

Lang2LTL-2: Grounding Spatiotemporal Navigation Commands Using Large Language and Vision-Language Models



Jason Xinyu Liu (xinyu_liu@brown.edu), Ankit Shah, George Konidaris, Stefanie Tellex, and David Paulius

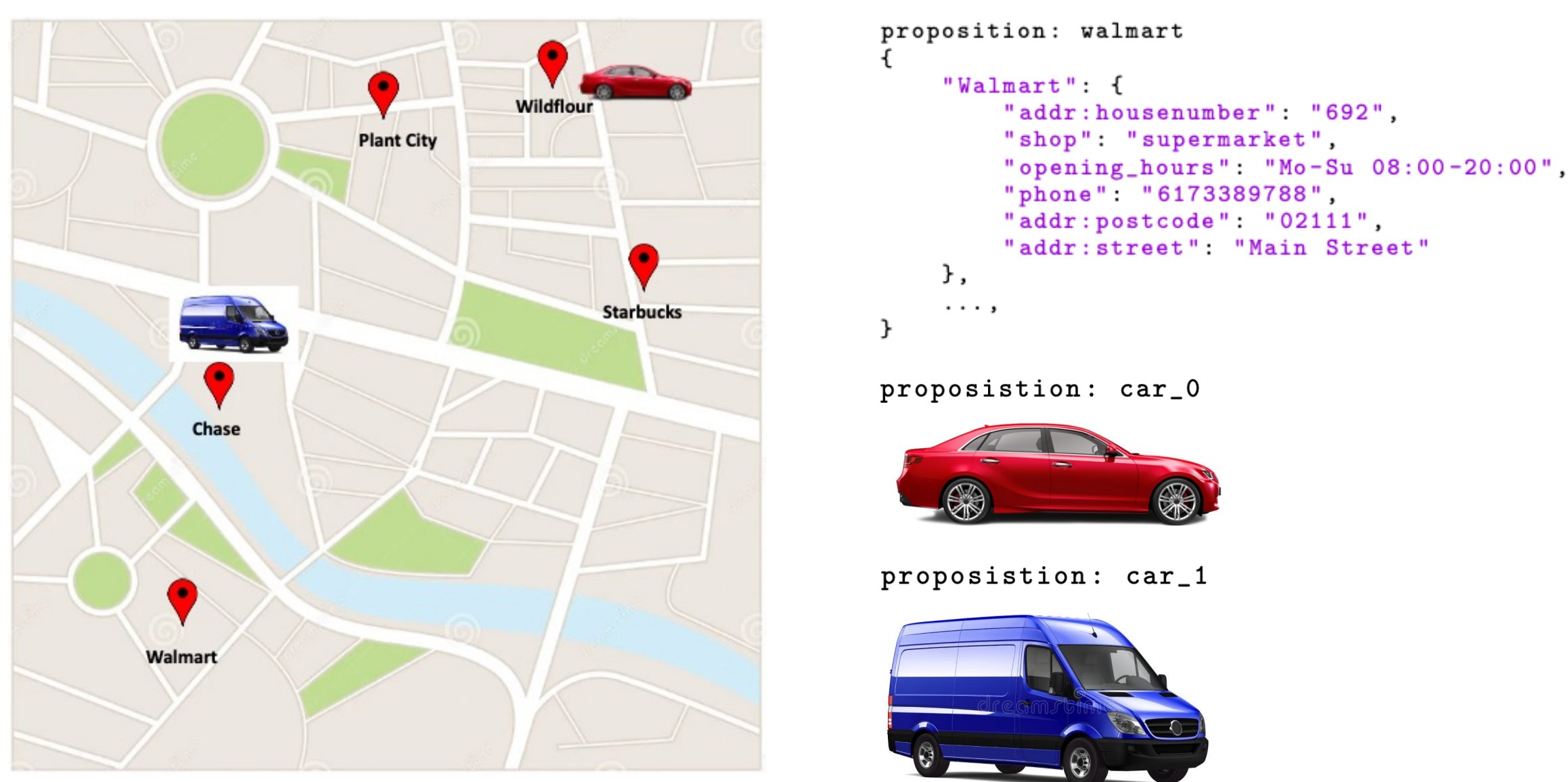
Problem Definition

- Input**
- spatiotemporal navigation command
 - semantic database: {proposition: (semantic info, labeling function)}
- Output**
- LTL formula whose propositions are grounded to real-world landmarks

Main Contributions

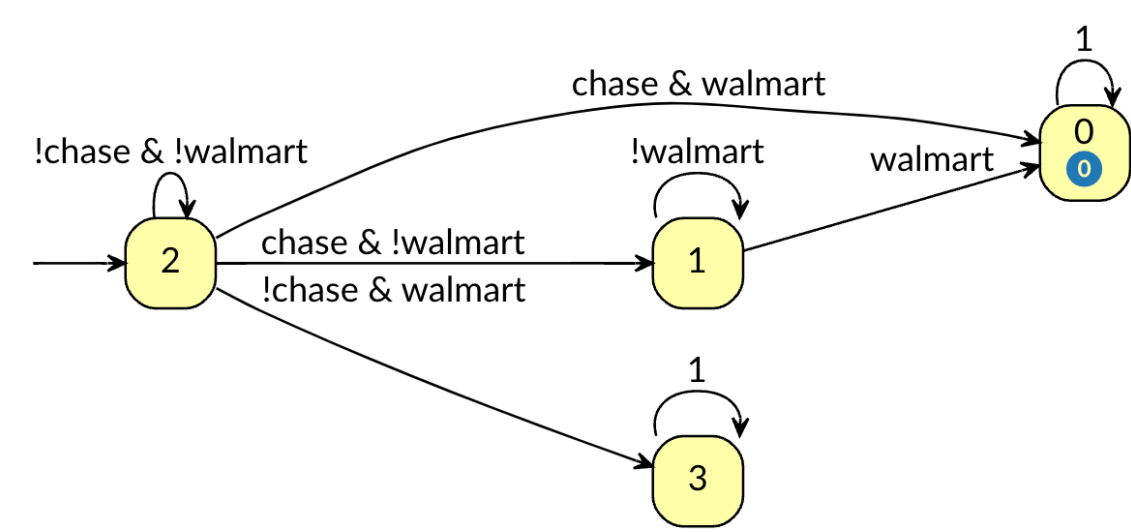
- A modular system Lang2LTL-2 that grounds spatiotemporal navigation commands in novel environments without training on language data
- Use multimodal semantic database
- Abort execution when detect infeasible task specification

Multimodal Semantic Database



Linear Temporal Logic (LTL)

- LTL can represent non-Markovian tasks
 - LTL = propositional logic + temporal operators
 - Syntax
- $$\varphi := \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \mathbf{X}\varphi \mid \mathbf{F}\varphi \mid \mathbf{G}\varphi \mid \varphi_1 \mathbf{U}\varphi_2$$
- Labeling function maps MDP states to propositions
 - Translate LTL to Büchi automaton to track task progress

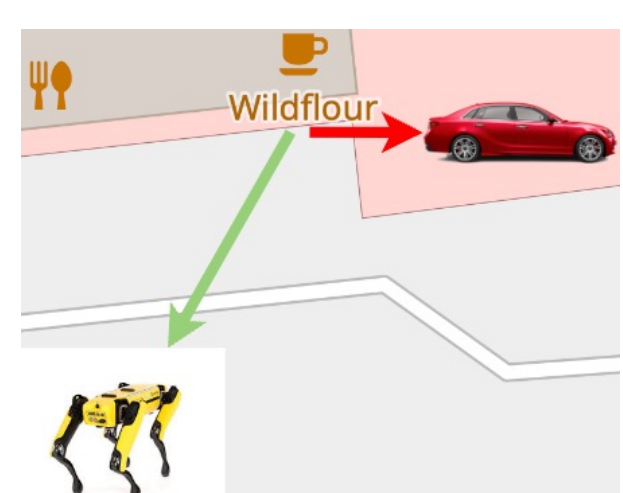


Example Temporal Patterns

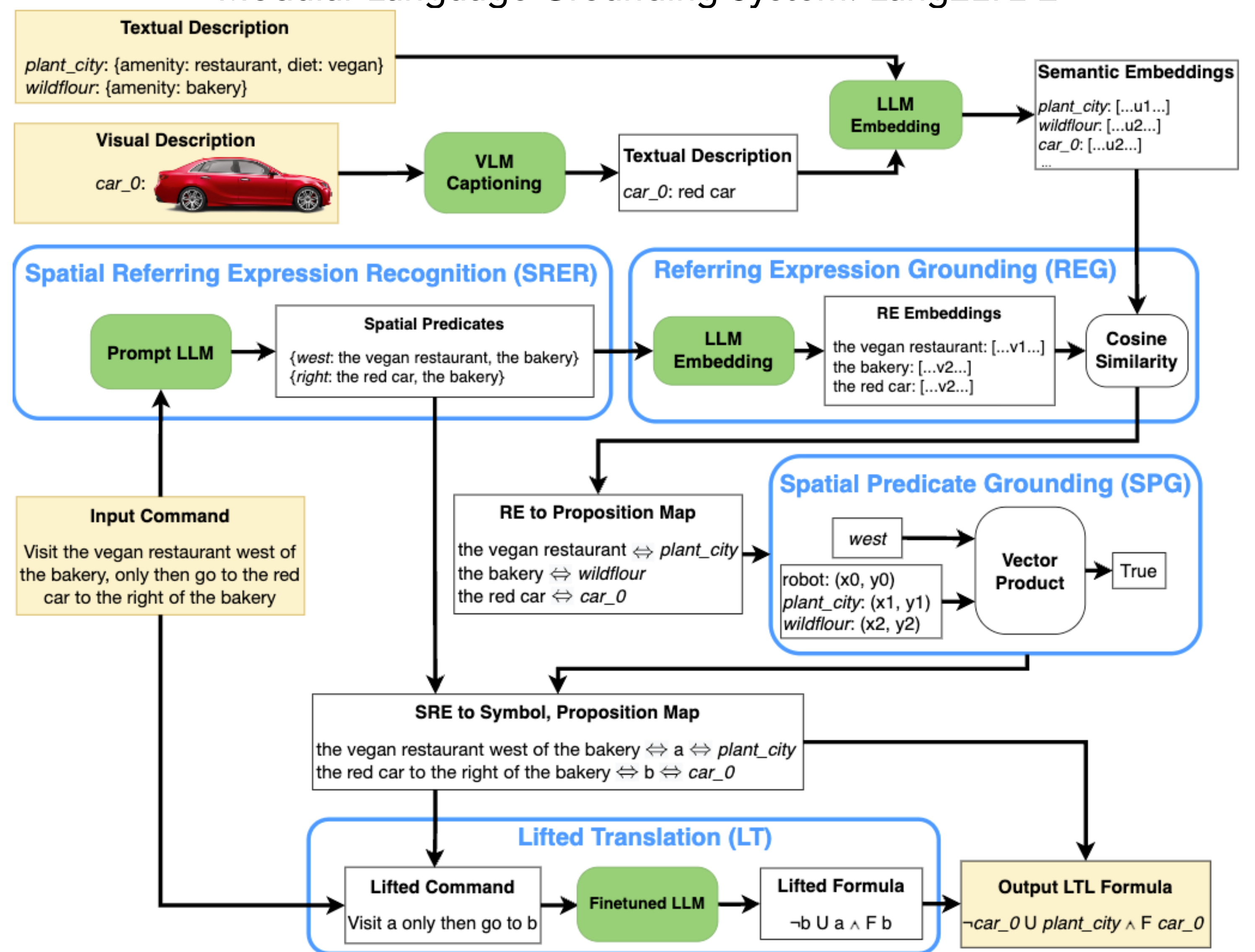
Type	Command	LTL
Visit	go to a and b in any order	$Fa \wedge Fb$
Sequenced Visit	move to a then b	$F(a \wedge Fb)$
Ordered Visit	visit b but only after a	$\neg b \mathbf{U} a \wedge Fb$
Patrolling	keep visiting a and b	$\mathbf{G}Fa \wedge \mathbf{G}Fb$
Global Avoidance	never visit a or b	$\mathbf{G} \neg a \wedge \mathbf{G} \neg b$
Lower Restricted Avoidance	visit a at least twice	$F(a \wedge (a \mathbf{U} (\neg a \wedge (\neg a \mathbf{U} Fa))))$

Spatial Relations

left, right, in front of, opposite to, behind, near, next to, adjacent to, close to, by, between, north of, south of, east of, west of, northeast of, northwest of, southeast of, southwest of



Modular Language Grounding System: Lang2LTL-2

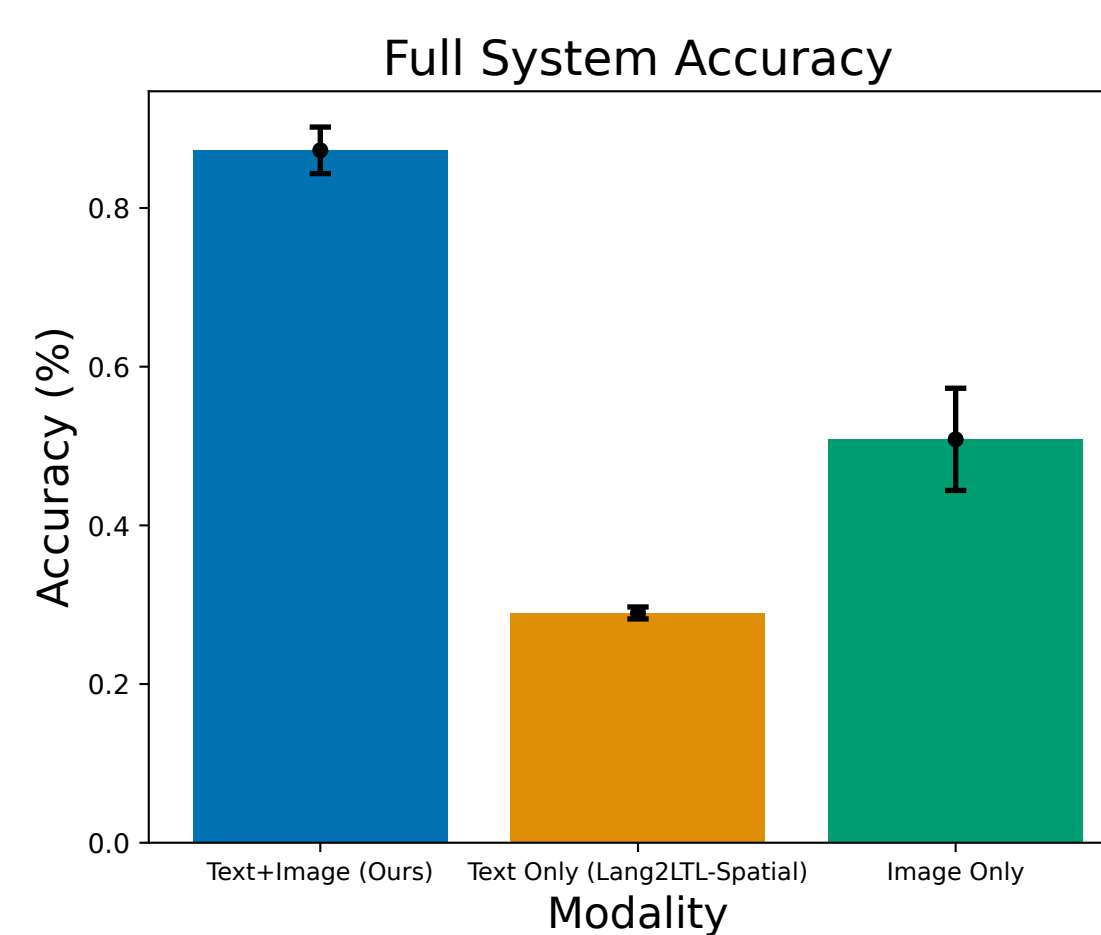


Modular Evaluation

Module	Accuracy				
	City 1 (9 landmarks)	City 2 (34 landmarks)	City 3 (44 landmarks)	City 4 (175 landmarks)	Average
SRER	99.45 ± 0.12%	99.43 ± 0.26%	99.56 ± 0.63%	99.39 ± 0.21%	99.46 ± 0.34%
REG	Top-1	99.68 ± 0.72%	97.98 ± 1.07%	88.74 ± 2.14%	91.19 ± 8.84%
	Top-5	100.00 ± 0.00%	100.00 ± 0.00%	99.56 ± 0.24%	99.15 ± 0.34%
	Top-10	100.00 ± 0.00%	100.00 ± 0.00%	99.70 ± 0.17%	99.98 ± 0.05%
SPG	100.00 ± 0.00%	100.00 ± 0.00%	99.53 ± 0.33%	99.35 ± 1.46%	99.72 ± 0.75%
LT	Finetuned T5-base	99.45 ± 0.00%	99.45 ± 0.00%	99.45 ± 0.00%	99.45 ± 0.00%
	RAG-10	69.33 ± 0.25%	70.34 ± 0.13%	69.65 ± 0.58%	70.39 ± 0.84%
	RAG-50	83.79 ± 0.06%	83.93 ± 0.12%	83.75 ± 0.52%	83.93 ± 0.65%
	RAG-100	88.20 ± 0.58%	88.25 ± 1.04%	87.79 ± 0.39%	87.70 ± 0.13%

Full System

- Generalizes to 4 novel cities
- Accuracy: 93.53% ± 4.33%



<https://spatiotemporal-ground.github.io/>

Robot Demonstration

Spatiotemporal Navigation Command	Grounding Result
navigate to the kitchen counter between the blue couch and the refrigerator	Success
walk to the chair in front of the bookshelf but only after the kitchen counter	Success
visit the couch left of the counter, in addition never go to the TV in front of the couch	Success
visit the stairs between the apartment and the silver car at most three times	Success
visit the silver car on the left side of the apartment exactly twice	Success
go to doorway exactly two times, in addition always avoid the table	Success
go to the blue couch next to the kitchen counter but never go to the blue couch next to the kitchen counter	Abort Correctly (contradiction)
find the dumpster near the white car, in addition avoid visiting the apartment and the red brick wall	Abort Correctly (environment)
visit counter at least six times	Abort Correctly (incorrect grounding)