Aligning with Your Own Voice: Self-Corrected Preference Learning for Hallucination Mitigation in LVLMs

Byeonggeuk Lim[1], **JungMin Yun**[2], **JuneHyoung Kwon**[2], **Kyeonghyun Kim**[2], **YoungBin Kim**[1,2]
[1]Graduate School of Advanced Imaging Sciences, Multimedia and Film, Chung-Ang University
[2]Department of Artificial Intelligence, Chung-Ang University
{banggeuk, cocoro357, dirchdmltnv, khyun8072, ybkim85}@cau.ac.kr

Abstract

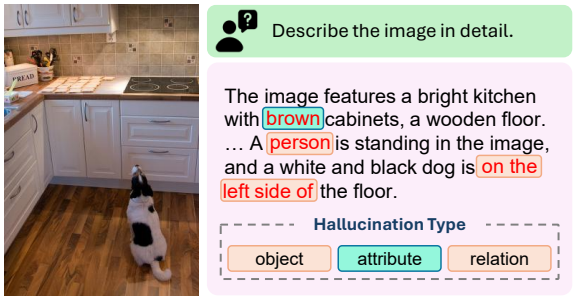
Large Vision-Language Models (LVLMs) frequently suffer from hallucinations. Existing preference learning-based approaches largely rely on proprietary models to construct preference datasets. We identify that this reliance introduces a distributional mismatch between the proprietary and target models that hinders efficient alignment. To address this, we propose Alignment via VErified Self-correction DPO (AVES-DPO), a framework that aligns LVLMs using in-distribution data derived from the model's intrinsic knowledge. Our approach employs a consensus-based verification mechanism to diagnose diverse hallucinations and guides the model to self-correct, thereby generating preference pairs strictly compatible with its internal distribution. Extensive experiments demonstrate that AVES-DPO surpasses existing baselines in hallucination mitigation while requiring only 5.2k samples.

1 Introduction

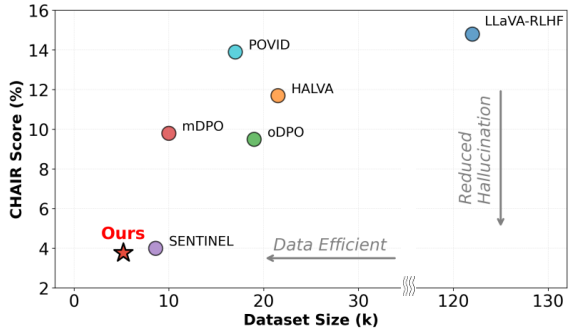
Large Vision-Language Models (LVLMs) have achieved impressive performance across various multimodal tasks (Li et al., 2023a; Liu et al., 2023, 2024b). However, existing LVLMs still suffer from hallucination, generating content inconsistent with the input images, such as non-existent objects or misrepresented attributes and relationships as illustrated in Figure 1(a) (Rohrbach et al., 2018; Li et al., 2023b). To mitigate this issue, preference learning has emerged as a promising approach, which optimizes models to prefer factual responses over hallucinated ones (Xie et al., 2024; Liu et al., 2024a; Lee et al., 2024; Sun et al., 2024). Among these methods, Direct Preference Optimization (DPO) has become mainstream, enabling efficient policy optimization using offline datasets without explicit reward modeling (Rafailov et al., 2023).

Building upon this foundation, recent research has actively applied DPO to enhance the factuality of LVLMs (Yu et al., 2024; Yang et al., 2025; Sun et al., 2024). The primary strategy involves constructing preference datasets where factual responses are favored over hallucinated ones. To achieve this, existing methods typically employ advanced proprietary models, such as GPT-4V (Achiam et al., 2023), to generate high-quality positive samples or to synthesize negative counterparts by manipulating ground-truth captions (Zhao et al., 2023; Sun et al., 2024; Yu et al., 2024; Yang et al., 2025; Ouali et al., 2025; Peng et al., 2025). By aligning LVLMs with these constructed preferences, these approaches have successfully demonstrated significant improvements in reducing hallucination rates across various benchmarks.

However, existing DPO-based approaches face



(a) An Example of Hallucination in LVLMs



(b) Comparison of Data Efficiency and Hallucination Mitigation

Figure 1: Overview of hallucination types and the effectiveness of the proposed method. (a) An example of hallucinations in LVLMs. (b) Our proposed AVES-DPO achieves the lowest CHAIR score with only 5.2k training samples, demonstrating strong data efficiency.

Findings of the Association for Computational Linguistics: ACL 2026, pages 35829–35849

two key limitations. First, constructing preference pairs relies heavily on external supervision from proprietary models such as GPT-4V (Sun et al., 2024; Yang et al., 2025; Sarkar et al., 2025). This dependence introduces a distribution mismatch, as the generation patterns of these external models often diverge from the target model's intrinsic distribution, thereby hindering effective preference alignment. Second, existing datasets and training objectives are disproportionately biased towards object hallucinations, which primarily concern determining object presence (Peng et al., 2025; Park et al., 2025; He et al., 2025). Although subtle hallucinations, such as misrepresented attributes or relationships, current methods remain limited in addressing these diverse and fine-grained error types. To address these challenges, we explore the self-correction strategy on diverse hallucination types, mitigating the distribution mismatch.

In this paper, we introduce **A**lignment via **VE**rified **S**elf-correction DPO (AVES-DPO), a framework comprising two distinct stages: Hallucination Verification and Self-correction. In the verification stage, generated responses are scrutinized across object, attribute, and relationship levels to identify whether specific elements are hallucinated. Subsequently, the LVLMs self-refine their outputs based on these diagnoses, rectifying errors while enriching the response with missing visual details. Using these self-refined outputs as preferred data provides in-distribution training signals, effectively mitigating the distribution mismatch. Furthermore, this comprehensive process also overcomes the limitations of targeting only single-type hallucinations. Finally, we employ these refined outputs to construct preference pairs for alignment training, demonstrating that this approach yields more efficient and effective learning compared to models trained with responses generated by proprietary models.

Extensive evaluations across various hallucination benchmarks demonstrate that our method significantly outperforms existing baselines. Notably, the 7B model achieves state-of-the-art (SOTA) performance on over 60% of these diverse benchmarks. Through this framework, as illustrated in Figure 1(b), AVES-DPO achieves superior performance with only 5.2k samples, representing approximately 25 times less data than LLaVA-RLHF, highlighting its exceptional data efficiency.

2 Related Work

Learning from Feedback. Early efforts to mitigate hallucination primarily relied on Supervised Fine-Tuning (SFT) with correction signals derived from proprietary LLMs. For instance, LRV-Instruction and ReCaption leverage proprietary LLMs like GPT-4 to generate counterfactual visual instructions or rewrite captions, training models to suppress hallucinations ranging from coarse object presence to fine-grained details (Liu et al., 2024a; Wang et al., 2024b).

Recent research has expanded toward preference learning to optimize models based on the divergence between positive and negative responses (Rafailov et al., 2023). LLaVA-RLHF employs Proximal Policy Optimization (PPO)-based reinforcement learning, mitigating hallucinations by incorporating factual information into the reward model (Schulman et al., 2017; Sun et al., 2024). However, to address the inherent training instability and complexity of PPO, various DPO-based methods have subsequently been proposed. For instance, POVID leverages GPT-4V to inject plausible hallucinations into ground-truth data and generate dispreferred responses via distorted images (Zhou et al., 2024a). To handle modality-specific challenges, mDPO (Wang et al., 2024a) introduces an image preference objective to address the unconditional preference problem where the model ignores visual inputs and relies solely on text, whereas oDPO utilizes object masks for fine-grained, object-level optimization (He et al., 2025). Distinct from the DPO paradigm, HALVA adopts a contrastive learning approach with generative data augmentation, offering an efficient alternative for mitigating hallucinations while preserving general model capabilities (Sarkar et al., 2025).

Despite these advances, most methods depend on proprietary models such as GPT-4V or Gemini Vision Pro (Team et al., 2023) to construct preference pairs, incurring substantial costs and potentially causing distribution mismatch between the external teacher and target models, which may degrade training efficiency and generalization performance (Zhou et al., 2024a; Compagnoni et al., 2025; Yang et al., 2025; Sarkar et al., 2025).

Inference-time Mitigation. Another line of research controls hallucinations at inference time without modifying model parameters, primarily through post-hoc verification. While some methods like LURE focus on rectifying hallucinations

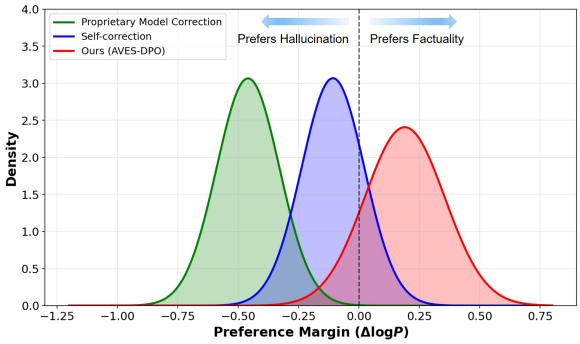


Figure 2: Distributional alignment of proprietary model vs self-correction. Unlike proprietary model correction (green) which suffers from distribution mismatch, self-correction (blue) aligns with the model's internal distribution. Our approach (red) successfully shifts the preference margin toward factuality.

based on statistical cues and uncertainty (Zhou et al., 2024b), recent work has increasingly adopted a decompose-and-verify paradigm using external experts. In this vein, Woodpecker validates generated key concepts using visual experts (e.g., object detectors) (Yin et al., 2024), a process further structured by Pelican, which decomposes claims into sub-claims to execute Program-of-Thought verification via composable tools (Sahu et al., 2024).

Another line of research focuses on inference time intervention through modified decoding strategies. VCD applies contrastive decoding by leveraging logit differences between original and distorted images (Leng et al., 2024). OPERA penalizes over-trust attention patterns and reallocates decoding when knowledge aggregation causes the model to neglect image tokens (Huang et al., 2024). However, these methods exhibit sensitivity to hyperparameter settings, limiting their generalizability across diverse domains. Furthermore, they predominantly focus on object existence verification, leaving fine-grained hallucinations involving attributes and relations inadequately addressed.

3 Empirical Analysis of Distributional Mismatch in Preference Learning

We conduct a preliminary empirical analysis to investigate the distribution mismatch issue arising from reliance on external models. This pilot study provides the rationale for adopting a self-correction mechanism.

Settings. We conduct experiments using LLaVA-1.5-7B on 200 hallucinated samples identified from the COCO dataset (Lin et al., 2014). We generate responses using the prompt "Describe the image in detail." and employ our hallucination verification module to filter samples, resulting in initial responses y_{init} containing verified hallucinations.

To investigate the distributional discrepancy arising from different correction sources, we prepare two types of corrected responses y_{corr}. For Proprietary Model Correction, we instruct GPT-4V to correct and remove hallucinations and subsequently enrich the description, generating corrections external to the target model's distribution. For Self-correction, LVLMs themselves rectify hallucinations and enrich the responses using hallucination information derived from our verification framework. We then measure the preference margin $\mathcal{M}$ to quantify the distributional shift, defined as the difference in response-averaged log probabilities $\frac{1}{L}\sum_i \log P_\theta(y_i|x, y_{<i})$ between the corrected response y_{corr} and the initial response y_{init}.

Results. Figure 2 illustrates the distribution of these preference margins. Proprietary Model Correction results in a relatively large negative margin, indicating that significant divergence between the proprietary model's generation patterns and the target model's intrinsic distribution. Learning from such data forces the model to adapt to distinct stylistic patterns, potentially leading to inefficient optimization and limited generalization. In contrast, Self-correction exhibits a margin distribution centered near zero. This observation confirms that corrections generated by the model itself reside within its internal distribution, providing natural and compatible learning signals. These findings suggest that utilizing self-corrected data minimizes distributional discrepancy, offering a more effective foundation for preference alignment compared to external supervision. This insight motivates the design of our proposed framework, AVES-DPO, which is detailed in the following section.

4 Method: AVES-DPO

4.1 Preliminaries

Large Vision-Language Models. Typically, LVLMs comprise three core components: a visual encoder that processes visual inputs, a modality alignment module that matches dimensions between modalities, and a LLM responsible for text generation. Given an input image X_v and text instruction X_t, the visual encoder converts the image into visual features H_v. These features are projected via the alignment module into visual tokens

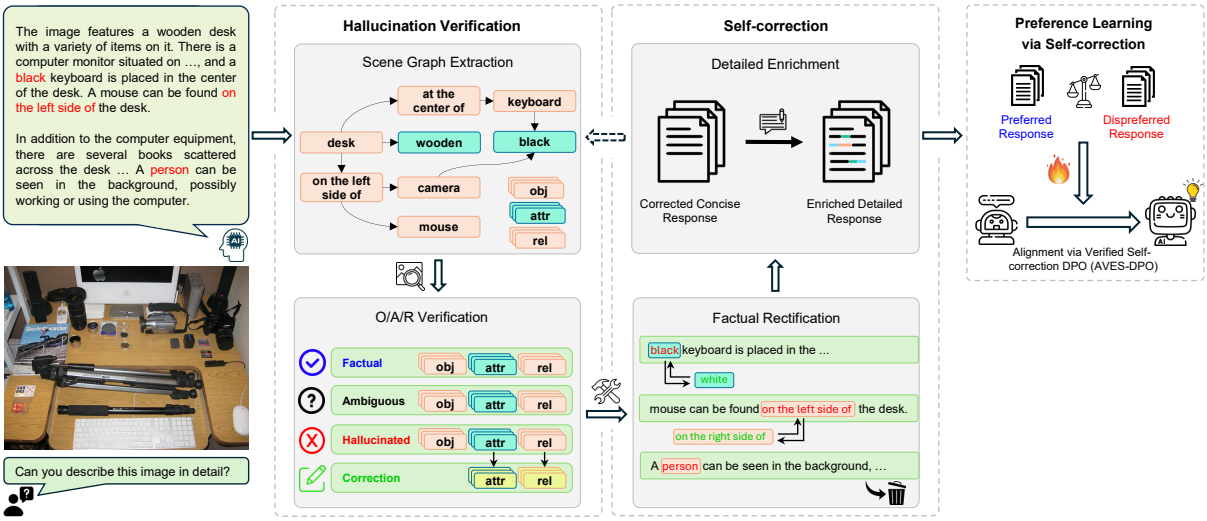


Figure 3: The overall framework of the proposed AVES-DPO.

H'_v, which are mapped to the embedding space of the LLM. The LLM then takes the projected visual tokens H'_v and the tokenized text instructions X'_t as a combined input $x = \{H'_v, X'_t\}$ to generate a response y in an auto-regressive manner, modeling the conditional probability distribution $\pi_\theta(y|x)$.

Direct Preference Optimization. To align LVLMs with human preferences, we adopt DPO, which directly optimizes the policy using preference data without requiring an explicit reward model. While traditional RLHF approaches require training a separate reward model, DPO derives the optimal policy by expressing the implicit reward function $r(x, y)$ in terms of the policy π_θ and a reference model π_{ref}:

$$r(x, y) = \beta \log \frac{\pi_\theta(y|x)}{\pi_{ref}(y|x)} + \beta \log Z(x). \quad (1)$$

Here, x denotes the multimodal input, $Z(x)$ represents the partition function, and β is a coefficient controlling the deviation. Under the Bradley-Terry model (Bradley and Terry, 1952), we can obtain the DPO loss function that directly optimizes the policy using a preference dataset $\mathcal{D} = \{x, y_w, y_l\}$:

$$\begin{aligned}
&\mathcal{L}_{DPO}(\pi_\theta; \pi_{ref}) \\
&= -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}} \left[\log \sigma \left(r(x, y_w) - r(x, y_l)\right)\right] \\
&= -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w|x)}{\pi_{ref}(y_w|x)} \right.\right. \\
&\qquad \left.\left. -\beta \log \frac{\pi_\theta(y_l|x)}{\pi_{ref}(y_l|x)}. \right)\right] \quad (2)
\end{aligned}$$

In this study, we utilize Eq. (2) to mitigate hallucinations by optimizing the likelihood difference between the model's self-corrected response y_w and the initial hallucinated response y_l.

4.2 Construction of Self-corrected Data

To construct a high-quality preference dataset, we propose a two-stage framework: Hallucination Verification and Self-correction. As illustrated in Figure 3, the Hallucination Verification stage consists of Scene Graph Extraction and O/A/R Verification, while the Self-correction stage consists of Factual Rectification and Detailed Enrichment.

4.2.1 Hallucination Verification

Scene Graph Extraction. To analyze the initial response y, we convert it into a scene graph $G = (O, A, R)$, comprising a set of objects $O = \{o_i\}$, attributes $A = \{a_{ij}\}$, and relationships $R = \{r_{ijk}\}$. This structured representation enables fine-grained hallucination detection. For precise correction, we utilize a vocabulary derived from the GQA dataset, organized into semantic subtypes to efficiently retrieve appropriate replacement candidates within the same category (details provided in Appendix A.1). The vocabulary is organized into semantic subtypes to enable the efficient retrieval of appropriate correction candidates within the same category when a hallucination is detected.

O/A/R Verification. To verify the factuality of each extracted element, we adopt a consensus-based verification strategy utilizing two independent verification models, V_1 and V_2. To ensure reliability, a label is finalized only when both mod-

els agree; otherwise, it is classified as *Ambiguous*. The verification result $L(x)$ for an arbitrary element $x \in O \cup A \cup R$ is defined as follows:

$$L(x) = \begin{cases} l & \text{if } V_1(x) = V_2(x) = l, \\ \textit{Ambiguous} & \text{otherwise,} \end{cases} \tag{3}$$

where $l \in \{\textit{Factual}, \textit{Hallucinated}\}$.

(i) Object Verification. First, we assess the presence of objects within the image. As the primary verifiers (V_1, V_2), we employ YOLO (Cheng et al., 2024) and Grounding DINO (Liu et al., 2025), known for their robust object detection. Given the potential limitations of detection models, objects initially classified as *Ambiguous* undergo a secondary verification phase using the Qwen3-VL series[1] (Bai et al., 2025). In this stage, the label $L(o_i)$ is updated to the consensus label (i.e., *Factual* or *Hallucinated*) only if both LVLM models independently produce the identical output.

(ii) Attribute & Relation Verification. Attributes and relationships are verified exclusively for objects confirmed as $L(o_i) = \textit{Factual}$. We utilize the two models from the Qwen3-VL series as verifiers under the same consensus rule. If an element is classified as *Hallucinated*, the system searches the pre-constructed vocabulary for a correct attribute or relationship matching the element's semantic subtype to serve as a correction candidate.

4.2.2 Self-correction

Guided by the diagnoses from the Hallucination Verification stage, Self-correction generates the final preferred response through two phases: Factual Rectification and Detailed Enrichment.

Factual Rectification. This phase ensures factuality by reconstructing the text based on verification results. Specifically, elements identified as hallucinations are either removed or replaced with correct content. This process yields a concise, factually accurate response free from hallucinated content.

Detailed Enrichment. To prevent oversimplification caused by mere error removal, this phase restores descriptive capability by incorporating missing visual details. Given the risk of introducing new hallucinations during generation, we employ a cyclic iterative filtering pipeline. Specifically, the enriched text is fed back into the Hallucination Verification module. If no hallucinations are detected, the response passes; if hallucinations are found, it is sent back to the Self-Correction stage for regeneration. We repeat this verification-correction loop up to 3 times for the 7B model and 5 times for the 13B model. Samples failing to yield a valid response within these attempts are discarded to guarantee high data quality. This rigorous process produces a final preferred response y^+ that balances factual accuracy with visual richness, which is then paired with the initial hallucinated response y^- for Preference Learning.

[1]Qwen3-VL-30B-A3B-Instruct, Qwen3-VL-32B-Instruct

4.3 Preference Learning via Self-correction

Following the procedures in Section 4.2, we construct a dataset $\mathcal{D}_{SC}$ comprising pairs of the model's initial hallucinated responses and enriched response obtained through precise refinement. Each sample is defined as (x, y^+, y^-), consisting of the input image and text $x = \{X_v, X_t\}$, the preferred response y^+, and the dispreferred response y^-. Here, y^- corresponds to the initial response identified as hallucinated during the verification process in Section 4.2.1, while y^+ denotes the final enriched response generated via the Self-correction process in Section 4.2.2.

Our training objective is to induce the model, conditioned on input x, to maximize the likelihood of generating the rectified response y^+, while minimizing the probability of generating the hallucinated response y^-. To achieve this, we employ the DPO loss function. Based on this objective, our proposed AVES-DPO loss is formulated as follows:

$$\begin{aligned} &\mathcal{L}_{\text{AVES-DPO}}(\pi_\theta; \pi_{ref}) \\ &= -\mathbb{E}_{(x,y^+,y^-)\sim\mathcal{D}_{SC}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y^+|x)}{\pi_{ref}(y^+|x)} \right.\right. \\ &\qquad \left.\left. -\beta \log \frac{\pi_\theta(y^-|x)}{\pi_{ref}(y^-|x)} \right) \right]. \end{aligned} \tag{4}$$

5 Experiment

5.1 Experimental Setup

Models and datasets. To evaluate the effectiveness of the proposed method, we employ two representative LVLMs with differing parameter scales: LLaVA-1.5-7B and LLaVA-1.5-13B (Liu et al., 2024b). Both models utilize CLIP-ViT-L-336px (Radford et al., 2021) as the visual encoder, with Vicuna-7B and Vicuna-13B (Chiang et al., 2023) as the respective language models. Using MS-COCO train set as the image source, we construct

Method	Data Size	Feedback	Object-Hal		AMBER						MMHal-Bench	
			$CHAIR_S\downarrow$	$CHAIR_I\downarrow$	CHAIR↓	Cover↑	Hal Rate↓	Cog↓	Acc↑	F1↑	Score↑	Hal Rate↓
LLaVA-1.5-7B†	-	-	51.4	14.8	7.1	50.5	32.5	3.8	78.8	83.0	2.22	0.57
+ VCD§ (Leng et al., 2024)	-	-	53.3	15.7	6.9	50.6	32.2	3.7	72.0	74.8	2.12	0.54
+ LLaVA-RLHF† (Sun et al., 2024)	122k	Self-Reward	53.6	14.8	8.4	52.3	42.5	4.6	70.2	76.5	2.06	0.64
+ HALVA‡ (Sarkar et al., 2025)	21.5k	GPT-4V	41.4	11.7	6.6	53.0	32.2	3.4	-	83.4	2.25	0.54
+ POVID† (Zhou et al., 2024a)	17k	GPT-4V	45.8	13.9	7.1	50.2	31.5	3.6	78.4	82.8	2.18	0.57
+ oDPO§ (He et al., 2025)	19k	GPT-4V	34.3	9.5	4.6	**53.4**	25.1	2.4	80.2	84.1	**2.50**	**0.49**
+ mDPO* (Wang et al., 2024a)	10k	GPT-4V	35.7	9.8	4.4	52.4	24.5	2.4	-	-	2.39	0.54
+ SENTINEL† (Peng et al., 2025)	8.6k	Self	12.6	4.0	**2.9**	43.7	14.6	1.2	80.3	85.2	2.04	0.58
+ AVES-DPO (Ours)	5.2k	Self	**12.2**	**3.9**	3.3	40.8	**12.6**	**0.9**	**82.3**	**86.9**	2.35	0.53
LLaVA-1.5-13B†	-	-	50.4	14.3	6.7	51.3	31.0	3.5	80.4	83.7	2.30	0.54
+ VCD§ (Leng et al., 2024)	-	-	47.7	13.2	6.7	51.3	31.0	3.5	71.5	73.5	2.40	0.51
+ LLaVA-RLHF† (Sun et al., 2024)	122k	Self-Reward	47.0	12.4	6.9	51.7	35.7	3.3	81.0	86.4	2.14	0.66
+ HALVA‡ (Sarkar et al., 2025)	21.5k	GPT-4V	45.4	12.8	6.4	**52.6**	30.4	3.2	-	**86.5**	**2.58**	**0.45**
+ oDPO§ (He et al., 2025)	19k	GPT-4V	34.7	9.8	4.3	52.1	23.1	2.2	79.3	82.2	2.74	**0.45**
+ SENTINEL† (Peng et al., 2025)	7k	Self	11.8	3.3	**2.8**	45.3	**13.2**	1.0	**81.4**	84.7	2.48	0.49
+ AVES-DPO (Ours)	4.6k	Self	**11.3**	**3.0**	3.9	40.4	17.6	**0.9**	80.4	83.3	2.36	0.58

Table 1: Comparison of hallucination mitigation performance across various benchmarks. For baseline algorithms with available official checkpoints, we re-evaluate the models, and these results are marked with †. Results sourced from (He et al., 2025) are denoted by §, (Sarkar et al., 2025) by ‡, and (Wang et al., 2024a) by *.

Method	Existence	Attribute	State	Number	Action	Relation	Overall
LLaVA-1.5-7B	87.4	77.7	75.3	80.7	84.2	58.5	78.8
+ LLaVA-RLHF (Sun et al., 2024)	67.4	73.2	72.0	72.0	83.8	64.7	70.2
+ POVID (Zhou et al., 2024a)	87.8	77.3	75.4	79.1	84.1	55.5	78.4
+ SENTINEL (Peng et al., 2025)	**91.4**	80.0	77.4	**82.9**	**87.6**	48.7	80.3
+ AVES-DPO (Ours)	90.6	**80.4**	**79.1**	81.0	87.0	**66.6**	**82.3**
LLaVA-1.5-13B	83.8	81.9	**80.0**	84.8	85.7	63.0	80.4
+ LLaVA-RLHF (Sun et al., 2024)	**93.8**	77.1	73.7	81.8	85.4	60.5	81.0
+ SENTINEL (Peng et al., 2025)	85.5	**82.2**	79.9	**86.1**	85.5	65.6	**81.4**
+ AVES-DPO (Ours)	82.6	79.6	76.0	84.8	**87.8**	**77.5**	80.4

Table 2: Detailed evaluation results on the discriminative tasks of the AMBER benchmark.

5.2k and 4.6k preference pairs for the 7B and 13B models, respectively.

Evaluation metrics. We evaluate on four benchmarks: Object HalBench (Rohrbach et al., 2018) for standard object hallucination metrics, AMBER (Wang et al., 2023) for comprehensive analysis covering both generative and discriminative tasks, MMHal-Bench (Sun et al., 2024) for GPT-4-based quality assessment in open-ended evaluations, and the MME Benchmark (Fu et al., 2025) to assess multimodal generalization capabilities. Details are provided in the Appendix C.

Baselines. We compare our proposed method against several state-of-the-art (SOTA) techniques. Specifically, we selected VCD (Leng et al., 2024) as a representative method for decoding strategy intervention. Additionally, we included LLaVA-RLHF (Sun et al., 2024), HALVA (Sarkar et al., 2025), oDPO (He et al., 2025), mDPO (Wang et al., 2024a), and SENTINEL (Peng et al., 2025) as comparative baselines, representing prominent approaches in preference optimization.

Implementation details. We apply Low-Rank Adaptation (LoRA) and use AdamW as the optimizer. For both the 7B and 13B models, we set the LoRA rank to 128 and alpha to 256, training for 1 epoch. The parameter β in AVES-DPO, which controls the deviation from the reference model, is set to 0.1. All experiments are conducted on a single NVIDIA RTX A6000 Ada GPU. Further details are provided in the Appendix B.

5.2 Main Results

Findings 1: Superior Performance with High Data Efficiency. Table 1 presents results on three standard hallucination benchmarks. Despite utilizing a smaller dataset than baseline methods, AVES-DPO demonstrates strong performance across multiple evaluation dimensions. For the 7B model, AVES-DPO achieves the best performance on 60% of the evaluation metrics. For the 13B model, it achieves the lowest Object-Hal scores with only

Method	Existence	Count	Position	Color	Posters	Celebrity	Scene	Landmark	Artwork	OCR	Overall
LLaVA-1.5-7B	**195.0**	**158.3**	123.3	155.0	128.6	133.2	155.5	161.0	**120.8**	125.0	1455.7
+ POVID (Zhou et al., 2024a)	190.0	**158.3**	123.3	150.0	128.6	133.5	154.8	161.8	**120.8**	125.0	1446.0
+ SENTINEL (Peng et al., 2025)	185.0	**158.3**	**133.3**	**165.0**	129.6	135.0	158.8	**162.5**	120.3	125.0	1472.8
+ AVES-DPO (Ours)	**195.0**	153.3	123.3	155.0	**144.6**	**138.8**	**164.3**	159.5	117.3	**140.0**	**1491.1**
LLaVA-1.5-13B	**195.0**	**153.3**	**120.0**	170.0	**158.2**	**168.2**	158.8	147.8	124.3	**132.5**	**1528.0**
+ SENTINEL (Peng et al., 2025)	**195.0**	143.3	**120.0**	**180.0**	156.8	167.1	161.0	138.0	**126.8**	**132.5**	1520.4
+ AVES-DPO (Ours)	190.0	140.0	100.0	155.0	153.7	127.4	**164.0**	**159.3**	101.8	72.5	1363.6

Table 3: Evaluation results on the MME benchmark for general multimodal perception and recognition capabilities.

Method	Object-Hal	
	$CHAIR_S\downarrow$	$CHAIR_I\downarrow$
LLaVA-1.5-7B	51.4	14.8
+ Hallucination Verification (Ours)	40.2	12.2
LLaVA-1.5-13B	50.4	14.3
+ Hallucination Verification (Ours)	25.4	7.6
GPT-4V	29.0	7.9
+ Hallucination Verification (Ours)	23.9	5.6

Table 4: Effectiveness and robustness analysis of the proposed Hallucination Verification strategy.

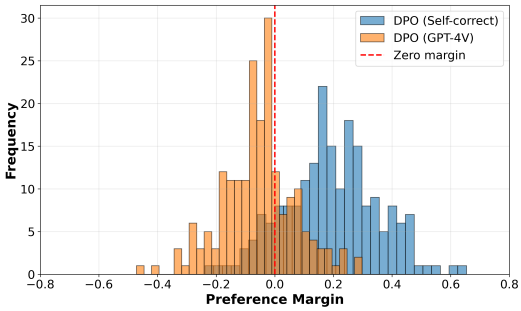


Figure 4: Distribution of preference margins. External GPT-4V supervision (orange) results in margins centered near zero due to distribution mismatch. In contrast, our approach (blue) leverages in-distribution data to induce a significant positive shift.

Method	Object-Hal		AMBER		
	$CHAIR_S\downarrow$	$CHAIR_I\downarrow$	CHAIR ↓	Hal Rate ↓	Cog ↓
LLaVA-1.5-7B	51.4	14.8	7.1	32.5	3.8
+ GPT-4V	38.4	11.4	7.6	31.7	3.6
+ AVES-DPO (Ours)	**12.2**	**3.9**	**3.3**	**12.6**	**0.9**

Table 5: Performance comparison between external supervision from proprietary models and our self-correction approach.

4.6k training samples. These results validate that AVES-DPO eliminates the dependency on proprietary API calls while achieving superior performance with minimal data, making it a practical and scalable solution for hallucination mitigation.

Findings 2: Improvements in Fine-grained Performance, especially in relation. Table 2 presents the fine-grained evaluation results on the AMBER discriminative task. For the 7B model, AVES-DPO demonstrates consistently superior performances, particularly achieving substantial gains in the relation category where existing methods often struggle. While SENTINEL shows performance degradation compared to the base model, AVES-DPO achieves a robust score of 66.6. For the 13B model, although overall score remains competitive, our method achieves the highest scores in relation category with a score of 77.5, mirroring the trend observed in the 7B model. These consistent improvements in relational reasoning across both model scales verify that our proposed verification and self-correction framework effectively captures complex inter-object relationships.

Findings 3: Enhancement of General Multimodal Capabilities. To assess generalization capability across diverse domains, we evaluate on the MME Benchmark, which comprises various subtypes. As shown in Table 3, for the 7B model, AVES-DPO achieves an overall score of 1491.1, demonstrating a significant performance improvement compared to existing methods. This result confirms that our framework not only mitigates hallucinations but also enhances fundamental multimodal perception abilities. However, we observed a performance trade-off in the 13B model. Given that MME includes subtasks heavily reliant on external knowledge, this pattern suggests that the 13B model becomes more conservative during the hallucination suppression, prioritizing factual grounding over broad knowledge retrieval. This highlights a challenge in balancing the hallucination mitigation with the preservation of extensive multimodal capabilities when scaling to larger models.

5.3 Ablation Study and Analysis

Robustness of verification strategy. To assess the efficacy and robustness of our Hallucination Verifi-

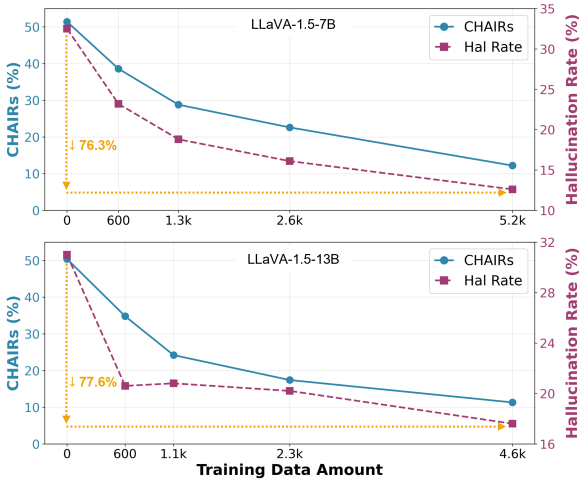


Figure 5: Impact of training data size on hallucination mitigation performance.

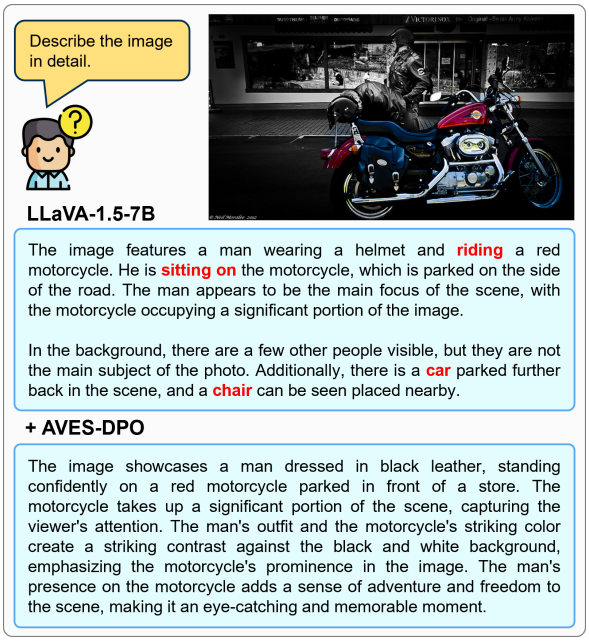


Figure 6: Qualitative comparison demonstrating the effectiveness of AVES-DPO in mitigating hallucinations.

cation process, we conduct experiments on LLaVA-1.5-7B, LLaVA-1.5-13B, and GPT-4V. Each model detects and rectifies hallucinations in initial responses, and we measure the resulting reduction in hallucination rates. As shown in Table 4, our approach consistently reduces both sentence- and object-level hallucination rates across all models. Notably, LLaVA-1.5-13B achieves relative reductions of 49.6% and 46.9% in CHAIRs and CHAIRi, respectively. Furthermore, even for GPT-4V, which already exhibits relatively low hallucination rates, our verification strategy yields additional gains. These results demonstrate that our verification strategy is model-agnostic and robust across varying model capacities.

Effect of distribution mismatch mitigation. To demonstrate the impact of utilizing in-distribution data, we train a comparative model following the setup in Section 3. Specifically, we construct an additional baseline dataset of 5.2k samples using responses generated by proprietary models. We then visualize the preference margin $\mathcal{M}$ to quantify the distributional alignment of each correction source. As shown in Figure 4, the distribution for proprietary models concentrates predominantly in the negative region. In contrast, our self-correction approach exhibits a significant positive shift, demonstrating that LVLMs can successfully align with preference signals within their own distribution. This disparity translates into empirical performance: as shown in Table 5, supervision from proprietary models yields only marginal improvement, while our in-distribution training achieves a significant reduction in hallucination rates. These results confirm that leveraging in-distribution data enables more effective preference alignment than external supervision. This finding is further supported by the preference margin distribution of AVES-DPO in Figure 2, which shifts into the positive region, indicating that the model successfully learns to prefer factual responses.

Effect of data scale. To validate the data efficiency of AVES-DPO, we analyze performance across varying training data sizes, as illustrated in Figure 5. Notably, even with only 600 samples, AVES-DPO achieves a significant reduction in hallucination rates. We attribute this efficiency to our self-generated responses, which ensure the training data remains in-distribution, enabling the model to learn effective hallucination mitigation patterns even from small-scale datasets. Detailed quantitative results across larger dataset sizes are provided in Appendix D.1.

Qualitative case study. To further validate our proposed method, we present a case study in Figure 6. The baseline model exhibits severe hallucinations, generating incorrect relationships and non-existent objects such as 'car' and 'chair'. In contrast, AVES-DPO accurately grounds the visual content, providing a more detailed and factually correct description. This demonstrates that our approach effectively mitigates hallucination while enhancing descriptive capabilities.

6 Conclusion

In this study, we proposed AVES-DPO, a novel framework that mitigates diverse hallucinations in LVLMs via verified self-correction. By generating preference pairs derived from the models' intrinsic knowledge, our approach effectively resolves the distribution mismatch problem inherent in external feedback methods. Extensive experiments demonstrate that AVES-DPO significantly outperforms baselines on multiple benchmarks, achieving exceptional data efficiency with only 5.2k samples.

Limitations

Despite the effectiveness of AVES-DPO, certain limitations remain that provide avenues for future research. First, our verification of attributes and relationships was strictly constrained to the single-object level to ensure rigorous evaluation, leaving complex scenarios involving multiple co-occurring objects for subsequent investigation. Second, while the 13B model successfully achieved a substantial reduction in hallucination rates, this came at the cost of a slight performance decline on general benchmarks. This phenomenon suggests that as model scale increases, the trade-off between suppressing hallucinations and preserving extensive pre-trained knowledge becomes more pronounced. Consequently, our findings point toward a promising direction for future work in developing scale-aware alignment strategies. Such approaches would aim to mitigate the potential issues inherent in learning from self-generated data while ensuring more comprehensive and robust performance enhancements across models of varying sizes.

Acknowledgements

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [RS-2021-II211341, Artificial Intelligence Graduate School Program (Chung-Ang University)] and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00556246). This research was supported by the "Regional Innovation System & Education (RISE)" through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government. (2025-RISE-01-024-06).

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, and 1 others. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Elmira Amirloo, Jean-Philippe Fauconnier, Christoph Roesmann, Christian Kerl, Rinu Boney, Yusu Qian, Zirui Wang, Afshin Dehghan, Yinfei Yang, Zhe Gan, and Peter Grasch. 2024. Understanding alignment in multimodal llms: A comprehensive study. *Preprint*, arXiv:2407.02477.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, and 45 others. 2025. Qwen3-vl technical report. *Preprint*, arXiv:2511.21631.

Ralph Allan Bradley and Milton E. Terry. 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, and Ying Shan. 2024. Yolo-world: Real-time open-vocabulary object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16901–16911.

Wei-Lin Chiang, Zhuohan Li, Ziqing Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, and 1 others. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*, 2(3):6.

Alberto Compagnoni, Davide Caffagni, Nicholas Moratelli, Lorenzo Baraldi, Marcella Cornia, and Rita Cucchiara. 2025. Mitigating hallucinations in multimodal llms via object-aware preference optimization. *arXiv preprint arXiv:2508.20181*.

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. 2025. MME: A comprehensive evaluation benchmark for multimodal large language models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Yixiao He, Haifeng Sun, Pengfei Ren, Jingyu Wang, Huazheng Wang, Qi Qi, Zirui Zhuang, and Jing Wang. 2025. Evaluating and mitigating object hallucination in large vision-language models: Can they still see removed objects? In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of*

the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 6841–6858, Albuquerque, New Mexico. Association for Computational Linguistics.

Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, and Nenghai Yu. 2024. Opera: Alleviating hallucination in multi-modal large language models via over-trust penalty and retrospection-allocation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13418–13427.

Seongyun Lee, Sue Hyun Park, Yongrae Jo, and Minjoon Seo. 2024. Volcano: Mitigating multimodal hallucination through self-feedback guided revision. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 391–404, Mexico City, Mexico. Association for Computational Linguistics.

Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, and Lidong Bing. 2024. Mitigating object hallucinations in large vision-language models through visual contrastive decoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13872–13882.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023a. BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 19730–19742. PMLR.

Yifan Li, Yifan Du, Kun Zhou, Jinpeng Wang, Xin Zhao, and Ji-Rong Wen. 2023b. Evaluating object hallucination in large vision-language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 292–305, Singapore. Association for Computational Linguistics.

Shaoqing Lin, Chong Teng, Fei Li, Donghong Ji, Lizhen Qu, and Zhuang Li. 2025. Discosg: Towards discourse-level text scene graph parsing through iterative graph refinement. *arXiv preprint arXiv:2506.15583*.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Computer Vision – ECCV 2014*, pages 740–755, Cham. Springer International Publishing.

Fuxiao Liu, Kevin Lin, Linjie Li, Jianfeng Wang, Yaser Yacoob, and Lijuan Wang. 2024a. Mitigating hallucination in large multi-modal models via robust instruction tuning. In *The Twelfth International Conference on Learning Representations*.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024b. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26296–26306.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36, pages 34892–34916. Curran Associates, Inc.

Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. 2025. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *Computer Vision – ECCV 2024*, pages 38–55, Cham. Springer Nature Switzerland.

Yassine Ouali, Adrian Bulat, Brais Martinez, and Georgios Tzimiropoulos. 2025. Clip-dpo: Vision-language models as a source of preference for fixing hallucinations in lvlms. In *Computer Vision – ECCV 2024*, pages 395–413, Cham. Springer Nature Switzerland.

Eunkyu Park, Minyeong Kim, and Gunhee Kim. 2025. Halloc: Token-level localization of hallucinations for vision language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 29893–29903.

Shangpin Peng, Senqiao Yang, Li Jiang, and Zhuotao Tian. 2025. Mitigating object hallucinations via sentence-level early intervention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 635–646.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741. Curran Associates, Inc.

Anna Rohrbach, Lisa Anne Hendricks, Kaylee Burns, Trevor Darrell, and Kate Saenko. 2018. Object hallucination in image captioning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4035–4045, Brussels, Belgium. Association for Computational Linguistics.

Pritish Sahu, Karan Sikka, and Ajay Divakaran. 2024. Pelican: Correcting hallucination in vision-LLMs via claim decomposition and program of thought verification. In *Proceedings of the 2024 Conference on*

Empirical Methods in Natural Language Processing, pages 8228–8248, Miami, Florida, USA. Association for Computational Linguistics.

Pritam Sarkar, Sayna Ebrahimi, Ali Etemad, Ahmad Beirami, Sercan O Arik, and Tomas Pfister. 2025. Mitigating object hallucination in MLLMs via data-augmented phrase-level alignment. In *The Thirteenth International Conference on Learning Representations*.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Zhiqing Sun, Sheng Shen, Shengcao Cao, Haotian Liu, Chunyuan Li, Yikang Shen, Chuang Gan, Liangyan Gui, Yu-Xiong Wang, Yiming Yang, Kurt Keutzer, and Trevor Darrell. 2024. Aligning large multimodal models with factually augmented RLHF. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13088–13110, Bangkok, Thailand. Association for Computational Linguistics.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, and 1 others. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

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

Junyang Wang, Yuhang Wang, Guohai Xu, Jing Zhang, Yukai Gu, Haitao Jia, Jiaqi Wang, Haiyang Xu, Ming Yan, Ji Zhang, and 1 others. 2023. Amber: An llm-free multi-dimensional benchmark for mllms hallucination evaluation. *arXiv preprint arXiv:2311.07397*.

Lei Wang, Jiabang He, Shenshen Li, Ning Liu, and Ee-Peng Lim. 2024b. Mitigating fine-grained hallucination by fine-tuning large vision-language models with caption rewrites. In *MultiMedia Modeling*, pages 32–45, Cham. Springer Nature Switzerland.

Yuxi Xie, Guanzhen Li, Xiao Xu, and Min-Yen Kan. 2024. V-DPO: Mitigating hallucination in large vision language models via vision-guided direct preference optimization. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13258–13273, Miami, Florida, USA. Association for Computational Linguistics.

Zhihe Yang, Xufang Luo, Dongqi Han, Yunjian Xu, and Dongsheng Li. 2025. Mitigating hallucinations in large vision-language models via dpo: On-policy data hold the key. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10610–10620.

Shukang Yin, Chaoyou Fu, Sirui Zhao, Tong Xu, Hao Wang, Dianbo Sui, Yunhang Shen, Ke Li, Xing Sun, and Enhong Chen. 2024. Woodpecker: Hallucination correction for multimodal large language models. *Science China Information Sciences*, 67(12):220105.

Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun, and Tat-Seng Chua. 2024. 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 (CVPR)*, pages 13807–13816.

Zhiyuan Zhao, Bin Wang, Linke Ouyang, Xiaoyi Dong, Jiaqi Wang, and Conghui He. 2023. Beyond hallucinations: Enhancing lvlms through hallucination-aware direct preference optimization. *arXiv preprint arXiv:2311.16839*.

Yiyang Zhou, Chenhang Cui, Rafael Rafailov, Chelsea Finn, and Huaxiu Yao. 2024a. Aligning modalities in vision large language models via preference fine-tuning. In *ICLR 2024 Workshop on Reliable and Responsible Foundation Models*.

Yiyang Zhou, Chenhang Cui, Jaehong Yoon, Linjun Zhang, Zhun Deng, Chelsea Finn, Mohit Bansal, and Huaxiu Yao. 2024b. Analyzing and mitigating object hallucination in large vision-language models. In *The Twelfth International Conference on Learning Representations*.

A Method Details

A.1 Details for Hallucination Verification

Scene Graph Extraction. We employ the DiscoSG framework (Lin et al., 2025) to parse textual responses into structured scene graphs. Specifically, we utilize flan-t5-large-VG-factual-sg as the generator and DiscoSG-Refiner-Large-t5-only as the refiner, executing two iterative refinement rounds to ensure graph quality.

Vocabulary Construction. To overcome the limited category coverage of MS-COCO, we expand our vocabulary by integrating frequent concepts from the GQA dataset. The final vocabulary comprises the original 80 COCO objects augmented with 68 GQA objects, along with 38 attributes and 17 relations. Crucially, all terms are organized into semantic subtypes to facilitate the precise retrieval of correction candidates. The detailed vocabulary lists are provided in Tables 6, 7, and 8.

O/A/R Verification. *(i) Object Verification.* Our pipeline relies exclusively on open-source models. For primary object detection, we utilize Grounding DINO-base and YOLOv8x-worldv2 with default configurations. To minimize interference between semantically similar queries, we implement a grouping strategy for object categories exhibiting a text embedding cosine similarity of 0.5 or higher, calculated using sentence-transformers. Furthermore, to ensure rigorous verification, scenes containing multiple instances of the same object class are excluded. Finally, discrepancies in detection results are resolved using the FP8-quantized Qwen3-VL series (30B and 32B) as a secondary verification step. The specific prompt employed to resolve these ambiguous objects is presented in Table 16.

(ii) Attribute & Relation Verification. We employed the Qwen series models to verify the factual consistency of attributes and relationships. The specific prompt templates employed for attribute and relationship verification are presented in Table 17 and Table 18, respectively. Upon detecting hallucinations, the models were instructed to generate replacements from the corresponding semantic subtypes within our predefined vocabulary. To ensure data reliability, we enforced a strict consensus mechanism. Only corrections where both models yielded identical outputs were retained, while discrepant samples were excluded to minimize noise.

A.2 Details for Self-correction

Factual Rectification. Guided by the error signals derived from the verification phase, the model self-refines its initial response to eliminate or rectify hallucinations. The specific prompt template utilized for this correction is presented in Table 19. This approach allows for flexible sentence reconstruction while preserving the model's intrinsic linguistic style and vocabulary.

Detailed Enrichment. While factually accurate, the verified captions may become monotonous or lack sufficient visual detail. To address this and construct high-quality preferred data for DPO training, we implement a Detailed Enrichment phase. In this step, the LVLMs expand the corrected concise response by incorporating missing visual details grounded in the image. Crucially, to prevent the recurrence of errors, the model is explicitly prohibited from mentioning objects previously identified as hallucinations. The specific prompt template used for enriching the caption with visual details while avoiding hallucinations is presented in Table 20. To secure a sufficient number of high-quality samples, we iterated this generation and filtering pipeline. Specifically, the process was repeated 3 times for the LLaVA-1.5-7B model and 5 times for the LLaVA-1.5-13B model.

B Training Details

B.1 Training Dataset

Dataset Statistics. To ensure robust verification signals, we construct distinct training datasets for the 7B and 13B models.

- *7B dataset*: comprises 5.2k samples, encompassing the verification of 19,610 objects, 2,562 attributes, and 897 relationships.
- *13B dataset*: consists of 4.6k samples, covering 16,983 objects, 2,862 attributes, and 1,098 relationships.

Hallucination Distribution. We analyze the prevalence of hallucinations across semantic categories, as illustrated in Figure 7. The distribution varies by model scale. The 7B model exhibits a higher object hallucination rate of 37.9% compared to 34.3% for the 13B model. Conversely, the 13B model shows increased hallucination rates for fine-grained details, recording 20.5% for attributes and

Source	Count	Object List
MS-COCO	80	person, bicycle, car, motorcycle, airplane, bus, train, truck, boat, traffic light, fire hydrant, stop sign, parking meter, bench, bird, cat, dog, horse, sheep, cow, elephant, bear, zebra, giraffe, backpack, umbrella, handbag, tie, suitcase, frisbee, skis, snowboard, sports ball, kite, baseball bat, baseball glove, skateboard, surfboard, tennis racket, bottle, wine glass, cup, fork, knife, spoon, bowl, banana, apple, sandwich, orange, broccoli, carrot, hot dog, pizza, donut, cake, chair, couch, potted plant, bed, dining table, toilet, tv, laptop, mouse, remote, keyboard, cell phone, microwave, oven, toaster, sink, refrigerator, book, clock, vase, scissors, teddy bear, hair drier, toothbrush
GQA	68	tree, building, sky, pole, window, table, door, fence, wheel, floor, jacket, hat, shoe, leaf, letter, plate, flower, bag, helmet, rock, boy, cloud, roof, cap, girl, bush, mirror, box, shelf, pillow, trunk, plant, lamp, wing, seat, house, counter, street light, glove, flag, cabinet, bike, child, container, sock, towel, mountain, basket, phone, animal, sticker, lady, license plate, cheese, wire, beach, desk, curtain, dress, tower, stone, blanket, drawer, ocean, t-shirt, television, trash can, computer

Table 6: List of the 148 object classes used for hallucination verification.

Subtype	Attributes
Color	white, black, green, blue, brown, red, gray, yellow, orange, silver, pink, purple, blond, gold, beige, light brown, light blue, dark brown, cream colored, maroon, dark blue, black and white, khaki
Material	wooden, metal, brick
Posture	standing, sitting, lying, sleeping
Condition/State	open, closed, empty, full, wet, dry, cut, uncut

Table 7: List of the 38 attribute terms used for hallucination verification.

41.9% for relationships. Notably, relationships consistently represent the most challenging category across both model scales, highlighting the difficulty of capturing inter-object dynamics.

B.2 Training Setup

We fine-tuned LLaVA-1.5 (7B, 13B) using LoRA. The detailed training hyperparameters are presented in Table 9.

C Evaluation Details

C.1 Evaluation Benchmarks

In this section, we elaborate on the specific evaluation metrics and calculation methods for each benchmark used to measure the model's hallucination mitigation performance.

Object HalBench. Object HalBench is widely adopted benchmark for evaluating object hallucinations in image descriptions. In this study, we generate captions using the standard image captioning prompt: "Describe the image in detail." The evaluation is performed on 500 samples randomly sampled from the MS-COCO validation set, quantified by two key metrics: sentence-level (CHAIR_S) and object-level (CHAIR_I). Note that object detection within LVLMs' responses is performed using an exact matching strategy.

- *CHAIR_S (Sentence-level Hallucination Rate):* Calculated as the ratio of captions containing at least one hallucinated object among all generated captions.

$$\text{CHAIR}_S = \frac{\#\{\text{captions with hallucinated objects}\}}{\#\{\text{all captions}\}}$$

- *CHAIR_I (Object-level Hallucination Rate):* Calculated as the ratio of hallucinated objects

Subtype	Relations
Spatial	on, under, behind, in front of, next to, on the left side of, on the right side of
Pose-relation	sitting on, standing on, leaning on, sitting next to, standing next to, standing in front of, standing to the right of, standing to the left of, standing in front of, standing behind

Table 8: List of the 17 relation predicates used for hallucination verification.

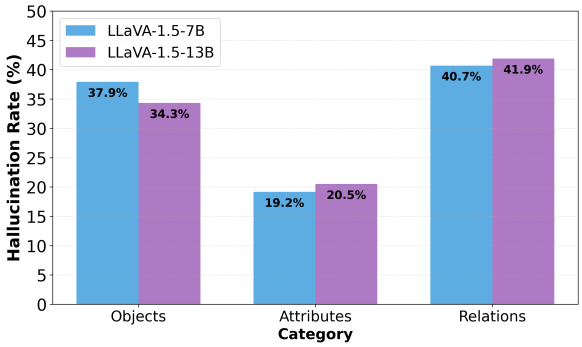


Figure 7: Comparative Analysis of Hallucination Rates. Comparison of hallucination rates across objects, attributes, and relations for LLaVA-1.5-7B and LLaVA-1.5-13B models.

Setting \ Model	LLaVA-1.5-7B	LLaVA-1.5-13B
LLM	Vicuna-1.5-7B	Vicuna-1.5-13B
Vision encoder	CLIP ViT-L 336px/14	
Projector	mlp2x_gelu	
Learning rate	2e-6	3.2e-6
Global batch size	16	8
Trainable parameters	LORA trains only LLM's linear layers	
LoRA rank r	128	
LoRA alpha	256	
Optimizer	AdamW	
Epochs	1	
Memory optimization	ZeRO stage 2	

Table 9: Detailed training hyperparameters for AVES-DPO.

to the total number of objects mentioned by the model.

$$\text{CHAIR}_I = \frac{\#\{\text{hallucinated objects}\}}{\#\{\text{all mentioned objects}\}}$$

AMBER. AMBER is a comprehensive benchmark encompassing both generative and discriminative tasks to analyze hallucination phenomena from multiple perspectives.

(Generative Task) This task evaluates the frequency of hallucinations in model-generated responses using four metrics:

- *CHAIR*: Measures the accuracy by calculating the intersection ratio between the set of objects mentioned by the model and the set of objects actually present in the image.
- *Cover*: Measures the object coverage by calculating the proportion of actual objects present in the image that are successfully mentioned by the model.
- *Hal Rate*: Measures the proportion of responses containing hallucinations. A response is considered hallucinated if its CHAIR score exceeds 0.
- *Cog*: Evaluates the alignment between the model's hallucinations and human cognitive tendencies. It is calculated as the probability that a pre-defined set of hallucination-prone objects appears in the generated output.

(Discriminative Task) This task assesses the severity of hallucinations across six dimensions: existence, attributes, relationships, state, number, and actions. We report the model's overall performance on these aspects using Accuracy and F1 score.

MMHal-Bench. MMHal-Bench is a Question-Answering benchmark employing GPT-4 (`gpt-4-0613`) for evaluation. Following the official protocol, model responses are scored on a scale of 0 to 6 based on factual accuracy and the hallucination presence. To calculate the final Hallucination Rate, we determine the proportion of responses that receiving a score of less than 3.

MME Benchmark. MME is a comprehensive evaluation benchmark designed to assess the perception and cognition capabilities of Multi-modal Large Language Models (MLLMs). To verify the generalization ability of our proposed method be-

Method	Overall ↑	Hall Rate ↓	Score in Each Question Type ↑							
			Attribute	Adv	Comp	Count	Rel	Env	Holistic	Other
LLaVA-1.5-7B (Liu et al., 2024b)	2.22	0.57	2.33	2.00	1.09	**2.33**	2.17	**2.09**	**3.08**	2.58
+LLaVA-RLHF (Sun et al., 2024)	2.06	0.64	**2.75**	**2.42**	1.17	1.42	**3.17**	1.75	2.00	1.83
+POVID (Zhou et al., 2024a)	2.18	0.57	2.58	2.33	1.25	1.25	1.92	1.42	2.92	**3.75**
+SENTINEL (Peng et al., 2025)	2.04	0.58	2.33	2.17	1.08	1.08	1.92	1.67	2.50	3.58
+AVES-DPO (Ours)	**2.35**	**0.53**	2.67	2.33	**2.17**	0.83	2.75	1.83	2.83	3.42
LLaVA-1.5-13B (Liu et al., 2024b)	2.30	0.54	**3.33**	**2.83**	1.83	1.58	1.92	1.42	2.08	3.42
+LLaVA-RLHF (Sun et al., 2024)	2.14	0.66	2.42	1.67	1.83	1.08	**2.42**	**2.33**	1.75	3.58
+SENTINEL (Peng et al., 2025)	**2.48**	**0.49**	**3.33**	2.42	**3.08**	**2.00**	1.67	1.50	2.50	3.33
+AVES-DPO (Ours)	2.36	0.58	2.33	2.58	2.42	1.67	1.58	1.83	**2.83**	**3.67**

Table 10: Fine-grained results on MMHal-Bench.

Method	Random		Popular		Adversarial		Overall	
	Acc ↑	F1 ↑	Acc ↑	F1 ↑	Acc ↑	F1 ↑	Acc ↑	F1 ↑
LLaVA-1.5-7B (Liu et al., 2024b)	**87.3**	**86.3**	85.6	**84.6**	82.9	**82.2**	85.3	**84.3**
+LLaVA-RLHF (Sun et al., 2024)	84.2	82.3	81.9	80.2	78.9	77.6	81.7	80.0
+SENTINEL (Peng et al., 2025)	86.9	85.6	85.5	84.2	82.7	81.8	85.0	83.8
+AVES-DPO (Ours)	86.8	85.0	**85.7**	84.0	**83.7**	**82.2**	**85.4**	83.7
LLaVA-1.5-13B (Liu et al., 2024b)	**90.1**	**89.6**	**88.4**	**88.1**	84.4	**84.6**	**87.6**	**87.4**
+LLaVA-RLHF (Sun et al., 2024)	84.4	81.9	82.9	80.5	81.0	78.8	82.8	80.4
+SENTINEL (Peng et al., 2025)	89.7	89.1	88.1	87.6	**84.8**	**84.6**	87.5	87.0
+AVES-DPO (Ours)	89.2	88.9	87.3	87.1	84.0	84.2	86.8	86.7

Table 11: POPE evaluation benchmark. Accuracy denotes the accuracy of predictions. "Yes" represents the probability of the model outputting a positive answer.

yond simple hallucination mitigation, we conduct evaluations on subsets relevant to visual grounding and detailed understanding, including existence, count, position, and color. The total score is derived from the aggregated accuracy across these subtasks, serving as an indicator of the model's overall multimodal robustness.

C.2 Evaluation Results

Results on MMHal-Bench. Table 10 summarizes the quantitative results on MMHal-Bench. For the 7B model, our method exhibits generally superior performance compared to the baselines. Notably, it demonstrates significant gains in the 'Attribute', 'Comparison', and 'Relation' categories. This improvement aligns well with the specific types of hallucinations targeted by our proposed framework. Conversely, the LLaVA-1.5-13B model achieves the second-highest performance among the compared methods. While the improvement margin was narrower than that of the 7B model, it maintained a robust performance level, indicating competitive stability.

Results on POPE Benchmark. Table 11 presents the quantitative results on the POPE benchmark. Consistent with the fine-grained results observed in MMHal-Bench, the performance dynamics vary by model scale. The 7B model demonstrates clear improvements over the baseline, achieving the highest overall accuracy of 85.4 and notably excelling in the Adversarial setting. On the other hand, while the 13B model shows a marginal performance trade-off compared to the strong baseline, it maintains high stability and outperforms other alignment method such as LLaVA-RLHF.

D Additional Ablations and Analysis

D.1 Extended Scalability Analysis

To further evaluate the scalability of our pipeline, we expanded the preference dataset up to 10.9k pairs by incorporating the GQA dataset. As shown in Table 12, while our automated pipeline scales effectively, we identify approximately 5.2k pairs as a practical sweet spot.

Beyond this scale, hallucination mitigation exhibits diminishing returns (e.g., CHAIR_S decreas-

Dataset Size	Object-Hal			AMBER			
	$CHAIR_S \downarrow$	$CHAIR_I \downarrow$	F1↑	CHAIR↓	Cover↑	Hal Rate↓	Cog↓
600	38.6	12.1	78.1	5.3	48.3	23.2	2.3
1.3k	28.8	9.2	75.6	4.3	46.3	18.8	1.9
2.6k	22.6	7.0	72.9	3.8	44.4	16.1	1.4
5.2k	12.2	3.9	72.7	3.3	40.8	12.6	0.9
8.1k	10.4	3.5	67.0	2.8	38.7	11.0	0.7
10.9k	10.6	3.5	65.3	2.8	37.8	10.1	0.6

Table 12: Effect of dataset size on hallucination mitigation performance for LLaVA-1.5-7B.

Category	y_w Win Rate ↑	y_l Win Rate ↓	Tie Rate
Object	88.3	9.1	2.6
Attribute	86.9	12.3	0.8
Relation	85.3	14.3	0.3
Overall	87.1	11.3	1.6

Table 13: Win, loss, and tie rates by category.

ing only marginally from 12.2 to 10.4). More importantly, this over-optimization toward factual precision comes at the expense of descriptive richness, evidenced by a decline in the F1 score (from 72.7 to 67.0) and AMBER Cover metric (from 40.8 to 37.8). Thus, 5.2k samples establish the optimal, data-efficient balance for mitigating hallucinations while actively preserving meaningful and informative content.

D.2 Quality Assessment of Preference Pairs

To empirically verify that our hierarchical consensus mechanism is not bottlenecked by the performance limits of individual auxiliary models, we conducted an independent evaluation using GPT-4o on 500 randomly sampled preference pairs. As shown in Table 13, our self-corrected responses achieve overwhelmingly high win rates over the initial responses across all fine-grained categories (objects, attributes, and relations). This result confirms that, despite relying on multiple open-source models, our rigorous verification pipeline successfully filters out individual model noise, consistently yielding high-quality, reliable preference pairs.

D.3 Comprehensive Evaluation with Multimodal Judges

To overcome the limitations of relying solely on text-based judgments for visual tasks, we expanded our assessment to include the multimodal judge-based MMHal-Bench-V (Amirloo et al., 2024). As demonstrated in Table 14, AVES-DPO achieves competitive performance in both hallucination reduction and overall response quality.

Method	MMHal-Bench-V	
	Score↑	Hal Rate↓
LLaVA-1.5-7B†	2.22	0.58
+ LLaVA-RLHF† (Sun et al., 2024)	1.55	0.78
+ POVID† (Zhou et al., 2024a)	2.30	0.57
+ SENTINEL† (Peng et al., 2025)	**2.38**	**0.54**
+ AVES-DPO (Ours)	2.32	0.56
LLaVA-1.5-13B†	2.47	0.52
+ LLaVA-RLHF† (Sun et al., 2024)	1.35	0.86
+ SENTINEL† (Peng et al., 2025)	**2.57**	**0.49**
+ AVES-DPO (Ours)	2.41	0.51

Table 14: Comparison on MMHal-Bench-V. For baseline algorithms with available official checkpoints, we re-evaluate the models, and these results are marked with †.

Verification	Object-Hal		AMBER			
	$CHAIR_S \downarrow$	$CHAIR_I \downarrow$	CHAIR↓	Cover↑	Hal Rate↓	Cog↓
Two-Phase	12.2	3.9	3.3	40.8	12.6	0.9
Phase-1 Only	16.6	4.8	3.5	40.9	12.8	1.0
Phase-2 Only	19.8	6.1	3.9	44.7	17.0	1.2

Table 15: Ablation results of verification strategies.

D.4 Necessity of Two-Phase Object Verification

Our two-phase object verification strategy is designed to balance precision and recall through hierarchical consensus. In the first phase, object detectors filter out explicitly hallucinated objects, forwarding only ambiguous cases to the second phase. There, two Qwen3-VL models perform strict semantic verification, effectively distinguishing actual hallucinations from simple detector failures.

To validate this mechanism, we compared the full approach against single-phase baselines on datasets of identical size. As shown in Table 15, relying on a single phase falls short: Phase-1 Only yields lower descriptive richness, and Phase-2 Only suffers from significantly higher hallucination rates. By accurately resolving ambiguous cases, our two-phase strategy achieves the lowest overall hallucination rates (e.g., a $CHAIR_S$ of 12.2) while maintaining competitive object coverage, proving its necessity for a robust verification pipeline.

E AI Assistant Usage

We have used Claude Code during the development of our research work.

System Prompt for Object Verification

You are a precise vision assistant analyzing objects in the image.

CRITICAL INSTRUCTIONS

1. Focus on the target object specified in the task.
2. Determine if the target object exists and is correctly identified in the image.

DECISION GUIDELINES

- **CORRECT**: The target object is clearly present and correctly identified in the image.
- **INCORRECT**: The target object is NOT present in the image.
- **UNCLEAR**: The object is too blurry, too small, too dark to identify, or it is ambiguous whether it matches the target object.

YOUR TASK

Based on the provided image, verify the target object below.

Object: "{object_name}"

Output:

Table 16: The prompt template used for verifying the existence of objects in the image.

System Prompt for Attribute Verification

You are a precise vision assistant analyzing object attributes in the image.

CRITICAL INSTRUCTIONS

1. Focus on the target object specified in the task.
2. Verify if the target attributes accurately describe the target object.
3. If an attribute is clearly INCORRECT:

- TRY to select a correction from the provided "ATTRIBUTE TYPE GUIDANCE" if a suitable option exists.

- If NO suitable option exists in the guidance, mark as INCORRECT without providing a correction.

DECISION GUIDELINES

- **CORRECT**: The attribute visually matches the target object.

- **INCORRECT**: The attribute does NOT match. Provide a correction from the "ATTRIBUTE TYPE GUIDANCE" list if possible. If no suitable correction exists, leave the correction empty.

- **UNCLEAR**: The attribute cannot be determined due to occlusion, ambiguity, or lack of visual evidence. Do NOT provide a correction.

YOUR TASK

Based on the provided image, verify the target attributes below.

Object: "{object_name}"

Attributes to verify: {attributes}

ATTRIBUTE TYPE GUIDANCE:

{attribute_guidance}

Output:

Table 17: The prompt template used for verifying attributes of the target object.

System Prompt for Relation Verification

You are a precise vision assistant analyzing the relationship between two objects in the image.

CRITICAL INSTRUCTIONS

1. Focus STRICTLY on the relationship between the subject and object specified in the task.
2. Verify if the target relation semantically and accurately describes the visual relationship.
3. If the relation is clearly INCORRECT:

- TRY to select a correction from the provided "RELATION TYPE GUIDANCE" if a suitable option exists.

- If NO suitable option exists in the guidance, mark as INCORRECT without providing a correction.

DECISION GUIDELINES

- **CORRECT**: The relation accurately and semantically describes the visual relationship between the objects. Consider semantic equivalence. If the relation meaning matches what you see, mark it as CORRECT.

- **INCORRECT**: The relation is clearly wrong and contradicts what you see in the image. Provide a correction from the "RELATION TYPE GUIDANCE" list if possible. If no suitable correction exists, leave the correction empty.

- **UNCLEAR**: The relationship cannot be determined due to occlusion, ambiguity, or lack of visual evidence. Do NOT provide a correction.

YOUR TASK

Based on the provided image, verify the target relation below.

Subject: "{subject_name}"

Relation: "{relation}"

Object: "{object_name}"

Relation to verify: "{subject_name} {relation} {object_name}"

RELATION TYPE GUIDANCE:

{relation_guidance}

Output:

Table 18: The prompt template used for verifying relationships between objects.

System Prompt for Caption Correction

You are an intelligent text editor. Fix the errors in the caption based on the ISSUES.

RULES

1. Remove objects: Remove ONLY the object mention and its related phrase completely. Do NOT delete an entire sentence unless the whole sentence is only about removed objects. If an object listed in ISSUES is not found in the caption, IGNORE it.

2. Replace: When 'A' -> 'B' is provided, REPLACE 'A' with 'B'.

3. Remove: Delete ONLY the specified adjective or relation phrase entirely.

4. Grammar & Style: Fix ONLY the items listed in ISSUES. NEVER add sentences stating what is missing. Output ONLY the final fixed caption directly.

YOUR TASK

ORIGINAL: {original_caption}

ISSUES:

{hallucination_info}

FIXED:

Table 19: The prompt template used for correcting captions based on identified issues.

System Prompt for Caption Enrichment

You are a precise vision assistant. Your task is to enrich the provided 'Basic Description' into a detailed, natural paragraph based on the image.

RULES

1. You must keep all the existing facts from the Basic Description exactly as they are, maintaining the original sentence structure as much as possible.

2. You should actively identify and include other objects or details that are clearly visible in the image but are missing from the Basic Description.

3. You must strictly utilize only visual evidence. Do not infer emotions or intentions.

4. You must combine the original facts and new visual details into a single, cohesive, and natural-sounding paragraph.

5. Describe ONLY what is visible. NEVER mention what is missing (e.g., strictly avoid phrases like "there is no", "not present", "does not contain", "not visible", "no visible").

Basic Description

{refined_caption}

CRITICAL WARNING (Negative Constraints)

The following objects have been confirmed as NOT present in the image. You must NEVER mention or describe them:

- {hallucinated_objects}

Enriched Description

Table 20: The prompt template used for enriching the caption with visual details while avoiding hallucinations.