Insight-V: Exploring Long-Chain Visual Reasoning with Multimodal Large Language Models

Supplementary Material

The supplementary material is organized as follows: Section 1 provides detailed implementation information about the Insight-V system to enhance reproducibility and promote a deeper understanding of our approach. Section 2 presents additional analytical experiments, emphasizing the significance of the multi-agent system design and offering an intuitive perspective on its necessity. Finally, Section 3 provides an in-depth discussion of our approach, acknowledges its limitations, and outlines future directions for developing o1-like reasoning models.

1. More Implementation Details

1.1. Training Details

In this section, we provide a detailed explanation of the implementation of the DPO strategy. To collect preference data, we sample 16 outputs for each image-text pair to ensure diversity and maintain data quality. Each question, along with its ground truth answer and corresponding reasoning processes, is then presented to advanced LLMs such as Qwen2.5-72B. The model evaluates all reasoning paths in a single forward pass, assigning scores to each. Reasoning paths with scores above 85 are selected as positive examples. To increase the task's complexity, we do not use the lowest-scoring reasoning path as the rejected example. Instead, we choose a reasoning path with a score around 25, ensuring that the DPO-trained model does not overfit specific data patterns. During DPO training, the parameter β is set to 0.1, and a standard supervised fine-tuning loss is incorporated to stabilize the training process.

2. Analysis Experiments of Multi-agent System

To further validate the effectiveness of the proposed multi-agent system, we conduct additional analysis experiments highlighting the superior performance of Insight-V.

Insight-V Generates More Accurate Reasoning Paths and Demonstrates Robustness to Flawed Reasoning. To demonstrate that the summary model of Insight-V effectively evaluates the quality of reasoning paths and selectively answers questions based on these paths, we conduct analysis experiments on MMStar. These experiments highlight why Insight-V benefits from the integration of the summary agent, leading to improvements in performance.

As illustrated in Figure 1, we compute the confusion matrix for the reasoning path and the final answer. A reasoning path is classified as `True Positive` if both the reasoning path and the final answer are correct, represented in the bottom-right corner of the matrix. Conversely, if the reasoning path is incorrect but the final answer is correct, it is categorized as `False Negative`, shown in the upper-left corner of the matrix. The results clearly demonstrate that Insight-V generates more accurate reasoning paths, as depicted in the figure. Moreover, even when the reasoning path is incorrect, Insight-V is still capable of producing the correct final answer, showcasing its superior ability to selectively utilize reasoning paths compared to direct fine-tuning with Chain-of-Thought data.

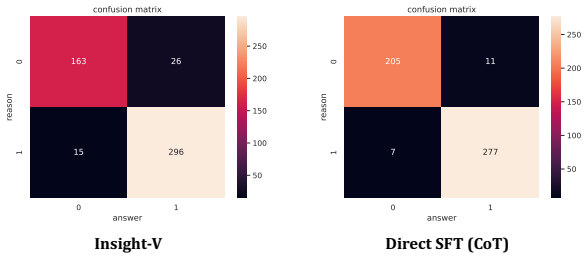

Figure 1. **Analysis of Multi-agent System.** Insight-V enhances reasoning capabilities while enabling the ability to selectively answer questions based on the provided reasoning process.

3. Discussion and Limitations

Insight-V represents an initial exploration into building models capable of o1-like reasoning. Our findings indicate that leveraging MLLMs to perform single-step reasoning, and organizing these steps into structured, long-chain reasoning paths, is a promising approach. After fine-tuning on this dataset, the model demonstrates the ability to perform long-chain reasoning. Additionally, we implement a multi-agent system to decompose the question-answering process into distinct reasoning and summarization stages, enabling the system to focus on reasoning while selectively incorporating its results into the summarization process.

As an early attempt to develop robust reasoning models, we acknowledge several limitations that warrant future improvement. First, enhancing sampling efficiency is critical. Currently, the process depends on other models for multi-granularity assessment, which could be made more efficient by evaluating reasoning results at each step and pruning redundant samples. This would streamline the system and improve its overall efficiency.

Furthermore, training two models of the same size may not be scalable. Improving the reasoning agent could allow for training a smaller, cost-effective summarization agent, as summarization is inherently a less complex task than reasoning. This adjustment would not only reduce resource requirements but also improve the system's scalability.

In conclusion, we hope our method serves as a foundational attempt to inspire and guide future research in this emerging and exciting field.