Mitigating Hallucination in Multimodal Large Language Model via Hallucination-targeted Direct Preference Optimization

Yuhan Fu[1,2*], **Ruobing Xie**[2*], **Xingwu Sun**[2,3], **Zhanhui Kang**[2], **Xirong Li**[1†]
[1]Renmin University of China
[2]Large Language Model Department, Tencent [3]University of Macau
{fuyuhan, xirong}@ruc.edu.cn ruobingxie@tencent.com

Abstract

Multimodal Large Language Models (MLLMs) are known to hallucinate, which limits their practical applications. Recent works have attempted to apply Direct Preference Optimization (DPO) to enhance the performance of MLLMs, but have shown inconsistent improvements in mitigating hallucinations. To address this issue more effectively, we introduce Hallucination-targeted Direct Preference Optimization (HDPO) to reduce hallucinations in MLLMs. Unlike previous approaches, our method tackles hallucinations from their diverse forms and causes. Specifically, we develop three types of preference pair data targeting the following causes of MLLM hallucinations: (1) insufficient visual capabilities, (2) long context generation, and (3) multimodal conflicts. Experimental results demonstrate that our method achieves superior performance across multiple hallucination evaluation datasets, surpassing most state-of-the-art (SOTA) methods and highlighting the potential of our approach. Ablation studies and in-depth analyses further confirm the effectiveness of our method and suggest the potential for further improvements through scaling up.

1 Introduction

Large Language Models (LLMs) have been verified in various fields, demonstrating their potential (OpenAI, 2024; Grattafiori et al., 2024; Sun et al., 2024), while they encounter challenges such as hallucination. Multimodal Large Language Models (MLLMs) are also known to hallucinate (Bai et al., 2024). Specifically, they often produce unfaithful content that does not align with the visual input, which undermines their reliability and practicality, particularly in critical applications such as autonomous driving (Cui et al., 2024) or medical tasks (Liu et al., 2023a). Hence, addressing MLLM hallucination (**M-hallu**) is essential.

Recently, some pioneer preference optimization methods like Direct Preference Optimization (DPO) (Rafailov et al., 2024) have been introduced, which encourages the model to learn from comparisons between positive and negative samples, alleviating hallucinations (Zhao et al., 2023; Pi et al., 2024). However, most current methods cannot deliver consistent improvements across all types of M-hallu tasks (e.g., VQA and captioning tasks, as shown in our experiments of Table 1). Additionally, it appears that the model's improvement on specific tasks is closely related to the format of the training data. For instance, the data of SeVa (Zhu et al., 2024) primarily consists of VQA, which explains its strong performance on VQA-related hallucination evaluation. However, its results on captioning tasks are relatively unsatisfactory. Moreover, these methods do not explicitly consider diverse sources of M-hallu. Hence, we argue that if we focus on mitigating multimodal hallucinations, we should be able to address diverse types of hallucination causes and tasks, and design hallucination-targeted preference pairs for DPO accordingly. Our goal is to comprehensively alleviate all multimodal hallucination problems, including both discriminative tasks (e.g., VQA) and generative tasks (e.g., image captioning).

Different from the hallucinations in LLMs, M-hallu primarily arises from the following three aspects: (1) ***Insufficient visual capability***: This occurs when the MLLM's visual encoder lacks the necessary strength, being distracted by relatively unimportant visual information, leading to hallucinations; (2) ***Incapable long-context generation***: We observe that hallucinations become more pronounced as the generated content grows longer,

*Equal contribution
†Corresponding author

Findings of the Association for Computational Linguistics: ACL 2025, pages 16563–16577
July 27 - August 1, 2025

similar to long-range forgetting, which needs to be addressed in practical applications; (3) ***Multimodal conflicts***: Multimodal conflicts frequently arise in real-world scenarios due to the inevitable noises in texts and images. MLLMs are more prone to hallucinations with conflicting information existing between text and image (Liu et al., 2025).

To address the aforementioned challenges, we propose **Hallucination-targeted Direct Preference Optimization (HDPO)** to mitigate M-hallu. Our approach constructs hallucination-targeted preference pairs, specifically designed to address various forms and causes of hallucinations. Specifically, we design three types of DPO data reflecting the corresponding hallucination causes as follows: (1) For *insufficient visual capability*, during the model's autoregressive decoding, we preserve only some visual tokens with the lowest attention scores to produce targeted negative responses that reflect incorrect visual information distraction, urging MLLMs to pay attention to more effective visual information. (2) For *incapable long context generation*, we specifically select positive examples from high-quality long-form captions, while creating negative examples where the latter part of the response deviates from the image content, simulating long-form hallucinations. (3) For *multimodal conflicts*, we add conflicting information with images into prompts to generate negative examples. We provide positive and negative pairs with questions featuring conflicting prefixes to train the model to correctly respond to the question even containing conflicting information.

We conduct extensive experiments to evaluate our approach across various types of M-hallu tasks. The results demonstrate that our HDPO framework achieves the overall best performance in effectively mitigating MLLM hallucinations on various tasks. Our contributions are summarized as follows:

- We analyze three key causes behind MLLM hallucinations from visual capability, long-context generation, and multimodal conflicts aspects, offering valuable insights to guide future advancements.

- Based on these analyses, we propose a novel HDPO, aiming to jointly address all types of M-hallu tasks. To the best of our knowledge, we are the first to adopt hallucination-targeted DPO from diverse aspects with our novel DPO data construction strategies.

- Through extensive experiments on different datasets, HDPO demonstrates consistent improvements in all types of M-hallu tasks.

2 Related Work

Hallucinations in MLLMs. Recently, the rapid progress of LLMs has accelerated the MLLMs, demonstrating impressive visual understanding ability. However, they still encounter hallucinations. Lots of works have explored various approaches to mitigate M-hallu. Some training-free methods are proposed including enhancing models' decoding process (Leng et al., 2024; Huang et al., 2024; Chen et al., 2024b) and utilizing external feedback to reduce hallucinations (Yin et al., 2024; Wu et al., 2024), while other training methods enhance datasets' quality (Liu et al., 2023b). Our work belongs training category. And we will elaborate more on related preference optimization methods for improving MLLMs below.

Preference Optimization on MLLMs. Recently, preference optimization like DPO has been used to enhance models. HA-DPO (Zhao et al., 2023) views hallucinations as models' preferences. By leveraging ChatGPT (Achiam et al., 2023) alongside ground truth annotations from existing image datasets, it generates positive examples aligned with image content, while the model's original outputs serve as negative examples for direct preferences optimization. Although effective, the construction of negative examples is suboptimal, as it may not fully capture the diverse forms of M-hallu. SeVa (Zhu et al., 2024) generates negative examples by introducing noise into images and treats the model's original outputs as positive examples, constructing pairs for DPO. In addition to adding noise, BPO (Pi et al., 2024) injects errors into positive examples via the LLM backbone of MLLMs to construct negative examples. However, our experiments indicate that while these methods demonstrate strong capabilities, their performance in hallucination-related evaluations is not particularly impressive. Nonetheless, these works demonstrate the superiority of DPO in enhancing models' capabilities. Inspired by these findings, we aim to develop methods to further mitigate M-hallu from its diverse forms with hallucination-targeted direct preference optimization.

Additionally, there exist other DPO-based methods that differ in application scenario: MIA-DPO (Liu et al., 2024d) for multi-image, VideoDPO (Liu et al., 2024c) for video, and SymDPO (Jia et al., 2024) for in-context learning. All these works bring valuable insights to our community.

HDPO differs from existing methods. Unlike other existing preference optimization approaches, we primarily focus on hallucination-targeted preference optimization. We analyze and address hallucinations in MLLMs from diverse causes and forms. During the preference optimization process, the model learns to distinguish between positive and negative examples. HA-DPO enables the model to be aware of hallucinated content in its original outputs, though its effectiveness is limited to capturing the diverse range of hallucinations as the data is insufficient. In contrast, other works use general preference data, which improves overall model capability but shows inconsistency across different hallucination benchmarks. Therefore, we aim to enhance the effectiveness of DPO by constructing examples that reflect a wider range of hallucination forms and characteristics, allowing the model to align better to make fewer hallucinations.

Causes of hallucinations in MLLMs. There are substantial works exploring M-hallu, offering insightful perspectives. VCD suggests that language prior within MLLM is a key factor in inducing hallucinations. The Less is More (Yue et al., 2024) highlights that hallucinations are more prevalent in longer texts. In contrast, Eyes Wide Shut (Tong et al., 2024) identifies limitations in the current CLIP-based visual encoders used in MLLMs, showing that they fail to capture fine-grained details. Furthermore, SID (Huo et al., 2025) points out that tokens with lower weights in the early layers can trigger subsequent hallucinations. Meanwhile, PhD (Liu et al., 2025) demonstrates that M-hallu stems from conflicts between multimodal information, and counterintuitive images particularly prone to causing hallucinations. Collectively, these studies provide valuable insights into understanding and addressing M-hallu.

3 Proposed HDPO Method

3.1 Preliminaries

MLLMs utilize LLMs to predict the probability distribution of the next token for each textual input. Given a prompt x that includes both an image and a text query, MLLMs generate a corresponding text response y. By incorporating visual information, MLLMs enhance the capabilities of LLMs, enabling multimodal understanding.

Direct Preference Optimization. To better align LLMs with human preferences, preference optimization methods have been developed. Among these, Reinforcement Learning from Human Feedback (RLHF) is a widely recognized method, though it involves training a reward model, which can be quite challenging. In contrast, Direct Preference Optimization (DPO) (Rafailov et al., 2024) utilizes preferences data directly, without the need for a reward model. This makes DPO the approach we employ. Given a pre-processed preference dataset D containing x, y_c, and y_r, where x represents the input prompt, y_c is the preferred response, and y_r is the rejected response, DPO optimizes the language model through the following loss function:

$$\mathcal{L}_{\text{d}} = -\mathbb{E}_{\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y_c|x)}{\pi_{\text{ref}}(y_c|x)} - \beta\log\frac{\pi_\theta(y_r|x)}{\pi_{\text{ref}}(y_r|x)}\right)\right] \quad (1)$$

where $\pi_{\text{ref}}(y|x)$ denotes the reference policy, i.e., the language model after supervised fine-tuning, with θ as the trainable parameter.

In our approach, we apply DPO to adjust the MLLM's parameters. We propose HDPO, which constructs high-quality preference pairs related to the major causes of MLLM hallucinations with DPO to alleviate M-hallu. Note that the main contributions of HDPO lie in the discovery, analysis, and appropriate sample constructions of three representative types of M-hallu.

3.2 Overview of HDPO

The primary goal of HDPO is to broadly tackle various M-hallu issues by constructing hallucination-targeted preference pairs, rather than relying on DPO data of specific tasks. Without loss of generality, we adopt a generalized data format: image-descriptive text data, which we believe more effectively captures various forms of hallucination.

For DPO in MLLMs, we require a preference dataset D, denoted as (I, q, y_c, y_r), where I is the image, q is the question, y_c is the preferred (positive) response, and y_r is the rejected (negative) response. Currently, there are already many high-quality positive examples available, such as the refined positive examples in HA-DPO for the VG dataset, which leverage ChatGPT to enhance image annotations, and a vast number of positive examples labeled by GPT-4V in ShareGPT4V (Chen et al., 2024a). These high-quality datasets have a strong alignment with the image content, making them suitable for use as positive examples in DPO. Therefore, our focus going forward is on how to

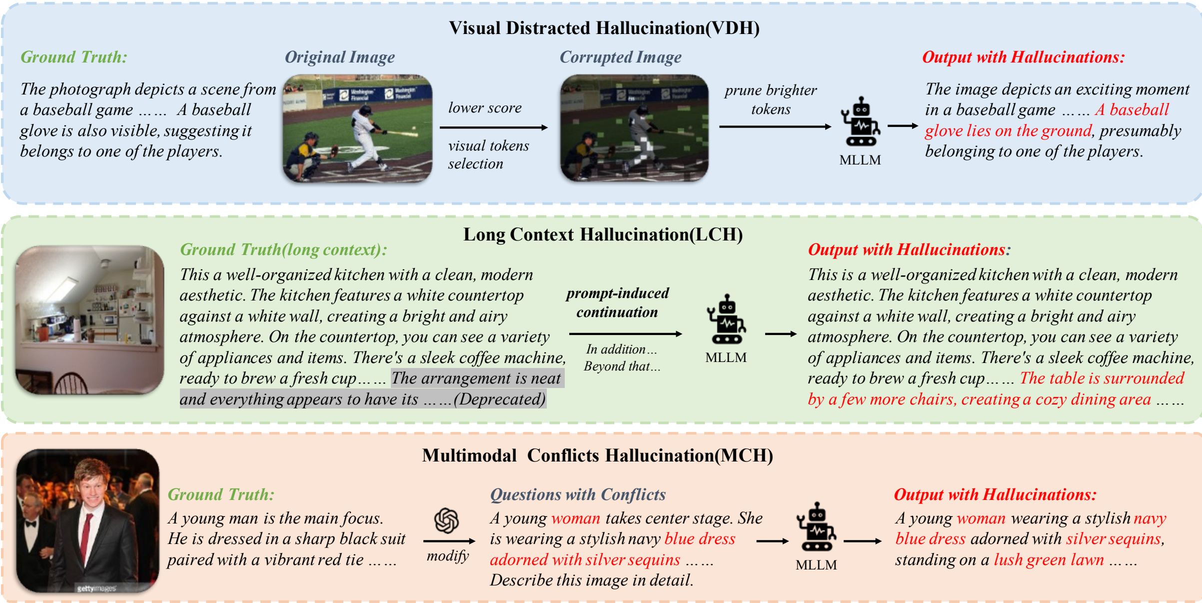

Figure 1: Illustrations of our three kinds of Hallucination-targeted Preference data. Best viewed on screen.

construct more valuable and informative negative examples, particularly those that target hallucination, which will help the model learn preferences and reduce hallucination occurrences.

To this end, we develop three types of pairwise samples specifically targeting hallucination issues: Visual Distracted Hallucination (**VDH**), Long Context Hallucination (**LCH**), and Multimodal Conflict Hallucination (**MCH**). An overview of each data type is provided in Figure 1, and further details are outlined in the sections below.

3.3 Visual Distracted Hallucination

Previous works generate negative samples by adding noise to create blurred images, while it may not always produce sufficiently effective negative samples, as indicated in appendix B. A more straightforward way is to construct negative samples using prompts, but the negative samples generated under prompt interference may fail to accurately reflect the issues related to the visual capabilities of MLLMs.

Therefore, to more precisely capture the insufficient visual capabilities of MLLMs, we propose more carefully designed novel approaches from attention perspective. Inspired by SID, we induce vision-and-text association hallucinations by leveraging vision tokens with low attention scores in the self-attention module. Formally, for the transformer block in the auto-regressive decoder, text instructions, vision inputs, and generated tokens are concatenated and projected into three vectors: Q, K, and V. The self-attention computes the relevance of each element to the others as follows to get the attention matrix:

$$\mathbf{A} = \mathrm{softmax}((\mathbf{Q} \cdot \mathbf{K}^T)/\sqrt{d} + \mathbf{M}) \tag{2}$$

where d represents the dimension of $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$, M represents the casual mask. $\mathbf{A} \in R^{(b,h,n,n)}$, where b, h, and n denote batch size, number of key-value heads, and total token number, respectively. We denote the $\mathbf{A}_i$ as the attention matrix after Layer i of MLLMs. Then we calculate vision token importance scores ($\mathrm{Score}_i(v)$) based on $\mathbf{A}_i$:

$$\mathrm{Score}_i(v) = \frac{1}{h} \sum_{j=1}^{h} \mathbf{A}_i^{(\cdot,j,\cdot,\cdot)}[-1] \tag{3}$$

During the model's auto-regressive decoding process, we retain the K vision tokens with the lowest importance scores, and the resulting decoded response serves as negative samples. By removing the most important visual token from the model in this way, it amplifies the influence of relatively irrelevant visual tokens, thus constructing visual information distracted hallucinations as negative samples, urging MLLMs to pay attention to more important visual information.

3.4 Long Context Hallucination

As previously mentioned, the occurrence of hallucinations tends to increase as models generate longer responses. To illustrate this more clearly, we present CHAIR scores by varying the 'max new tokens' parameter. As shown in Figure 2, the

CHAIR score of LLaVA-v1.5-7B exhibits a clear positive correlation with the 'max new tokens', indicating that more hallucinations are produced as the generated content increases. This issue has also been highlighted in recent studies (Yue et al., 2024).

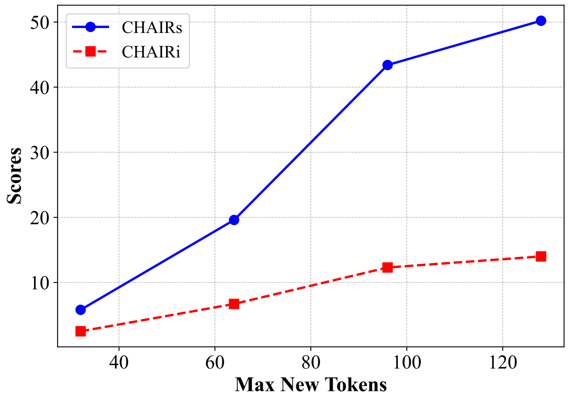

Figure 2: CHAIR scores under different max new tokens

This phenomenon is both logical and explainable. As the model generates longer texts, the proportion of text tokens gradually increases while the proportion of image tokens decreases. This shift causes the model to increasingly neglect visual tokens, resulting in descriptions that appear reasonable but fail to accurately align with the visual content. Our aim is to construct preference data that guides the model to better align its generated content with the visual input and the given question, even when generating long responses. However, existing datasets lack sufficient positive and negative pairs for long-form content and often contain noise with other factors, making them difficult to directly apply for training. *To address this, we firstly propose approach for constructing positive and negative preference pairs for long-form content, ensuring the long text hallucinations while maintaining minimal semantic divergence.*

Given our focus on relatively long-form content, the responses need to be sufficiently lengthy (high-quality long responses). For negative examples, we truncate the last two sentences from a positive example and use the preceding portion as a prefix. The MLLM then continues generating text from this prefix, which compels the model to produce common errors associated with extended text generation. This process is repeated by concatenating the newly generated content to the prefix for three iterations in a loop.

Hint Phrase. Simply providing the prefix and instructing the model to continue often results in unexpected behavior, as the model tends to conclude the response quickly, generating low-information descriptions. To address this issue, we append a 'hint phrase' to the prefix, guiding the model toward producing more informative and detailed responses. Besides that, we also modify the system prompt. Details can be seen in appendix D.2. It helps produce responses prone to more likely errors when generating long texts. By creating positive and negative pairs in this manner, we aim to use DPO to teach the model how to minimize hallucinations in long-form responses and improve alignment.

3.5 Multimodal Conflicts Hallucination

One of the more challenging yet often overlooked scenarios in mainstream evaluation tasks involves conflicts between modalities. In such cases, models may naturally favor textual content due to their autoregressive generating manner and the larger proportion of the language model component, leading to incorrect outputs. *In this paper, we bring this issue to the forefront to address and firstly use preference optimization to mitigate it.*

To be specific, we construct positive and negative pairs with conflicting prefixes and apply DPO to optimize the model. Specifically, we utilize GPT-4o-mini to rewrite details of the positive examples through prompting, generating information conflicting with the image contents. These conflicting informations are then placed at the beginning of normal questions, prompting the model to produce incorrect responses. As shown in Figure 3, the model is indeed prone to being hallucinated by the conflicting prefixes. We take the model's incorrect outputs as negative examples. Further details on the prompts can be found in Figure 9. Unlike previous types of data, the questions for training of MCH contain conflicting prefixes, as we aim for the model to generate correct responses in the query even when presented with conflicting information.

3.6 Implement details

For LCH, which requires longer responses, we sampled 6k examples with over 300 tokens from ShareGPT4V. For MCH, we randomly sampled 6k examples from ShareGPT4V. For VDH, we obtain 6k examples from ShareGPT4V and 4k examples from VG with positive examples from HA-DPO to enhance data diversity; the preserved K is 500, with other settings aligned with SID (e.g., i = 2). Details of data can be found in appendix D.

	POPE	CHAIR		AMBER				
	F1 Score ↑	CHAIR_s ↓	CHAIR_i ↓	CHAIR ↓	HalRate ↓	Cog. ↓	F1 Score ↑	AMBER-S ↑
LLaVA-v1.5-7B	86.1	51.2	14.2	7.6	35.1	4.3	74.5	83.5
Vlfeedback †	83.7	40.3	13.2	–	–	–	–	–
POVID †	86.9	35.2	8.3	–	–	–	–	–
HA-DPO	86.9	37.2	10.0	6.4	29.9	3.2	78.2	85.9
SeVa	86.8	54.6	15.9	7.4	35.6	3.2	84.1	88.3
BPO	83.1	42.2	10.1	5.0	33.5	2.0	**84.5**	89.7
CSR	**87.0**	19.6	5.4	3.8	16.9	1.4	76.0	86.1
HDPO (ours)	86.8	**16.6**	**5.1**	**3.3**	**15.8**	**0.8**	84.1	**90.4**

Table 1: Experimental results of HDPO on LLaVA-v1.5-7B compared with baselines applied on LLaVA-v1.5-7B. The best result for each metric is in bold. Some results† are referenced from Zhou et al. (2024b). The F1 of POPE and AMBER are discriminative metrics, AMBER-s is a comprehensive metric, and the others are generative metrics.

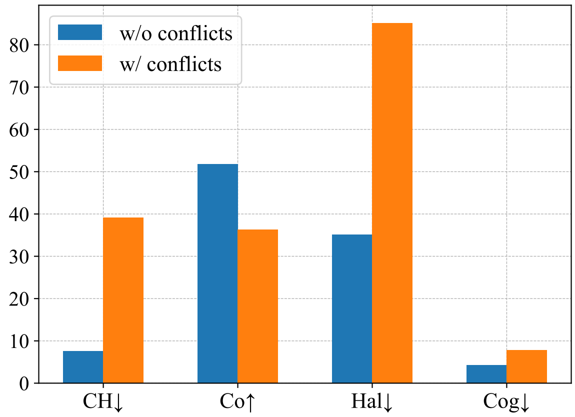

Figure 3: Performance of LLaVA-v1.5-7B w/ and w/o conflicts on AMBER, details in Section 4.4.2.

4 Experiments

In this section, we empirically investigate the evaluation of HDPO. We begin by describing the experimental settings, including the evaluation datasets and training details. Next, we present the results on various hallucination evaluation datasets, demonstrating the promising performance of HDPO. Additionally, we validate the expected functions of LCH and MCH. Finally, we provide ablation studies and conduct in-depth analyses in more detail.

4.1 Experimental Settings

Evaluation Datasets. We evaluate the effectiveness of HDPO in mitigating hallucinations across both captioning tasks and simplified visual question answering (VQA) tasks using three evaluation datasets as follows: (1) CHAIR is an evaluation method used in image captioning tasks to assess object hallucinations in model responses. There are two metrics: CHAIR_s and CHAIR_i. CHAIR_s measures hallucinations at the sentence level, while CHAIR_i measures them at the image level respectively. (2) POPE is a popular dataset for evaluating object hallucinations in MLLMs. We calculate and report the average F1 score on different splits. (3) AMBER is an LLM-free multi-dimensional benchmark, offering a cost-effective and efficient evaluation pipeline. It supports the evaluation of both generative and discriminative tasks including hallucinations related to existence, attributes, and relations. For all details of datasets and metrics used can be seen in appendix A.

Training Details. As most related works (Chen et al., 2024a; Zhu et al., 2024; Pi et al., 2024) are carried on LLaVA-v1.5 (Liu et al., 2024a), we select it as our base model for experiments, which allows for easy comparison with other existing works. Models' weights are pretrained and further fine-tuned using supervised fine-tuning (SFT) before applying HDPO. During the training phase, we employ Zero stage-3 optimization and use Vicuna-7B/13B and CLIP-VIT-L-336px as our LLM and vision encoder, respectively. The training is conducted with 2 epochs with a batch size of 64, a learning rate of 2e-6, weight decay as 0, LoRA rank as 64, and a beta value of 0.1. Besides, we also validate HDPO on InstructBLIP, further demonstrating effectiveness in Section 4.3.

Competitor. We first compare HDPO with its base model. We also select several preference learning methods, including Vlfeedback (Li et al., 2024), POVID (Zhou et al., 2024a), CLIP-DPO (Ouali et al., 2024), HA-DPO (Zhao et al., 2023), SeVa (Zhu et al., 2024), BPO (Pi et al., 2024), and CSR (Zhou et al., 2024b). Furthermore, we compare HDPO on AMBER with other MLLMs in appendix D.5.

4.2 Results on Diverse Hallucination Tasks

HDPO achieves SOTA level on both generative and discriminative hallucination tasks. The results indicate that HDPO performs well in mitigating hallucinations, achieving almost SOTA level,

	POPE	CHAIR		AMBER				
	F1 Score ↑	CHAIR_s ↓	CHAIR_i ↓	CHAIR ↓	HalRate ↓	Cog. ↓	F1 Score ↑	AMBER-S ↑
LLaVA-v1.5-13B	85.8	48.0	13.6	6.6	31.0	3.3	73.0	83.2
HA-DPO	87.3	46.0	12.1	6.0	30.7	3.0	79.1	86.6
SeVa	86.9	59.8	17.4	9.0	43.3	3.7	**84.8**	87.9
CSR	87.3	24.0	5.6	**3.6**	19.0	1.8	73.1	84.8
HDPO (ours)	**87.6**	**15.4**	**5.3**	3.8	**16.5**	**0.8**	81.2	**88.7**

Table 2: Experimental results of HDPO on LLaVA-v1.5-13B compared with baselines applied on LLaVA-v1.5-13B. More details of baselines can be seen in appendix C.

especially on generative tasks. This outcome is natural, as our data contains only descriptive content, leading to relatively strong performance on generative tasks. Since we don't specifically construct data tailored for discriminative tasks, the improvement in these tasks is not substantial. However, the overall performance remains strong, indicating that our approach, which targets the sources of hallucinations rather than specific tasks, is more effective for mitigating hallucinations. Notably, HDPO achieves **67.6%** improvement on CHAIR_s, **64.1%** improvement on CHAIR_i, **55%** enhancement on HalRate, best performance on AMBER-S. Besides, we also evaluate HDPO on a comprehensive benchmark, MM-Vet (Yu et al., 2024), where we observe a slight improvement. This aligns with our expectations, as the model is not fine-tuned on a wide range of tasks and data types, but focused on reducing hallucinations.

Brief analyses on other baselines. Some baselines lack comprehensive performance on hallucination evaluation. SeVa, though effective on AMBER's discriminative tasks, shows no improvement on generative tasks, likely due to its reliance on VQA-type data. Similarly, BPO underperforms on CHAIR. In contrast, CSR excels in generative tasks but struggles with AMBER's discriminative tasks. This indicates that while these methods enhance model performance, they do not fully optimize for hallucination, and their ability to mitigate hallucinations remains inconsistent and incomplete, while HDPO demonstrates strong performance in hallucination evaluation, as evidence of its 'hallucination-targeted' design.

Advantages of our HDPO Data. The size of our dataset also provides a relative advantage. For instance, with nearly 12% data amount compared with BPO, HDPO significantly improves model's performance on hallucination, achieving better performance than BPO on generative tasks by a large margin. Moreover, we did not construct VQA data for discriminative tasks. Nevertheless, the results are already impressive, demonstrating that our HDPO is universally effective.

4.3 Universality on Different Base Models

We also conduct experiments across different base models to verify our HDPO's universality. Specifically, we apply HDPO to the widely-used LLaVA-v1.5-13B for MLLM hallucination evaluation. The results are shown in Table 2, demonstrating that the model's performance remains consistent with expectations, with improvements in hallucination mitigation. It also implies that our generated hallucination-targeted DPO data is effective for different LLM sizes. To further validate the generalization capabilities of other MLLMs, we also conduct experiments on InstructBLIP (Liu et al., 2024b). The results in Table 5 also show consistent improvement on the overall performance.

4.4 Analyses on Different Hallucinations

The results from above experiments demonstrate our method's superior performance in mitigating hallucinations. However, do they truly work effectively in the scenarios we claim? Below, we briefly design two more challenging sub-tasks of hallucination that align with our claims, aiming to further showcase the effectiveness of our data construction of LCH and MCH. We also conduct experiments to compare VDH with adding noise in appendix B, further demonstrating effectiveness of VDH.

4.4.1 Long Context Hallucination

To evaluate the effectiveness of LCH on longer responses, we conduct an extended experiment on the AMBER generative task. Specifically, when the model is asked the question "Describe this image in detail", we append the instruction "answer in 800 words" to encourage longer responses. As indicated in Table 3, HDPO shows good and stable performance in handling longer responses, with

	CHAIR ↓	HalRate ↓	Cog. ↓
LLaVA-v1.5-7B	9.0	45.1	5.7
HA-DPO	7.5	37.6	4.4
SeVa	7.5	43.4	4.3
BPO	6.4	55.3	4.8
HDPO	**3.4**	**21.4**	**1.3**
w/o LCH	4.6	26.4	1.8

Table 3: Results of long context hallucination.

	CHAIR ↓	HalRate ↓	Cog. ↓
LLaVA-v1.5-7B	39.1	85.1	7.8
HA-DPO	40.3	86.1	8.1
SeVa	39.1	86.1	7.8
BPO	22.3	81.2	7.7
HDPO	**14.3**	**52.0**	**5.2**
w/o MCH	39.8	84.7	6.7

Table 4: Results of multimodal conflict hallucination.

the lowest HalRate, CHAIR$_s$, and Cog. It demonstrates that our construction for LCH works as expected in longer responses.

4.4.2 Multimodal Conflicts Hallucination

In real-world scenarios, multimodal conflicts are common when using MLLMs. To better evaluate the model's performance under such conditions, we design a more challenging task. Specifically, we randomly select 200 questions from the generative task in the AMBER dataset. First, LLaVA-1.5-7B is used to generate answers for these questions to get coarse-grained image descriptions. Next, GPT-4o-mini rewrites the details in the descriptions, following the construction method of MCH. We then introduce the incorrect information as a prefix to the question and ask the model to describe the image while influenced by the conflicting context.

The experimental results are shown in Table 4, demonstrating that despite encountering conflicting prefixes, our HDPO maintains promising performance. Compared to other baselines, HDPO achieves the best scores in CHAIR$_s$, HalRate, and Cog. It reveals that our HDPO shows significant improvement in the model's performance under this more difficult setting, highlighting the effectiveness of MCH. Additionally, we also make a comparison between the effects of adding noise and preserving visual tokens with lower scores. Further details can be seen in the appendix B.

4.5 Ablation Study

To demonstrate the contributions of VDH, LCH, and MCH to overall performance, we progressively remove each component and report the results. (1) As shown in Table 6, the performance declines as we remove each data type. The model achieves the best performance when all three data types are included. These experimental results confirm the individual contributions of each component. (2) It can also be observed that after incorporating MCH, there is no improvement in CHAIR$_s$ and CHAIR$_i$. However, the inclusion of both positive and negative examples for training leads to improvement in F1 of discriminative task (**4.8%**↑). (3) With the addition of LCH, F1 of the discriminative task shows minimal change, whereas the generative task demonstrates a substantial improvement, with CHAIR$_s$ (**44.5%**↓) and CHAIR$_i$ (**50.3%**↓) showing marked gains. This indicates that LCH is particularly effective for generative tasks. (4) Finally, incorporating VDH enhances model's performance across all tasks, and the combination of all three categories achieves the best results. The significance of LCH and MCH is also verified in Section 4.4 with the corresponding tasks.

	POPE ↑	CHAIR$_s$ ↓	CHAIR$_i$ ↓	AMBER-S ↑
InstructBLIP	83.7	57.0	16.1	82.5
HA-DPO	**85.6**	56.6	15.5	84.3
HDPO (ours)	84.8	**34.8**	**10.9**	**85.9**

Table 5: Results of HDPO on InstructBLIP-13B.

	CHAIR		AMBER	
	CHAIR$_s$↓	CHAIR$_i$↓	CHAIR↓	F1↑
LLaVA-v1.5-7B	51.4	14.2	7.6	74.5
+VDH +LCH +MCH	**16.6**	**5.1**	**3.3**	**84.1**
+LCH +MCH	28.4	7.5	4.8	78.9
+MCH	51.2	15.1	7.6	78.1

Table 6: Results of ablation study.

4.6 Scalability of HDPO

We analyze the impact of data size on our method. The performance of LLaVA-v1.5-7B fine-tuned on datasets of varying sizes with the same proportions is shown in Figure 4. As the data size increases, the effectiveness of our approach also improves, highlighting the potential for scaling up. This demonstrates the superior performance of HDPO.

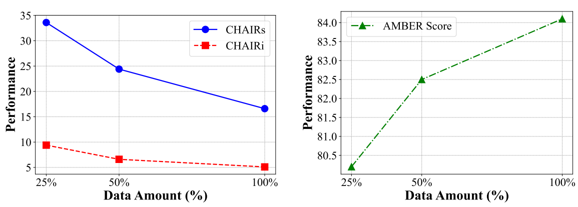

Figure 4: Scalability of HDPO with different data sizes.

5 Conclusions

In this paper, we present HDPO, a novel approach designed to effectively mitigate hallucinations in MLLMs. We analyze three types of hallucinations observed in MLLMs and create hallucination preference data based on the identified causes. Extensive experiments across different benchmarks demonstrate the ability of HDPO to effectively reduce hallucinations in MLLMs.

Limitations

In this paper, we introduce HDPO, which effectively mitigates the hallucination problem in current multimodal large language models. However, several issues remain unresolved. Specifically, we have not yet developed distinct strategies for controlling data quality, whilst the auto-generated negative examples require further verification and optimization. How to enhance the quality of positive examples also deserves extra investigation. Moreover, our construction methods and strategies could potentially be integrated with other techniques for processing more high-quality preference data, which may further improve the model's performance. Fine-tuning larger models with extensive, integrated datasets may not only enhance overall reasoning capabilities but also increase the model's robustness against hallucinations. This represents a promising area for further investigation, and we leave these open questions for future research.

Ethics Statement

This work mitigates hallucinations of multimodal large language models to enhance their reliability and practicality. We have carefully considered the ethical implications of our work. The models and datasets we used are publicly available and commonly used, and our findings may inherit the biases and limitations carried in these resources.

Acknowledgments

We appreciate the valuable discussions with Qidong Huang and Fushuo Huo, and their helpful suggestions. This work was supported by NSFC (62172420), Tencent Marketing Solution Rhino-Bird Focused Research Program, and the Young Elite Scientists Sponsorship Program by CAST (2023QNRC001).

References

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

Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. 2023. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv preprint arXiv:2308.12966*.

Zechen Bai, Pichao Wang, Tianjun Xiao, Tong He, Zongbo Han, Zheng Zhang, and Mike Zheng Shou. 2024. Hallucination of multimodal large language models: A survey. *arXiv preprint arXiv:2404.18930*.

Lin Chen, Jinsong Li, Xiaoyi Dong, Pan Zhang, Conghui He, Jiaqi Wang, Feng Zhao, and Dahua Lin. 2024a. Sharegpt4v: Improving large multi-modal models with better captions. In *European Conference on Computer Vision*, pages 370–387. Springer.

Zhaorun Chen, Zhuokai Zhao, Hongyin Luo, Huaxiu Yao, Bo Li, and Jiawei Zhou. 2024b. HALC: Object hallucination reduction via adaptive focal-contrast decoding. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 7824–7846. PMLR.

Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, Yang Zhou, Kaizhao Liang, Jintai Chen, Juanwu Lu, Zichong Yang, Kuei-Da Liao, et al. 2024. A survey on multimodal large language models for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 958–979.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

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*, pages 13418–13427.

Fushuo Huo, Wenchao Xu, Zhong Zhang, Haozhao Wang, Zhicheng Chen, and Peilin Zhao. 2025. Self-introspective decoding: Alleviating hallucinations for large vision-language models. In *The Thirteenth International Conference on Learning Representations*.

Hongrui Jia, Chaoya Jiang, Haiyang Xu, Wei Ye, Mengfan Dong, Ming Yan, Ji Zhang, Fei Huang, and Shikun Zhang. 2024. Symdpo: Boosting in-context learning of large multimodal models with symbol demonstration direct preference optimization. *arXiv preprint arXiv:2411.11909*.

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*, pages 13872–13882.

Lei Li, Zhihui Xie, Mukai Li, Shunian Chen, Peiyi Wang, Liang Chen, Yazheng Yang, Benyou Wang, Lingpeng Kong, and Qi Liu. 2024. 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*, pages 6227–6246, Miami, Florida, USA. Association for Computational Linguistics.

Yifan Li, Yifan Du, Kun Zhou, Jinpeng Wang, Xin Zhao, and Ji-Rong Wen. 2023. 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.

Fenglin Liu, Tingting Zhu, Xian Wu, Bang Yang, Chenyu You, Chenyang Wang, Lei Lu, Zhangdaihong Liu, Yefeng Zheng, Xu Sun, et al. 2023a. A medical multimodal large language model for future pandemics. *NPJ Digital Medicine*, 6(1):226.

Fuxiao Liu, Kevin Lin, Linjie Li, Jianfeng Wang, Yaser Yacoob, and Lijuan Wang. 2023b. Mitigating hallucination in large multi-modal models via robust instruction tuning. In *The Twelfth International Conference on Learning Representations*.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2024a. Visual instruction tuning. *Advances in neural information processing systems*, 36.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2024b. Visual instruction tuning. *Advances in neural information processing systems*, 36.

Jiazhen Liu, Yuhan Fu, Ruobing Xie, Runquan Xie, Xingwu Sun, Fengzong Lian, Zhanhui Kang, and Xirong Li. 2025. Phd: A chatgpt-prompted visual hallucination evaluation dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Runtao Liu, Haoyu Wu, Zheng Ziqiang, Chen Wei, Yingqing He, Renjie Pi, and Qifeng Chen. 2024c. Videodpo: Omni-preference alignment for video diffusion generation. *arXiv preprint arXiv:2412.14167*.

Ziyu Liu, Yuhang Zang, Xiaoyi Dong, Pan Zhang, Yuhang Cao, Haodong Duan, Conghui He, Yuanjun Xiong, Dahua Lin, and Jiaqi Wang. 2024d. Mia-dpo: Multi-image augmented direct preference optimization for large vision-language models. *arXiv preprint arXiv:2410.17637*.

OpenAI. 2023. GPT-4V(ision) system card.

OpenAI. 2024. Hello GPT-4o.

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

Renjie Pi, Tianyang Han, Wei Xiong, Jipeng Zhang, Runtao Liu, Rui Pan, and Tong Zhang. 2024. Strengthening multimodal large language model with bootstrapped preference optimization. In *European Conference on Computer Vision*, pages 382–398. Springer.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

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.

Min Shi, Fuxiao Liu, Shihao Wang, Shijia Liao, Subhashree Radhakrishnan, Yilin Zhao, De-An Huang, Hongxu Yin, Karan Sapra, Yaser Yacoob, et al. 2024. Eagle: Exploring the design space for multimodal llms with mixture of encoders. *arXiv preprint arXiv:2408.15998*.

Xingwu Sun, Yanfeng Chen, Yiqing Huang, Ruobing Xie, Jiaqi Zhu, Kai Zhang, Shuaipeng Li, Zhen Yang, Jonny Han, Xiaobo Shu, et al. 2024. Hunyuan-large: An open-source moe model with 52 billion activated parameters by tencent. *arXiv preprint arXiv:2411.02265*.

Shengbang Tong, Zhuang Liu, Yuexiang Zhai, Yi Ma, Yann LeCun, and Saining Xie. 2024. Eyes wide shut? exploring the visual shortcomings of multimodal llms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9568–9578.

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

Weihan Wang, Qingsong Lv, Wenmeng Yu, Wenyi Hong, Ji Qi, Yan Wang, Junhui Ji, Zhuoyi Yang, Lei Zhao, Song XiXuan, et al. 2024. Cogvlm: Visual expert for pretrained language models. *Advances in Neural Information Processing Systems*, 37:121475–121499.

Junfei Wu, Qiang Liu, Ding Wang, Jinghao Zhang, Shu Wu, Liang Wang, and Tieniu Tan. 2024. Logical closed loop: Uncovering object hallucinations in large vision-language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6944–6962, Bangkok, Thailand. Association for Computational Linguistics.

Qinghao Ye, Haiyang Xu, Jiabo Ye, Ming Yan, Anwen Hu, Haowei Liu, Qi Qian, Ji Zhang, and Fei Huang. 2024. mplug-owl2: Revolutionizing multi-modal large language model with modality collaboration. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 13040–13051.

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.

Weihao Yu, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Kevin Lin, Zicheng Liu, Xinchao Wang, and Lijuan Wang. 2024. MM-vet: Evaluating large multimodal models for integrated capabilities. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 57730–57754. PMLR.

Zihao Yue, Liang Zhang, and Qin Jin. 2024. Less is more: Mitigating multimodal hallucination from an EOS decision perspective. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11766–11781, Bangkok, Thailand. Association for Computational Linguistics.

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. *arXiv preprint arXiv:2402.11411*.

Yiyang Zhou, Zhiyuan Fan, Dongjie Cheng, Sihan Yang, Zhaorun Chen, Chenhang Cui, Xiyao Wang, Yun Li, Linjun Zhang, and Huaxiu Yao. 2024b. Calibrated self-rewarding vision language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. 2023. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*.

Ke Zhu, Liang Zhao, Zheng Ge, and Xiangyu Zhang. 2024. Self-supervised visual preference alignment. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 291–300.

Zhuofan Zong, Bingqi Ma, Dazhong Shen, Guanglu Song, Hao Shao, Dongzhi Jiang, Hongsheng Li, and Yu Liu. 2024. Mova: Adapting mixture of vision experts to multimodal context. In *Advances in Neural Information Processing Systems*, volume 37, pages 103305–103333. Curran Associates, Inc.

A Details of Datasets and Metrics

We evaluate the effectiveness of HDPO in mitigating hallucinations across both captioning tasks and simplified visual question answering (VQA) tasks using three evaluation datasets as follows:

CHAIR (Rohrbach et al., 2018): The Caption Hallucination Assessment with Image Relevance (CHAIR) is an evaluation method used in image captioning tasks to assess object hallucinations in model responses. There are two metrics: CHAIR_s and CHAIR_i. CHAIR_s measures hallucinations at the sentence level, while CHAIR_i measures them at the image level respectively. We conduct the CHAIR evaluation on the MSCOCO dataset following the setting in OPERA(Huang et al., 2024) with 500 random images. For each image, the model is prompted with: "Please describe this image in detail." to obtain their descriptions. By default, we set the 'max new tokens' to 512. More specifically, the calculation for the CHAIR_s and CHAIR_i metrics are as follows:

$$\text{CHAIR}_s = \frac{|\{\text{hallucinated objects}\}|}{|\{\text{all mentioned objects}\}|} \quad (4)$$

$$\text{CHAIR}_i = \frac{|\{\text{captions w/ hallucinated objects}\}|}{|\{\text{all captions}\}|} \quad (5)$$

POPE (Li et al., 2023): The Polling-based Object Probing Evaluation (POPE) is a popular dataset for evaluating object hallucinations in MLLMs. The evaluation is asking the model questions in the format: "Is there a <object> in the image?". It can be divided into three splits: popular, adversarial, and random. In the popular split, the evaluation targets the most frequently occurring objects in the dataset. In the adversarial split, it assesses the MLLM's ability to identify objects that are highly relevant to those present in the image. We evaluate the metrics for all splits, and calculate and report the average F1 score. POPE can be constructed on different datasets, and we evaluate models on the POPE dataset built on COCO.

AMBER (Wang et al., 2023): An Automated Multi-dimensional Benchmark for Multi-modal Hallucination Evaluation (AMBER) is an LLM-free multi-dimensional benchmark, offering a cost-effective and efficient evaluation pipeline. It supports the evaluation of both generative and discriminative tasks including hallucinations related to existence, attributes, and relations. Its generative evaluation aligns with our desired assessment of long descriptions, while the other dimensions provide insights into the model's performance on relatively simple VQA tasks, thereby reflecting the model's hallucination comprehensively. For its generative task, three metrics are used: CHAIR, Hal, and Cog. **CHAIR** measures the frequency of hallucinatory objects in the responses, **Hal** represents the proportion of responses containing hallucinations, and **Cog** assesses whether the hallucinations produced by MLLMs resemble those found in human cognition. For its discriminative task, we calculate and report the average F1 score. We also calculate **AMBER Score** denoted as AMBER-S, which reflects overall performance, and it's calculated as follows:

$$\textit{AMBER Score} = \frac{1}{2} \times (1 - \textit{CHAIR} + \textit{F1}) \quad (6)$$

B Comparison of noise and token preservation

We also conduct experiments to compare the impact of adding noise versus preserving visual tokens. Specifically, we use 6k samples from ShareGPT4V to construct negative samples by introducing diffusion noise and preserving visual tokens, and train the LLaVA-v1.5-7B model by direct preference optimization. The results of these experiments are presented in Table 7. As the experimental results show, using visual token preservation can achieve better performance on hallucination evaluation.

C Baseline Selection of 13B

For the experiments on the 13B model, we select several recent strong baselines, including SeVa and CSR, using their open-sourced checkpoints for evaluation. Additionally, we reimplement HA-DPO on LLaVA-v1.5-13B, as the original repository does not provide this checkpoint.

D Details about Our data

D.1 Visual Disctracted Hallucination

We obtain positive examples for our dataset from two sources: VG(with positive examples in HA-DPO) and ShareGPT4V. After extracting positive examples from ShareGPT4V, we found them to be too long. To mitigate length bias, we used GPT4o-mini to rewrite them to match the length of the negative examples. The prompt used is shown in Figure 7. For positive examples sourced from HA-DPO, after generating negative examples, we followed the original approach by rewriting the

	POPE	CHAIR		AMBER				
	F1 Score ↑	$CHAIR_s$ ↓	$CHAIR_i$ ↓	CHAIR ↓	HalRate ↓	Cog. ↓	F1 Score ↑	AMBER-S ↑
LLaVA-v1.5-7B	86.1	51.2	14.2	7.6	35.1	4.3	74.5	83.5
+ Diffu_{6k}	86.2	62.8	18.4	9.2	47.5	4.3	78.1	84.5
+ VDH $_{6k}$	**87.1**	**48.2**	**13.7**	**6.1**	**32.0**	**2.7**	**80.2**	**87.1**

Table 7: Experimental results of LLaVA-v1.5-7B trained with two ways to construct preference pairs: adding noise and preserving visual tokens. The diffusion noise step is 800. The best result for each metric is in bold.

negative examples using GPT4o-mini. The prompt used is shown in Figure 6. Also, we can adopt the method in HA-DPO to create more data. For k and i, we make an empirical choice based on performance and original settings.

D.2 Long Context Hallucination

We use LLaVA-1.5-7B to continue generating text for the positive examples, with the system prompt in Figure 5, and the hint phrases in Figure 8. By excluding the last two sentences, we aim to increase the concentration of hallucinated content in the tail of the response. Generating three continuations at a time maintains an approximate balance in the average length between positive and negative examples.

D.3 Multimodal Conflicts Hallucination

We utilize GPT-4o-mini to modify the details of the positive examples, following the prompt shown in Figure 9. This approach introduces conflicting information that deviates from the image content.

D.4 Effect of data ratio

We did not conduct detailed experiments comparing different data type ratios. However, throughout the experiments, all tested ratios showed significant improvements over the original model. We report the best-performing dataset from our experiments. Determining the optimal ratio of different data types is inherently a more challenging and general problem, which goes beyond the scope of this paper.

D.5 Comparison on AMBER with other MLLMs

We also report the hallucination evaluation results on AMBER for both generative and discriminative tasks of HDPO on LLaVA-1.5-7B compared with other MLLMs including mPLUG-Owl2 (Ye et al., 2024), MiniGPT4 (Zhu et al., 2023), CogVLM (Wang et al., 2024), Qwen-VL (Bai et al., 2023) and GPT4V (OpenAI, 2023) in Table 9.

	Len	Cover	Co. / Len ↑	CHAIR ↓
LLaVA-1.5-7B	75.0	51.8	0.69	7.6
BPO	148.0	58.8	0.40	5.0
SeVa	76.0	53.4	0.70	7.4
CSR	64.0	45.0	0.70	3.8
HDPO	69.0	50.2	**0.73**	**3.3**

Table 8: Analysis of Cover. on AMBER

	CHAIR ↓	Hal↓	Cog.↓	F1↑	AMBER-S↑
mPLUG-Owl	21.6	76.1	11,5	18.9	48.7
LLaVA	11.5	48.8	5.5	32.7	60.6
MiniGPT4	13.6	65.3	11.3	64.7	75.6
CogVLM	5.6	23.6	1.3	72.3	83.4
mPLUG-Owl2	10.6	39.9	4.5	78.5	84.0
Qwen-VL	5.5	23.6	1.9	84.9	89.7
GPT-4V	4.6	30.7	2.6	**87.4**	**91.4**
HDPO	**3.3**	**15.8**	**0.8**	84.1	90.4

Table 9: Comparison on AMBER with more MLLMs, most results are source from(Wang et al., 2023).

D.6 Computational cost and efficiency

As computational efficiency is crucial for real-world applications, we briefly discuss the training efficiency of our method. Training time is primarily influenced by the size of the training dataset. Except for BPO, which requires a relatively longer training period, the training cost and duration of the other methods are within a comparable range. Our method and the baselines (SeVa and CSR) exhibit similar computational overhead. Therefore, we believe our method holds strong potential for practical deployment.

D.7 More Analysis of Cover

There is another Cover metric in AMBER, represents object coverage. It's related to the length of generated content. We calculate the Cover / Length and report it in Table 8. It shows that HDPO's outputs are more precise and of higher quality with the highest Co./ Len. Additionally, we have conducted experiments showing that generating longer outputs improves Cover while maintaining good hallucination performance.

D.8 Further Discussion of Limitation

Although HDPO enjoys promising performance in Mitigating Hallucination, there are still some potential boundaries we meet as follows:

(1) For relatively long content generation, HDPO may still struggle to fully address the issue. As the generated content becomes longer, hallucinations may persist. To completely resolve this problem, the model's intrinsic long-context processing capabilities might first need to be enhanced. However, the current long-text abilities of MLLMs are not as advanced as those of LLMs, which presents an intriguing direction for future exploration.

(2) Additionally, as highlighted by (Tong et al., 2024; Zong et al., 2024; Shi et al., 2024), the visual encoder in current MLLMs operates at a relatively coarse granularity, resulting in insufficient or suboptimal visual features. These limitations cannot be fully addressed by HDPO and will likely require either more powerful visual encoders or improved MLLM architectures, both of which are also promising directions for future research.

(3) What's more, fine-tuning larger models on extensive, integrated datasets could improve reasoning capabilities and robustness against hallucinations. These open questions remain promising directions for future research.

(4) In data generation, it is beneficial to enhance data quality. As datasets scale, maintaining diversity and broad content coverage becomes increasingly challenging. In the future, effective post-processing and improved data synthesis strategies could further enhance data quality and support automated generation.

We think the above additional discussion clarifies the limitations of HDPO and outlines potential directions for addressing these challenges.

System Prompt:

You should describe in detail all elements in the image.
Be thorough in addressing aspects such as color, shape, size, position, quantity, actions, emotions, and more.
Your response should be as much as possible.

Figure 5: System Prompt used in LCH

Rewrite Prompt:

Help me rewrite the given sentence.
Don't change any detail and information in the original sentence. Don't add any new information.

The sentence you need to rewrite: %s
Directly give the rewritten sentence:

Figure 6: Rewrite Prompt used in VDH

Adjust Length Prompt:

Please adjust the length of the Description to approximately %s words.

Ensure all essential details and meanings are preserved, with clear, concise, and accurate expression.
Provide the modified Description directly.

Original Description: "%s"
Modified Description:

Figure 7: Adjust Length Prompt used in VDH

Hint Phrases:

"In addition",
"Moreover",
"Furthermore",
"Besides that",
"Additionally",
"What's more",
"As well as that",
"Beyond that",
"There is something else that needs to be mentioned",
"Not only that",
"It should also be noted that”

Figure 8: Hint Phrases used in LCH

Modify Prompt:

I will give you a description of an image, and you need to modify various details of the description, such as the number of objects, types of objects, their positions, colors, behaviors, and so on.

Description: %s
Modified Description:

Figure 9: Modify Prompt used in MCH