

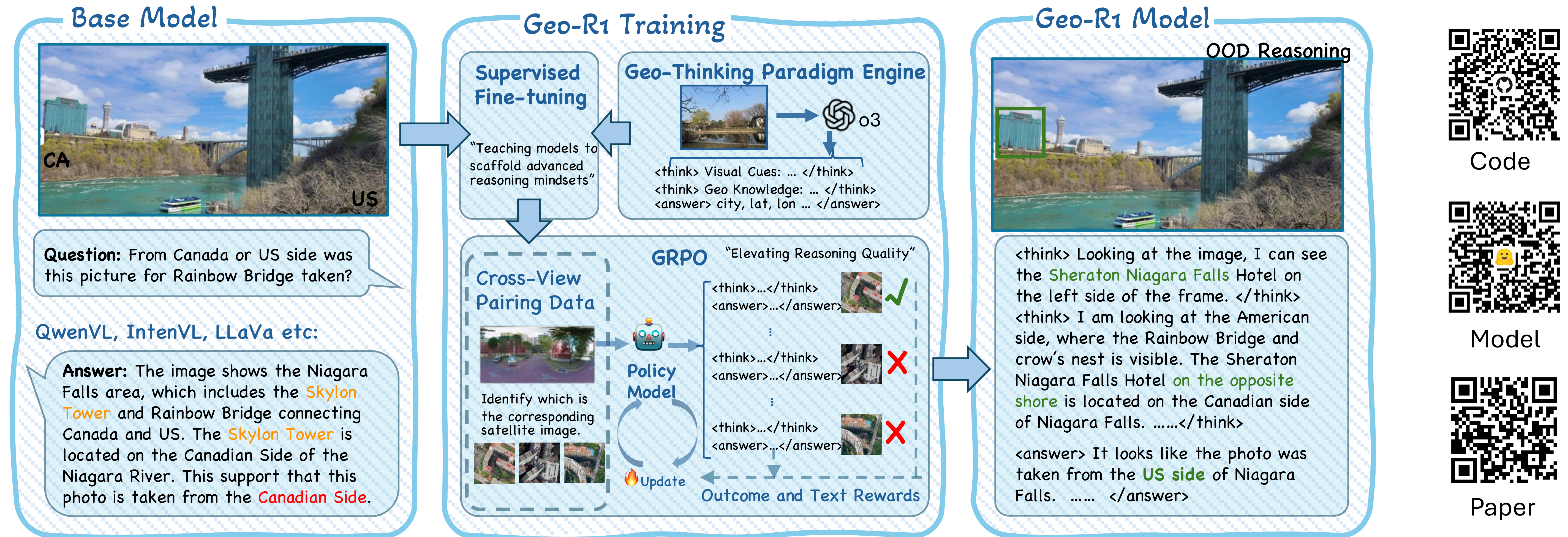
TL;DR: Our Geo-R1 demonstrates that indirect metadata rewards can unlock zero-shot geospatial reasoning in VLMs without task-specific supervision.

Background and Motivation

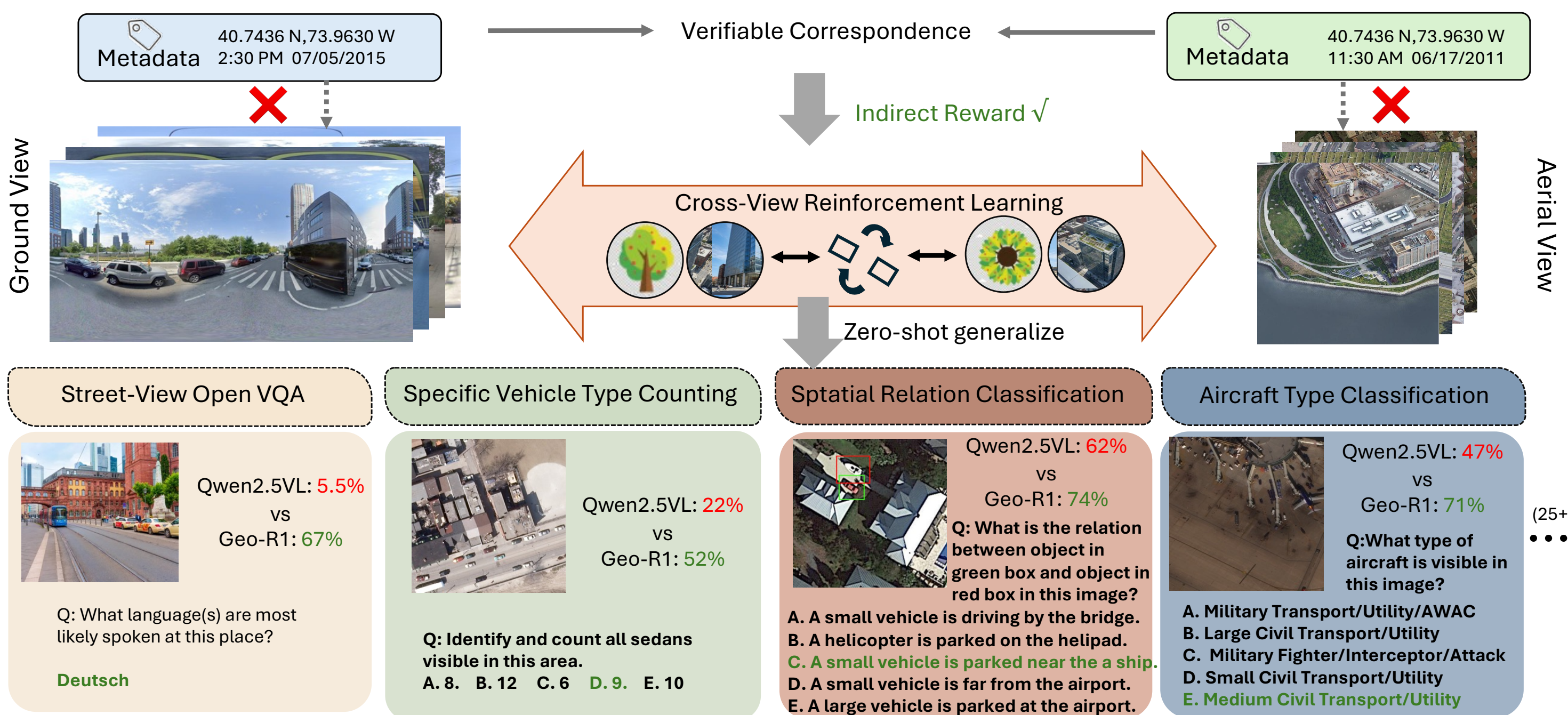
- Geospatial VLMs suffer from scarcity of direct annotated training data.
- Raw imagery is abundant, but high-quality reasoning labels are hard to access.
- Metadata such as GPS coordinates and timestamps are scalable but can not directly supervise the training of autoregressive VLMs.

Key Question: How can we train reasoning geospatial vision language models using massive amounts of unlabeled ground and aerial imagery?

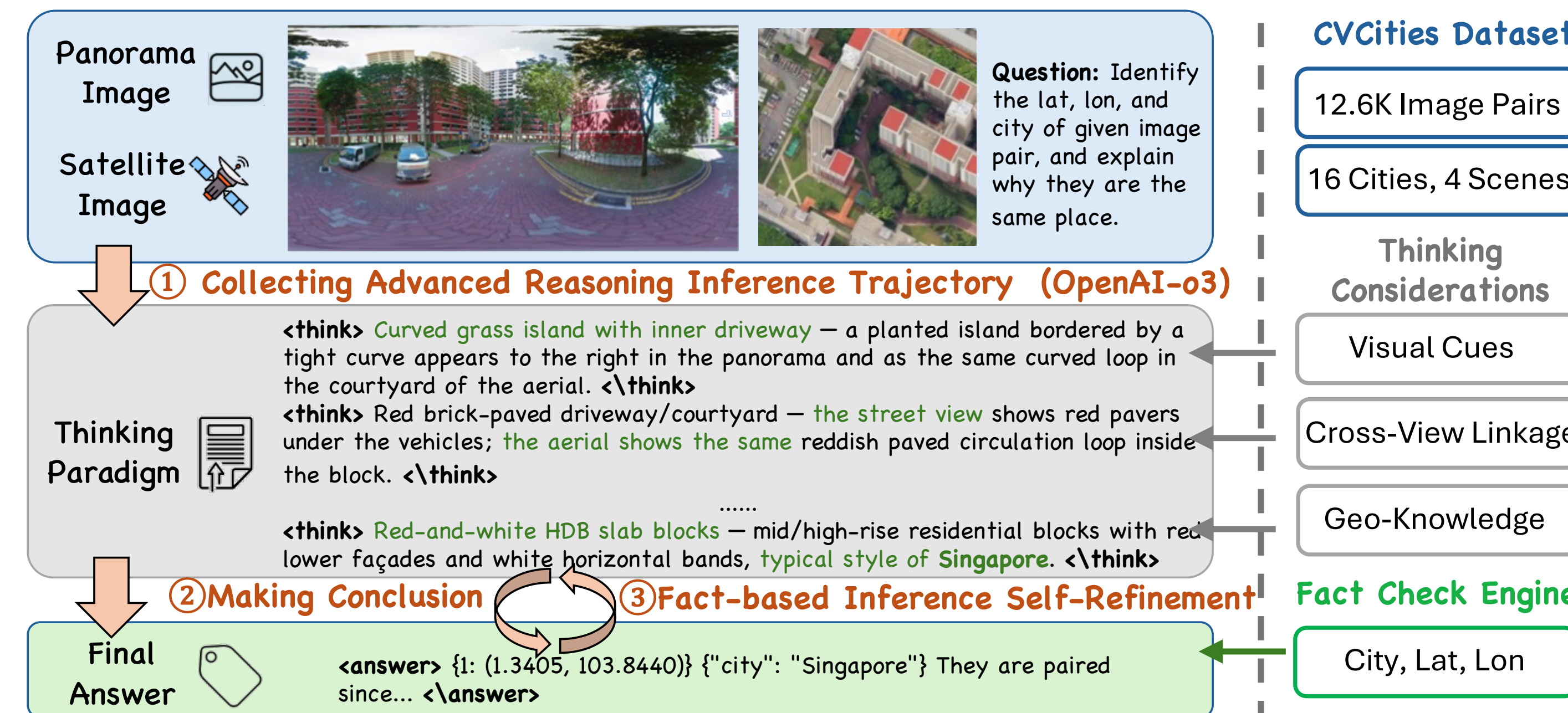
Geo-R1 Framework



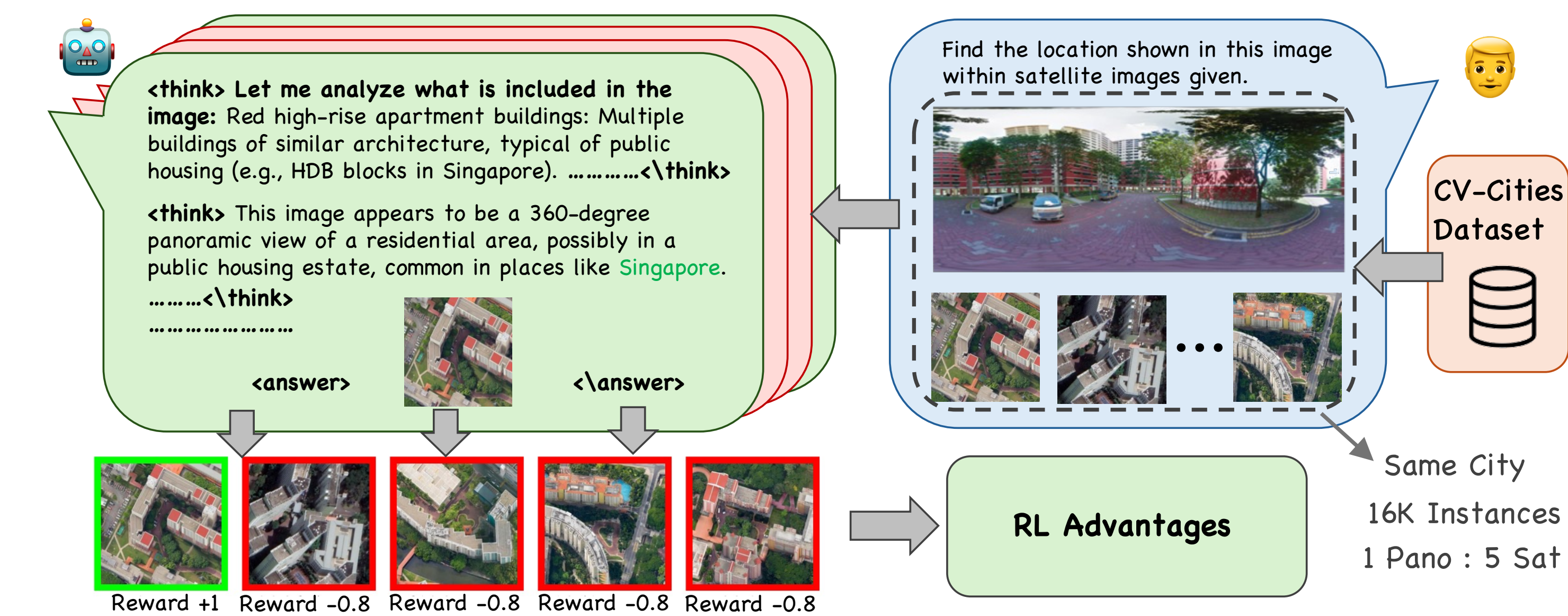
Core Idea



Geospatial CoT Data Engine



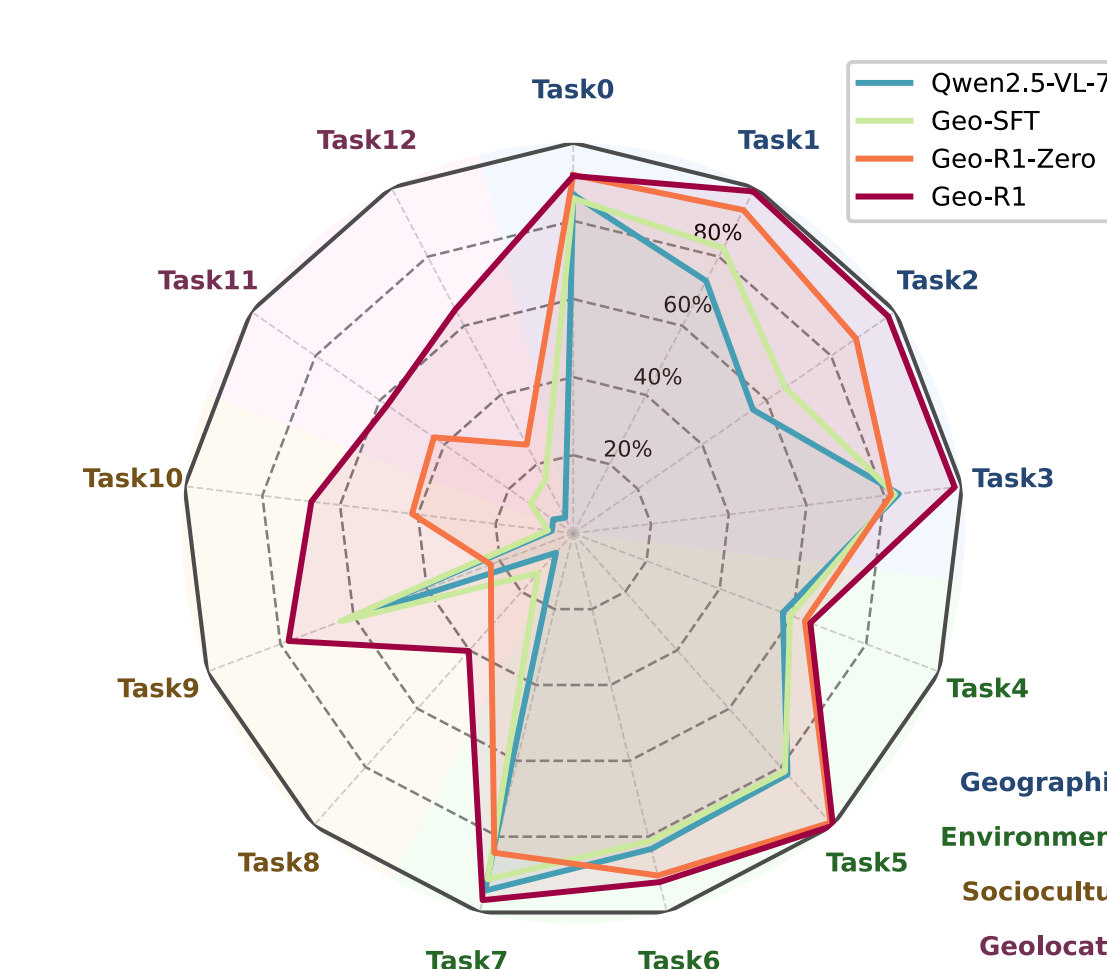
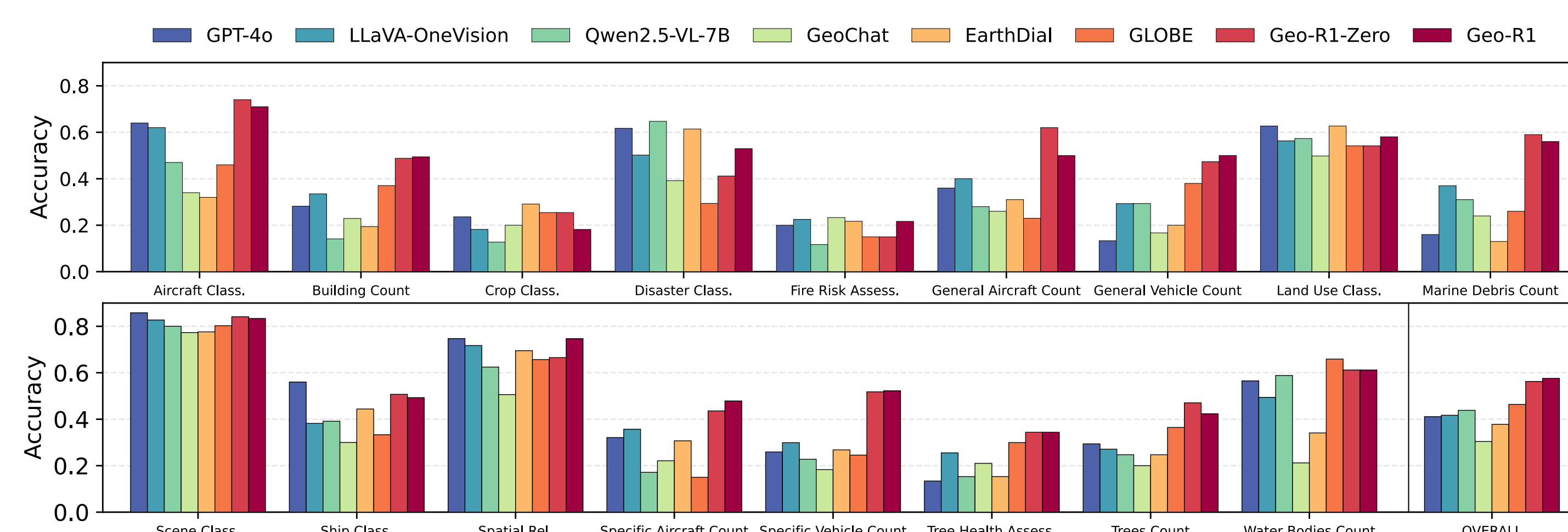
Cross-view Pairing via GRPO



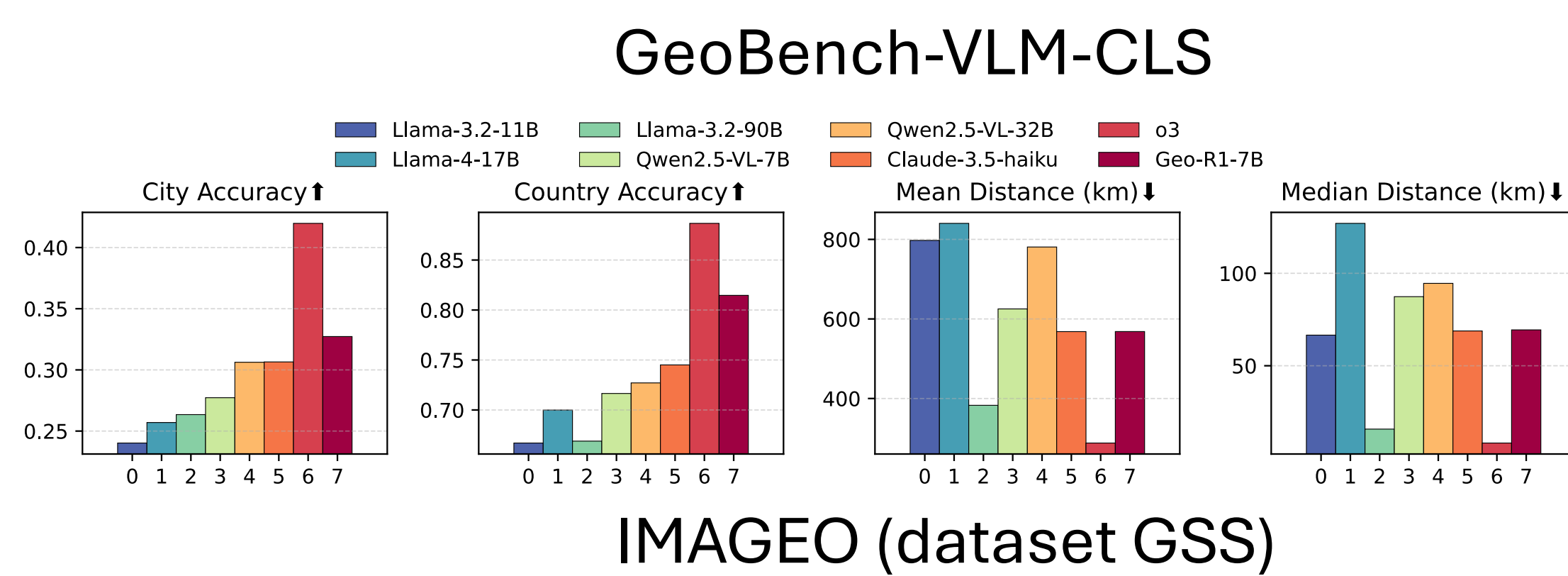
Cross-view Alignment between ground-view and aerial-view images provides scalable indirect rewards for the reinforcement learning that try to reason out this correspondence. This RL process enhances the model's basic geospatial element understanding, and as a result, can generalize to multiple geospatial tasks.

- Use metadata correspondence as a verifiable reward.
- Train the model to bridge ground-to-air evidence
- Achieve strong out-of-task generalization

Experimental Results



Method	Street 1km	City 25km	Region 200km	Country 750km	Continent 2500km
ISNs (Muller-Budack et al.)	26.24	47.38	55.88	68.48	80.92
GeoCLIP (Vivanco Pedra et al.)	29.28	52.52	66.85	84.07	93.33
Hybrid (Astruc et al.)	0.97	16.53	28.72	50.31	71.47
RFM-YFCC (Dufour et al.)	11.72	46.64	60.46	77.97	91.96
Qwen2.5-VL-7B (Bai et al.)	15.42	52.72	62.86	75.11	83.47
InternVL3-8B (Zhu et al.)	12.01	44.17	55.66	75.36	86.98
Task-specific reasoning supervision					
GeoReasoner-7B (Li et al.)	10.06	40.44	50.91	68.01	79.68
GLOBE-7B (Li et al.)	17.99	62.85	73.83	86.68	92.52
Geo-R1 (Ours)	17.98	61.02	73.91	86.78	93.56



Model	Accuracy	Completion Length
Qwen2.5-VL-7B	19.0%	204.6
Geo-SFT	23.1%	1127.6
Geo-R1-Zero	78.1%	587.4
Geo-R1	82.4%	378.8

