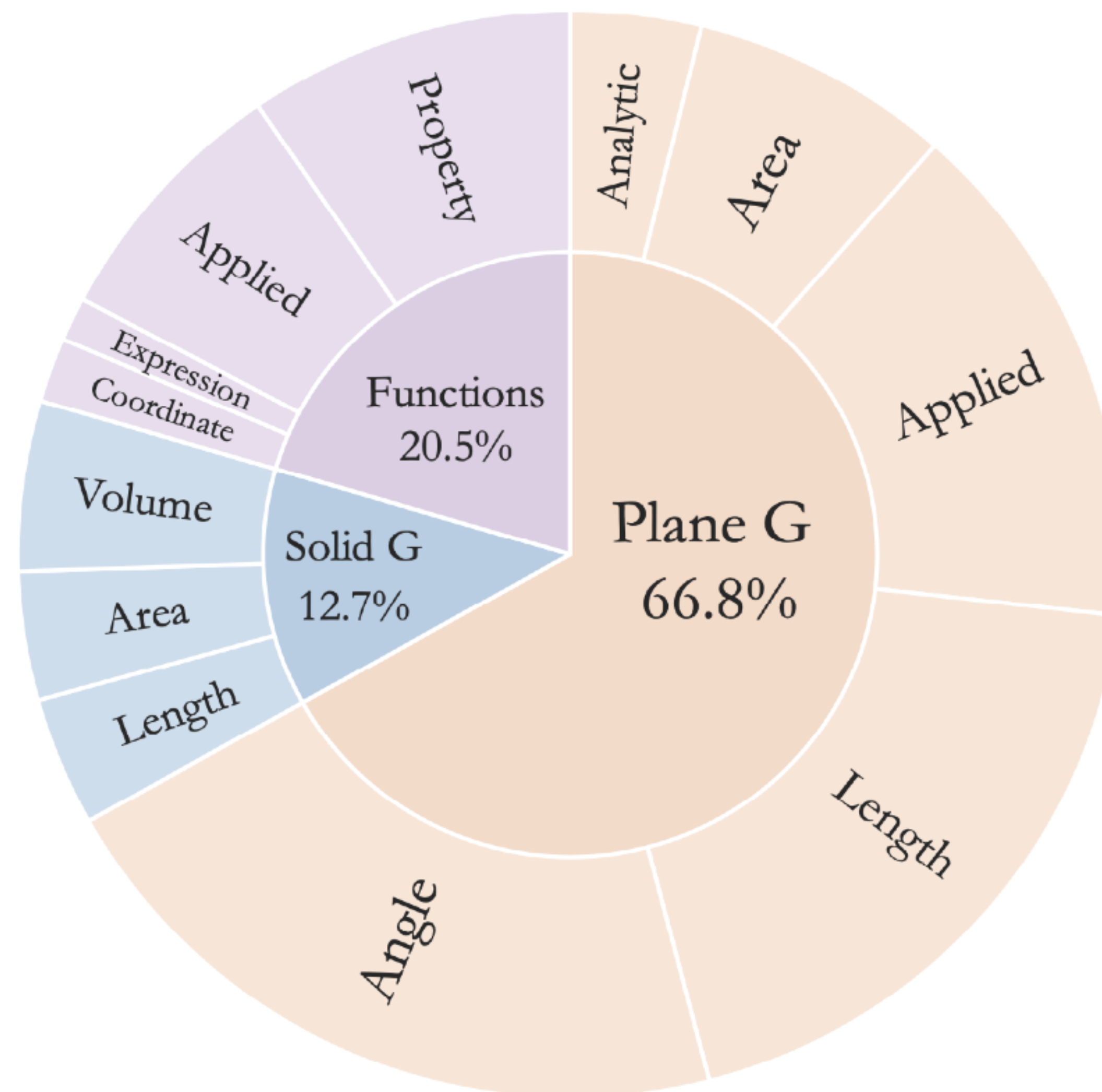


MATHVERSE

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 our ablation study, we find MLLMs' performance without text redundancy is worse than its performance without diagram.

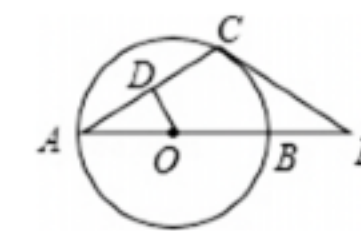
GeoQA



Question:

As shown in the figure, AB is parallel to CD, and a straight line EF intersects AB at point E, intersects CD at point F, EG bisects angle BEF, and it intersects CD at point G, angle 1 = 50° , angle 2 is equal to ()

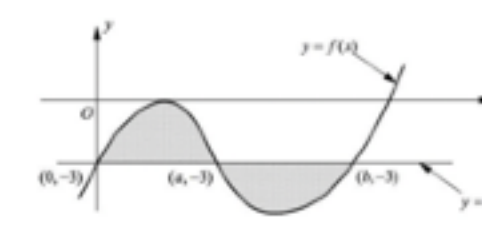
MathVista



Question:

AB is the diameter of $\odot O$, C is the point on $\odot O$, passing point C is the tangent of $\odot O$ and intersects the extended line of AB at point E, $OD \perp AC$ at point D, if $\angle E = 30^\circ$, $CE = 6.0$, the value of OD is ()

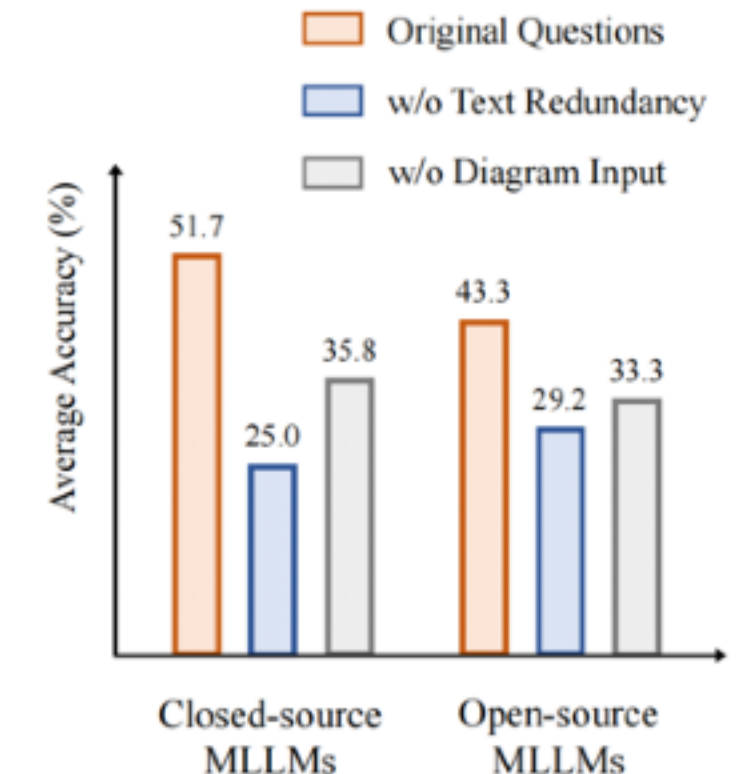
MMMU



Question:

The curve $y = f(x)$ and the line $y = -3$, as shown in the figure, intersect at the points $(0, -3)$, $(a, -3)$, and $(b, -3)$. The sum of the area of the shaded region enclosed by the curve and the line is given by ()

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



(b) Ablation Study