Multi-modal Preference Alignment Remedies Degradation of Visual Instruction Tuning on Language Models

Shengzhi Li
TIFIN Inc
alex.li@tifin.com

Rongyu Lin
KAUST
rongyu.lin@kaust.edu.sa

Shichao Pei*
University of Massachusetts Boston
shichao.pei@umb.edu

ACL 2024
Bangkok, Thailand

Motivation

Multi-modal large language models (MLLMs) are expected to support multi-turn queries of interchanging image and text modalities in production.

However, the current MLLMs trained with visual-question-answering (VQA) datasets could suffer from degradation, as VQA datasets lack the diversity and complexity of the original text instruction datasets with which the underlying language model was trained.

Image

+

Question

What do you see happening in this image?

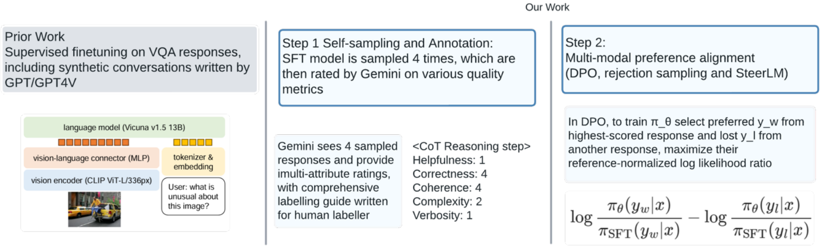
Contribution

- **Exploration of Modality Degradation**
  This work is **the first to identify and address modality degradation** in MLLMs, a phenomenon where visual instruction tuning detrimentally impacts language instruction capabilities.
- **Efficient and scalable preference alignment pipeline as remedy**

- Our data collection strategy employs a granular quality metric annotation format, leveraging cost-effective commercial APIs.
- This scalable approach enables the efficient production of high-quality datasets.
- We are able to surpass LLaVA and Vicuna's language instructionfollowing capability **with DPO on a 6k dataset**.

Method

We propose to harness alignment methods that utilize **self-sampled responses and preference annotations** in addition to Supervised Fine-Tuning (SFT) as a baseline.



From a visual-instruction-tuned pre-trained model, **we generate 4 completions for a given image-question prompt**. These answers are then **presented to Gemini** to obtain granular annotations given a labeling guide. **We construct a preference dataset** of (image-text prompt, preferred completion) and (image-text prompt, rejected completion). **We benchmarked DPO, Rejection sampling, and SteerLM alignment methods**, in addition to a pure SFT baseline using Gemini-provided answers directly.

5 Metric

Helpfulness I Correctness
Coherence I Complexity
Verbosity

3 Alignment Methods

- Direct Preference Optimization(DPO)
- Self-sampled SteerLM
- Rejection Sampling

2 Data

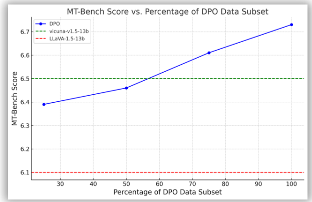
Data Type	Data Name	Size
VQA	LRV-Instruct	2562
	SciGraphQA	2522
Total		5084

Our SFT method relied on answers from the respective datasets directly, while DPO, SteerLM, and rejection sampling methods use self-generated instead.

Experiments

Model Name	Visual Instruction Benchmark		Visual Multi-Choice Benchmark		Text Instruction Benchmark	
	MM-Vet	LLaVA-bench	PoPe	MM-Bench	MT-bench	AlpacaEval
Vicuna-1.5-13b (Chiang et al., 2023)	-	-	-	-	6.57	81.4
LLaVA-1.5-13b (Liu et al., 2023b)	36.3	73.1	0.859	67.4	5.99	79.3
LLaVA-RLHF-13b (Sun et al., 2023)	37.2	76.8	0.869	60.1	6.18	81.0
Alignment method we benchmarked, finetuning LLaVA-1.5-13b						
Standard SFT	36.5	63.7	0.850	65.4	5.01	50.2
SteerLM	35.2	67.0	0.878	65.1	5.70	68.8
Rejection-sampling	38.0	70.6	**0.883**	**67.6**	6.22	74.9
DPO	**41.2**	**79.1**	0.870	66.8	**6.73**	**86.4**

Result: LLaVA's performance on MT-Bench had dipped from Vicuna's 6.57 to 5.99, whereas **our DPO model advanced to 6.73**. DPO also bolstered performance on multi-modal benchmarks, enhancing **accuracy by 4.9% on MM-Vet and 6% on LLaVABench**.



The performance surpasses that of the Vicuna-v1.5-13b benchmark **using less than 75% or 4.2K of the DPO data**, underlining DPO's data efficiency.

Discussion

1 Limitation of collecting multi-modal preference data manually

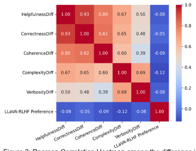


Figure 3: Pearson Correlation Heatmap among the difference in Gemini-Annotated data attributes and LLaVARLHF human annotated preference (n=500).

2 Empirical validation of data quality

Data Noising Level	MT-Bench Score
No flip (baseline)	6.73
25% flip	6.35
50% flip	6.26
75% flip	5.99

Table 5: Impact of data noising on model performance.

3 Robustness of the model with noisy context

Model	LLM	Noisy-image MT-Bench
Vicuna 13B v1.5	Vicuna_{13B}	6.57
Vicuna 7B v1.5	Vicuna_{7B}	6.17
BLIP-2	FlanT5	1.93
InstructBLIP	Vicuna_{7B}	4.73
LLaVA-v1.5-13b	Vicuna_{13B}	5.92
DPO (ours)	Vicuna_{13B}	6.63

Table 7: This benchmark adds an irrelevant image in context when benchmarking MT-Bench, testing for robustness in a real-world condition. Vicuna-7B and Vicuna-13B are baseline references.

4 Cross-model transfer ability: Can preference dataset generated by one model be transferred to other models?

Model Name	MT-Bench Score	MM-Vet Score
Vicuna-7B-v1.5	6.17	N/A
LLaVA-V1.5-7b	5.87	30.5
DPO-7b (ours)	6.228	39.8

Table 6: Performance improvements with multi-modal preference data application.

5 Multi-modal preference alignment as a data-efficient remedy to instruction tuning capabilities

Model Name	Visual Instruction Benchmark		Visual Multi-Choice Benchmark		Text Instruction Benchmark	
	MM-Vet	LLaVA-bench	PoPe	MM-Bench	MT-bench	AlpacaEval
Vicuna-1.5-13b (Chiang et al., 2023)	-	-	-	-	6.57	81.4
LLaVA-1.5-13b (Liu et al., 2023b)	36.3	73.1	0.859	67.4	5.99	79.3
LLaVA-RLHF-13b (Sun et al., 2023)	37.2	76.8	0.869	60.1	6.18	81.0
Alignment method we benchmarked, finetuning LLaVA-1.5-13b						
Standard SFT	36.5	63.7	0.850	65.4	5.01	50.2
SteerLM	35.2	67.0	0.878	65.1	5.70	68.8
Rejection-sampling	38.0	70.6	**0.883**	**67.6**	6.22	74.9
DPO	**41.2**	**79.1**	0.870	66.8	**6.73**	**86.4**

Table 4: Performance comparison among alignment strategies.