UNLOCKING WHAT MLLMS ALREADY SEE FOR HALLUCINATION MITIGATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Hallucinations remain a critical challenge for Multimodal Large Language Models (MLLMs), which often generate statements inconsistent with the input image. In this work, we uncover a **generation-verification gap**: although an MLLM may hallucinate during generation, it can often recognize these hallucinations when its output is decomposed into atomic claims and each claim is independently verified against the image. Building on this observation, we propose a self-supervised framework called **SVG-OPD** (***S**elf-**V**erification **G**uided **O**n-**P**olicy **D**istillation*), which turns this self-verification capability into more reliable generation behavior. SVG-OPD first elicits complementary fact-oriented and probing responses, decomposes them into atomic claims, and self-verifies each claim to identify hallucinated and non-hallucinated content. These verified claims are then provided as privileged information to a frozen teacher, whose token-level distributions supervise student-generated trajectories through on-policy distillation. Unlike existing training methods, SVG-OPD requires neither external knowledge nor a stronger teacher model, instead exploiting the MLLM's own visual verification capability for hallucination-aware post-training. Extensive experiments across generative and discriminative hallucination benchmarks, model families, and scales show that SVG-OPD substantially mitigates hallucinations while broadly preserving general multimodal capabilities.

1 INTRODUCTION

Multimodal Large Language Models (MLLMs) have achieved impressive progress in visual understanding, open-ended question answering, and multimodal reasoning (Achiam et al., 2023; Liu et al., 2023; Bai et al., 2025a; Liu et al., 2026). Nevertheless, even state-of-the-art MLLMs frequently generate statements that are inconsistent with the visual content, a phenomenon commonly referred to as visual hallucination (Li et al., 2023; Wang et al., 2024b). Such hallucinations undermine the reliability of model outputs and hinder their deployment in real-world settings (Liu et al., 2024a).

Existing learning-based hallucination mitigation methods typically mitigate hallucination by introducing a stronger source of supervision, including external teacher models (Li et al., 2024), preference annotations (Yu et al., 2024a), or additional visual feedback (Chen et al., 2026). Despite their effectiveness, these approaches largely share an implicit assumption: correcting hallucinations requires external supervision that is better than what the model itself can provide. This raises a less explored question:

> ***Can the MLLM identify its own hallucinations without acquiring external knowledge?***

Inspired by the aforementioned works, we conduct an empirical analysis and find that MLLMs can often accurately identify hallucinations in their own outputs. As shown in Figure 1, we first let the MLLM generate a response, then decompose it into atomic visual claims and ask the same MLLM to check each claim against the image. Despite producing hallucinations during generation, the MLLM achieves high verification accuracy on both hallucinated and non-hallucinated claims. Thus, the MLLM can often recognize hallucinations in its own output even when it fails to avoid them during generation. We refer to this discrepancy as the generation-verification gap. This gap suggests a different view of visual hallucination: hallucination does not always mean that the model lacks

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

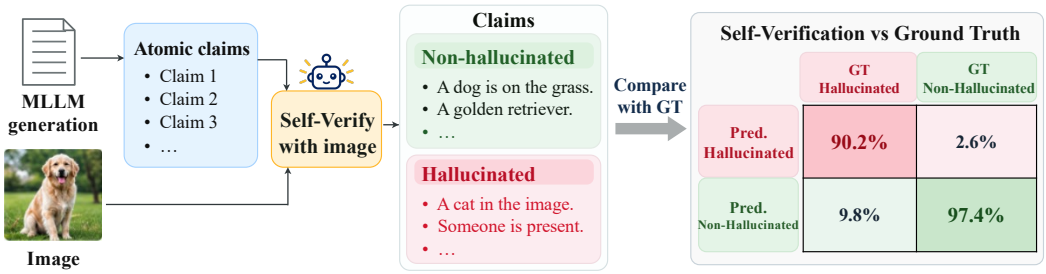


Figure 1: **MLLMs exhibit a generation-verification gap.** Although hallucinations may arise during MLLM generation, the MLLM can often identify them through atomic claim verification against the image. We evaluate this self-verification capability using Qwen3-VL-8B-Instruct on 622 atomic claims from 150 randomly sampled AMBER-Gen examples (Wang et al., 2024b). Self-verification correctly detects 90.2% of hallucinated claims and correctly accepts 97.4% of non-hallucinated claims. Ground-truth labels are derived from the original AMBER-Gen annotations when applicable and manually annotated otherwise.

the relevant visual knowledge; instead, the model may already know the correct answer but fail to reliably express it during generation.

Motivated by this generation-verification gap, we propose **SVG-OPD**, a self-supervised post-training framework for MLLMs that extracts the model's latent visual judgments and internalizes them into its generation policy through two components. **(1) Atomic visual self-verification:** To address the difficulty of verifying complex open-ended responses, we introduce an atomic self-verification procedure in Section 3.2. We decompose sampled responses into independent atomic visual propositions and let the MLLM re-examine the image to distinguish non-hallucinated claims from hallucinated ones. This converts response-level factual assessment into localized verification decisions, allowing the model's existing visual knowledge to be extracted more reliably. **(2) Self-verification guided on-policy distillation:** To incorporate the above self-verified atomic evidence into the model's generation policy, we introduce an evidence-aware teacher and perform on-policy distillation in Section 3.3. The teacher is built from the MLLM itself but additionally conditions on the above self-verified atomic evidence. It then supervises responses sampled from the student's current policy through on-policy distillation (Agarwal et al., 2024), keeping supervision aligned with the student's evolving generation distribution. By internalizing the post-generation verification evidence, SVG-OPD enables the model to move from recognizing hallucinations after generation to reducing them during generation.

Contributions. We explore a self-supervised approach to hallucination mitigation: improving generation by exploiting the model's own verification capability. We identify a generation-verification gap: MLLMs can often recognize hallucinations in their own responses when visual claims are isolated and explicitly verified against the image. Building on this finding, we propose SVG-OPD, a self-supervised framework that extracts self-verified atomic visual evidence and internalizes the resulting guidance through on-policy distillation, without external supervision or evidence at inference time. Extensive experiments across multiple benchmarks, including 7 main evaluation criteria and diverse backbone architectures, demonstrate substantial reductions in hallucination while maintaining strong general multimodal capabilities.

2 Related Work

Internal Model Signals for Hallucination Mitigation. Hallucinations in MLLMs often arise from an imbalance between visual evidence and language priors (Favero et al., 2024; Liu et al., 2024c). Recent approaches therefore seek to recover or amplify visually grounded signals already present in the model. VCD (Leng et al., 2023) contrasts predictions under clean and distorted visual inputs to reduce reliance on language priors, while ICD (Wang et al., 2024c) constructs contrastive distributions through perturbed instructions. DeCo (Wang et al., 2025) further shows that visually grounded predictions can emerge in intermediate layers even when the final output hallucinates, and leverages

these intermediate signals to guide decoding. These methods reveal that useful hallucination-related signals are already encoded within the model, but primarily exploit them through inference-time interventions, introducing additional computational overhead during generation. In contrast, our approach turns the model's own visual judgments into a training-time objective, internalizing these signals directly into the generation policy and incurring no additional inference-time overhead.

Training-Time Hallucination Mitigation. Hallucination in MLLMs has also been actively addressed through training-time interventions that improve visual grounding by refining supervision signals (Yu et al., 2025) or optimization objectives (Wang et al., 2024a). A prominent line of work constructs training signals that explicitly distinguish grounded from hallucinated content. HA-DPO (Zhao et al., 2024) aligns models with preferred non-hallucinated responses through preference optimization, while HALVA (Sarkar et al., 2025) uses fine-grained contrastive supervision between grounded and hallucinated content. RLSF-V (He et al., 2026) instead derives self-feedback from uncertainty in the model's generation distribution and uses it to improve visual grounding. These approaches demonstrate the effectiveness of training-time alignment, but rely on preference pairs, constructed contrastive supervision, or uncertainty-based feedback. Our method instead directly elicits the model's own visual knowledge through atomic claim verification and distills the verification signals into the generation policy via on-policy training.

3 METHOD

We introduce a self-supervised teacher-student framework that turns visual judgments elicited after generation into guidance for generation. We first formulate the task and learning objective (Section 3.1). We then elicit responses under two complementary views, decompose them into atomic visual claims, and verify these claims against the image to construct non-hallucinated and hallucinated claims (Section 3.2). These self-verified claims serve as privileged information for a frozen teacher, which provides token-level guidance along the student's own trajectories (Section 3.3). We distill this guidance into the student through an on-policy objective (Eq. 9), enabling generation from the image and instruction alone at inference time. Figure 2 illustrates the overall framework. We next describe each component in detail.

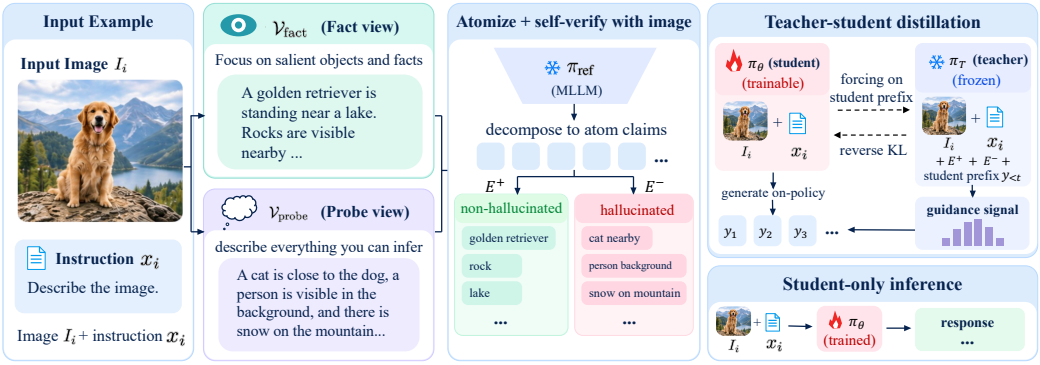


Figure 2: Overview of SVG-OPD. The reference model π_{ref} generates responses under fact-oriented and probing views, decomposes them into atomic visual claims, and verifies each claim against the image to construct non-hallucinated claims E_i^+ and hallucinated claims E_i^-. The frozen teacher π_T conditions on this evidence to provide next-token distributions along student-generated trajectories. The student π_θ learns from these distributions through on-policy reverse-KL distillation and uses only the image input and instruction at inference time.

3.1 PROBLEM FORMULATION

Given an image-instruction pair (I_i, x_i), an MLLM with policy π_θ generates a response $y_i = (y_{i,1}, \ldots, y_{i,T_i})$ of length $T_i = |y_i|$ autoregressively:

$$y_{i,t} \sim \pi_\theta(\cdot \mid I_i, x_i, y_{i,<t}), \qquad t = 1, \ldots, T_i, \tag{1}$$

where $y_{i,<t} = (y_{i,1}, \ldots, y_{i,t-1})$ denotes the previously generated tokens. Visual hallucination occurs when the response contains claims that are not supported by the image. Our objective is to learn a generation policy that reduces visual hallucinations while preserving relevant, visually grounded content, using supervision derived from the model's own visual judgments rather than human corrections or an external teacher. The learned policy must generate directly from the original image and instruction, without requiring atomic evidence or a teacher at inference time.

3.2 Eliciting and Verifying Atomic Claims

Motivated by the generation-verification gap discussed in Section 1, we first elicit hallucinated and non-hallucinated claims through two complementary problem views: a *probing view* $\mathcal{V}_{\text{probe}}$ and a *fact-oriented view* $\mathcal{V}_{\text{fact}}$. Specifically, $\mathcal{V}_{\text{probe}}$ encourages the MLLM to explore the image as broadly as possible, thereby increasing the chance of hallucinated claims as longer responses shift the model's attention away from image tokens and toward language priors. In contrast, $\mathcal{V}_{\text{fact}}$ instructs the model to report only salient objects and facts that are clearly supported by the image. (The full instructions are provided in Appendix.) Based on these two complementary views $\mathcal{V}_{\text{fact}}$ and $\mathcal{V}_{\text{probe}}$, we use the original MLLM before post-training π_{ref} to generate responses for each image I_i:

$$y_{i,t}^{\text{fact}} \sim \pi_{\text{ref}}(\cdot \mid I_i, x_i, \mathcal{V}_{\text{fact}}, y_{i,<t}^{\text{fact}}), \quad y_{i,t}^{\text{probe}} \sim \pi_{\text{ref}}(\cdot \mid I_i, x_i, \mathcal{V}_{\text{probe}}, y_{i,<t}^{\text{probe}}). \tag{2}$$

From responses to atomic claims. The above two views shape the responses generated by the model, but do not yet indicate which parts should provide supervision. As shown in Figure 3, a response may contain both hallucinated and non-hallucinated visual propositions: a single sentence may describe one visual fact correctly while hallucinating an object, attribute, or relation. Because language models struggle to attend to relevant details in long contexts (Liu et al., 2024b), and a single response can contain both non-hallucinated and hallucinated content, response-level verification may overlook localized hallucinations (Min et al., 2023).[1] We therefore decompose each response into atomic visual claims, allowing each claim to be verified independently against the image.



Question: Describe this image.

Response: ... The artwork features bold vertical stripes of red, orange, and gray.

Direct verification: non-hallucinated

Atomic claim: a.The artwork has bold vertical stripes.

b. The artwork features stripes of red, orange, and gray.

Atomic verification: a. hallucinated b. non-hallucinated

Figure 3: An example of atomic claim verification. Direct verification labels the response as non-hallucinated, overlooking the incorrect description of stripe orientation. Decomposing the response into atomic claims allows the hallucinated orientation claim and the non-hallucinated color claim to be correctly verified separately.

Specifically, we use $\text{Atomize}_{\pi_{\text{ref}}}(\cdot)$ to represent the claim-level decomposition performed by the reference model π_{ref}:

$$\mathcal{A}_i^{\text{fact}} = \text{Atomize}_{\pi_{\text{ref}}}\left(y_i^{\text{fact}}\right) = \{a_{i,j}^{\text{fact}}\}, \quad \mathcal{A}_i^{\text{probe}} = \text{Atomize}_{\pi_{\text{ref}}}\left(y_i^{\text{probe}}\right) = \{a_{i,j}^{\text{probe}}\}. \tag{3}$$

where $a_{i,j}^{\text{fact}}$ and $a_{i,j}^{\text{probe}}$ are the j-th atomic visual claims extracted from the corresponding responses, each serving as an independent unit for subsequent verification against the corresponding image I_i.

From atomic claims to self-verification. We next reuse the reference model π_{ref} as an atomic verifier to assess each claim against the image. Specifically, $\text{Verify}_{\pi_{\text{ref}}}(I_i, a)$ re-examines the image I_i and individually judges the visual validity of an atomic claim:

$$s_i(a) = \text{Verify}_{\pi_{\text{ref}}}(I_i, a), a \in \mathcal{A}_i^{\text{fact}} \cup \mathcal{A}_i^{\text{probe}}, \tag{4}$$

[1] See Appendix C.1 for a comparison of atomic-level and response-level verification.

where $s_i(a)$ labels the claim as *non-hallucinated*, *hallucinated*, or *uncertain*, as judged by π_{ref} (see the verifier prompt in Appendix). This verification step turns the two generation views into complementary evidence sources:

$$\begin{aligned} E_i^+ &= \{a \in \{a_{i,j}^{\text{fact}}\} \mid s_i(a) = \text{non-hallucinated}\}, \\ E_i^- &= \{a \in \{a_{i,j}^{\text{probe}}\} \mid s_i(a) = \text{hallucinated}\}. \end{aligned} \tag{5}$$

The *fact-oriented view* provides verified visual facts, while the *probing view* provides claims that the model generated but the verifier subsequently identifies as hallucinated; uncertain claims are discarded. We intentionally construct E_i^+ from the fact-oriented view and E_i^- from the probing view, as the two views are designed to preferentially elicit image-consistent and hallucination-prone claims, respectively.

3.3 Self-Verification Guided On-Policy Distillation

Self-verification shows that an MLLM can reliably distinguish hallucinated from non-hallucinated visual claims produced by its own generation policy. To bridge this generation–verification gap, we introduce ***S**elf-**V**erification **G**uided **O**n-**P**olicy **D**istillation* (**SVG-OPD**), which transfers post-generation verification signals back into the generation policy. SVG-OPD uses self-verified atomic evidence as privileged information to construct a frozen teacher that supervises the student's own trajectories, while the student itself observes only the input image and instruction (I_i, x_i).

Sampling from the evolving student policy. At each training iteration, we sample trajectories from the current student policy π_θ:

$$y_{i,t} \sim \pi_\theta(\cdot \mid I_i, x_i, y_{i,<t}), \tag{6}$$

where π_θ denotes the student policy being continuously updated during training. The student itself never observes the self-verified atomic evidence (E_i^+, E_i^-) and conditions only on the input image and instruction (I_i, x_i), matching the information available at inference time.

Evidence-aware teacher guidance. We reuse the original MLLM before post-training as the teacher π_T, and provide it with the self-verified atomic evidence. Under teacher forcing on each prefix of the student-generated trajectory, the teacher produces the next-token score distribution:

$$\pi_T\big(\cdot \mid I_i,\, x_i, E_i^+, E_i^-,\, y_{i,<t}\big)\,, \tag{7}$$

where π_T remains frozen throughout training, and E_i^+ and E_i^- are the verified non-hallucinated and hallucinated claims obtained in the previous section, respectively. Notably, the teacher receives no external knowledge and simply converts the model's own self-verification into token-level generation guidance at each student-generated prefix by conditioning on the self-verified atomic evidence.

On-Policy Distillation. To internalize the teacher's verification guidance into the student policy, we follow prior work on on-policy distillation (Agarwal et al., 2024; Gu et al., 2024) and minimize the reverse KL divergence from the student to the evidence-conditioned teacher along student-generated trajectories. For each sampled response y_i, we optimize

$$D_{\text{KL}}(y_i;\theta) = \frac{1}{T_i}\sum_{t=1}^{T_i} D_{\text{KL}}\Big(\pi_\theta(\cdot \mid I_i, x_i, y_{i,<t}) \| \, \pi_T\big(\cdot \mid I_i, x_i, E_i^+, E_i^-, y_{i,<t}\big)\Big). \tag{8}$$

Finally, the overall SVG-OPD objective minimizes the expected divergence under the student's on-policy trajectory distribution:

$$\mathcal{L}_{\text{SVG-OPD}}(\theta) = \mathbb{E}_{(I_i,x_i,E_i)\sim\mathcal{D}_E}\left[\mathbb{E}_{y_i\sim\pi_\theta(\cdot|I_i,x_i)}\left[D_{\text{KL}}(y_i;\theta)\right]\right]. \tag{9}$$

where $E_i = (E_i^+, E_i^-)$ and $\mathcal{D}_E$ denotes the training set of image-instruction pairs augmented with their self-verified atomic evidence. Only the student policy π_θ is updated during optimization, whereas the teacher π_T is frozen throughout training and provides privileged supervision conditioned on the self-verified evidence. Importantly, both teacher and student policies operate on the

same student-generated prefix $y_{i,<t}$. This keeps the teacher supervision aligned with states visited by the evolving student, which is intended to alleviate potential distribution mismatch between training and inference (Ross et al., 2011). [2]

Taken together, SVG-OPD provides a closed loop from self-verification to generation: atomic evidence exposes visual judgments that the MLLM can verify, the privileged teacher converts these judgments into trajectory-level supervision, and on-policy distillation internalizes this supervision into a student that requires no privileged information at inference time.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Baselines. We compare our approach with several representative hallucination mitigation methods, spanning both decoding-time and training-time strategies. The decoding-time baselines include VCD (Leng et al., 2023), ICD (Wang et al., 2024c), and DeCo (Wang et al., 2025), while the training-time baselines consist of HA-DPO (Zhao et al., 2024), HALVA (Sarkar et al., 2025), and RLSF-V (He et al., 2026). For a fair comparison, all baselines are implemented based on Qwen3-VL-8B, and their hyperparameters follow the settings specified in the corresponding original papers. Beyond the main experiments on Qwen3-VL-8B, we also apply our method to Qwen3-VL-4B-Instruct and Qwen2.5-VL-7B-Instruct (Bai et al., 2025b), demonstrating its applicability across different model families and scales.

Implementation Details. For a fair comparison, we follow the experimental setup of RLSF-V (He et al., 2026) and train on 10K unlabeled image-instruction pairs from the RLAIF-V dataset (Yu et al., 2025). We use the VeRL training framework and train our method with AdamW using a learning rate of 1×10^{-6} and a global batch size of 64. For each training input, the student model performs a single rollout, with both the temperature and top-p set to 1.0. At inference time, all methods use the same prompt for a fair comparison and adopt greedy decoding to eliminate stochasticity introduced by sampling. All experiments are conducted on 8 NVIDIA H100 80GB GPUs.

Benchmarks. We evaluate our SVG-OPD from three perspectives: generative hallucination, discriminative hallucination, and general multimodal capability. **(1)** For generative hallucination, we use AMBER-Gen (Wang et al., 2024b) and Object HalBench (Rohrbach et al., 2019). **(2)** For discriminative hallucination, we use AMBER-Dis (Wang et al., 2024b), which evaluates binary visual judgments of object existence, attributes, and relationships. **(3)** To assess whether hallucination mitigation preserves general multimodal ability, we further evaluate on MMStar (Chen et al., 2024), which measures multimodal understanding across diverse visual reasoning tasks, as well as RealWorldQA (xAI, 2024), MMMU-val (Yue et al., 2024), MME (Fu et al., 2025), MathVista-testmini (Lu et al., 2024), MMBench-Dev Circular (Liu et al., 2024d), and MM-Vet v2 (Yu et al., 2024b).

Evaluation Metrics. On AMBER-Gen, we report CHAIR, HalRate, and Cog, which measure the proportion of hallucinated object mentions, the proportion of responses containing hallucinated objects, and the frequency of mentioning predefined hallucination-prone objects, respectively. On Object HalBench, we report response-level (Resp.) and mention-level (Ment.) hallucination rates. On AMBER-Dis, we report accuracy and F1 score. For the general multimodal benchmarks, we follow their standard evaluation protocols and report accuracy or benchmark-specific scores as appropriate. Details of each metric and benchmark are provided in Appendix A.

4.2 MAIN RESULTS

As shown in Table 1, SVG-OPD consistently improves hallucination-related performance across both generative and discriminative settings, while preserving strong general multimodal capability.

(1) Hallucination Mitigation on Generative Tasks. On AMBER-Gen, SVG-OPD achieves the lowest CHAIR and Cog scores of 5.7 and 1.7, substantially improving over the vanilla model's 7.7

[2] See Appendix C.3 for an empirical comparison of on-policy and off-policy distillation.

Table 1: Comparison on generative hallucination, discriminative hallucination, and MMStar using the Qwen3-VL-8B-Instruct backbone.

	AMBER-Gen ↓			Object HalBench ↓		AMBER-Dis ↑		MMStar ↑
Model	CHAIR	HalRate	Cog	Resp.	Ment.	Acc.	F1	Acc.
Qwen3-VL-8B-Instruct (Bai et al., 2025a)	7.7	61.2	3.8	18.8	8.9	87.3	90.3	62.0
VCD (Leng et al., 2023)	9.1	65.8	4.3	21.2	10.4	86.5	89.6	55.3
ICD (Wang et al., 2024c)	8.1	65.6	4.2	19.0	9.6	87.0	90.1	62.0
DeCo (Wang et al., 2025)	7.5	60.6	3.9	17.8	8.5	87.6	90.6	55.3
HA-DPO (Zhao et al., 2024)	7.6	59.1	3.6	15.5	7.3	87.2	90.2	56.3
RLSF-V (He et al., 2026)	6.5	47.3	2.3	13.2	6.2	88.3	91.2	62.0
HALVA (Sarkar et al., 2025)	6.0	**38.8**	2.8	18.8	9.0	87.2	90.4	59.6
SVG-OPD (ours)	**5.7**	41.2	**1.7**	**11.6**	**5.5**	**88.5**	**91.3**	**62.3**

and 3.8, respectively. It also reduces HalRate from 61.2 to 41.2. On Object HalBench, SVG-OPD obtains the best response- and mention-level hallucination rates of 11.6 and 5.5, outperforming all competing methods and demonstrating more reliable generation grounded in visual evidence.

(2) Generalization to Discriminative Tasks. On AMBER-Dis, SVG-OPD achieves the best 88.5% accuracy and 91.3% F1 score, exceeding both the vanilla model and all hallucination mitigation baselines. These results show that the improved visual grounding learned during generation transfers effectively to discriminative object-existence judgments.

(3) Generalization to General Multimodal Tasks. On MMStar, SVG-OPD reaches 62.3% accuracy, slightly improving upon the vanilla model's 62.0% and achieving the best result among all compared methods. This trend further holds across the broader general multimodal evaluation in Table 2, where SVG-OPD consistently preserves general capabilities on more general multimodal benchmarks. These results show that our SVG-OPD mitigates hallucinations without sacrificing broader multimodal capabilities.

Table 2: General multimodal benchmark performance before and after applying SVG-OPD to Qwen3-VL-8B-Instruct.

Benchmark	Vanilla	+SVG-OPD	Change
RealWorldQA (xAI, 2024)	69.28	**69.41**	+0.13
MMMU-val (Yue et al., 2024)	51.89	**52.33**	+0.44
MME (Fu et al., 2025)	2401.38	**2411.24**	+9.86
MathVista-testmini (Lu et al., 2024)	44.70	**46.10**	+1.40
MMBench (Liu et al., 2024d)	84.98	**85.14**	+0.16
MM-Vet v2 (Yu et al., 2024b)	67.49	**68.84**	+1.35

4.3 Ablation Analysis

Cross-backbone generalization. To examine whether the benefits of SVG-OPD extend beyond a specific backbone, we further evaluate it across different MLLM families and scales. Table 3 shows that SVG-OPD generalizes consistently across different MLLM backbones and scales. Across all three backbones, SVG-OPD improves every hallucination metric on AMBER-Gen and Object HalBench, while also improving AMBER-Dis and maintaining MMStar performance. These results demonstrate that the effectiveness of SVG-OPD is not tied to a particular model family or scale.

Non-hallucinated and hallucinated claims. We ablate non-hallucinated and hallucinated claims separately to assess the contribution of each evidence type. As shown in Table 4, removing either type consistently degrades performance across hallucination benchmarks and MMStar, with a larger drop observed when non-hallucinated claims are removed. This indicates that both types of evidence contribute to the final model, while non-hallucinated claims play a more prominent role under the current training configuration.

Table 3: Cross-backbone results. For each backbone, we compare the corresponding vanilla model with the model trained using SVG-OPD.

		AMBER-Gen ↓			Object HalBench ↓		AMBER-Dis ↑		MMStar ↑
Backbone	Model	CHAIR	HalRate	Cog	Resp.	Ment.	Acc.	F1	Acc.
Qwen3-VL-4B-Instruct (Bai et al., 2025a)	Vanilla	8.0	61.6	3.7	21.3	10.2	89.8	92.2	59.5
	SVG-OPD	**5.9**	**41.3**	**1.9**	**14.8**	**7.2**	**90.1**	**92.6**	**59.6**
Qwen2.5-VL-7B-Instruct (Bai et al., 2025b)	Vanilla	5.0	26.2	1.8	11.3	6.0	86.5	90.3	60.9
	SVG-OPD	**4.3**	**23.2**	**1.5**	**8.7**	**4.5**	**87.7**	**91.2**	**61.3**
Qwen3-VL-8B-Instruct (Bai et al., 2025a)	Vanilla	7.7	61.2	3.8	18.8	8.9	87.3	90.3	62.0
	SVG-OPD	**5.7**	**41.2**	**1.7**	**11.6**	**5.5**	**88.5**	**91.3**	**62.3**

Table 4: Ablation on non-hallucinated and hallucinated claims with the Qwen3-VL-8B-Instruct.

	AMBER-Gen ↓			Object HalBench ↓		AMBER-Dis ↑		MMStar ↑
Model	CHAIR	HalRate	Cog	Resp.	Ment.	Acc.	F1	Acc.
SVG-OPD (ours)	**5.7**	**41.2**	**1.7**	**11.6**	**5.5**	**88.5**	**91.3**	**62.3**
w/o non-hallucinated claims	7.7	57.5	3.1	16.3	7.7	87.6	90.6	61.9
w/o hallucinated claims	6.2	45.3	2.1	12.0	5.8	88.3	91.0	62.2

4.4 Case Study

As shown in Figure 4, hallucination mitigation can sometimes lead to conservative generations that omit valid visual details. HALVA and RLSF-V correctly describe the donut, plate, fork, and napkin, but fail to mention the visible knife. SVG-OPD instead retains this grounded detail, illustrating its ability to improve faithfulness without unnecessarily suppressing informative content.



Question: Describe this image.
Ground truth: donut, fork, napkin, knife, paper plate

Qwen3-VL-8B-Instruct + HALVA: This is a close-up photo of a sugar-coated donut on a white paper plate. ... A fork and a napkin are visible on the right side of the plate. The background is dark and out of focus.

Qwen3-VL-8B-Instruct + RLSF-V: This is a close-up, eye-level photograph of a single, golden-brown sugar-coated donut resting on a white, fluted paper plate. ... The Setting: To the right of the donut, a silver fork and the edge of a white napkin are visible, suggesting a meal or snack setting.

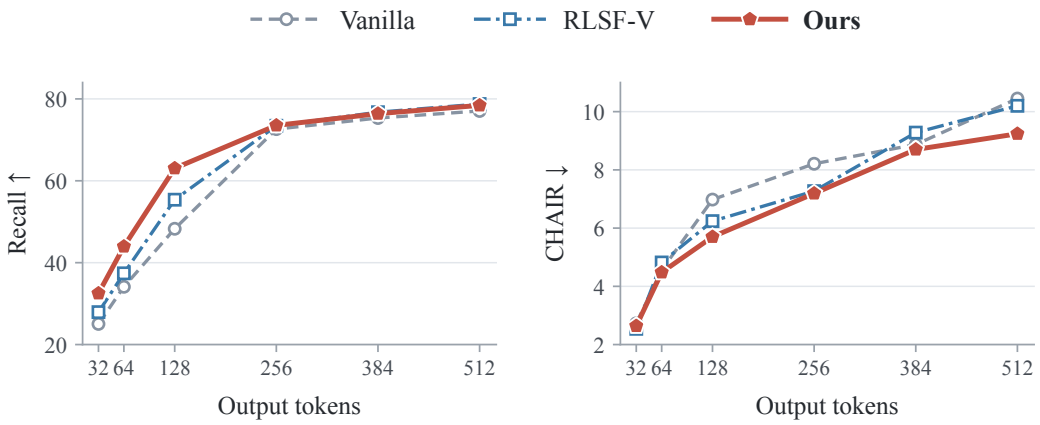
Qwen3-VL-8B-Instruct + SVG-OPD (Ours): This image features a single, golden-brown donut resting on a white paper plate. ... To the right of the donut, a silver fork and a portion of a silver knife are placed on the plate, with the fork partially tucked under the donut. A white paper napkin is also visible in the upper right corner, tucked behind the cutlery.

Figure 4: **Qualitative comparison of advanced hallucination mitigation methods.** Compared with HALVA and RLSF-V, SVG-OPD preserves more visually grounded details, successfully identifying the knife in addition to the donut, fork, napkin, and paper plate.

5 Further Discussion

Coverage–hallucination behavior under controlled generation budgets. We further examine how hallucination changes as models are given increasing opportunities to describe visual content. Specifically, we evaluate Vanilla, RLSF-V, and SVG-OPD on AMBER-Gen with Qwen3-VL-8B-Instruct under controlled output-token budgets ranging from 32 to 512 tokens. Following the budget-forcing strategy of s1 (Muennighoff et al., 2025), when a generation terminates before reaching the target budget, we append "Wait" to encourage continued generation and repeat this procedure as needed, after which the response is truncated at the specified budget. This controlled setting allows us to examine the evolution of object coverage and hallucination as the available generation budget

increases. Figure 5 reports object Recall and CHAIR across different budgets. SVG-OPD achieves higher Recall than both baselines for budgets up to 256 tokens, while attaining lower CHAIR at 64–256 tokens. At larger budgets, the three methods achieve comparable Recall, while SVG-OPD maintains lower CHAIR, indicating a consistent advantage in the coverage–hallucination trade-off.



(a) Recall across output lengths

(b) CHAIR across output lengths

Figure 5: Recall (↑) and CHAIR (↓) on AMBER-Gen using Qwen3-VL-8B-Instruct at different output-token budgets. If generation ends before the budget is reached, we append "Wait" to prompt continued generation, repeating as needed. Responses are truncated at the specified token budget.

6 Limitations

A practical limitation of our method is that, although the underlying formulation is model-agnostic, training requires access to the model's output score distributions and optimizable parameters. As a result, the current framework is primarily applicable to open-source MLLMs that expose logits and support parameter updates, and cannot be directly applied to closed-source API models with restricted internal access. Although our experiments demonstrate consistent gains across several MLLM backbones and scales, the current evaluation remains restricted to open-source models.

7 Conclusion

We propose SVG-OPD, a self-supervised framework that exploits the generation-verification gap to mitigate hallucinations in MLLMs. By decomposing generations into atomic claims and self-verifying them against the image, SVG-OPD extracts self-verified atomic evidence to guide the model itself as the teacher, whose token-level supervision is internalized through on-policy distillation for more reliable generation. Experiments across multiple benchmarks, model families, and scales show that SVG-OPD consistently reduces generative and discriminative hallucinations while preserving general multimodal capabilities. These results demonstrate the potential of self-verification to convert latent visual knowledge into more reliable generation.

AI use statement

In this work, we used generative AI tools to polish and improve the clarity of our writing and to assist with literature surveys. We did not use generative AI tools to generate ideas for the research methods. All AI-assisted text and literature-related outputs were carefully reviewed and verified by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

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

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

Lin Chen, Jinsong Li, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Zehui Chen, Haodong Duan, Jiaqi Wang, Yu Qiao, Dahua Lin, and Feng Zhao. Are we on the right way for evaluating large vision-language models?, 2024. URL https://arxiv.org/abs/2403.20330.

Wei Chen, Xin Yan, Bin Wen, Fan Yang, Tingting Gao, Di Zhang, and Long Chen. Decoupling contrastive decoding: Robust hallucination mitigation in multimodal large language models. *Advances in Neural Information Processing Systems*, 38:166734–166756, 2026.

Alessandro Favero, Luca Zancato, Matthew Trager, Siddharth Choudhary, Pramuditha Perera, Alessandro Achille, Ashwin Swaminathan, and Stefano Soatto. Multi-modal hallucination control by visual information grounding. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14303–14312. IEEE, 2024.

Chaoyou Fu, Peixian Chen, Yunhang Shen, Yulei Qin, Mengdan Zhang, Xu Lin, Jinrui Yang, Xiawu Zheng, Ke Li, Xing Sun, Yunsheng Wu, Rongrong Ji, Caifeng Shan, and Ran He. Mme: A comprehensive evaluation benchmark for multimodal large language models, 2025. URL https://arxiv.org/abs/2306.13394.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Minillm: Knowledge distillation of large language models. In *International Conference on Learning Representations*, volume 2024, pp. 32694–32717, 2024.

Changhao He, Shuhao Yan, Shuxian Li, Xi Peng, and Peng Hu. Rlsf-v: Mitigating hallucinations in mllms via fuzzy semantic self-feedback. In *Forty-third International Conference on Machine Learning*, 2026.

Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, and Lidong Bing. Mitigating object hallucinations in large vision-language models through visual contrastive decoding, 2023. URL https://arxiv.org/abs/2311.16922.

Lei Li, Zhihui Xie, Mukai Li, Shunian Chen, Peiyi Wang, Liang Chen, Yazheng Yang, Benyou Wang, Lingpeng Kong, and Qi Liu. Vlfeedback: A large-scale ai feedback dataset for large vision-language models alignment. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 6227–6246, 2024.

Yifan Li, Yifan Du, Kun Zhou, Jinpeng Wang, Xin Zhao, and Ji-Rong Wen. Evaluating object hallucination in large vision-language models. In *Proceedings of the 2023 conference on empirical methods in natural language processing*, pp. 292–305, 2023.

Hanchao Liu, Wenyuan Xue, Yifei Chen, Dapeng Chen, Xiutian Zhao, Ke Wang, Liping Hou, Rongjun Li, and Wei Peng. A survey on hallucination in large vision-language models. *arXiv preprint arXiv:2402.00253*, 2024a.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023.

Jiaxi Liu, Yifeng Yang, Xinbing Wang, Qinying Gu, and Nanyang Ye. Anchor-final self-supervision drives hallucination-aware optimization in large vision-language models. In *Forty-third International Conference on Machine Learning*, 2026.

Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *Transactions of the association for computational linguistics*, 12:157–173, 2024b.

Shi Liu, Kecheng Zheng, and Wei Chen. Paying more attention to image: A training-free method for alleviating hallucination in lvlms. In *European Conference on Computer Vision*, pp. 125–140. Springer, 2024c.

Yuan Liu, Haodong Duan, Yuanhan Zhang, Bo Li, Songyang Zhang, Wangbo Zhao, Yike Yuan, Jiaqi Wang, Conghui He, Ziwei Liu, Kai Chen, and Dahua Lin. Mmbench: Is your multi-modal model an all-around player?, 2024d. URL https://arxiv.org/abs/2307.06281.

Pan Lu, Hritik Bansal, Tony Xia, Jiacheng Liu, Chunyuan Li, Hannaneh Hajishirzi, Hao Cheng, Kai-Wei Chang, Michel Galley, and Jianfeng Gao. Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts, 2024. URL https://arxiv.org/abs/2310.02255.

Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. Factscore: Fine-grained atomic evaluation of factual precision in long form text generation. In *Proceedings of the 2023 conference on empirical methods in natural language processing*, pp. 12076–12100, 2023.

Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel Candès, and Tatsunori B Hashimoto. s1: Simple test-time scaling. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 20286–20332, 2025.

Anna Rohrbach, Lisa Anne Hendricks, Kaylee Burns, Trevor Darrell, and Kate Saenko. Object hallucination in image captioning, 2019. URL https://arxiv.org/abs/1809.02156.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 627–635. JMLR Workshop and Conference Proceedings, 2011.

Pritam Sarkar, Sayna Ebrahimi, Ali Etemad, Ahmad Beirami, Sercan Ö. Arık, and Tomas Pfister. Mitigating object hallucination in mllms via data-augmented phrase-level alignment, 2025. URL https://arxiv.org/abs/2405.18654.

Chenxi Wang, Xiang Chen, Ningyu Zhang, Bozhong Tian, Haoming Xu, Shumin Deng, and Huajun Chen. Mllm can see? dynamic correction decoding for hallucination mitigation, 2025. URL https://arxiv.org/abs/2410.11779.

Fei Wang, Wenxuan Zhou, James Y Huang, Nan Xu, Sheng Zhang, Hoifung Poon, and Muhao Chen. mdpo: Conditional preference optimization for multimodal large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 8078–8088, 2024a.

Junyang Wang, Yuhang Wang, Guohai Xu, Jing Zhang, Yukai Gu, Haitao Jia, Jiaqi Wang, Haiyang Xu, Ming Yan, Ji Zhang, and Jitao Sang. Amber: An llm-free multi-dimensional benchmark for mllms hallucination evaluation, 2024b. URL https://arxiv.org/abs/2311.07397.

Xintong Wang, Jingheng Pan, Liang Ding, and Chris Biemann. Mitigating hallucinations in large vision-language models with instruction contrastive decoding, 2024c. URL https://arxiv.org/abs/2403.18715.

xAI. Realworldqa: A benchmark for real-world spatial understanding. https://huggingface.co/datasets/xai-org/RealworldQA, 2024.

Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun, et al. Rlhf-v: Towards trustworthy mllms via behavior alignment from fine-grained correctional human feedback. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13807–13816, 2024a.

Tianyu Yu, Haoye Zhang, Qiming Li, Qixin Xu, Yuan Yao, Da Chen, Xiaoman Lu, Ganqu Cui, Yunkai Dang, Taiwen He, et al. Rlaif-v: Open-source ai feedback leads to super gpt-4v trustworthiness. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19985–19995. IEEE, 2025.

Weihao Yu, Zhengyuan Yang, Lingfeng Ren, Linjie Li, Jianfeng Wang, Kevin Lin, Chung-Ching Lin, Zicheng Liu, Lijuan Wang, and Xinchao Wang. Mm-vet v2: A challenging benchmark to evaluate large multimodal models for integrated capabilities. *arXiv preprint arXiv:2408.00765*, 2024b.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, and Wenhu Chen. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi, 2024. URL https://arxiv.org/abs/2311.16502.

Zhiyuan Zhao, Bin Wang, Linke Ouyang, Xiaoyi Dong, Jiaqi Wang, and Conghui He. Beyond hallucinations: Enhancing lvlms through hallucination-aware direct preference optimization, 2024. URL https://arxiv.org/abs/2311.16839.

A Evaluation Metrics and Benchmarks

AMBER. AMBER (Wang et al., 2024b) evaluates generative (AMBER-Gen) and discriminative (AMBER-Dis) hallucinations on 1,004 images annotated for object existence, attributes, and relationships without relying on LLM judges.

For AMBER-Gen, CHAIR and HalRate measure hallucination at the object-mention and description levels:

$$\text{CHAIR} = \frac{|\{\text{hallucinated object mentions}\}|}{|\{\text{all evaluated object mentions}\}|}, \tag{10}$$

$$\text{HalRate} = \frac{|\{\text{descriptions with hallucinated objects}\}|}{|\{\text{all descriptions}\}|}. \tag{11}$$

HalRate is the original Hal metric; Cog measures coverage of image-specific, predefined hallucination-prone objects under AMBER's matching procedure. All three are percentages; lower is better.

AMBER-Dis uses binary visual questions. We report accuracy and F1, the harmonic mean of precision and recall for "No" answers, assessing rejection of unsupported statements. Both are percentages; higher is better.

Object HalBench. Object HalBench evaluates object hallucinations using Caption Hallucination Assessment with Image Relevance (CHAIR) (Rohrbach et al., 2019), with MSCOCO captions and object annotations as references. Following RLHF-V (Yu et al., 2024a), we report response-level (Resp.) and mention-level (Ment.) rates:

$$\text{Resp.} = \text{CHAIR}_S = \frac{|\{\text{responses with hallucinated objects}\}|}{|\{\text{all responses}\}|}, \tag{12}$$

$$\text{Ment.} = \text{CHAIR}_I = \frac{|\{\text{hallucinated object mentions}\}|}{|\{\text{all object mentions}\}|}. \tag{13}$$

Both are percentages; lower is better. AMBER's CHAIR shares the mention-level definition of CHAIR_I, while HalRate corresponds to the response-level CHAIR_S. However, AMBER uses its own annotations and matching pipeline, whereas Object HalBench relies on MSCOCO references. Differences in evaluation data, annotations, and object matching prevent direct comparison of scores across benchmarks.

MMStar. MMStar (Chen et al., 2024) contains 1,500 visually dependent multiple-choice questions, evenly covering six capabilities: coarse perception, fine-grained perception, instance reasoning, logical reasoning, science and technology, and mathematics. Each capability has 250 questions, selected to reduce text-only shortcuts. We report overall accuracy.

RealWorldQA. RealWorldQA (xAI, 2024) assesses real-world visual and spatial understanding through 765 images, including anonymized vehicle-captured images and everyday scenes. Each image has a question and verifiable reference answer targeting scene details or spatial configurations in practical settings. We report answer accuracy.

MMMU. MMMU (Yue et al., 2024) evaluates college-level multimodal reasoning requiring specialized subject knowledge across 30 subjects in six disciplines: art and design, business, science, health and medicine, humanities and social science, and technology and engineering. Visual inputs include diagrams, charts, tables, and scientific illustrations. We report standard overall accuracy on the 900-question validation split (MMMU-val).

MME. MME (Fu et al., 2025) comprises 10 perception and 4 cognition tasks, covering recognition, OCR, scene understanding, and reasoning. Cognition includes commonsense reasoning, calculation, translation, and code reasoning. Each task sums question-level accuracy (Acc) and paired-question accuracy (Acc+), which requires both questions for an image to be answered correctly. We report MME Total, the sum of task scores across Perception and Cognition, with a maximum of 2,800.

MathVista. MathVista (Lu et al., 2024) tests mathematical reasoning over geometric diagrams, charts, scientific figures, and natural images through multiple-choice and free-form questions. These tasks require joint visual interpretation and mathematical operations. We evaluate on the 1,000-example testmini split (MathVista-testmini), reporting accuracy after answer extraction and normalization under the standard protocol.

MMBench. MMBench (Liu et al., 2024d) tests multimodal perception and reasoning with multiple-choice questions across a range of capabilities. We report circular accuracy on its development split (MMBench-Dev Circular). Circular Evaluation cyclically shifts answer options to reduce position bias; a question counts as correct only if answered correctly under every circular permutation.

MM-Vet v2. MM-Vet v2 (Yu et al., 2024b) contains 517 open-ended questions combining recognition, knowledge, OCR, spatial awareness, language generation, mathematics, and image-text sequence understanding. It supports interleaved images and text, extending the original benchmark beyond single-image question answering. An LLM-based grader uses reference answers and few-shot examples to assign partial credit. We report the overall score on a 0–100 scale.

B PROMPTS

All instructions are written in English. For each sample, the placeholders {`question`}, {`caption`}, and {`proposition`} are filled with the corresponding question, caption, and proposition. The image is provided separately as a multimodal input.

B.1 TRAINING PROMPTS

The prompts presented below are used for reference-model response generation, atomic decomposition, verification, and the construction of teacher prompts.

Fact-oriented view $\mathcal{V}_{\text{fact}}$ for VQA

```
Answer the original question using only high-confidence evidence directly visible
in the image. Consider only objects, attributes, counts, actions, spatial relations,
scene context, and visible text that are clearly established by the image. Omit
uncertain details, hidden causes, identities, and assumptions based on common sense.

ORIGINAL QUESTION:
{question}
```

For image-description samples, we use the following task-specific form.

Fact-oriented view $\mathcal{V}_{\text{fact}}$ for image description

```
Describe high-confidence facts directly visible in the image.
Omit uncertain attributes, exact counts, identities, text, or relations. Do not fill
missing details using common sense.
```

Probing view $\mathcal{V}_{\text{probe}}$ for VQA

```
Answer the original question after broadly inspecting the image. In the analysis,
consider all relevant visible objects, attributes, counts, actions, spatial
relations, scene context, and legible text. The analysis is detailed.

ORIGINAL QUESTION:
{question}
```

Probing view $\mathcal{V}_{\text{probe}}$ for image description

```
Give a broad, detailed description of the visible image. Cover objects, attributes,
counts, actions, spatial relations, scene context, and legible text when relevant.
Include as many specific visual details as reasonably possible.
Return only a natural-language caption.
```

Atomic decomposition

```
Convert the caption below into independently image-verifiable atomic visual
propositions. Split object existence, each attribute, count, action, spatial
relation, and visible text into separate facts. Remove subjective opinions and
background knowledge, and merge semantic duplicates.
Prefer propositions directly relevant to answering the question.
Never repeat a proposition.
```

```
Return strict JSON only, with this schema:
{"propositions": ["one fact", "another fact"]}

QUESTION:
{question}

CAPTION:
{caption}
```

Atomic verification

```
Judge one proposition using only directly visible image evidence. Return strict
JSON only, with no reason:
{"label":"entailed|contradicted|unknown"}
Use entailed only when the image directly establishes the complete claim. Use
contradicted only when the image directly establishes that the complete claim is
false. Use unknown when the image does not provide enough evidence to establish
either the claim or its contradiction.
CLAIM: {proposition}
```

The verifier labels entailed, contradicted, and unknown correspond to the non-hallucinated, hallucinated, and uncertain labels used in 3.2, respectively.

Evidence-conditioned teacher

```
Use the following visual notes to answer the user's request accurately.
Details confirmed to exist in the image:
- {non-hallucinated claim 1}
- {non-hallucinated claim 2}
These are model hallucinations and not actually exist in the image:
- {hallucinated claim 1}
- {hallucinated claim 2}
```

B.2 Evaluation Prompt

At inference time, all methods use the same prompt to ensure a fair comparison. We use the following prompts for multiple-choice VQA and image description, respectively:

VQA

```
{question}
Output using exactly the following format:
<think>Analysis based on visual evidence </think>
<answer>final answer</answer>
Base both fields only on directly visible image evidence.
```

Image Description

```
{question}
```

C Additional Results

C.1 About Atomic Decomposition

We investigate whether atomic decomposition is essential for reliable self-verification. We compare atomic-level verification, which first decomposes a response into atomic visual claims and then verifies each claim independently, with response-level verification, which asks the same reference model to directly extract and verify claims from the complete responses. Table 5 shows that atomic-level verification produces more reliable self-verified evidence. Compared with response-level verification, atomic-level verification reduces the object-level hallucination rate of positive evidence (CHAIR) from 5.5 to 4.6. A more pronounced difference arises in the extraction of hallucinated claims. After discarding claims judged as uncertain, atomic-level verification yields 365 hallucinated claims, whereas response-level verification extracts only one. This result suggests that assessing only the complete response can easily miss localized errors. In contrast, explicit atomic

decomposition exposes atomic visual statements as independent verification units, making it easier to separate potential hallucinated claims from complex responses while also improving the reliability of the retained non-hallucinated evidence.

Table 5: Comparison of different evidence-extraction pipelines, evaluated with Qwen3-VL-8B-Instruct on 150 randomly sampled AMBER-Gen examples.

Pipeline	CHAIR ↓	Hallucinated claims (#) ↑
Direct paragraph extraction	5.5	1
Atomic decomposition + verification	**4.6**	365

C.2 Effect of Atomic Self-Verification

We test whether verifying atomic claims improves distillation compared with using them directly. We compare SVG-OPD with *Atomic-Unverified OPD*, which passes atomic claims from the same fact-oriented and probing views to the frozen teacher without verification or filtering, replacing the verified evidence E_i^+ and E_i^-. All other pipeline components and training settings remain unchanged.

As shown in Table 6, explicit self-verification reduces all reported generative hallucination rates and improves AMBER-Dis accuracy, while maintaining the same MMStar accuracy. These results suggest that providing atomic evidence verified in advance by the reference model helps the frozen teacher focus on guiding the student's generation.

Table 6: Effect of atomic self-verification on Qwen3-VL-8B-Instruct.

	AMBER-Gen			Object HalBench		AMBER-Dis		MMStar
Method	CHAIR ↓	HalRate ↓	Cog ↓	Resp. ↓	Ment. ↓	Acc. ↑	F1 ↑	Acc. ↑
Atomic-Unverified OPD	7.6	60.2	3.0	14.2	7.1	87.9	90.7	**62.3**
SVG-OPD (ours)	**5.7**	**41.2**	**1.7**	**11.6**	**5.5**	**88.5**	**91.3**	**62.3**

C.3 Off-Policy versus On-Policy Distillation

We compare SVG-OPD with an off-policy distillation variant. Both approaches use self-verified evidence E_i^+ and E_i^- to guide a frozen teacher instantiated from the same base model. In the off-policy variant, the teacher generates corrected responses, and the student minimizes a token-level KL objective against the teacher's distributions on these responses. Thus, the student is trained on teacher-generated prefixes rather than its own sampled prefixes.

As shown in Table 7, this variant achieves comparable hallucination mitigation performance, while its general multimodal performance remains slightly below that of SVG-OPD. One possible explanation is the mismatch between teacher-generated training prefixes and student-generated inference prefixes. On-policy distillation instead provides teacher feedback on prefixes sampled from the current student, making the supervision more directly relevant to its generation behavior. Related benefits of on-policy training have also been reported in language-model distillation (Agarwal et al., 2024).

Table 7: Comparison of on-policy and off-policy distillation.

	Object HalBench ↓		AMBER-Dis ↑		MMStar ↑	MME ↑	MMBench ↑
Method	Resp.	Ment.	Acc.	F1	Acc.	Total	Acc.
SVG-OPD (ours)	11.62	**5.49**	**88.53**	**91.31**	**62.33**	**2411.24**	**85.14**
Off-policy distillation	**10.75**	5.58	88.33	91.19	61.80	2405.54	84.98