, it struggles with global Aesthetic coherence. Conversely, while Model β is highly adept at capturing holistic aesthetic harmony, its lack of zoom capabilities makes it struggle with local Shadow verification.

To bridge these complementary capability gaps without inducing gradient conflicts, we employ a mutual on-policy distillation (OPD) process. Instead of forcing one model to passively mimic the other's fixed trajectories (Off-Policy), the model being distilled (the student) generates its own on-policy reasoning trajectory. The other model (the teacher) then provides step-by-step, token-level supervision based on the states the student actually visits.

Specifically, Model α acts as the teacher to refine Model β's understanding of *Shadows*, while Model β acts as the teacher to refine Model α's understanding of *Aesthetics*. We introduce domain-specific binary masks, $M_{shadow}(t)$ and $M_{aesthetic}(t)$, to activate the distillation loss exclusively during the generation of the respective evaluation dimensions.

When Model β is the student learning *Shadow* reasoning, it samples an on-policy trajectory $y \sim \pi_\beta(\cdot|I, q)$. We minimize the reverse KL-divergence guided by Model α:

$$\mathcal{L}_{OPD}(\theta_\beta) = \mathbb{E}_{I,q,y\sim\pi_\beta}\Bigg[\frac{1}{|y|}\sum_{t=1}^{|y|} M_{shadow}(t) \times D_{KL}\Big(\pi_\beta(\cdot|I,q,y_{<t};\theta_\beta) \parallel \pi_\alpha(\cdot|I,q,y_{<t})\Big)\Bigg] \quad (3)$$

Conversely, when Model α is the student learning *Aesthetics*, it samples its own trajectory $y \sim \pi_\alpha(\cdot|I, q)$, and minimizes the reverse KL-divergence guided by Model β:

$$\mathcal{L}_{OPD}(\theta_\alpha) = \mathbb{E}_{I,q,y\sim\pi_\alpha}\Bigg[\frac{1}{|y|}\sum_{t=1}^{|y|} M_{aesthetic}(t) \times D_{KL}\Big(\pi_\alpha(\cdot|I,q,y_{<t};\theta_\alpha) \parallel \pi_\beta(\cdot|I,q,y_{<t})\Big)\Bigg] \quad (4)$$

Through this mutual on-policy "teaching" process, the optimization signals are absorbed with high efficiency because the teachers evaluate the exact token states generated by the students. Ultimately, Model β internalizes the agentic logic for shadows, and Model α absorbs the holistic aesthetic judgment, achieving unified multidimensional capabilities. A case is shown in Fig2.

3.4 Final Unified Joint Optimization

After mutual distillation, Model β successfully absorbs the fine-grained physical scrutiny capabilities of Model α while maintaining its strong global aesthetic perception.

In the final phase of PCPD, we discard Model α and exclusively utilize Model β (which operates without explicit tools or external skill prompts, ensuring high inference efficiency). We subject Model β to a final, comprehensive GRPO fine-tuning stage. During this stage, the reward function is a unified aggregate of all four dimensions:

$$\mathcal{R}_{total} = \lambda_1\mathcal{R}_{Aesthetics} + \lambda_2\mathcal{R}_{Lighting} + \lambda_3\mathcal{R}_{Shadow} + \lambda_4\mathcal{R}_{Realism} \quad (5)$$

Because the dimensional conflicts have already been untangled and structurally internalized during the partial complementary distillation phase, Model

β can now smoothly optimize across all four rewards simultaneously without suffering from attention hijacking or gradient interference, achieving SOTA multi-dimensional evaluation performance.

4 Experiments

4.1 Experimental Setup

Training details: We instantiate both the Agentic Expert (Model α) and the Global Expert (Model β) using the Qwen-VL architecture (Bai et al., 2025) as our base Multimodal Large Language Model (MLLM). Optimization employs AdamW with a weight decay of 0.1 and a warmup ratio of 0.03. The learning rate is set to 2×10^{-5}. Training is conducted on 8 NVIDIA A800 GPUs.

Datasets and metrics: Since traditional IQA datasets lack high-fidelity AIGC images and multi-dimensional annotations, we construct and utilize **AIGC-IQA-Bench** as our primary training and evaluation benchmark. It utilizes a 4:1 train/test split with matched data distributions. We report results for the four individual dimensions explicitly addressed by our framework—Aesthetics, Lighting, Shadow, and Realism—as well as an "Overall" score derived from the pooled test set. The primary evaluation metrics are the Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank-Order Correlation Coefficient (SRCC) between our model's predicted scores and human Mean Opinion Scores (MOS).

4.2 IQA Comparisons

We systematically evaluate various IQA approaches on AIGC-IQA-Bench, categorized into: (i) Handcrafted methods, (ii) Deep Learning models, and (iii) MLLMs-based approaches. Table 1 presents the PLCC/SRCC correlations with human MOS. Within each category block, the best result is bolded and the second best is underlined.

Traditional Handcrafted and Deep Learning methods fail on high-fidelity AIGC images due to their inability to perceive semantics or physical geometric rules, resulting in extremely poor shadow and realism assessments. Early MLLMs (e.g., LIQE, SA-IQA) significantly boost aesthetic perception (0.58–0.76) but fall victim to the "Halo Effect"; their strong stylistic priors mask local optical anomalies, causing shadow evaluations to stagnate (e.g., SA-IQA achieves 0.761 for aesthetics but only 0.351 for shadows). While advanced RL-MLLMs (VisualQuality-R1, Q-Probe) elevate overall correlations (0.76–0.83), they face an inherent dimensionality conflict within their shared parameter space, compromising fine-grained physics for global aesthetics and capping shadow accuracy at $\sim$0.68. In stark contrast, our **PCPD** framework fundamentally untangles this conflict. By distilling local agentic reasoning into a structurally isolated global policy, PCPD achieves peak aesthetic correlation (0.866) while radically boosting shadow correlation to 0.821, providing unprecedented physical scrutiny without sacrificing holistic artistic appreciation.

4.3 Ablation Studies

To validate the core mechanisms of our Partial Complementary Policy Distillation framework, we performed extensive ablation studies on the AIGC-IQA-Bench. The results are summarized in Table 2.

Effectiveness of Zoom-in/Zoom-out Tools: Comparing Configuration (1) and (2), Model β (relying on single-pass global evaluation) completely fails to evaluate Shadows accurately (0.512), hijacked by the "Halo Effect." Model α, equipped with the explicit `<zoom_in>` and `<zoom_out>` tools, achieves a significant boost in Shadow assessment (0.731). However, forced to split its attention mechanically, Model α's Aesthetic score drops to 0.711, demonstrating the "Masking Effect."

Impact of Self-Evolving Skills: Configuration (3) adds the GRPO-driven skill reflection to Model α. By explicitly summarizing logical rules to avoid aesthetic-shadow biases, Shadow correlation jumps further to 0.814. This confirms that guiding the model to autonomously articulate and prepend "skills" significantly enhances physical anomaly detection.

Complementary Distillation Validation: Configuration (4) is our final proposed method: discarding Model α at inference time and utilizing a unified Model β that has learned from α via KL-divergence distillation. Without needing explicit tool-calling at inference, the final model effectively internalizes the implicit spatial attention and sequential reasoning. It achieves the best of both worlds: peak Aesthetic correlation (0.866) and peak Shadow correlation (0.821).

5 Conclusion

In this work, we introduced Partial Complementary Policy Distillation (PCPD), a novel framework

Methods	Aesthetics	Lighting	Shadow	Realism	Overall
Handcrafted					
BRISQUE (Mittal et al., 2012a)	**0.124/0.138**	0.151/0.162	**0.017/0.023**	**0.058/0.059**	0.086/0.091
NIQE (Mittal et al., 2012b)	0.102/0.117	**0.164/0.177**	0.022/0.017	0.044/0.052	**0.093/0.084**
Deep Learning					
MUSIQ (Ke et al., 2021)	0.356/0.342	0.411/0.422	0.124/0.103	0.184/0.189	0.318/0.322
UNIQUE (Zhang et al., 2021)	0.383/0.396	0.456/0.444	0.156/0.167	0.223/0.216	0.353/0.366
MANIQA (Yang et al., 2022)	**0.414/0.418**	**0.484/0.491**	**0.178/0.172**	**0.254/0.261**	**0.387/0.391**
MLLMs-based					
LIQE (Zhang et al., 2023)	0.583/0.589	0.521/0.539	0.282/0.266	0.309/0.327	0.479/0.496
Q-Align (Wu et al., 2023)	0.614/0.626	0.551/0.562	0.318/0.323	0.384/0.388	0.521/0.536
DeQA-Score (You et al., 2025)	0.652/0.644	0.588/0.596	0.342/0.336	0.418/0.404	0.561/0.574
SA-IQA (Gao and Song, 2025)	0.761/0.756	0.692/0.688	0.351/0.344	0.454/0.441	0.679/0.673
UnifiedReward-T (Wang et al., 2025b)	0.784/0.792	0.713/0.706	0.483/0.496	0.569/0.581	0.729/0.733
VisualQuality-R1 (Wu et al., 2025)	0.813/0.822	0.747/0.756	0.511/0.527	0.609/0.606	0.766/0.774
Qwen3-VL-8B (Standard GRPO)	0.833/0.821	0.712/0.699	0.531/0.508	0.557/0.489	0.814/0.807
Q-Probe (Li et al., 2026)	0.848/0.836	0.788/0.776	0.683/0.671	0.693/0.689	0.833/0.828
PCPD (Ours)	**0.866/0.851**	**0.808/0.801**	**0.821/0.811**	**0.783/0.776**	**0.868/0.859**

Table 1: PLCC/SRCC performance comparison on AIGC-IQA-Bench. Best results are bolded, and second-best results are underlined.

Model Configuration	Aesthetics	Lighting	Shadow	Realism	Overall
(1) Model β Only (No Distillation)	0.834/0.821	0.701/0.694	0.512/0.486	0.541/0.522	0.802/0.796
(2) Model α (Agentic, No Skills)	0.711/0.689	0.726/0.711	0.731/0.716	0.691/0.681	0.744/0.729
(3) Model α (Agentic + Skills)	0.742/0.732	0.756/0.741	0.814/0.801	0.766/0.754	0.786/0.772
(4) PCPD Unified Model β (Ours)	**0.866/0.851**	**0.808/0.801**	**0.821/0.811**	**0.783/0.776**	**0.868/0.859**

Table 2: Ablation of PCPD components. Model β alone suffers from shadow evaluation deficits. Model α excels at shadows but compromises aesthetics. Our unified PCPD successfully merges these capabilities.

designed to resolve the inherent dimensionality conflict in multi-dimensional AIGC Image Quality Assessment. By identifying the "Local-to-Global Masking" and "Global-to-Local Halo" effects, we revealed why traditional unified training fails to balance mutually exclusive dimensions like global aesthetics and local physical constraints. PCPD structurally decouples these dimensions, employing an Agentic Expert equipped with self-evolving `Zoom-in`/`Zoom-out` skills to perform fine-grained physical scrutiny. Through a rigorous mutual distillation process, this multi-scale reasoning capability is successfully internalized into a highly efficient unified global policy. Extensive experiments demonstrate that PCPD achieves state-of-the-art performance across diverse evaluation dimensions and serves as a highly reliable reward signal for generative model optimization. We believe this framework paves a new path for comprehensive and human-aligned visual evaluation.

Limitations

While PCPD demonstrates remarkable performance, it is not without limitations. First, the two-branch training paradigm—requiring the simultaneous initialization and optimization of both an Agentic Expert and a Global Expert—introduces higher computational overhead and memory requirements during the training phase compared to standard single-model pipelines. Second, the effectiveness of the self-evolving skills inherently depends on the base MLLM's zero-shot capacity for instruction following and spatial understanding. If the base model lacks sufficient visual acuity, the initial skill accumulation phase may yield suboptimal reasoning trajectories. Finally, while PCPD effectively resolves spatial feature conflicts (e.g., shadows versus overall aesthetics), its efficacy on highly abstract, culturally-dependent semantic conflicts remains to be fully explored.

References

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, and 1 others. 2025. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*.

Zhuoxuan Cai, Jian Zhang, Xinbin Yuan, Peng-Tao Jiang, Wenxiang Chen, Bowen Tang, Lujian Yao, Qiyuan Wang, Jinwen Chen, and Bo Li. 2025. Q-ponder: A unified training pipeline for reasoning-based visual quality assessment. *arXiv preprint arXiv:2506.05384*.

Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. 2020. Perceptual quality assessment of smartphone photography. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3677–3686.

Yuan Gao and Jin Song. 2025. Sa-iqa: Redefining image quality assessment for spatial aesthetics with multi-dimensional rewards. *arXiv preprint arXiv:2512.05098*.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. 2021. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157.

Weiqi Li, Xuanyu Zhang, Shijie Zhao, Yabin Zhang, Junlin Li, Li Zhang, and Jian Zhang. 2025. Q-insight: Understanding image quality via visual reinforcement learning. *arXiv preprint arXiv:2503.22679*.

Xiang Li, XueHeng Li, Yu Wang, XuanHua He, ZhangChi Hu, WeiWei Yu, and ChengJun Xie. 2026. Q-probe: Scaling image quality assessment to high resolution via context-aware agentic probing. *arXiv preprint arXiv:2601.15356*.

Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012a. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708.

Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012b. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212.

OpenAI. 2025. Thinking with images. https://openai.com/index/thinking-with-images/.

Avinab Saha, Sandeep Mishra, and Alan C Bovik. 2023. Re-iqa: Unsupervised learning for image quality assessment in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5846–5855.

Haozhe Wang, Alex Su, Weiming Ren, Fangzhen Lin, and Wenhu Chen. 2025a. Pixel reasoner: Incentivizing pixel-space reasoning with curiosity-driven reinforcement learning. *arXiv preprint arXiv:2505.15966*.

Yibin Wang, Zhimin Li, Yuhang Zang, Chunyu Wang, Qinglin Lu, Cheng Jin, and Jiaqi Wang. 2025b. Unified multimodal chain-of-thought reward model through reinforcement fine-tuning. *arXiv preprint arXiv:2505.03318*.

Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Kaixin Xu, Chunyi Li, Jingwen Hou, Guangtao Zhai, and 1 others. 2024. Q-instruct: Improving low-level visual abilities for multi-modality foundation models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25490–25500.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, and 1 others. 2023. Q-align: Teaching lmms for visual scoring via discrete text-defined levels. *arXiv preprint arXiv:2312.17090*.

Tianhe Wu, Jian Zou, Jie Liang, Lei Zhang, and Kede Ma. 2025. Visualquality-r1: Reasoning-induced image quality assessment via reinforcement learning to rank. *arXiv preprint arXiv:2505.14460*.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. 2022. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1191–1200.

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

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

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. 2023. 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*, pages 14071–14081.

Shijie Zhao, Xuanyu Zhang, Weiqi Li, Junlin Li, Li Zhang, Tianfan Xue, and Jian Zhang. 2025. Reasoning as representation: Rethinking visual reinforcement learning in image quality assessment. *arXiv preprint arXiv:2510.11369*.

Ziwei Zheng, Michael Yang, Jack Hong, Chenxiao Zhao, Guohai Xu, Le Yang, Chao Shen, and Xing Yu. 2025. Deepeyes: Incentivizing" thinking with images" via reinforcement learning. *arXiv:2505.14362*.

Supplementary Details for PCPD

This appendix provides comprehensive supplementary details regarding the human annotation protocols, the LLM-driven skill summarization mechanism using state-of-the-art agentic prompt templates, the precise ReAct-style prompt designs utilized during training and inference, hyperparameters for the Partial Complementary Policy Distillation (PCPD), and an extended analysis of cognitive biases in IQA.

A Human Evaluation Protocol for AIGC-IQA-Bench

To establish authoritative ground-truth labels and explicitly prevent dimensional convergence, a panel of five professional digital artists manually scored the AIGC images. Evaluators were forced to utilize a spatial-masking interface to assess mutually exclusive dimensions in separate cognitive passes. This interface physically blocks out global contextual information when evaluating local physics, actively counteracting the "Halo" and "Masking" effects.

Human Annotation Guidelines: Multi-Dimensional AIGC-IQA
Phase 1: Global Aesthetics & Lighting Assessment • *Aesthetics (1-5):* Evaluate artistic composition, color harmony, and visual appeal. **Constraint:** You must actively ignore structural anomalies (e.g., floating objects, extra fingers). 5 = Flawless artistic expression; 1 = Visually repulsive. • *Lighting (1-5):* Assess the global illumination mood and tonal transitions.
Phase 2: Local Physical Scrutiny (Shadow & Realism) **Constraint:** The UI will apply a heavy blur to the global image, forcing you to focus solely on the magnified bounding boxes of specific object-surface contacts. • *Shadow (1-5):* Deduct points for geometrical inconsistency, incorrect projection angles, or physically impossible cast shadows (e.g., a shadow cast towards the primary light source). 5 = Geometrically perfect; 1 = Severe optical violation. • *Realism (1-5):* Penalize structural hallucinations regardless of how "beautiful" the overall style is.

Table 3: Strict two-phase human scoring rubric requiring sequential UI masking to decouple Aesthetics from Physical Realism.

B Agentic LLM Skill Summarization Mechanism

During the GRPO training phase, when the discrepancy detector identifies a severe conflict between the model's dimensional scores and the human ground truth ($\mathcal{D} > \delta$), the content are reserved. The LLM acts as a "Cognitive Bias Reflection Engine." Using XML-based prompt structures, it diagnoses its own flawed trajectory and outputs a JSON-formatted generalizable skill to append to the dynamic Skill Bank.

C Training Reasoning Prompt (Model α)

During training, the Agentic Expert (Model α) adopts a ReAct-style (Reasoning and Acting) workflow. It is explicitly equipped with spatial tools (`zoom_in`, `zoom_out`) and the dynamic Skill Bank to generate high-quality Chain-of-Thought (CoT) trajectories.

D Inference Prompt (Unified Model β)

After the Partial Complementary Policy Distillation process, the Unified Model (Model β) has successfully internalized the spatial reasoning logic. During inference, it simulates the tool-use logic via an internal monologue without requiring explicit external APIs.

E Extended Bad Case Studies: Cognitive Biases in IQA

To further illustrate the necessity of PCPD, we present detailed conceptual examples of the "Attention Hijacking" phenomena observed in standard baseline models.

Prompt 1: Cognitive Bias Reflection Engine

System Role & Incident Report

<Role>
You are an Expert IQA Cognitive Reflection Agent. Your task is to diagnose cognitive biases in multi-dimensional evaluation trajectories.
</Role>

<Incident_Report>
```
- Dimension A (Aesthetics): Predicted 4.8 (High)
- Dimension B (Shadow): Predicted 4.5, Ground Truth 1.5 (Severe Overestimation)
- System Diagnosis: "Global-to-Local Halo Effect"
```
</Incident_Report>

Reflection Logic & JSON Output Formatting

<Instructions>
1. Analyze the flawed trajectory: Why did the global stylistic prior override the local physical geometry check?
2. Formulate a generalizable 'Skill' (maximum 3 sentences) to immunize the policy against this specific failure mode.
3. Format your output strictly in JSON:

```
{
  "analysis": "...",
  "extracted_skill": "..."
}
```
</Instructions>

Figure 4: The reflection prompt utilizing XML tags and JSON structuring. The full-width stacked design ensures code blocks and lengthy instructions render without forced hyphenation.

Failure Mode I: The Global-to-Local Halo Effect. As illustrated in Figure 7, in highly stylized AIGC generations (e.g., Cyberpunk neon scenes or ethereal fantasy landscapes), the vibrant color palette and striking global composition create a strong positive prior. Standard models, dominated by this global feature map, will output an Aesthetic score of 4.8. When prompted to evaluate Shadows, the model relies on the global context rather than local high-frequency details. Consequently, it fails to notice that a character's shadow is cast towards the streetlamp rather than away from it, erroneously assigning a Shadow score of 4.5 instead of the human-annotated 1.5. PCPD prevents this by utilizing Model α's localized scrutiny to decouple the shadow geometry from the global aesthetic tone.

Failure Mode II: Local-to-Global Masking. As shown in Figure 8, in scenes with extremely high-fidelity rendering, an AIGC model might generate a geometrically perfect, sharply defined shadow with excellent texture (e.g., capturing the exact silhouette of a bicycle). A standard IQA model focused on physical metrics recognizes this high-quality local feature. However, if this perfect shadow is hallucinated indoors, cast onto a living room wall where no bicycle exists, it severely damages the logical composition and overall aesthetic of the image. Standard models, "masked" by the high local quality of the shadow patch, fail to penalize the global Aesthetic score. Our unified Model β, having learned through PCPD to

Prompt 2: Agentic Expert Training (ReAct Workflow)

Role Assignment & Active Skills Injection

<Role>
Agentic IQA Evaluator
</Role>

<Active_Skills>
`- [Skill 012]: Do not let vibrant neon color palettes distract from checking the angular consistency of ground shadows.`
</Active_Skills>

ReAct Workflow & Agentic Constraints

<Workflow>
You must follow this ReAct loop to prevent cognitive bias:

- **<Step_1> (Global Sweep):** Output initial aesthetic impression based on the full image.
- **<Step_2> (Local Probing):** You MUST invoke `<tool name="zoom_in" args="[x1,y1,x2,y2]" />` to inspect suspicious structural contacts.
- **<Step_3> (Context Verify):** Invoke `<tool name="zoom_out" />` to verify local anomalies against the global light source.
- **<Step_4> (Scoring):** Provide scores strictly formatted as: `<score dimension="...">value</score>`.

</Workflow>

Figure 5: Training prompt enforcing ReAct-style tool usage and skill integration. The stacked layout cleanly separates the dynamic skill context from the rigid agentic execution steps.

conduct sequential verification (*Local Shadow Scrutiny* → *Global Entity Verification*), correctly identifies the semantic mismatch and heavily penalizes the Aesthetic and Realism scores.

Prompt 3: Unified Model Deployment (Internalized Saccade)

System Role & Saccade Instruction

<Role>
Unified Multi-Dimensional IQA Expert
</Role>

<Instruction>
Evaluate the provided AIGC image across Aesthetics, Lighting, Shadow, and Realism.
Constraint: Even without explicit cropping tools, you must simulate a "Global-to-Local-to-Global" visual saccade in your reasoning.
</Instruction>

Internal Monologue Formatting

<Format>
You must structure your response as follows:
<Internal_Monologue>

- `[Phase 1: Global Aesthetic prior acknowledgment]`
- `[Phase 2: Simulated local scrutiny - explicitly describe the geometry of specific shadows and structural joints]`
- `[Phase 3: Conflict resolution - balance stylistic intent with physical laws]`

</Internal_Monologue>

<Final_Scores>
`<Aesthetics>...</Aesthetics>`
`<Lighting>...</Lighting>`
`<Shadow>...</Shadow>`
`<Realism>...</Realism>`
</Final_Scores> </Format>

Figure 6: The streamlined inference prompt utilizing internalized reasoning. The structured internal monologue explicitly forces the unified model to decouple spatial features before scoring.

Figure 7: **Failure Mode I (Global-to-Local Halo Effect):** In highly stylized AIGC generation, the strong positive aesthetic prior "hijacks" the model's attention. Baseline models assign high scores to geometrically incorrect shadows (e.g., shadows cast towards the light source) simply because they match the global atmosphere. PCPD isolates local scrutiny to correctly identify these physical violations.

Figure 8: **Failure Mode II (Local-to-Global Masking):** When an image contains high-fidelity local features (like perfectly textured, sharp shadows), the model can become fixated on this local quality. It fails to recognize when these shadows are logically misplaced (e.g., cast onto empty air or indoors where they do not belong), thus failing to penalize the global aesthetic and realism scores. PCPD breaks this masking via sequential context verification.