

A Quality of Point-CoT

To verify the quality of Point-CoT, we invited 10 volunteers and GPT-4o itself to validate 1,000 randomly selected samples, calculating the proportion of high-quality samples. We define high-quality samples as those that do not contain reasoning errors and have essentially correct points. As shown in Table 6, we found that the proportion of high-quality samples in Point-CoT approaches 95%, demonstrating the effectiveness of our Point-CoT.

B Comparison of Models

Table 7 extends the main paper’s Table 2 by comparing two additional baselines, Molmo-7B and Qwen2.5-VL-7B evaluated without the CoT prompting. Superficially, Qwen2.5-VL-7B appears to regress when verbose CoT reasoning is enabled. This drop, however, stems primarily from the model’s in-domain short-answer post-training, not from an intrinsic weakness of long-form reasoning—a phenomenon echoed in other RFT studies [Deng et al. \(2025\)](#); [Wang et al. \(2025a\)](#); [Huang et al. \(2025\)](#); [Meng et al. \(2025\)](#); [Peng et al. \(2025\)](#); [Chen et al. \(2025\)](#); [Zheng et al. \(2025\)](#). Nonetheless, our Point-RFT, which couples grounded point-level reasoning with CoT, consistently outperforms both versions of Qwen2.5-VL-7B as well as Molmo-7B. These results confirm that integrating point-based reasoning substantially boosts multimodal reasoning performance, validating the design of both the Point-CoT dataset and the Point-RFT model.

C Training Steps

As shown in Table 8, SFT requires careful balancing of training steps, reaching peak accuracy at 500 steps (56.72%). Training beyond this point leads to overfitting and performance degradation. For RL, performance monotonically improves with increasing steps, peaking at 100 steps (81.04%). This indicates that extended reward shaping continually refines the reasoning patterns grounded in visual pointing.

D Training Dataset

Table 9 reveals that mixing SFT datasets harms performance (81.04% drops to 75.20%). This suggests that targeted format learning on high-quality grounded CoT data is more effective than training on broader but noisier datasets. Notably, Point-RFT even outperforms models trained on pure ChartQA data for both SFT and RL, demonstrating its ability to generalize.

E Broader Impact

We present Point-RFT, a multimodal reasoning framework that bridges textual and visual reasoning through visually grounded Chain-of-Thought (CoT). By explicitly anchoring rationales to visual elements, our approach mitigates hallucinations and enhances multimodal integration, advancing the reliability of AI systems in real-world document analysis. This innovation holds significant potential for human-AI collaboration. Point-RFT paves the way for trustworthy AI assistants capable of seamless multimodal reasoning across diverse domains.

F Limitation

Despite the promising results, our approach has several limitations. First, the two-stage training pipeline (format finetuning and reinforcement finetuning) requires substantial computational resources, particularly when scaling to larger datasets. Moreover, like most autoregressive models, the sequential reasoning process in Point-RFT introduces significant inference latency and resource consumption, especially for complex visual documents. These challenges highlight the need for more efficient training paradigms and inference acceleration techniques. In future work, we will explore lightweight training strategies and parallelizable decoding mechanisms to address these limitations while maintaining reasoning accuracy.

Table 6: Statistics of dataset.

Evaluation	Samples	Ratio
Human	1000	94.5%
Machine	1000	96.8%

Table 7: Overall results among different datasets.

Method	Setting CoT	In-Domain	Out-of-Domain					
		ChartQA	CharXiv	PlotQA	IconQA	TabMWP	Counting	Avg.
Qwen2.5-VL-7B		87.30	34.40	22.30	59.10	56.40	15.00	39.06
Qwen2.5-VL-7B	✓	70.88	26.50	17.80	53.40	61.00	21.00	35.94
Molmo-7B	✓	20.48	7.60	3.95	16.50	20.70	0.00	9.75
Point-RFT	✓	90.04	36.20	20.40	59.80	70.90	78.50	53.16

Table 8: Ablation studies of training steps.

Method	Steps		ChartQA		
	SFT	RL	Overall	Inner	Format
Base	-	-	79.28	82.17	96.48
Point-SFT	300	-	50.00	71.88	69.56
	500	-	56.72	76.20	74.44
	600	-	45.56	75.48	60.36
	700	-	37.00	74.71	49.52
	1000	-	20.44	75.45	25.68
Point-RFT	500	5	80.92	82.14	98.52
	500	10	83.36	83.66	99.64
	500	20	84.72	84.96	99.72
	500	50	85.48	85.58	99.88
	500	100	86.24	86.52	99.68

Table 9: Ablation studies of training dataset.

Method	Dataset		ChartQA		
	SFT	RL	Overall	Inner	Format
Base	-	-	79.28	82.17	96.48
Base-SFT	ChartQA	-	3.88	74.62	5.20
Base-RFT	ChartQA	ChartQA	80.92	81.51	99.28
Point-RFT	Point-CoT (w/o ChartQA)	ChartQA	81.04	82.42	98.32
Point-RFT	Point-CoT	ChartQA	86.24	86.52	99.68