

Mitigating Hallucinations in Multimodal LLMs via Object-aware Preference Optimization

Alberto Compagnoni
alberto.compagnoni@unimore.it

Davide Caffagni
davide.caffagni@unimore.it

Nicholas Moratelli
nicholas.moratelli@unimore.it

Lorenzo Baraldi
lorenzo.baraldi@unimore.it

Marcella Cornia
marcella.cornia@unimore.it

Rita Cucchiara
rita.cucchiara@unimore.it

University of Modena and Reggio
Emilia
Modena, Italy

Abstract

Multimodal Large Language Models (MLLMs) emerge as a unified interface to address a multitude of tasks, ranging from NLP to computer vision. Despite showcasing state-of-the-art results in many benchmarks, a long-standing issue is the tendency of MLLMs to hallucinate, that is to generate answers to the user’s query that are not reflected in the visual input. In this paper, we address the problem of hallucinations as an alignment problem, seeking to steer the MLLM so that it prefers generating content without hallucinations. In contrast to recent approaches that require complicated pipelines to build synthetic preference data for alignment training, often relying on proprietary models, we capitalize on the well-known CHAIR metric, originally proposed to gauge the degree of hallucinations in image captioning. Given a pair of generated answers, we leverage CHAIR to distinguish winner and loser options (*i.e.*, non-hallucinated and hallucinated samples) and fine-tune off-the-shelf MLLMs via Direct Preference Optimization (DPO). The resulting method, which we refer to as CHAIR-DPO, effectively diminishes the amount of hallucinated answers on several hallucination benchmarks, demonstrating the effectiveness of fine-tuning the MLLM with a CHAIR-based reward. Source code and trained models are publicly available at <https://github.com/aimagelab/CHAIR-DPO>.

1 Introduction

Research interest in Multimodal Large Language Models (MLLMs) is raging. By capitalizing on massive self-supervised pre-training on text, images [0, 21, 25, 26, 51], and possibly other modalities [13, 30, 42], they emerge as a unified interface the user can interact with to solve different problems [9]. Not only they are accessible to even non-expert users, as the

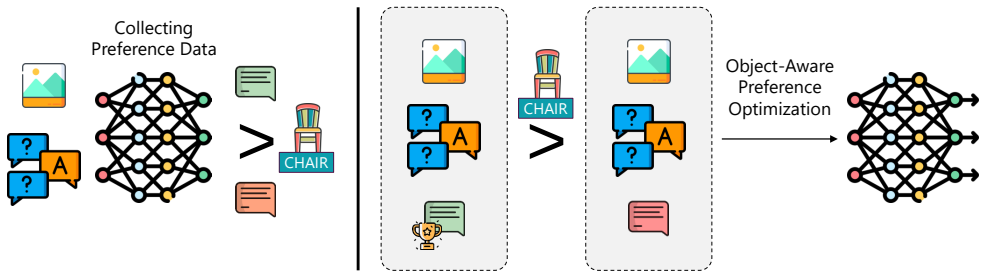


Figure 1: Overview of CHAIR-DPO. We begin by collecting preference data for optimization. Given two completions for a visual and textual prompt, we leverage CHAIR_i to rank them: the preferred and dispreferred answers are chosen based on the lowest and highest number of hallucinated objects. Next, we use these preference pairs to fine-tune an MLLM with DPO, resulting in an object-aware model conscious of the objects truly depicted in the image.

interaction happens via natural language, but they deliver strong performance on several tasks, often challenging specialized models tailored to a unique task alone. The capabilities of an MLLM goes beyond the production of text in response to a user query, spanning from visual grounding and referring [8, 31, 36], up to the generation of images and videos [22, 43, 44].

Despite their impressive capabilities, MLLMs still suffer from hallucinations – generating content that is unsupported by the input. Hallucination is a long-standing problem widely studied in the natural language processing community [16, 50], but having access to complementary data modalities other than text stretches it out and opens new paths for the model to hallucinate [58]. For instance, visual hallucinations [17, 37, 47, 52, 53] manifest whenever an MLLM mentions an object not depicted in the input image.

An intuitive framework for reducing hallucinations is to view them as a human alignment problem: as we humans reasonably prefer answers devoid of hallucinations, so should an MLLM properly aligned to human preference. Unfortunately, established techniques employed in developing MLLMs, such as visual instruction tuning [25, 26], Reinforcement Learning from Human Feedback (RLHF) [29], or Direct Preference Optimization (DPO) [54], often prioritize that the generated answer effectively fulfills the user query, overlooking whether it contains hallucinations or not. Since they are implemented as the last training stage of MLLMs, they play the most important role in aligning the model to human preference.

Recent research [19, 39, 46, 49, 57] has introduced alignment methods to steer MLLMs toward preferring non-hallucinated outputs, with a particular focus on DPO. DPO is especially appealing as it replaces the complexity of reinforcement learning with a more tractable supervised learning approach. A key challenge prior to its application, however, is to collect preference data about hallucinations. In other words, how can we know which answer generated by an MLLM is hallucinated and which is not? Resorting to human annotators is costly and does not scale, so a compelling alternative becomes to query cutting-edge proprietary MLLMs such as GPT-4 [49, 57, 58] and Gemini [39] to act as a judge.

In this work, we introduce CHAIR-DPO, a new preference optimization method to tackle visual hallucinations in MLLMs. CHAIR-DPO builds on top of CHAIR [57], a well-known metric proposed to assess the degree of hallucination of image captioning models [8, 37, 40]. The key idea is to leverage CHAIR to quantitatively measure hallucinations in the responses generated by an MLLM, and thus select as preferred the response with the lower hallucination rate. After collecting enough preference pairs, we apply DPO to fine-tune an MLLM,

enhancing its awareness concerning the presence or absence of objects in the input image, and drastically reducing visual hallucinations. An overview of the proposed approach is outlined in Figure 1.

We argue that CHAIR enables more efficient preference data collection for hallucination mitigation compared to existing approaches. Experiments on multiple hallucination benchmarks, such as AMBER [47], CHAIR-MSCOCO [55], and Object HalBench [57, 53], show that CHAIR-DPO achieves state-of-the-art performance, significantly reducing hallucination rates without degrading the original capabilities of the underlying MLLM.

2 Related Work

Multimodal Large Language Models. Building on the advancements of Large Language Models (LLMs), interest is surging to extend LLMs to the multimodal domain [0, 6, 20, 26, 36, 42, 43, 44, 50], with most of the effort devoted to the integration of visual understanding. MLLMs perceive different modalities thanks to external encoders (*e.g.*, CLIP visual encoder [53] for images) that transform (unimodal) data into modality-specific embeddings that can be understood by the LLM upon the application of an adapter module. In this work, we utilize open-source MLLMs [9, 26] built on the popular LLaVA [25, 26] framework. These models feature a CLIP-ViT-L/14@336 [83] as the visual encoder and a lightweight feed-forward network as the vision-to-language adapter. LLaVA initially trains only the adapter module on an image captioning dataset. Next, it also unfreezes the pre-trained LLM and jointly optimizes it on a visual instruction tuning dataset comprising multi-turn visual dialogues. The parameters of the CLIP visual encoder are kept frozen and never updated.

Direct Preference Optimization (DPO). DPO [54] has been established as a compelling alternative to Reinforcement Learning from Human Feedback (RLHF) [29] to align fine-tuned LLMs with human preferences. DPO presents two main advantages against RLHF. First, it saves the effort of training a reward model to mimic human preferences, and second, it completely bypasses the reinforcement learning stage, which is non-trivial to implement in practice. The intuition behind DPO is to frame the constrained minimization problem of RLHF such that the objective does not rely on the reward gained by the model anymore, but instead, it depends upon an implicit reward expressed in terms of the optimizing and reference policies. Originally born for LLMs only, DPO and its variants [12, 41, 48] are now successfully applied for aligning MLLMs as well [18, 56], not to mention different generative domains spanning from diffusion models [45] to image captioning [28].

Hallucination Mitigation. Despite the success of MLLMs, hallucinations remain an open challenge. Recent research [0] identifies the root of hallucinations in noisy or corrupted training samples, eventually exacerbated by the addition of synthetic training data, as well as an imbalance in the frequency of certain entities appearing more often than others during optimization. This is further compounded by the overwhelming dominance of the LLM compared to the size of the visual encoder, which may lead to an over-reliance on linguistic priors rather than visual grounding. To mitigate hallucinations, prior work explores training-free and training-based approaches. Training-free methods [8, 17, 22] modify decoding strategies or apply post-hoc corrections [52] using external models. Training-based approaches fine-tune MLLMs via variants of preference alignment losses [19, 59, 46, 57, 58] or modified cross-entropy objectives [49, 55]. Notably, HA-DPO [57] augments DPO with auxiliary language modeling loss component, while HALVA [39] uses a fine-grained feedback to only penalize hallucinated tokens at phrase level. Lastly, both mDPO [46] and MFPO [19]

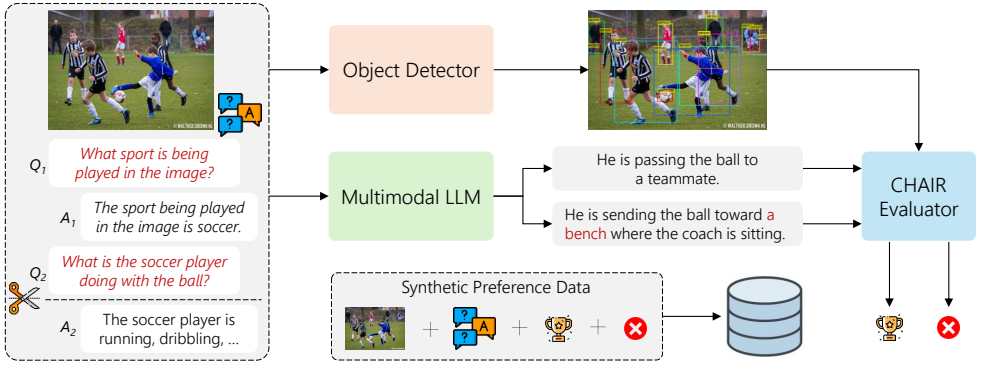


Figure 2: Details of the preference data collection pipeline. First, we ask an MLLM to generate two completions conditioned on a textual prompt x_T and an input image x_I . Next, we use a pre-trained object detector to obtain the set of class names of the objects appearing in x_I . We use this set to compute CHAIR_i to evaluate the completions: lower CHAIR_i score identifies the preferred one, as it contains fewer hallucinations.

incorporate an image preference loss and anchor terms to preserve visual grounding and prevent reward degradation. Crucially, most of these approaches create or use datasets derived from complicated pipelines, often leveraging proprietary MLLMs. In contrast, our method relies solely on (i) an open-source model, the same one being fine-tuned, used exclusively for sampling completions, and (ii) an off-the-shelf object detector, drastically simplifying preference data collection.

3 Proposed Method

3.1 Problem Formulation

In our settings, an MLLM conditioned on textual prompt x_T and image x_I outputs the continuation text y with probability $P(y | x_T, x_I)$. We call y a hallucinated answer and denote it as y_l , whenever y mentions any object not present in x_I . Building on the intuition that humans should prefer a non-hallucinated answer y_w more than a hallucinated one, we seek to align the model with this preference, so that the probability of generating y_w gets higher than y_l .

3.2 Collecting Preference Data for Object Hallucinations

To align our model with human preferences regarding object hallucinations, we need a supervised signal that can distinguish between non-hallucinated and hallucinated responses. In this work, we create this signal with CHAIR [57], a metric designed to assess the hallucination extent of image captioning models. We summarize in Figure 2 our data collection process. Given an answer y generated by an MLLM, we define its CHAIR_i score as the fraction of hallucinated object instances mentioned in y :

$$\text{CHAIR}_i(y) = \frac{|\text{[hallucinated objects]}_y|}{|\text{[all mentioned objects]}_y|}. \quad (1)$$

Note that the hallucinated and mentioned objects in Eq. 1 are not treated as a set, but rather as a *list*. It follows that CHAIR_i can penalize a sentence for hallucinating the same objects multiple times. Normally, CHAIR_i is applied on a supervised dataset where the objects mentioned in a ground-truth caption are known in advance. As we do not have access to similar annotations concerning hallucination, we propose to adapt publicly available datasets commonly used for visual instruction tuning of MLLMs [26]. Such a dataset can be seen as a collection of triplets $\{x_T, x_I, y\}$, where y is the ground-truth answer conditioned on the prompt x_T and image x_I . In case of multi-turn dialogs, we randomly truncate the conversation, and provide the model with prior turns along with the current question as textual context.

First of all, for each triplet in the dataset, we discard y , and sample a new pair of possible answers $\{y_1, y_2\}$ from a reference instruction-tuned MLLM π_{ref} . We then apply an off-the-shelf detector trained to recognize a predefined set of objects to the image x_I , treating the class name of the detected items as the ground-truth set of objects present in x_I . At this point, it is possible to compute the CHAIR_i score of the generated answers, considering as hallucinated any mentioned object that is not in the ground-truth set. To make the evaluation robust, we leverage a predefined set of synonyms to match the words in the generated answers with the class names of the ground-truth objects. We can now designate the non-hallucinated (*i.e.*, winner) and hallucinated (*i.e.*, loser) answers as:

$$y_w = \min_{y \in \{y_1, y_2\}} \text{CHAIR}_i(y), \quad y_l = \max_{y \in \{y_1, y_2\}} \text{CHAIR}_i(y), \quad \text{with } y_1, y_2 \sim \pi_{\text{ref}}(y \mid x_T, x_I). \quad (2)$$

3.3 Object-Aware Preference Optimization

With our preference data established, we apply Direct Preference Optimization (DPO) [52], an effective training approach for the human alignment of MLLMs. DPO bypasses the need for reinforcement learning to explicitly maximize a reward signal that scores the completions generated by the model. Instead, DPO simultaneously trains the policy model (*i.e.*, the chosen MLLM) along with an implicit reward model that assigns a higher score to a winner answer y_w compared to the loser answer y_l :

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}(x_T, x_I, y_w, y_l) \sim \mathcal{D} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w \mid x_T, x_I)}{\pi_{\text{ref}}(y_w \mid x_T, x_I)} - \beta \log \frac{\pi_\theta(y_l \mid x_T, x_I)}{\pi_{\text{ref}}(y_l \mid x_T, x_I)} \right) \right], \quad (3)$$

where β is a regularizing hyperparameter controlling the strength of the Kullback-Leibler divergence, expressing the degree to which the policy model π_θ must be tied to a frozen reference model π_{ref} . In practice, π_θ is initialized with the same weights as π_{ref} . Note that our method may be applied iteratively: after exhausting the dataset once, fresh preference data can be built following Sec. 3.2, upon updating π_{ref} with π_θ .

Minimizing Eq. 3 pulls up the probability of generating the non-hallucinated completion $\pi_\theta(y_w \mid x_T, x_I)$ to the detriment of the hallucinated one $\pi_\theta(y_l \mid x_T, x_I)$. Succeeding in that has the immediate consequence of making π_θ *aware* of what objects really appear in the presented image, and what are missing instead. We refer to the proposed method as CHAIR-DPO.

Data Filtering for Reliable Preference Supervision. A critical challenge we observe during training stems from the presence of several completion pairs with indistinguishable CHAIR_i scores, making it impossible to assign winner and loser labels. Incorporating such pairs into the optimization process introduces noisy supervision, as the model is forced to learn from randomly chosen preference labels. To address this, we introduce a simple yet effective filtering strategy: we discard all training instances where the CHAIR_i score difference between

completions is zero. This ensures that the retained supervision pairs reflect a meaningful distinction in object hallucination severity, leading to a more reliable optimization signal.

4 Experiments

4.1 Experimental Setting

Implementation and Training Details. We build hallucination preference data starting from LLaVA-Instruct-665k [26], an open-source and widely employed dataset for visual instruction tuning, comprising multi-turn dialogues. We ask our chosen MLLMs, namely LLaVA-1.5-7B [26] and LLaVA-MORE-8B [9]¹, to generate the candidate completions $\{y_1, y_2\}$ conditioned on the image and the first k human-assistant turns, where k is randomly chosen. Next, we identify winner and loser candidates by computing the CHAIR_i score, using as ground-truth objects those detected by DETR-DC5-R101 [6]. We employ class names from MSCOCO [24] using the list of synonyms provided in the reference implementation of CHAIR [57]. After filtering out all the instances with null CHAIR_i difference, we end up with 70k and 77k training samples for LLaVA-1.5-7B and LLaVA-MORE-8B, respectively. We refer to Appendix A.1 for more details on the DPO fine-tuning stage, carried out with LoRA [15] for efficiency and performance preservation.

Datasets and Evaluation Benchmarks. We evaluate the performance of CHAIR-DPO on three popular hallucination-oriented benchmarks: AMBER [47], CHAIR-MSCOCO [53], and Object HalBench [57, 53]. AMBER is designed to assess the visual hallucination tendencies of MLLMs without relying on external LLMs for response annotation or judgment. Following recent works [19, 46], we focus on the generative task, evaluating hallucination proneness using the CHAIR_i , Hallucination Rate (HalRate), and Cognition (Cog) metrics, and assessing object recall via the Coverage (Cover) metric. The CHAIR_i score is computed as described in Eq. 1, while HalRate corresponds to the CHAIR_s metric [57], defined as the percentage of responses that contain at least one hallucinated object. CHAIR-MSCOCO builds on similar principles but operates on a subset of the MSCOCO dataset [24]. It evaluates hallucinations using CHAIR_i and CHAIR_s , leveraging the predefined MSCOCO object classes as a reference vocabulary and extracting mentioned objects with a standard NLP toolkit, in line with the official CHAIR implementation [57]. Object HalBench also reports CHAIR_i and CHAIR_s scores but uses a different subset of the MSCOCO validation set. Unlike CHAIR-MSCOCO, it employs a proprietary LLM (*i.e.*, GPT-3.5) to extract mentioned objects, which improves the precision and recall of object identification. Further details about datasets and metrics are provided in Appendix A.2.

4.2 Experimental Results

Comparison with the State of the Art. We compare in Table 1 our fine-tuned models against a multitude of recent approaches to alleviate hallucinations in MLLMs. Many of them resort to proprietary MLLMs to collect preference data for fine-tuning [39, 49, 57] or design sophisticated data pipelines to craft preferred and dispreferred completions [19]. Conversely, CHAIR-DPO only requires the offline application of an off-the-shelf detector

¹Specifically, LLaVA-1.5-7B is based on Vicuna-7B [0], while LLaVA-MORE-8B is based on LLaMA-3.1-8B [10]. Both are trained following the two-stage training pipeline introduced in [26].

	FT	Ext. Support	AMBER				CHAIR-MSCOCO		Object HalBench	
			CHAIR _i ↓	Cover ↑	HalRate ↓	Cog ↓	CHAIR _s ↓	CHAIR _i ↓	CHAIR _s ↓	CHAIR _i ↓
LLaVA-1.5-7B [8]	-	-	7.6	51.7	35.0	4.2	50.6	13.9	-	-
+ DoLa [8]	-	-	7.6	51.6	36.0	4.0	51.6	14.1	-	-
+ VCD [52]	-	-	-	-	-	-	48.6	14.9	48.0	22.3
+ OPERA [53]	-	-	7.3	49.6	32.0	3.5	47.8	14.6	-	-
+ Woodpecker [54]	-	GPT-3.5	6.9	48.9	30.4	3.6	45.8	14.8	-	-
+ POVID [55]	✓	GPT-4	7.4	51.3	34.3	3.9	-	-	50.7	15.3
+ HA-DPO [56]	✓	GPT-4	6.7	49.8	30.9	3.3	38.2	11.0	39.9	19.9
+ HALVA [57]	✓	Gemini	6.6	<u>53.0</u>	32.2	3.4	41.4	11.7	-	-
+ EOS [58]	✓	-	5.1	49.1	22.7	2.0	36.8	11.3	-	-
+ mDPO [59]	✓	-	4.4	52.4	24.5	2.4	-	-	35.7	9.8
+ MFPO [60]	✓	-	4.1	55.7	22.5	1.9	-	-	13.4	<u>6.6</u>
+ REVERSE [61]	✓	GPT-4	4.0	26.9	10.2	0.9	13.6	6.1	-	-
+ CHAIR-DPO ($\beta=0.5$)	✓	-	3.8	48.7	18.4	1.8	20.8	5.8	20.1	10.9
+ CHAIR-DPO ($\beta=0.3$)	✓	-	<u>3.2</u>	47.4	16.2	<u>1.3</u>	16.4	<u>4.4</u>	<u>13.1</u>	6.7
+ CHAIR-DPO ($\beta=0.2$)	✓	-	3.0	46.6	<u>14.7</u>	<u>1.3</u>	<u>14.4</u>	3.6	8.6	4.6
LLaVA-MORE-8B [8]	-	-	8.1	53.2	38.4	4.0	51.2	14.4	50.5	24.7
+ DoLa [8]	-	-	7.9	<u>53.1</u>	38.4	4.1	51.8	13.8	-	-
+ Woodpecker [54]	-	GPT-3.5	7.4	50.7	36.7	3.7	51.0	14.3	-	-
+ REVERSE [61]	✓	GPT-4	5.1	38.9	20.8	2.1	25.2	8.4	-	-
+ CHAIR-DPO ($\beta=0.5$)	✓	-	3.4	50.7	17.1	<u>1.2</u>	21.0	<u>5.2</u>	19.7	10.0
+ CHAIR-DPO ($\beta=0.3$)	✓	-	<u>2.9</u>	49.4	<u>14.7</u>	1.0	<u>12.3</u>	6.2	<u>16.0</u>	<u>3.8</u>
+ CHAIR-DPO ($\beta=0.2$)	✓	-	2.6	49.7	14.2	1.0	11.8	3.1	9.2	4.5

Table 1: Comparative evaluation of hallucination mitigation methods for MLLMs, conducted on the AMBER, CHAIR-MSCOCO, and Object HalBench datasets. For each method, we indicate whether the underlying MLLM is fine-tuned (FT) and whether it relies on external support from proprietary LLMs, either via additional training data or post-hoc refinement. Bold-faced and underlined values indicate the best and second-best results.

and the computation of the CHAIR_i, which is widely known in the captioning literature and does not need any engineering.

We notice that both LLaVA-1.5-7B and LLaVA-MORE-8B combined with CHAIR-DPO always achieve the best results, or the second-best at worst, on metrics directly assessing the hallucination degree. Moreover, we can slightly adapt the behavior of CHAIR-DPO by varying the β hyperparameter, which controls the strength of the Kullback-Leibler divergence in DPO (cf. Eq. 3). Specifically, lower β values soften this regularization, achieving lower hallucination rates at the expense of slightly worse Coverage scores on AMBER. This pattern of trading off Coverage for hallucination is consistent in all the considered models. However, CHAIR-DPO manages to strike a good balance between the two metrics, in contrast to other methods. For instance, REVERSE reaches a lower HalRate than CHAIR-DPO $_{\beta=0.2}$ with LLaVA-1.5-7B on AMBER (10.2 vs 14.7), but its Coverage falls to 26.9, recording a severe -24.8 points drop with respect to the 51.7 points of the baseline model, while for CHAIR-DPO $_{\beta=0.2}$ the degradation is limited to -5.1 points.

The comparison with LLaVA-MORE-8B is even more favorable to CHAIR-DPO, in that it scores the best hallucination results. Training-free methods that do not modify the weights of LLaVA-MORE-8B, such as DoLa [8] and Woodpecker [54], are superior in terms of Coverage, but they greatly fall short against CHAIR-DPO in terms of hallucination metrics.

Ablation Studies. We assess the impact of filtering out instances where the CHAIR_i difference between the two candidate completions is null, presenting the results in Table 2. When data filtering is on, CHAIR-DPO records the best hallucination scores concerning CHAIR_i and HalRate on AMBER. The same effect is observable on the Cognition metric in AMBER and in general on CHAIR-MSCOCO, except for LLaVA-1.5-7B with $\beta = 0.2$, which performs

Data Filtering		AMBER				CHAIR-MSCOCO		Object HalBench	
		CHAIR _i ↓	Cover ↑	HalRate ↓	Cog ↓	CHAIR _s ↓	CHAIR _i ↓	CHAIR _s ↓	CHAIR _i ↓
LLaVA-1.5-7B [14]	-	7.6	51.7	35.0	4.2	50.6	13.9	47.4	25.6
+ CHAIR-DPO ($\beta=0.5$)	✗	4.4	50.1	22.8	2.0	25.6	7.1	19.7	11.1
+ CHAIR-DPO ($\beta=0.5$)	✓	3.8	48.7	18.4	1.8	20.8	5.8	20.1	10.9
+ CHAIR-DPO ($\beta=0.3$)	✗	4.0	49.0	20.2	1.4	17.0	4.9	16.6	8.0
+ CHAIR-DPO ($\beta=0.3$)	✓	3.2	47.4	16.2	1.3	16.4	4.4	13.1	6.7
+ CHAIR-DPO ($\beta=0.2$)	✗	3.4	48.4	17.2	1.2	14.0	3.3	11.3	5.4
+ CHAIR-DPO ($\beta=0.2$)	✓	3.0	46.6	14.7	1.3	14.4	3.6	8.6	4.6
LLaVA-MORE-8B [9]	-	8.1	53.2	38.4	4.0	51.2	14.4	50.5	24.7
+ CHAIR-DPO ($\beta=0.5$)	✗	3.7	51.5	20.1	1.7	22.8	5.8	19.6	9.3
+ CHAIR-DPO ($\beta=0.5$)	✓	3.4	50.7	17.1	1.2	21.0	5.2	19.7	10.0
+ CHAIR-DPO ($\beta=0.3$)	✗	3.4	51.0	19.0	1.4	18.6	5.0	14.0	7.0
+ CHAIR-DPO ($\beta=0.3$)	✓	2.9	49.4	14.7	1.0	16.0	3.8	12.3	6.2
+ CHAIR-DPO ($\beta=0.2$)	✗	2.9	50.1	16.4	1.4	15.8	4.4	12.6	6.3
+ CHAIR-DPO ($\beta=0.2$)	✓	2.6	49.7	14.2	1.0	11.8	3.1	9.2	4.5

Table 2: Ablation study of the application of data filtering. We assess the impact of filtering out any preference instance where the CHAIR_i difference is zero.

	MME		SEED			MMMU	Science-QA	AI2D
	Perception	Cognition	All	Video	Image	Acc	Acc	Acc
InstructBLIP-7B [10]	-	-	53.4	58.8	38.1	-	60.5	-
Qwen-VL-7B [8]	-	-	56.3	62.3	39.1	-	67.1	-
Qwen-VL-7B-Chat [8]	1487.5	-	58.2	65.4	37.8	-	68.2	-
LLaVA-1.5-LLaMA3-8B [63]	1544.4	330.3	64.3	42.0	70.1	37.3	74.2	60.7
LLaVA-1.5-7B [14]	1474.3	314.6	61.6	42.0	66.8	34.2	69.0	56.4
+ CHAIR-DPO ($\beta=0.5$)	1520.9	372.9	60.8	39.5	66.4	34.9	69.6	55.0
+ CHAIR-DPO ($\beta=0.3$)	1525.9	372.9	60.8	39.7	66.3	35.0	69.3	54.8
+ CHAIR-DPO ($\beta=0.2$)	1518.8	375.0	60.7	39.8	66.2	35.2	69.3	54.7
LLaVA-MORE-8B [9]	1531.5	353.3	64.1	42.4	69.8	39.4	76.3	61.8
+ CHAIR-DPO ($\beta=0.5$)	1417.3	327.5	64.0	43.0	69.5	37.1	74.4	59.3
+ CHAIR-DPO ($\beta=0.3$)	1414.0	340.7	64.1	43.5	69.5	36.1	74.4	58.9
+ CHAIR-DPO ($\beta=0.2$)	1412.6	335.4	64.1	43.8	69.5	36.8	74.2	59.0

Table 3: General cognition evaluation of CHAIR-DPO. We compare LLaVA-1.5-7B and LLaVA-MORE-8B with and without CHAIR-DPO to ensure performance preservation.

slightly better without the filter. With LLaVA-MORE-8B, the data filtering benefits are consistent on AMBER and CHAIR-MSCOCO, while we notice a minimal degradation on Object HalBench when β is set to 0.5. We impute it to the strong regularization imposed by $\beta = 0.5$ during the fine-tuning, which limits how much the model can learn from severe changes in the hallucination level measured by CHAIR_i. Finally, we notice the same pattern related to the Coverage metrics on AMBER as in Table 1, where models trade off better hallucination rates for lower Coverage. These results suggest that the proposed data filtering approach is well-suited for CHAIR-DPO, with the added benefit of significantly speeding up fine-tuning by eliminating nearly 90% of the dataset.

Performance Preservation Analysis. An important research question is whether CHAIR-DPO hurts the general cognitive capabilities of the reference MLLM, *i.e.* π_{ref} (cf. Eq. 3). To address this point, we test CHAIR-DPO on several benchmarks typically employed in a comprehensive evaluation of MLLMs [76]. These benchmarks include MME [10], which measures capabilities in 14 distinct multimodal interaction categories, SEED-Bench [23], which examines 12 cognitive aspects spanning from visual comprehension to text recognition

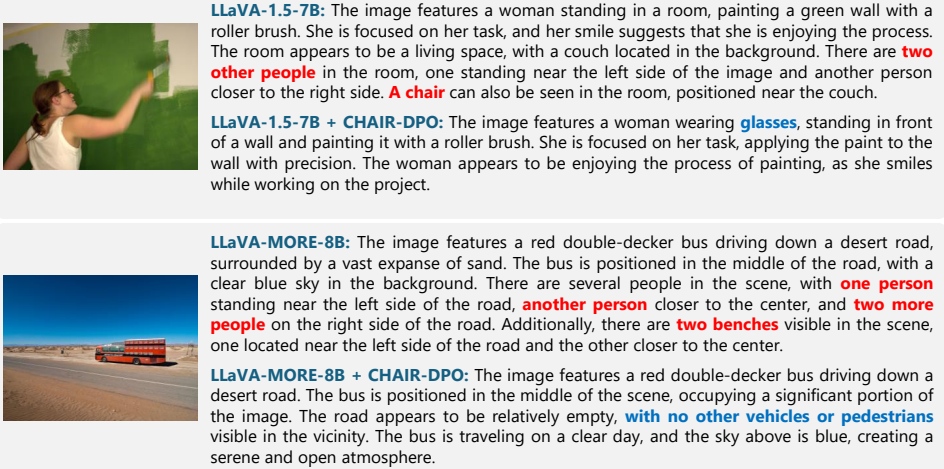


Figure 3: Image captioning qualitative results with LLaVA-1.5-7B and LLaVA-MORE-8B. When applied, CHAIR-DPO not only reduces hallucinations (*i.e.*, **red words**), but also incentivizes focusing on items overlooked by the baseline models, outlined in **blue**.

via 19k expert-developed questions, MMMU [64], which challenges models with specialized academic content drawn from collegiate curricula, Science-QA [77], which probes multidisciplinary reasoning across scientific domains with structured question formats, and AI2D [20], which assesses visual-scientific literacy through diagram interpretation exercises. The results are presented in Table 3. We highlight that CHAIR-DPO even improves LLaVA-1.5-7B on MME, MMMU, and Science-QA, while scoring comparable to the reference model and other open-source MLLMs on SEED and AI2D. Notably, this behavior is consistent no matter the strength of the β regularizer. On the other hand, LLaVA-MORE-8B records a reasonable penalty on most benchmarks, even though CHAIR-DPO enhances its performance on SEED-Video. We argue that this regression on general benchmarks is more than acceptable given the notable reduction in hallucinations testified by Table 1. Conversely, we conclude that CHAIR-DPO does not cause a catastrophic forgetting of the knowledge acquired by LLaVA models during visual instruction tuning. We credit for this the Kullback-Leibler regularization of DPO, controlled by β , as well as the efficient fine-tuning enabled by LoRA, which results in a minimal change to the parameters of the baseline model.

Qualitative Results. Finally, Figure 3 presents captions generated by LLaVA-1.5-7B and LLaVA-MORE-8B [9] conditioned on the prompt: *Describe the image*, before and after fine-tuning with CHAIR-DPO. As it can be seen, LLaVA-1.5-7B [45] generates a reasonable description of the first picture, but then mentions the erroneous presence of two other people and a chair in the same room. CHAIR-DPO effectively avoids such hallucinations, and rather adds the detail of the glasses worn by the woman. A similar pattern is repeated with LLaVA-MORE-8B, which hallucinates a total of four alleged people, as well as a pair of benches. Conversely, CHAIR-DPO correctly identifies that no other vehicles nor pedestrian appear in the image other than the double-decker bus, further confirming the effectiveness of the proposed method in mitigating hallucinations. Additional qualitative results of both models are shown in Appendix B.

5 Conclusion

In this work, we introduced CHAIR-DPO, a novel preference optimization method for mitigating hallucinations in Multimodal Large Language Models. By leveraging the well-established CHAIR metric to build preference data for DPO training, our approach achieves state-of-the-art performance across multiple hallucination benchmarks while requiring only an off-the-shelf object detector in the data collection stage. Unlike existing approaches that rely on complex pipelines or proprietary MLLMs to generate preference data, CHAIR-DPO provides a simpler yet effective alignment framework. Our experiments demonstrate that CHAIR-DPO not only reduces hallucinations significantly but also preserves the general capabilities of the baseline models, with minimal regression on standard benchmarks. The effectiveness of our method suggests that object awareness is a critical component in developing MLLMs that produce factually accurate responses grounded in visual inputs.

Acknowledgments

We acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources. This work has been supported by the EU Horizon projects “ELIAS” (GA No. 101120237) and “ELLIOT” (GA No. 101214398), by the EuroHPC JU project “MINERVA” (GA No. 101182737), and by the PRIN project “MUSMA” (CUP G53D23002930006 - M4C2 I1.1), funded by the EU - NextGenerationEU.

References

- [1] Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. Qwen-VL: A Versatile Vision-Language Model for Understanding, Localization, Text Reading, and Beyond. *arXiv preprint arXiv:2308.12966*, 2023.
- [2] Zechen Bai, Pichao Wang, Tianjun Xiao, Tong He, Zongbo Han, Zheng Zhang, and Mike Zheng Shou. Hallucination of Multimodal Large Language Models: A Survey. *arXiv preprint arXiv:2404.18930*, 2024.
- [3] Manuele Barraco, Sara Sarto, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. With a Little Help from your own Past: Prototypical Memory Networks for Image Captioning. In *ICCV*, 2023.
- [4] Davide Caffagni, Federico Cocchi, Luca Barsellotti, Nicholas Moratelli, Sara Sarto, Lorenzo Baraldi, Lorenzo Baraldi, Marcella Cornia, and Rita Cucchiara. The Revolution of Multimodal Large Language Models: A Survey. In *ACL Findings*, 2024.
- [5] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-End Object Detection with Transformers. In *ECCV*, 2020.
- [6] Keqin Chen, Zhao Zhang, Weili Zeng, Richong Zhang, Feng Zhu, and Rui Zhao. Shikra: Unleashing Multimodal LLM’s Referential Dialogue Magic. *arXiv preprint arXiv:2306.15195*, 2023.

- [7] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality, 2023. URL <https://lmsys.org/blog/2023-03-30-vicuna/>.
- [8] Yung-Sung Chuang, Yujia Xie, Hongyin Luo, Yoon Kim, James Glass, and Pengcheng He. DoLa: Decoding by Contrasting Layers Improves Factuality in Large Language Models. In *ICLR*, 2024.
- [9] Federico Cocchi, Nicholas Moratelli, Davide Caffagni, Sara Sarto, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. LLaVA-MORE: A Comparative Study of LLMs and Visual Backbones for Enhanced Visual Instruction Tuning. In *ICCV Workshops*, 2025.
- [10] Wenliang Dai, Junnan Li, Dongxu Li, Anthony Meng Huat Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale Fung, and Steven Hoi. InstructBLIP: Towards General-purpose Vision-Language Models with Instruction Tuning. *arXiv preprint arXiv:2305.06500*, 2023.
- [11] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The Llama 3 Herd of Models. *arXiv preprint arXiv:2407.21783*, 2024.
- [12] Chaoyou Fu, Peixian Chen, Yunhang Shen, Yulei Qin, Mengdan Zhang, Xu Lin, Jinrui Yang, Xiawu Zheng, Ke Li, Xing Sun, et al. MME: A Comprehensive Evaluation Benchmark for Multimodal Large Language Models. *arXiv preprint arXiv:2306.13394*, 2023.
- [13] Jiaming Han, Kaixiong Gong, Yiyuan Zhang, Jiaqi Wang, Kaipeng Zhang, Dahua Lin, Yu Qiao, Peng Gao, and Xiangyu Yue. OneLLM: One Framework to Align All Modalities with Language. In *CVPR*, 2024.
- [14] Jiwoo Hong, Noah Lee, and James Thorne. ORPO: Monolithic Preference Optimization without Reference Model. In *EMNLP*, 2024.
- [15] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models. *arXiv preprint arXiv:2106.09685*, 2021.
- [16] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Trans. on Information Systems*, 43(2):1–55, 2025.
- [17] Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, and Nenghai Yu. OPERA: Alleviating Hallucination in Multimodal Large Language Models via Over-Trust Penalty and Retrospection-Allocation. In *CVPR*, 2024.
- [18] Hongrui Jia, Chaoya Jiang, Haiyang Xu, Wei Ye, Mengfan Dong, Ming Yan, Ji Zhang, Fei Huang, and Shikun Zhang. SymDPO: Boosting In-Context Learning of Large Multimodal Models with Symbol Demonstration Direct Preference Optimization. *arXiv preprint arXiv:2411.11909*, 2024.

- [19] Songtao Jiang, Yan Zhang, Ruizhe Chen, Yeying Jin, and Zuozhu Liu. Modality-Fair Preference Optimization for Trustworthy MLLM Alignment. *arXiv preprint arXiv:2410.15334*, 2024.
- [20] Aniruddha Kembhavi, Mike Salvato, Eric Kolve, Minjoon Seo, Hannaneh Hajishirzi, and Ali Farhadi. A Diagram is Worth a Dozen Images. In *ECCV*, 2016.
- [21] Hugo Laurençon, Andrés Marafioti, Victor Sanh, and Léo Tronchon. Building and Better Understanding Vision-Language Models: Insights and Future Directions. In *NeurIPS Workshops*, 2024.
- [22] Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, and Lidong Bing. Mitigating Object Hallucinations in Large Vision-Language Models through Visual Contrastive Decoding. In *CVPR*, 2024.
- [23] Bohao Li, Rui Wang, Guangzhi Wang, Yuying Ge, Yixiao Ge, and Ying Shan. SEED-Bench: Benchmarking Multimodal LLMs with Generative Comprehension. *arXiv preprint arXiv:2307.16125*, 2023.
- [24] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *ECCV*, 2014.
- [25] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual Instruction Tuning. In *NeurIPS*, 2023.
- [26] Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved Baselines with Visual Instruction Tuning. In *CVPR*, 2024.
- [27] Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering. In *NeurIPS*, 2022.
- [28] Nicholas Moratelli, Davide Caffagni, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Revisiting Image Captioning Training Paradigm via Direct CLIP-based Optimization. In *BMVC*, 2024.
- [29] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training Language Models to Follow Instructions with Human Feedback. In *NeurIPS*, 2022.
- [30] Artemis Panagopoulou, Le Xue, Ning Yu, Junnan Li, Dongxu Li, Shafiq Joty, Ran Xu, Silvio Savarese, Caiming Xiong, and Juan Carlos Niebles. X-InstructBLIP: A Aramework for Aligning X-Modal Instruction-Aware Representations to LLMs and Emergent Cross-Modal Reasoning. *arXiv preprint arXiv:2311.18799*, 2023.
- [31] Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, and Furu Wei. KOSMOS-2: Grounding Multimodal Large Language Models to the World. *arXiv preprint arXiv:2306.14824*, 2023.
- [32] Suzanne Petryk, David M Chan, Anish Kachinthaya, Haodi Zou, John Canny, Joseph E Gonzalez, and Trevor Darrell. ALOHa: A New Measure for Hallucination in Captioning Models. In *NAACL*, 2024.

- [33] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning Transferable Visual Models from Natural Language Supervision. In *ICML*, 2021.
- [34] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *NeurIPS*, 2023.
- [35] Hanoona Rasheed, Muhammad Maaz, Salman Khan, and Fahad S. Khan. LLaVA++: Extending Visual Capabilities with LLaMA-3 and Phi-3, 2024.
- [36] Hanoona Rasheed, Muhammad Maaz, Sahal Shaji, Abdelrahman Shaker, Salman Khan, Hisham Cholakkal, Rao M Anwer, Eric Xing, Ming-Hsuan Yang, and Fahad S Khan. GLaMM: Pixel Grounding Large Multimodal Model. In *CVPR*, 2024.
- [37] Anna Rohrbach, Lisa Anne Hendricks, Kaylee Burns, Trevor Darrell, and Kate Saenko. Object Hallucination in Image Captioning. In *EMNLP*, 2018.
- [38] Pranab Sahoo, Prabhash Meharia, Akash Ghosh, Sriparna Saha, Vinija Jain, and Aman Chadha. A Comprehensive Survey of Hallucination in Large Language, Image, Video and Audio Foundation Models. In *EMNLP Findings*, 2024.
- [39] Pritam Sarkar, Sayna Ebrahimi, Ali Etemad, Ahmad Beirami, Sercan Ö Arık, and Tomas Pfister. Mitigating Object Hallucination in MLLMs via Data-augmented Phrase-level Alignment. In *ICLR*, 2025.
- [40] Sara Sarto, Marcella Cornia, and Rita Cucchiara. Image Captioning Evaluation in the Age of Multimodal LLMs: Challenges and Future Perspectives. In *IJCAI*, 2025.
- [41] Feifan Song, Bowen Yu, Minghao Li, Haiyang Yu, Fei Huang, Yongbin Li, and Houfeng Wang. Preference Ranking Optimization for Human Alignment. In *AAAI*, 2024.
- [42] Quan Sun, Yufeng Cui, Xiaosong Zhang, Fan Zhang, Qiying Yu, Yueze Wang, Yongming Rao, Jingjing Liu, Tiejun Huang, and Xinlong Wang. Generative Multimodal Models Are In-Context Learners. In *CVPR*, 2024.
- [43] Quan Sun, Qiying Yu, Yufeng Cui, Fan Zhang, Xiaosong Zhang, Yueze Wang, Hongcheng Gao, Jingjing Liu, Tiejun Huang, and Xinlong Wang. Emu: Generative Pretraining in Multimodality. In *ICLR*, 2024.
- [44] Chameleon Team. Chameleon: Mixed-Modal Early-Fusion Foundation Models. *arXiv preprint arXiv:2405.09818*, 2024.
- [45] Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caimeing Xiong, Shafiq Joty, and Nikhil Naik. Diffusion Model Alignment Using Direct Preference Optimization. In *CVPR*, 2024.
- [46] Fei Wang, Wenxuan Zhou, James Y Huang, Nan Xu, Sheng Zhang, Hoifung Poon, and Muhao Chen. mDPO: Conditional Preference Optimization for Multimodal Large Language Models. In *EMNLP*, 2024.

- [47] Junyang Wang, Yuhang Wang, Guohai Xu, Jing Zhang, Yukai Gu, Haitao Jia, Jiaqi Wang, Haiyang Xu, Ming Yan, Ji Zhang, et al. AMBER: An LLM-free Multi-dimensional Benchmark for MLLMs Hallucination Evaluation. *arXiv preprint arXiv:2311.07397*, 2023.
- [48] Junkang Wu, Yuexiang Xie, Zhengyi Yang, Jiancan Wu, Jinyang Gao, Bolin Ding, Xiang Wang, and Xiangnan He. β -DPO: Direct Preference Optimization with Dynamic β . In *NeurIPS*, 2024.
- [49] Tsung-Han Wu, Heekyung Lee, Jiaxin Ge, Joseph E Gonzalez, Trevor Darrell, and David M Chan. Generate, but Verify: Reducing Hallucination in Vision-Language Models with Retrospective Resampling. *arXiv preprint arXiv:2504.13169*, 2025.
- [50] Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli. Hallucination is Inevitable: An Innate Limitation of Large Language Models. *arXiv preprint arXiv:2401.11817*, 2024.
- [51] Qinghao Ye, Haiyang Xu, Jiabo Ye, Ming Yan, Anwen Hu, Haowei Liu, Qi Qian, Ji Zhang, and Fei Huang. mPLUG-Owl2: Revolutionizing Multi-modal Large Language Model with Modality Collaboration. In *CVPR*, 2024.
- [52] Shukang Yin, Chaoyou Fu, Sirui Zhao, Tong Xu, Hao Wang, Dianbo Sui, Yunhang Shen, Ke Li, Xing Sun, and Enhong Chen. Woodpecker: Hallucination Correction for Multimodal Large Language Models. *Science China Information Sciences*, 67(12): 220105, 2024.
- [53] Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun, et al. RLHF-V: Towards Trustworthy MLLMs via Behavior Alignment from Fine-grained Correctional Human Feedback. In *CVPR*, 2024.
- [54] Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, et al. MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI. In *CVPR*, 2024.
- [55] Zihao Yue, Liang Zhang, and Qin Jin. Less is More: Mitigating Multimodal Hallucination from an EOS Decision Perspective. In *ACL*, 2024.
- [56] Ruohong Zhang, Liangke Gui, Zhiqing Sun, Yihao Feng, Keyang Xu, Yuanhan Zhang, Di Fu, Chunyuan Li, Alexander Hauptmann, Yonatan Bisk, et al. Direct Preference Optimization of Video Large Multimodal Models from Language Model Reward. *arXiv preprint arXiv:2404.01258*, 2024.
- [57] Zhiyuan Zhao, Bin Wang, Linke Ouyang, Xiaoyi Dong, Jiaqi Wang, and Conghui He. Beyond Hallucinations: Enhancing LVLMs through Hallucination-Aware Direct Preference Optimization. *arXiv preprint arXiv:2311.16839*, 2023.
- [58] Yiyang Zhou, Chenhang Cui, Rafael Rafailov, Chelsea Finn, and Huaxiu Yao. Aligning Modalities in Vision Large Language Models via Preference Fine-tuning. *arXiv preprint arXiv:2402.11411*, 2024.