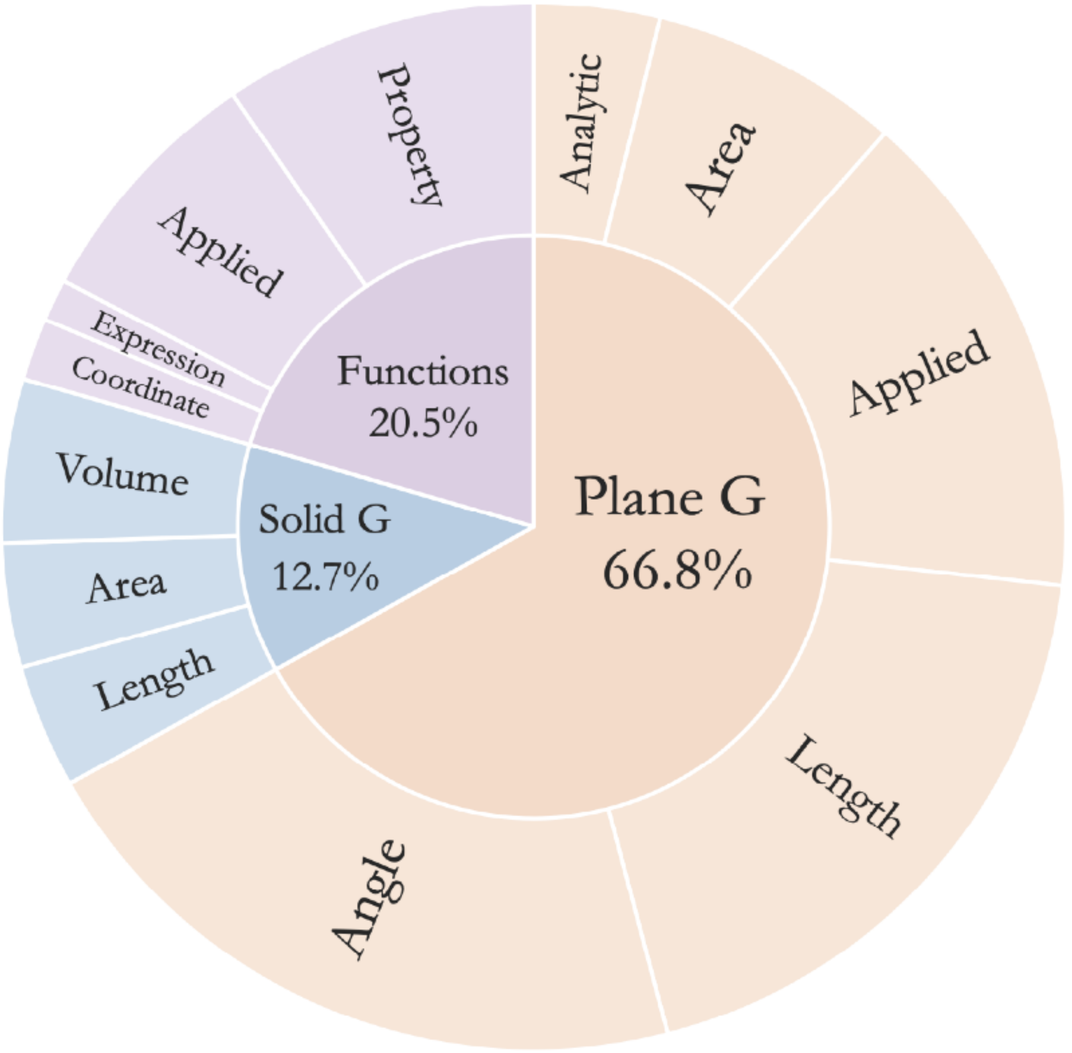
MATHVERSE

Does Your Multi-modal LLM Truly See the Diagrams in Visual Math Problems?

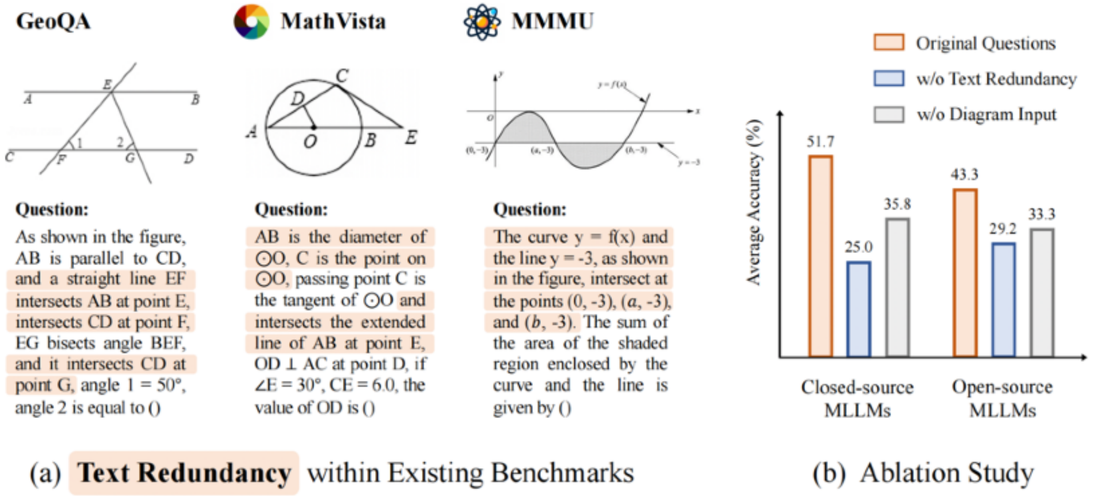
- Visual math benchmark
- 2612 multi-subject math problems with 6 distinct versions -> 15k test samples



MATHVERSE

Background, Motivation

- Do MLLMs truly see the math diagrams in evaluation?
  - Existing math benchmarks include text redundancy.
  - In out ablation study, we find MLLMs' performance without text redundancy is worse than its performance without diagram.



(a) **Text Redundancy** within Existing Benchmarks

(b) Ablation Study