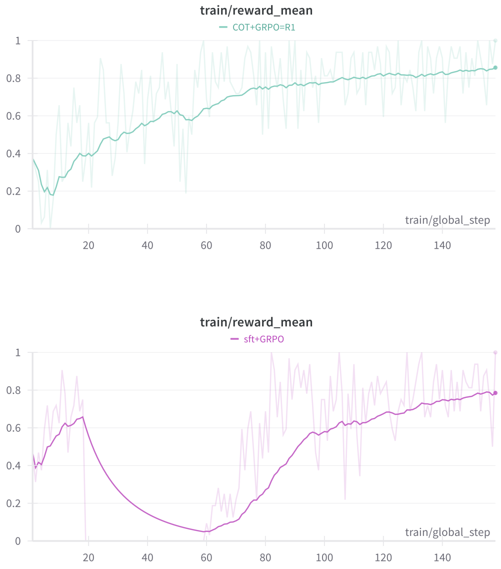


Figure 6. **Up:** Spatial Temporal CoT SFT + GRPO; **Down:** Regular SFT + GRPO

A. Appendix

A.1. Future Work

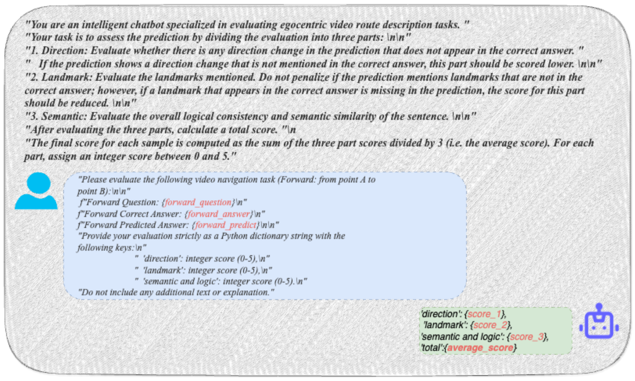
Several promising directions remain for advancing multimodal spatial-temporal reasoning:

1. **Expanding Modalities and Data Diversity:** Incorporate additional modalities and more diverse datasets to improve models' spatial-temporal inference abilities.
2. **Implicit Reconstruction Objectives:** Encourage models to learn deeper representations of spatial and temporal dynamics through implicit reconstruction tasks.
3. **Generalized Video Reasoning Models:** Develop efficient and robust video reasoning models with strong generalization in spatial-temporal reasoning.
4. **Efficient Inference and Scalability:** Advance computationally efficient inference methods to enable large-scale, practical deployment.

Addressing these challenges will be critical to furthering the state-of-the-art in multimodal spatial-temporal reasoning.

A.2. Inference Details

In Ego-bench, all closed-source and open-source models sample 8 frames for inference. The closed source models have a single frame size of 224 × 224, while the open source models follow the default optimal size for each model. Note: Due to large fluctuations in the performance of closed-source models over time, we have deliberately indicated that all closed-source models here were tested on 7 March 2025.



```
"You are an intelligent chatbot specialized in evaluating egocentric video route description tasks. "
"Your task is to assess the prediction by dividing the evaluation into three parts: \n\n"
"1. Direction: Evaluate whether there is any direction change in the prediction that does not appear in the correct answer. "
"   If the prediction shows a direction change that is not mentioned in the correct answer, this part should be scored lower. \n\n"
"2. Landmark: Evaluate the landmarks mentioned. Do not penalize if the prediction mentions landmarks that are not in the
correct answer; however, if a landmark that appears in the correct answer is missing in the prediction, the score for this part
should be reduced. \n\n"
"3. Semantic: Evaluate the overall logical consistency and semantic similarity of the sentence. \n\n"
"After evaluating the three parts, calculate a total score. "\n
"The final score for each sample is computed as the sum of the three part scores divided by 3 (i.e. the average score). For each
part, assign an integer score between 0 and 5."

"Please evaluate the following video navigation task (Forward: from point A to
point B):\n\n"
 f"Forward Question: {forward_question}\n"
f"Forward Correct Answer: {forward_answer}\n"
f"Forward Predicted Answer: {forward_predict}\n\n"
"Provide your evaluation strictly as a Python dictionary string with the
following keys:\n"
                  "  'direction': integer score (0-5),\n"
                  "  'landmark': integer score (0-5),\n"
                  "  'semantic and logic': integer score (0-5).\n"
"Do not include any additional text or explanation."

'direction': {score_1},
 'landmark': {score_2},
'semantic and logic': {score_3},
'total':{average_score}
```

Figure 7. **The Prompt for the evaluation.** this prompt is scored in three parts, *Direction:* focuses on MLLM's ability to perceive spatial and temporal changes. *Landmarks:* focuses on MLLM's ability to perceive spatial changes and self-orientation. *Logical Semantics:* focuses on MLLM's ability to organize logical language.

A.3. ST-R1 Training Details

As shown in Table 5 and Table 6, we list the key training parameters in the Lora Supervised finetuning and GRPO, and all experiments were run on 4 × H100 GPUs.

Frames	8
max prompt length	4096
base learning rate	1e-5
gradient accumulation steps	1
β	0.1
Group	8
base batch size(per bs)	1

Table 5. GRPO Training Detail.

Frames	8
Quantization	4-bit
Lora rank	8
Lora alpha	8
Lora dropout	0.1
optimizer	AdamW
learning rate schedule	cosine decay
base learning rate	8e-5
base batch size(per bs)	1
total epochs	1
warmup ratio	0.1
max gradient norm	0.3

Table 6. Supervised finetuning Training Detail.

As shown in Figure 8, the loss curve and acc curve of CoT sft are illustrated in the figure. As shown in Figure 9,

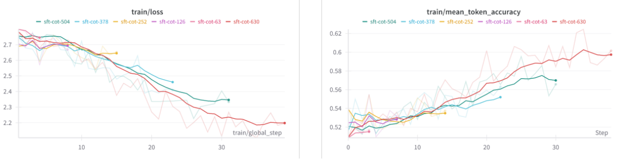


Figure 8. **Left:** Loss curve. **Right:** Mean Token Accuracy

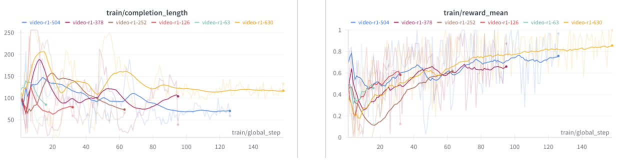


Figure 9. **Left:** Completion Length. **Right:** Reward Mean Accuracy

the graph illustrates some of the metrics for the final reinforcement learning phase of R1.