Improving Compositional Reasoning of Vision Language Models

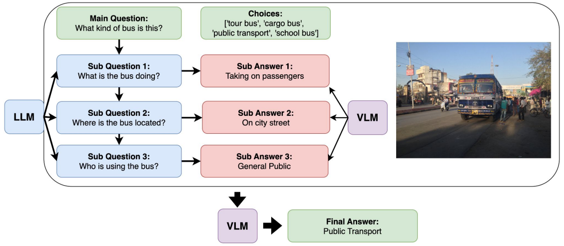
Team 13: Busra Tugce Gurbuz, Ba Luan Dang, Qian Yang

Compositional Reasoning & Multi-Agent Collaboration

Vision-Language Models struggle with **compositional reasoning** — breaking down complex visual tasks into simpler steps.

🧠 Like how humans talk through problems, we pair a **VLM** with an **LLM** as a collaborator.

🤝 The LLM becomes a **Decomposer Agent**
→ Breaks down complex questions
→ Guides the VLM step-by-step



Smarter Task Decomposition

⚠️ **Challenges:**

🟣 LLMs not trained specifically for task decomposition

🟣 LLMs unaware of VLM strengths & limits

🟣 Prior work [1] fine-tuned an LLM using DPO with VLM accuracy as the reward, but relied on preferences generated from a general-purpose LLM, limiting the decomposers ability to specialize for the VLM.

[1] Yang, Q., Yan, W., & Agrawal, A. (2024). Enhancing Multi-Agent Multi-Modal Collaboration with Fine-Grained Reward Modeling. In Adaptive Foundation Models: Evolving AI for Personalized and Efficient Learning

[2] Chen, C., Liu, Z., Du, C., Pang, T., Liu, Q., Sinha, A., ... & Lin, M. (2025). Bootstrapping language models with DPO implicit rewards. arXiv preprint arXiv:2406.09760.

Do Decomposed Questions Help VLMs?

Experiment-1: Idefics-2-8B + OpenHermes-2.5-Mistral-7B

Idefics-2-8B	SNLI-VE	VCR	MathVista	Average
Base MLLM	41.1	62.1	49.3	50.8
+ Chain of Thought	44.4	59.1	47.2	50.2
+ Pre-Decomposition	55	**63.9**	**49.8**	**56.2**
+ Interactive Decomposition	56.5	61.7	49.3	55.8
+ Interactive Decomposition with SF	56.5	63	48.3	55.9
+ Interactive Decomposition with DPO [1]	**57.9**	62.3	48.4	56.2

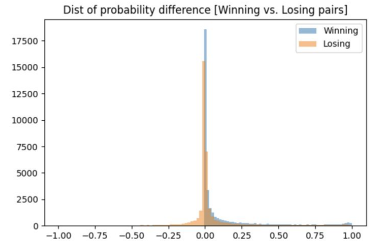
✅ Multi-agent collaboration helps to guide weaker VLMs

🔍 Can DPO Help More?

🟣 We analyzed sub-question quality from the LLM decomposer.

🟣 The **winning vs. losing sub-questions look similar**.

🟣 Sub-questions often **aren't informative enough** for the VLM.



🔍 Stronger VLMs Benefit More from LLM Decomposer?

Experiment-2: Idefics-3-8B / Qwen-VL-32B + OpenHermes-2.5-Mistral-7B

Idefics-3-8B / Qwen-VL-2.5-32B	SNLI-VE		VCR		MathVista		Average	
Base MLLM	**67.3**	**73.3**	**61.6**	**69.8**	50.9	74.8	**59.9**	72.6
+ Chain of Thought	55.2	71.3	46	71.5	50.2	**76.1**	50.5	**73**
+ Pre-Decomposition	62.3	69.1	58.8	67.1	48.8	70.4	56.6	68.9
+ Interactive Decomposition	60.3	-	58.1	-	**51.2**	-	56.5	-

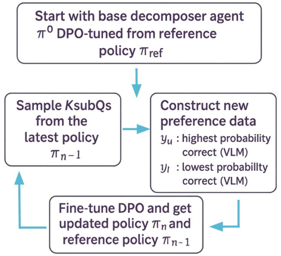
❌ Stronger VLMs perform reasoning better on their own

🔍 Stronger VLMs with Stronger LLM

Experiment-3: Qwen-VL-32B + DeepSeek-R1-Distill-Qwen-32B

✅ Paired with stronger LLM, performance rises to match the base VLM level (71.05 on SNLI-VE).

💡 VLM-Specialized Decomposition via Adaptive Fine-Tuning Loop



Shaping implicit DPO reward with mutual information [2]:

$$r_{\mathrm{MI}}(\mathbf{x}, \mathbf{y}) = \beta \log \frac{\pi_\theta(\mathbf{y} \mid \mathbf{x})}{\pi_{\text{ref}}(\mathbf{y} \mid \mathbf{x})} + \lambda \, \mathrm{MI}(\mathbf{y}, \mathbf{o} \mid \mathbf{x})$$

Summary

Key Insight:

➢ Multi-agent collaboration helps weaker VLMs

➢ Stronger VLMs require **better-tuned decomposers**.

Our Contributions:

✅ Show that Decomposition boosts mid-tier VLMs performance

✅ Analyze why naive DPO struggles: uninformative sub-questions

✅ Propose a **VLM-aware adaptive fine-tuning loop** for the LLMs

✅ Introduce **MI–shaped reward** for better alignment

Future Work:

⌛Apply proposed adaptive fine-tuning with the better starting point