

# Supplementary Material of Integrating Visual Interpretation and Linguistic Reasoning for Geometric Problem Solving

## 0.1. Visualization of More Test Cases

Here, we show more inference results of the basic foundation models in Figure 1 and Figure 2. Both cases show that directly solving the problem can lead to errors in the processing of visual information, even when the images are frustratingly simple. Using multimodal models to interpret images to obtain accurate fine-grained information extraction and then perform linguistic reasoning can improve the accuracy of the answer.

## 0.2. Used Prompt

For the compared methods, the test setting of lonely LVLM is zero-shot prompting, i.e., we directly instruct the LVLM to think and finally answer the question according to the image. The used prompts are shown in Figure 3.

For the test setting of decoupled reasoning, i.e., LVLM visual interpretation combined with LLM reasoning, we prompt LVLM with one in-context example because we found that prompting with an example can demonstrate the image interpretation task to LVLM better and make the output format of LVLM more stable. The prompt we use is shown in Figure 4.

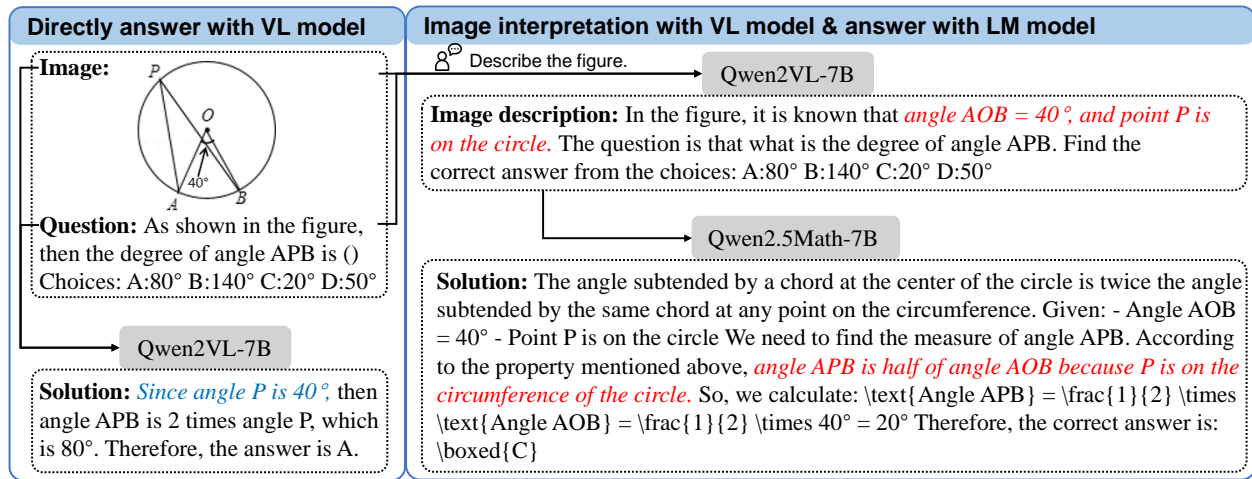


Figure 1. Another case of comparing the solving results of a geometry math problem by directly answering with a VL model or first interpreting the image with a VL model and then answering with an LM model.

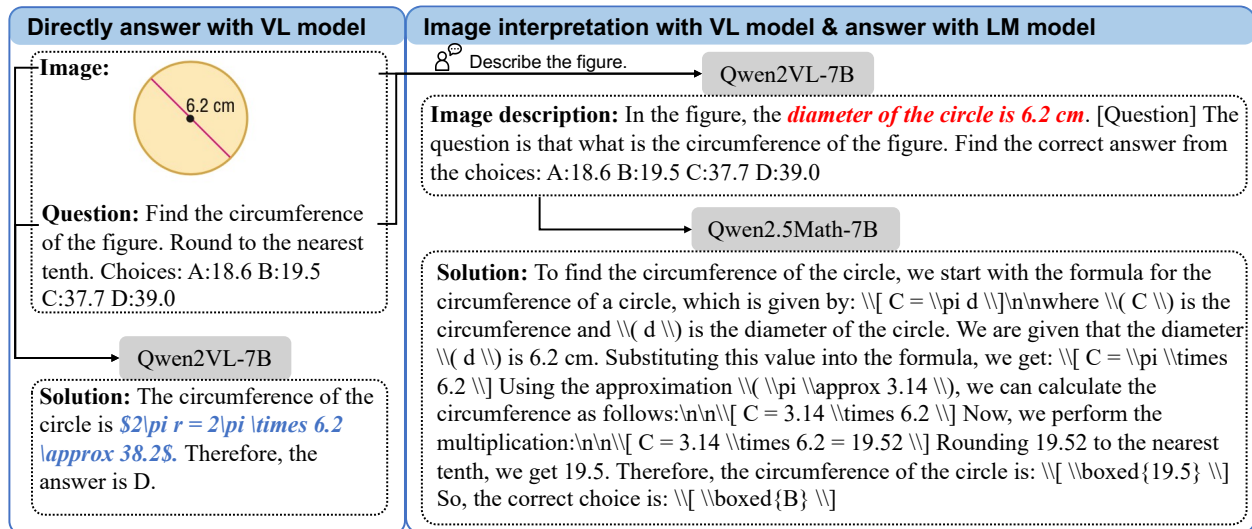
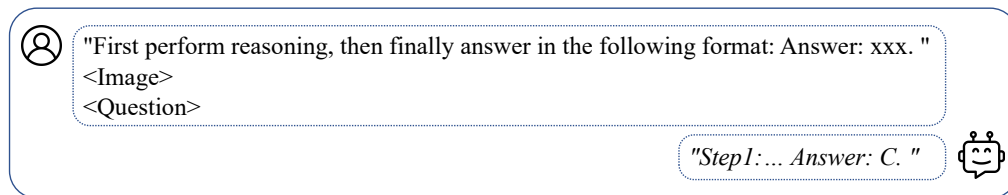


Figure 2. Another case of comparing the solving results of a geometry math problem by directly answering with a VL model or first interpreting the image with a VL model and then answering with an LM model.

Zero-shot instruction used for problem-solving.



Zero-shot prompt used for trained LVLM image interpretation.

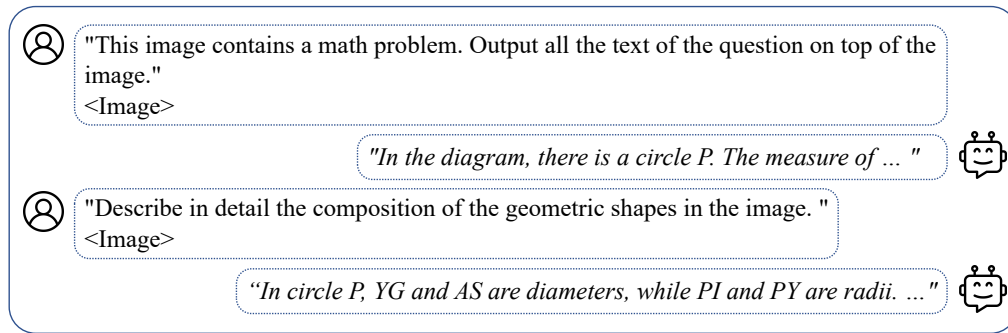
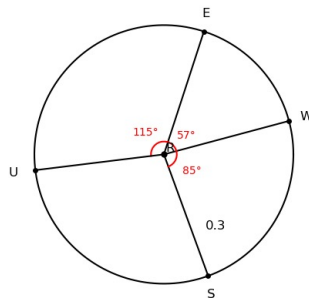


Figure 3. The zero-shot prompt used to test the base model and the trained model.



I will provide you with a geometry math problem with an image and the question. You need to summarize all the known conditions and the question (with options to choose from if available) to be answered based on the provided information and the image. Here is an example:

**Problem image:** In the diagram, there is a circle R with angles  $SRW = 85^\circ$  degrees,  $WRE = 57^\circ$  degrees, and  $ERU = 115^\circ$  degrees. What is the measure of angle  $SRU$ ?



**Question:** What is the measure of angle  $SRU$ ?

**Summarization:**

In circle R, RS, RW, RE, and RU are radii. The radius RS is 0.3 units long. The angles formed between these radii are:  $SRW$  at  $85^\circ$  degrees,  $WRE$  at  $57^\circ$  degrees, and  $ERU$  at  $115^\circ$  degrees.

Below is an image and description of a new geometry problem. Please give the summarized problem information according to the example above.

**Problem image:**

<image>

**Question:**

<question>

Summarization: ...



Figure 4. One-shot prompt used for image interpretation of foundation LVLm.