SFT 모델 성능 최적화를 위한 데이터셋 품질 중심 접근*

김도연[1] 하세인[O1] 이상민[2] 이다영[1] 김명섭[1]
[1]서강대학교 컴퓨터공학과
[2]인하대학교 전자공학과
jshine0914@sogang.ac.kr, gktpdls@sogang.ac.kr, smlee25@inha.edu, dalle3934@sogang.ac.kr,
kyh147159@sogang.ac.kr

Quality over Quantity: The Impact of Dataset Refinement on SFT Model Performance

Doyeon Kim[1] Sein Ha[O1] Sangmin Lee[2] Dayoung Lee[1] Myoungsub Kim[1]
[1]School of Computer Science and Engineering, Sogang University
[2]School of Electrical Engineering, Inha University

Abstract

In this study, we investigate the impact of dataset refinement on the performance of supervised fine-tuning(SFT) for multimodal reasoning models. While previous research has emphasized scaling up dataset size, our experiments show that data quality plays a more crucial role than quantity. We construct 2500 multimodal SFT trajectories using a multi-expert sampling strategy and refine them to 1700 high-quality samples based on strict filtering criteria. Using Qwen2.5-VL-7B-Instruct as the base model, we compare SFT models trained on both noisy and refined datasets across three benchmarks: SlideVQA, ViDoSeek, and MMLongBench. The refined 1300-sample model outperforms the unrefined 2500-sample model by 13.89% in overall accuracy, despite using fewer data. Furthermore, smaller subsets of 100-900 refined samples yield comparable results, indicating that small but clean datasets suffice for effective SFT. These findings highlight that dataset refinement, not expansion, is the key driver of reasoning performance in multimodal SFT.

1. Introduction

Supervised fine-tuning(SFT) is widely used to adapt large multimodal models to complex reasoning tasks, yet the belief that "more data is better" often ignores the harm caused by noisy or inconsistent samples. This work challenges that assumption by showing that refining SFT datasets produces greater performance gains than merely scaling them.

We study structured SFT for multimodal reasoning models, where reasoning steps are represented through tags such as <think>, <search>, <bbox>, and <search_complete>. This design yields interpretable reasoning traces and allows precise analysis of data quality. Through systematic filtering and quantitative evaluation, we demonstrate that dataset refinement substantially improves learning stability, reliability, and reasoning coherence.

2. Related Work

Retrieval-Augmented Generation

Retrieval-Augmented Generation(RAG) integrates retrieval and generation to support knowledge-intensive reasoning[1]. Beyond traditional text-based pipelines, recent work explores multimodal retrieval using OCR-free visual alignment, and agents that couple reasoning with visual perception for improved grounding[2]. Unlike end-to-end RAG models, our study isolates the multimodal search agent to examine how SFT dataset refinement influences its reasoning and retrieval performance.

Supervised Fine-Tuning

Supervised Fine-Tuning(SFT) adapts pre-trained models to task-specific objectives[3]. Although past research largely focused on scaling dataset size, studies such as LIMA[4] and SHED[5] show that high-quality, well-curated data can outperform large but noisy corpora. Motivated by this, we investigate how refining SFT trajectories, by removing inconsistent or low-quality reasoning steps, improves stability and retrieval accuracy in multimodal search agents.

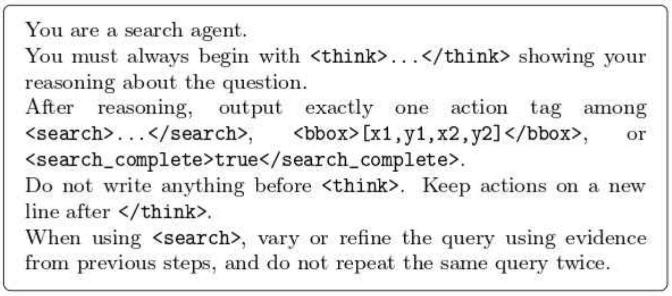
* 본 연구는 대동AILab의 "생성형AI 기반 농업영역 멀티모달 질의응답 에이전트 기술자문" 과제와 2025년 과학기술정보통신부 및 정보통신기획평가원의 "SW중심대학사업(2024-0-00043)"의 지원을 받아 수행되었습니다.

3. Method

3.1 SFT Prompt

In the supervised fine-tuning(SFT) stage, we employ a structured system prompt to explicitly guide the model's reasoning and retrieval behavior.

The system prompt is defined as follows:



You are a search agent.
You must always begin with <think>...</think> showing your reasoning about the question.
After reasoning, output exactly one action tag among <search>...</search>, <bbox>[x1,y1,x2,y2]</bbox>, or <search_complete>true</search_complete>.
Do not write anything before <think>. Keep actions on a new line after </think>.
When using <search>, vary or refine the query using evidence from previous steps, and do not repeat the same query twice.

Figure 1: SFT Prompt

This prompt enforces a structured reasoning-to-action pipeline. The model first generates an explicit chain of thought enclosed with <think> tags, then executes one of three possible actions, conducting a new search(<search>), identifying a visual region(<bbox>), or completing the search process(<search_complete>). After the reasoning process concludes with <search_complete>, a post-search generation step produces the final textual answer based on the retrieved visual evidence. By constraining the reasoning process with explicit tags, the prompt guides the model to follow a clear and goal directed reasoning path instead of unstructured generation. This design not only improves consistency and control in multi-turn reasoning but also enhances the interpretability of the model's decision-making process.

3.2 Dataset Construction

The initial 2500 SFT trajectories were generated using a multi-expert sampling strategy. Large-scale models guide the reasoning process and tool selections within a trajectory, while smaller expert models annotate coordinates under the guidance of large-scale models. This approach enhances trajectory diversity and reasoning richness but inevitably introduces noise, such as incorrect <search> queries, redundant <bbox> calls, and incomplete search terminations.

3.3 Dataset Refinement

From the initial 2500 trajectories, we refined and retained 1700 high-quality SFT samples by applying a set of carefully designed filtering criteria aimed at improving reasoning consistency and alignment with real evaluation settings.

First, we performed reference page validation to ensure that each <search> action successfully retrieved the reference page containing the ground-truth answer. By removing trajectories that failed to retrieve the reference page, we ensured that the SFT model learned meaningful retrieval behavior rather than memorizing incomplete or spurious completions.

Second, we applied controlled <bbox> usage filtering, discarding trajectories that exhibited redundant or arbitrary cropping actions. Excessive <bbox> generation often led to incorrect or premature termination of the search process, thereby distorting the intended reasoning sequence.

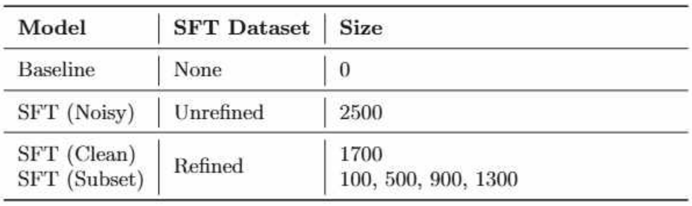
Third, we imposed a multi-turn depth constraint by excluding trajectories exceeding five reasoning turns. This decision aligns with the inference configuration of the multimodal search agent(MAX_TURNS = 5), ensuring that the model's learning behavior reflects the same structural constraints as during evaluation.

These refinements yield a dataset where each trajectory represents a clear, goal-oriented reasoning path, closely aligned with real evaluation settings.

3.4 Experimental Setup

We evaluated seven model configurations in total, including one non-trained baseline and six SFT-trained variants of the Qwen2.5-VL-7B-Instruct model.

Table 1: Dataset Refinement Result



Model	SFT Dataset	Size
Baseline	None	0
SFT (Noisy)	Unrefined	2500
SFT (Clean) SFT (Subset)	Refined	1700 100, 500, 900, 1300

Training Dataset

We constructed 2500 SFT trajectories using the SlideVQA dataset as the base source.

Evaluation Dataset

Model performance was evaluated on three benchmarks: SlideVQA[6], ViDoSeek[7], and MMLongBench[8], each containing 300 samples.

Evaluation Metric

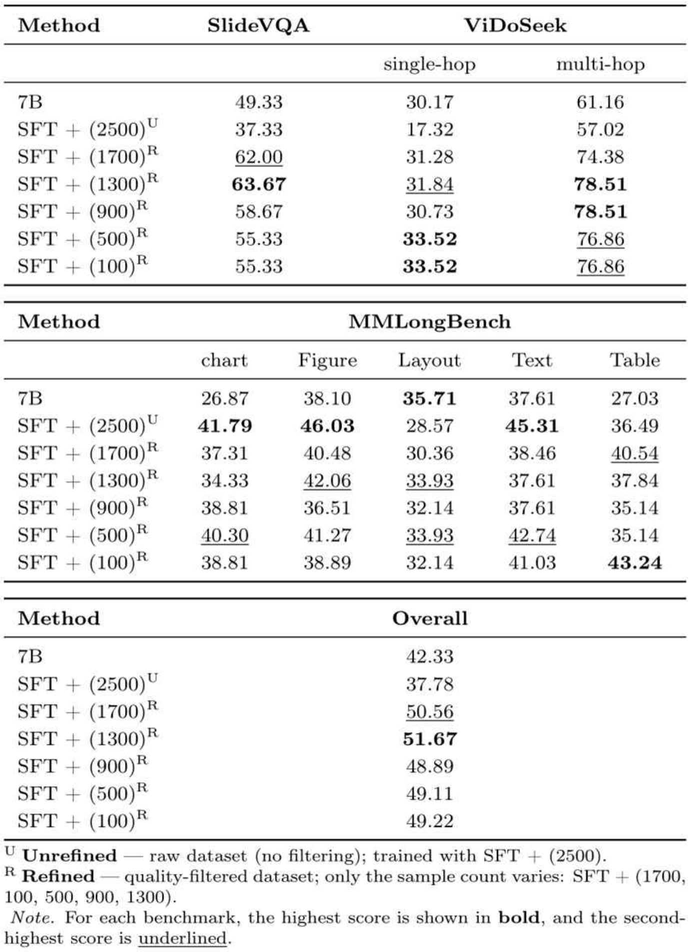
We adopted answer accuracy as the evaluation metric, assessed using the LLM-as-a-Judge framework, which measures the semantic equivalence between generated and reference answers.

3.5 Results

The model trained on 1300 refined samples achieved

13.89% higher accuracy than the one trained on 2500 unrefined samples.

Table 2: Accuracy by Dataset Quality and Quantity



Method	SlideVQA	ViDoSeek	
		single-hop	multi-hop
7B	49.33	30.17	61.16
SFT + (2500)^{U}	37.33	17.32	57.02
SFT + (1700)^{R}	62.00	31.28	74.38
SFT + (1300)^{R}	**63.67**	31.84	**78.51**
SFT + (900)^{R}	58.67	30.73	**78.51**
SFT + (500)^{R}	55.33	**33.52**	76.86
SFT + (100)^{R}	55.33	**33.52**	76.86

Method	MMLongBench				
	chart	Figure	Layout	Text	Table
7B	26.87	38.10	**35.71**	37.61	27.03
SFT + (2500)^{U}	**41.79**	**46.03**	28.57	**45.31**	36.49
SFT + (1700)^{R}	37.31	40.48	30.36	38.46	40.54
SFT + (1300)^{R}	34.33	42.06	33.93	37.61	37.84
SFT + (900)^{R}	38.81	36.51	32.14	37.61	35.14
SFT + (500)^{R}	40.30	41.27	33.93	42.74	35.14
SFT + (100)^{R}	38.81	38.89	32.14	41.03	**43.24**

Method	Overall
7B	42.33
SFT + (2500)^{U}	37.78
SFT + (1700)^{R}	50.56
SFT + (1300)^{R}	**51.67**
SFT + (900)^{R}	48.89
SFT + (500)^{R}	49.11
SFT + (100)^{R}	49.22

^{U} **Unrefined** — raw dataset (no filtering); trained with SFT + (2500).
^{R} **Refined** — quality-filtered dataset; only the sample count varies: SFT + (1700, 100, 500, 900, 1300).
Note. For each benchmark, the highest score is shown in **bold**, and the second-highest score is underlined.

4. Analysis

4.1 Data Refinement over Data Expansion

The most striking observation is that the unrefined 2500-sample model performs worse than the baseline. This demonstrates that noisy or inconsistent trajectories can degrade model reasoning, even when data size increases. In contrast, the refined 1300-sample model achieves 9.34% higher overall accuracy than the baseline, showing that data filtering amplifies effective learning signals and improves reasoning alignment.

4.2 Diminishing Returns of Data Quantity

Among models trained with refined subsets(100-1300), performance differences remain marginal. This suggests that once the data distribution is clean and representative, a small dataset suffices to capture trajectory structures and reasoning formats. In other words, quality saturation occurs: beyond a certain threshold, additional data contributes little to performance.

This observation further implies that if additional filtering mechanisms are introduced to improve dataset precision, even smaller SFT datasets could yield greater performance gains. Future studies could investigate several enhanced refinement strategies, such as difficulty-aware filtering based on problem complexity estimation and crop-accuracy filtering that evaluates the precision of visual region selection. These approaches aim to push the boundary of data efficiency in SFT, enabling higher reasoning performance with minimal data while maintaining strong generalization.

5. Conclusion

This study shows that dataset quality matters more than dataset size in improving the SFT performance of multimodal search agents. Controlled experiments reveal that 1700 refined samples surpass 2500 unrefined ones, and even 100-500 refined samples achieve competitive accuracy. These results underscore the importance of clean, consistent reasoning trajectories. Future work will explore extending this principle to reinforcement learning, leveraging the refined SFT policy for quality-driven curriculum optimization.

6. References

[1] Bowen Jin et al., "Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning", in COLM, 2025.

[2] Qiuchen Wang et al., "VRAG-RL: Empower Vision-Perception-Based RAG for Visually Rich Information Understanding via Iterative Reasoning with Reinforcement Learning", in arXiv preprint arXiv:2505.22019, 2025.

[3] Tianzhe Chu, Yuexiang Zhai et al., "SFT Memorizes, RL Generalizes: A Comparative Study of Foundation Model Post-training", in ICML, 2025.

[4] Chunting Zhou, Pengfei Liu et al., "LIMA: Less Is More for Alignment", in NeurIPS, 2023.

[5] Yexiao He et al., "SHED: Shapley-Based Automated Dataset Refinement for Instruction Fine-Tuning", in NeurIPS, 2024.

[6] Ryota Tanaka et al., "SlideVQA: A Dataset for Document Visual Question Answering on Multiple Images", in AAAI, 2023.

[7] Qiuchen Wang et al., "ViDoRAG: Visual Document Retrieval-Augmented Generation via Dynamic Iterative Reasoning Agents", in arXiv preprint arXiv:2502.18017, 2025.

[8] Yubo Ma et al., "Mmlongbench-doc: Benchmarking long-context document understanding with visualizations", in NeurIPS, 2024.