GOMAA-Geo: Goal Modality Agnostic Active Geo-localization (NeurIPS 2024)

Anindya Sarkar[1], Srikumar Sastry[1], Aleksis Pirinen[2, 3], Chongjie Zhang[1], Nathan Jacobs[1], Yevgeniy Vorobeychik[1]
[1]Department of Computer Science and Engineering, Washington University in St. Louis [2]RISE Research Institutes of Sweden [3]Climate AI Nordics

Code & models: *github.com/mvrl/GOMAA-Geo*

climateainordics.com

Increasing use of UAVs in environmental applications

Land use monitoring | Infrastructure inspection | Emission monitoring
Search-and-rescue ← our motivation, but the approach is general

Towards automating search-and-rescue

Intensifying climate change →increased frequency and severity of natural disasters

Thus also an increased need for search-and-rescue missions

Practical constraints (e.g. flight time, battery, expert pilots)
→ big potential for an automated approach

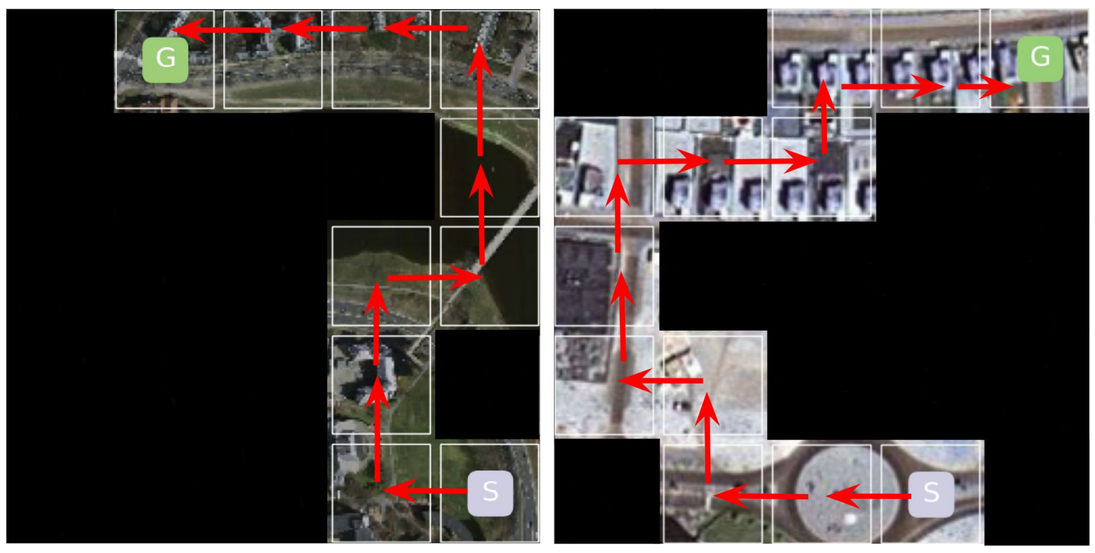
Proposed proxy task: *Active geo-Localization (AGL)*

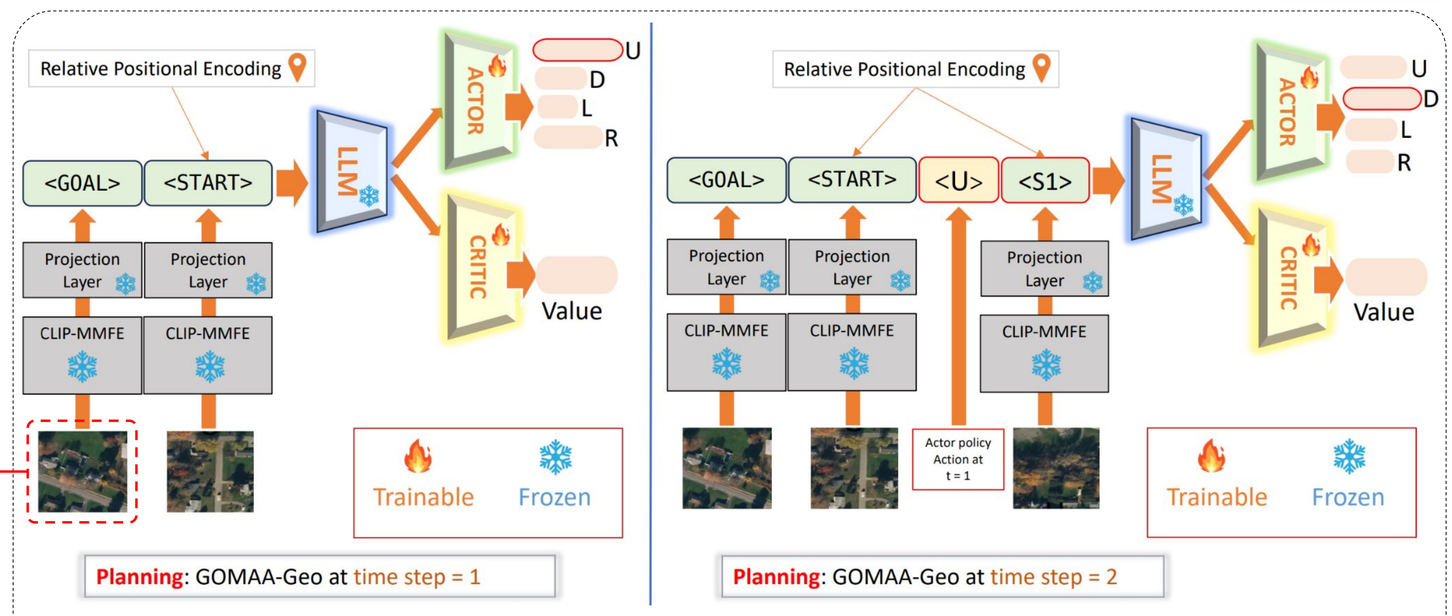
Enables **controllable** and **reproducible** experimentation

Goal locations specified using **visual / textual cues only** — no GNSS information assumed to be available (e.g. because disturbed)

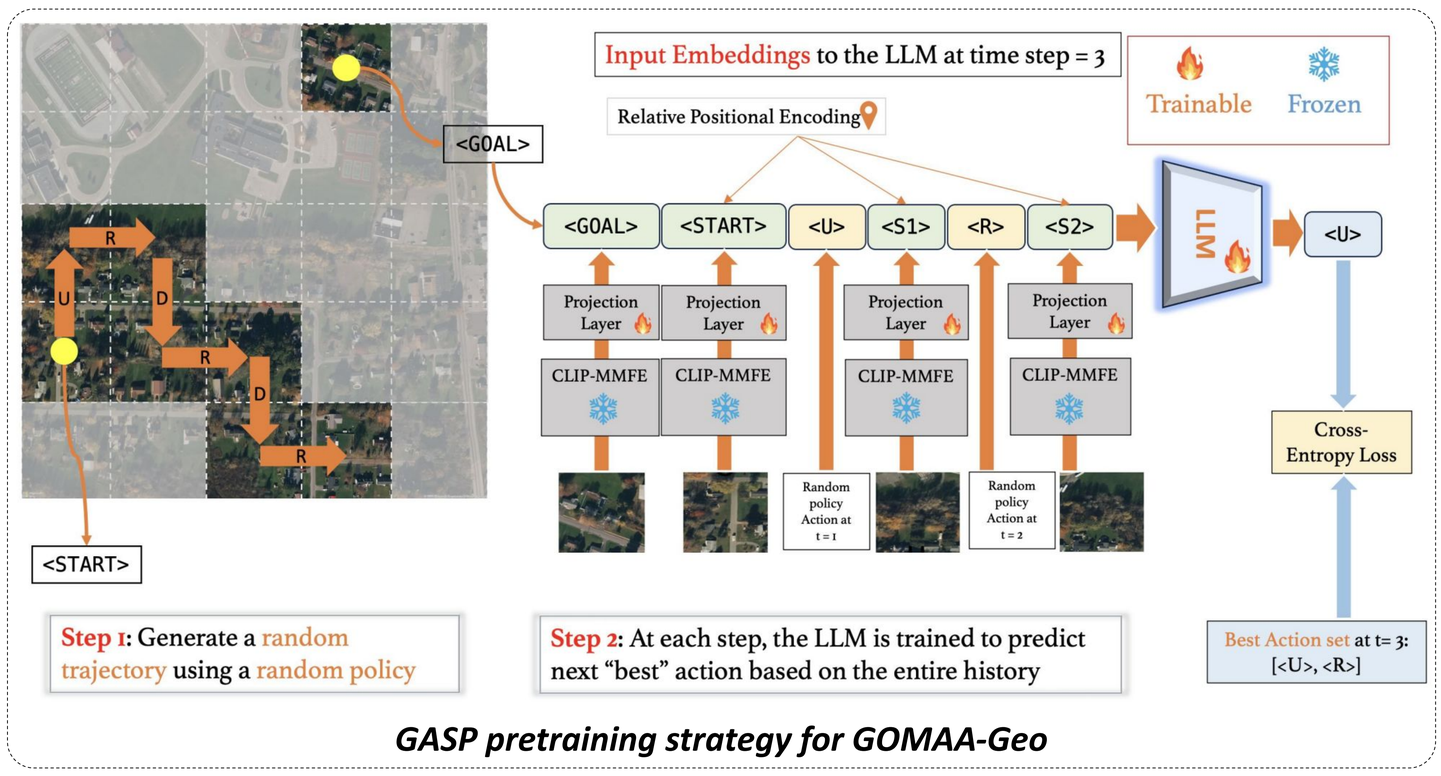
Goal modality agnostic: Zero-shot generalization between: aerial view | ground level view | natural language text

Limited field of view for agent (simulated UAV) → must perform efficient and accurate localization based on **partial observations**





Overview of proposed GOMAA-Geo agent



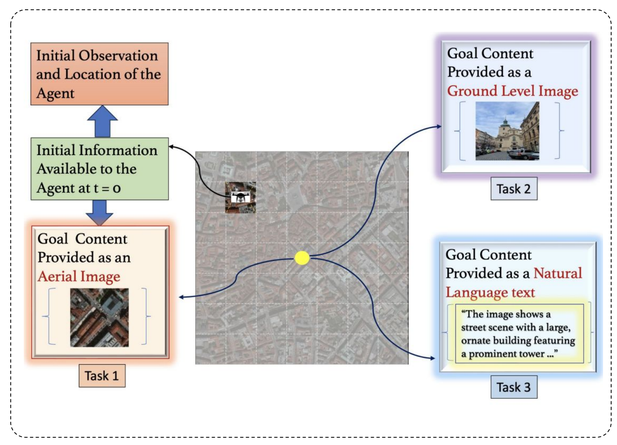
GASP pretraining strategy for GOMAA-Geo

Key steps and components of GOMAA-Geo

(1) Representation alignment across modalities

(2) RL-aligned representation learning using Goal Aware Supervised LLM Pretraining (*GASP*)

(3) Planning using RL (PPO)



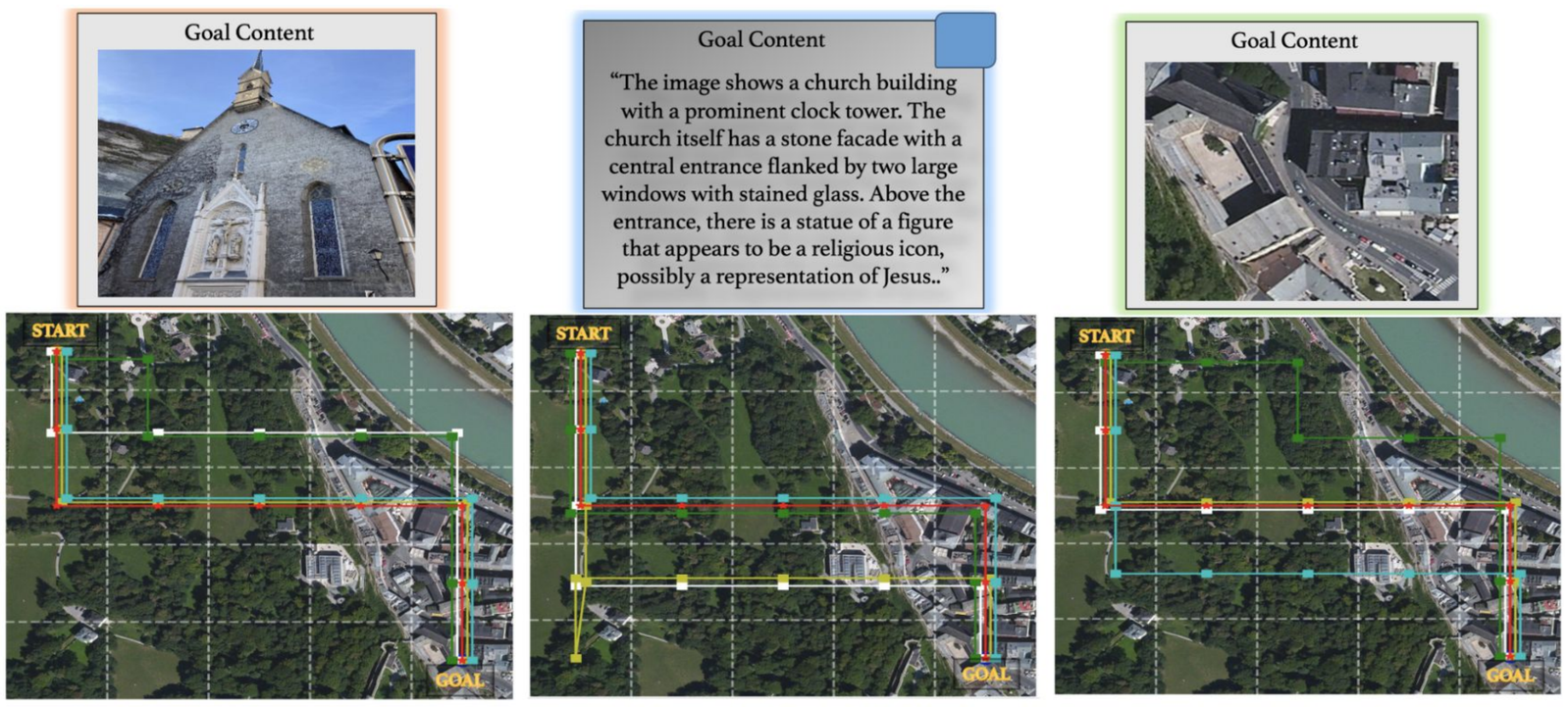
Experimental results

GOMAA-Geo generalizes well across goal modalities
(only trained using aerial images as goal modality!)

Goal Modality	$\mathcal{C} = 4$	$\mathcal{C} = 5$	$\mathcal{C} = 6$	$\mathcal{C} = 7$	$\mathcal{C} = 8$
Text	0.4000	0.4978	0.6766	0.7702	0.6595
Ground Image	0.4383	0.5150	0.6808	0.7489	0.6893
Aerial Image	0.4085	0.5064	0.6638	0.7362	0.7021

RL-based planning (3) is crucial (LLM-based predictions alone are not sufficient for solving the AGL task)

Method	$\mathcal{C} = 4$	$\mathcal{C} = 5$	$\mathcal{C} = 6$	$\mathcal{C} = 7$	$\mathcal{C} = 8$
LLM-Geo	0.2331	0.2591	0.3121	0.3967	0.4051
GOMAA-Geo	**0.4090**	**0.5056**	**0.7168**	**0.8034**	**0.7854**

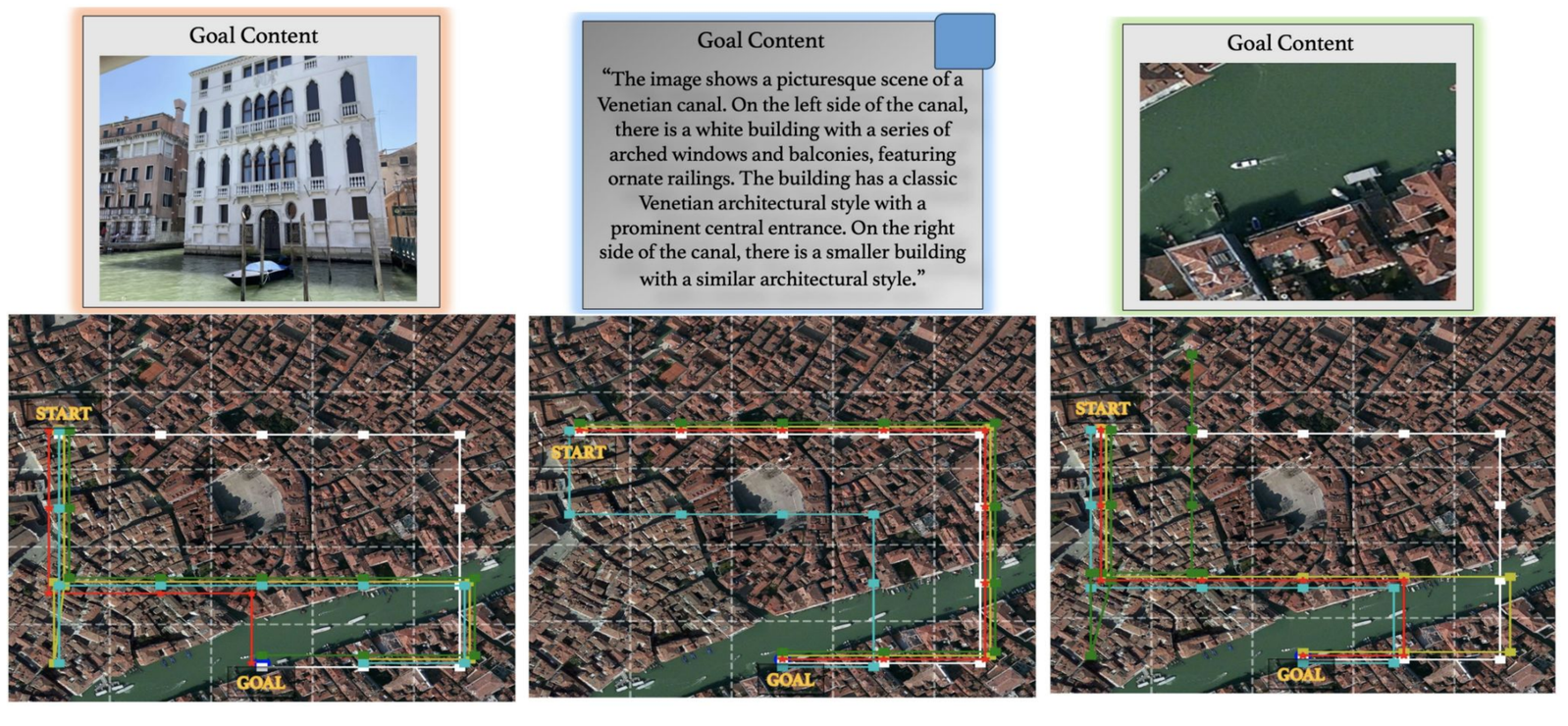


GOMAA-Geo outperforms alternative approaches
(results below are on post-disaster data, but trained on pre-disaster data)

Method	$\mathcal{C} = 4$	$\mathcal{C} = 5$	$\mathcal{C} = 6$	$\mathcal{C} = 7$	$\mathcal{C} = 8$
Random	0.1412	0.0584	0.0640	0.0247	0.0236
PPO	0.1132	0.1146	0.1292	0.1665	0.1953
DiT	0.1012	0.2389	0.3067	0.3390	0.3543
AiRLoc	0.1201	0.1298	0.1507	0.1631	0.1989
GOMAA-Geo	**0.4002**	**0.4632**	**0.6553**	**0.7391**	**0.6942**

The AGL task is challenging (GOMAA-Geo outperforms alternatives in larger grids, but further research needed)

Method	$\mathcal{C} = 12$	$\mathcal{C} = 13$	$\mathcal{C} = 14$	$\mathcal{C} = 15$	$\mathcal{C} = 16$
Random	0.0314	0.0280	0.0157	0.0112	0.0101
DiT	0.0923	0.1015	0.0876	0.0864	0.0932
GOMAA-Geo	**0.2427**	**0.2360**	**0.2438**	**0.2685**	**0.2685**



The AGL task is challenging (GOMAA-Geo outperforms alternatives in larger grids, but further research needed)