PANCAKE: Purpose and Context Activate Knowledge Efficiently

Yi, Seung (School: Asia Pacific International School)
Jung, Chaeyun (School: St. Johnsbury Academy)

This study proposes Purpose And Context Activate Knowledge Efficiently(PANCAKE), a novel structured Chain-of-Thought methodology for enhancing geometric problem-solving in Vision-Language Models (VLMs). Despite strong performance in general visual question answering, existing VLMs struggle to achieve precise alignment between visual perception and logical reasoning. We address these limitations in mathematics, particularly in geometry, which demands precise visual-grounded logical inference and serves as a rigorous testbed for generalizable multimodal reasoning. PANCAKE introduces a geometry-specific, stage-wise reasoning framework consisting of three stages: Purpose (goal identification), Description (visual extraction), and Think (logical reasoning). We also propose a two-stage training pipeline combining Supervised Fine-Tuning and Direct Preference Optimization to instill PANCAKE-structured reasoning into Qwen3-VL. Experiments on Geometry3K demonstrate that PANCAKE achieves a 16.4 percentage point accuracy improvement over the baseline, reaching 70% and outperforming existing models, including Inter-GPS (57.5%). Controlled experiments confirm that PANCAKE gains stem from structured reasoning design, outperforming the token-comparable Long-Think by 9.2 percentage points. Additional evaluation on UniGeo further confirms generalization, with PANCAKE achieving 79.0% accuracy and surpassing GAPS (67.8%) and GOLD (75.2%), confirming that PANCAKE generalizes robustly to geometric reasoning without overfitting to a specific dataset. PANCAKE's novel stage-wise reasoning framework effectively bridges the visual perception and logical reasoning gaps of existing VLMs, achieving substantial accuracy gains and establishing a foundation for generalizable multimodal reasoning.

Awards Won: