

# See, Think, Learn: A Self-Taught Multimodal Reasoner

## Supplementary Materials

### A. Visualization

Figure 8 illustrate the test set responses of domain-specific models trained using the proposed STL approach. As shown, the quality of reasoning significantly improves when preceded by high-quality image descriptions. During fine-tuning, the model is explicitly guided to enhance its perceptual understanding first. Subsequently, our structured rationale prompt directs the model to reason based on this refined perception. This sequential guidance, first enhancing perception and then structuring reasoning, leads to notable improvements in both components, ultimately resulting in more accurate answers. Furthermore, Figure 9 illustrates the positive and negative rationales produced by the model trained using the proposed STL framework.

### B. Additional Implementation Details

#### B.1. Structured Rationale Prompt: See-Before-Thinking

Visual question answering requires both understanding what's in the image and reasoning about it. The model needs to correctly interpret visual details and then make higher-level inferences to produce meaningful answers. However, standard chain-of-thought prompting methods, like the commonly used "Let's think step by step", often let the model skip proper visual grounding. In these cases, the model may rely on statistical patterns or shortcuts instead of focusing on the actual image, which can lead to plausible but incorrect answers.

To solve this problem, we introduce a structured rationale prompt that follows a "see before thinking" approach. As shown in Figure 3, our method has three steps: (1) **Caption**, which gives a detailed description of the image based on the question; (2) **Reasoning**, which involves a step-by-step thought process grounded in visual details; and (3) **Conclusion**, which gives the final answer based on the reasoning.

This step-by-step structure encourages the model to look at the image first before starting to reason. By anchoring the reasoning in visual evidence, the model is more likely to give accurate and relevant answers, reducing its tendency to rely on unrelated patterns and improving performance on vision-based tasks.

#### B.2. Training Setup

In the original STaR framework, few-shot prompting is used to reduce noise in the generated bootstrapped rationale set. Likewise, R3V mitigates such noise by building the implementation and evaluation pipeline on a GPT-distilled baseline, where GPT-4o supplies high-quality CoT rationales for a small subset of each dataset. In

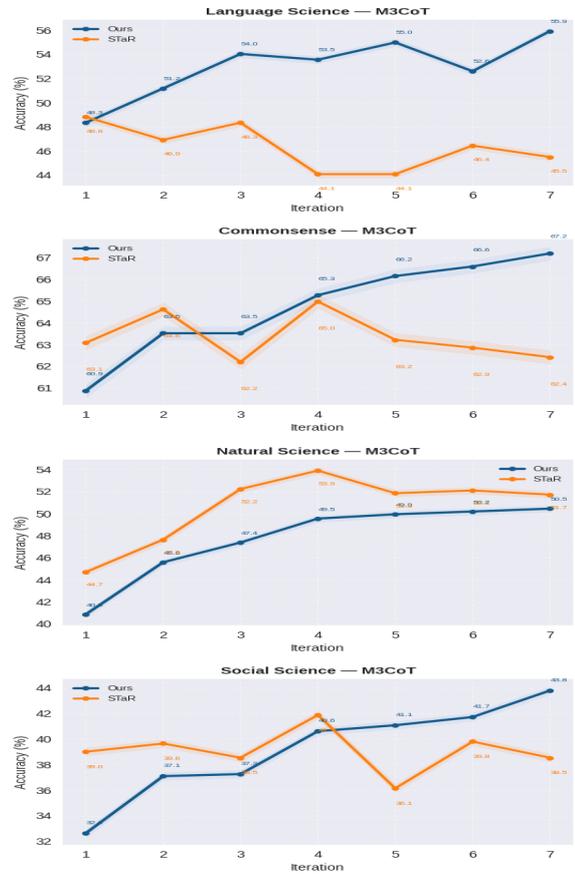


Figure 7. Comparison of test accuracy of STL (Ours) with STaR over 7 iterations.

contrast, we do not use few-shot prompting for STaR in our comparisons, nor do we adopt any GPT-based distillation as in R3V. This choice stems from the limited few-shot prompting capabilities of LVLMs and our core principle that *quality outweighs quantity*: careful selection of high-quality rationales is more effective than relying on expensive GPT-generated warm-up annotations for each domain.

#### B.3. Training Dynamics

Figure 7 compares the trend of test accuracy of our STL framework with STaR over 7 iterations on the four domain splits of M3CoT dataset.

Figure 11 illustrates the Weights & Biases (WandB) tracking data from the final iteration of LoRA fine-tuning applied to LLaVA-v1.5-7B within the proposed STL framework, specifically targeting the language science domain.

### C. Subjective Analysis

Figure 10 illustrates the interface developed to collect these annotations.



helium balloons      bubble gum      rubber band      melted marshmallow

**Question:**  
Which property do these four objects have in common?  
Select the correct answer from the following choices:  
(a) fragile  
(b) stretchy  
(c) sour

**Response:**  
**Caption:** The image shows four objects: a balloon, a bubble gum, a rubber band, and a marshmallow.  
**Reasoning:** The objects in the image have a common property of being stretchy. The balloon is stretchy when it is inflated, the bubble gum is stretchy when it is chewed, the rubber band is stretchy when it is stretched, and the marshmallow is stretchy when it is melted.  
**Conclusion:** (b)

---



(A) Falco tinnunculus      (B) Ardea goliath      (C) Strix uralensis

**Question:**  
Your objective is to recognize the scientific name of the organism based on the provided image.  
Select the correct answer from the following choices:  
(a) Falco tinnunculus  
(b) Ardea goliath  
(c) Strix uralensis

**Response:**  
**Caption:** The image features a bird standing on a grassy area near a body of water. The bird is large and has a long neck.  
**Reasoning:** The bird in the image is a heron, which is a type of wading bird known for its long neck and legs. The scientific name for this bird is Ardea goliath.  
**Conclusion:** (b)

---

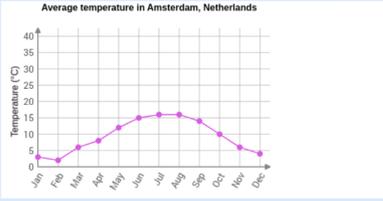
**Choices:**

- (A) Bella was stung by a bee, so her finger is slightly swollen.
- (B) Mr. Estrada's afternoon lecture was about art and culture in Germany before World War II.

**Question:**  
Which is a compound sentence?  
Select the correct answer from the following choices:  
(a) Option A in the image  
(b) Option B in the image  
(c) None of the options are correct.

**Response:**  
**Caption:** The image shows a text with two sentences, one of which is "Bella was stung by a bee, so her finger is slightly swollen." The other sentence is "Mr. Estrella attended a lecture on art and culture in Germany before World War II."  
**Reasoning:** The first sentence is a simple sentence, while the second sentence is a compound sentence. A compound sentence is a sentence that contains two or more independent clauses joined by a conjunction. In this case, the two sentences are independent clauses joined by the conjunction "so."  
**Conclusion:** (a)

---



Month	Temperature (°C)
Jan	2
Feb	1
Mar	4
Apr	7
May	11
Jun	15
Jul	16
Aug	15
Sep	11
Oct	7
Nov	4
Dec	2

**Question:**  
Which month has the lowest average temperature in this area?  
Select the correct answer from the following choices:  
(a) February  
(b) November  
(c) December

**Response:**  
**Caption:** The image shows a graph of average temperature in Amsterdam, Netherlands. The graph shows that the lowest average temperature occurs in February.  
**Reasoning:** The graph shows that the lowest average temperature occurs in February, which is the correct answer.  
**Conclusion:** (a)

Figure 8. **Response for Commonsense, Natural Science, Language Science, Social Science Domains. (top to bottom).** The response of our STL framework is a detailed caption and reasoning along with the answer.



Question:

Where is the woman in the picture located?

Select the correct answer from the following choices:

- (a) Her private garden
- (b) Public park
- (c) Library
- (d) Virgin forest

**Positive Rationale:**

Caption: A woman is sitting on a bench in a park, wearing sunglasses and a white shirt. She appears to be enjoying her time outdoors.

Reasoning: The woman is sitting on a bench, which is a common outdoor seating option in parks. She is also wearing sunglasses, which suggests that it is a sunny day. Additionally, she is wearing a white shirt, which is a casual and comfortable clothing option.

Conclusion: (b)

**Negative Rationale: Explain why this answer is wrong: (a).**

Caption: A woman is sitting on a bench in a park, wearing sunglasses and a white shirt.

Explanation: The answer (a) is incorrect because the woman is not in her private garden. The image shows her sitting on a bench in a park, which is a public outdoor space.



Question:

From the image provided, deduce the scientific name of the organism being shown.

Select the correct answer from the following choices:

- (a) *Camellia sasanqua*
- (b) *Trametes versicolor*
- (c) *Hyacinthus orientalis*

**Positive Rationale:**

Caption: The image shows a pink flower with green leaves, which is a *Camellia sasanqua*.

Reasoning: The flower in the image is a *Camellia sasanqua*, which is a type of flower known for its pink petals and green leaves. The scientific name for this flower is (a) *Camellia sasanqua*.

Conclusion: (a)

**Negative Rationale: Explain why this answer is wrong: (b).**

Caption: The image shows a pink flower with green leaves, which is a *Camellia sasanqua*.

Explanation: The answer (b) is incorrect because *Trametes versicolor* is a type of mushroom, not a flower.

Figure 9. Visualization of Positive and Negative Rationales from the proposed STL. The examples illustrate correct identification and reasoning for the chosen answer and rejection of an incorrect alternative.

### Question physics-5



Which property do these three objects have in common?  
Select the correct answer from the following choices.  
(a) bouncy  
(b) hard  
(c) soft

Rationale Position 1:

The objects in the image have a common property of being soft. The blue shorts are made of soft fabric, the red jelly is a soft, gel-like substance, and the purple apron is made of soft fabric as well. Therefore, the correct answer is (c).

Rank: 1

Rationale Position 2:

The three objects in the image are a soccer short, a jelly, and an apron. They all have a common property of being blue. Therefore, the correct answer is (c).

Rank: 2

Anonymous rationales for fair annotations

A drop-down for giving rank (out of two) to each rationale.

Figure 10. Screenshot of our subjective annotation GUI. Please note that the name of the rationale generator model has been kept anonymous to ensure a fair comparison.

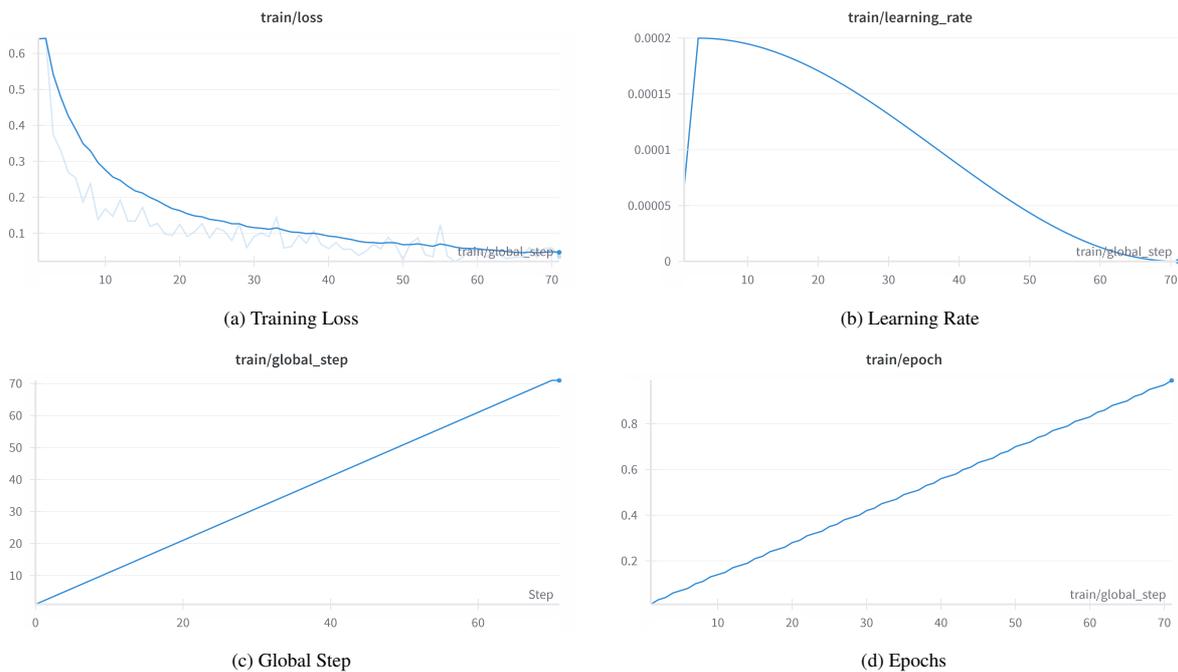


Figure 11. Tracking Data from WandB (Training). (a) Depicts the decay of the training loss over steps, indicating effective learning and convergence of the model. (b) Illustrates a learning rate schedule where the rate initially increases slightly before gradually decreasing, following a cosine strategy. (c) Shows a linear increase in the global training steps. (d) Displays the linear progression of training epochs with respect to global steps.