Learning to Navigate in Mazes with Novel Layouts using Abstract Top-down Maps

Linfeng Zhao, Lawson L.S. Wong

Khoury College of Computer Sciences, Northeastern University

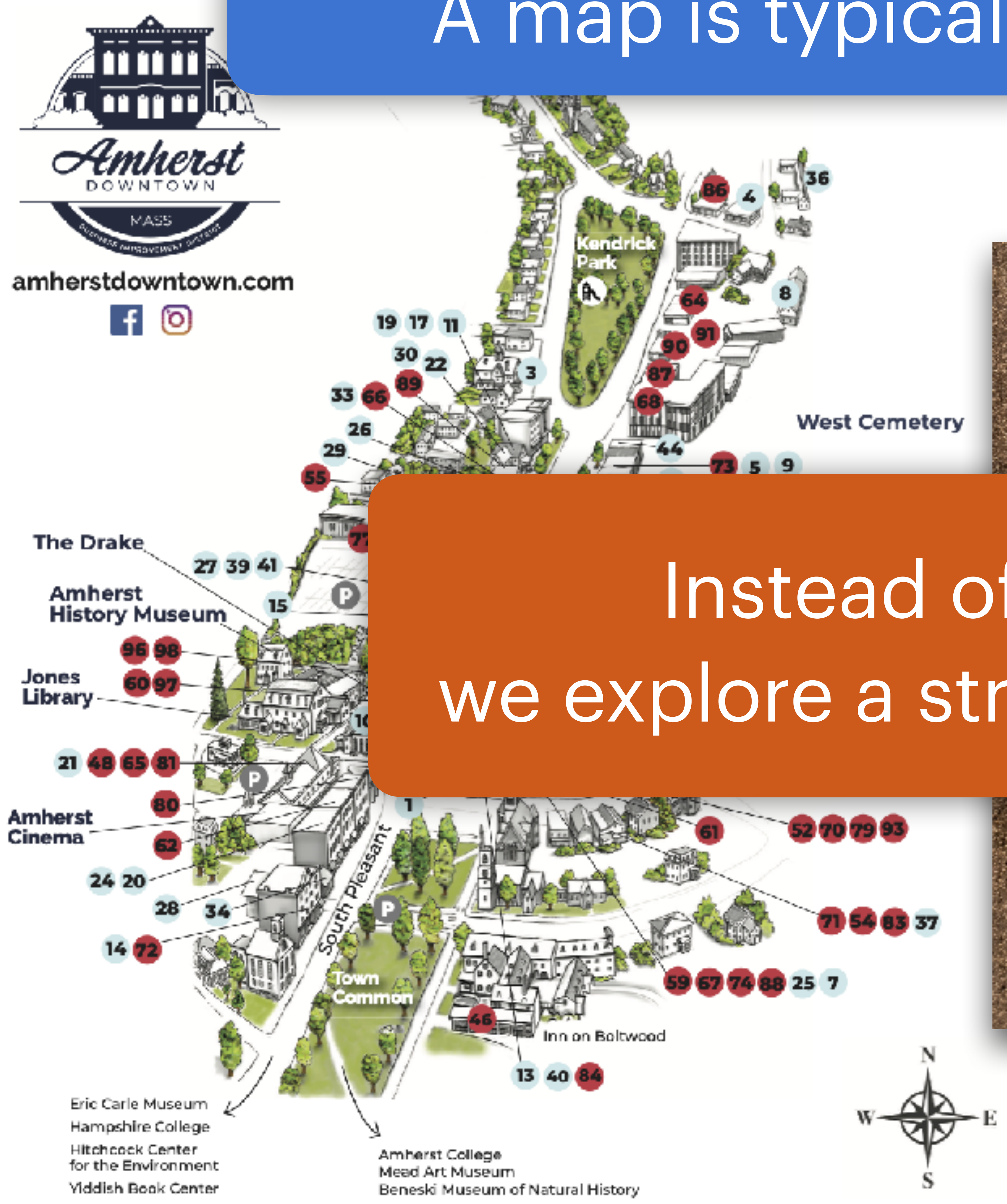
Reinforcement Learning Conference (RLC) 2024

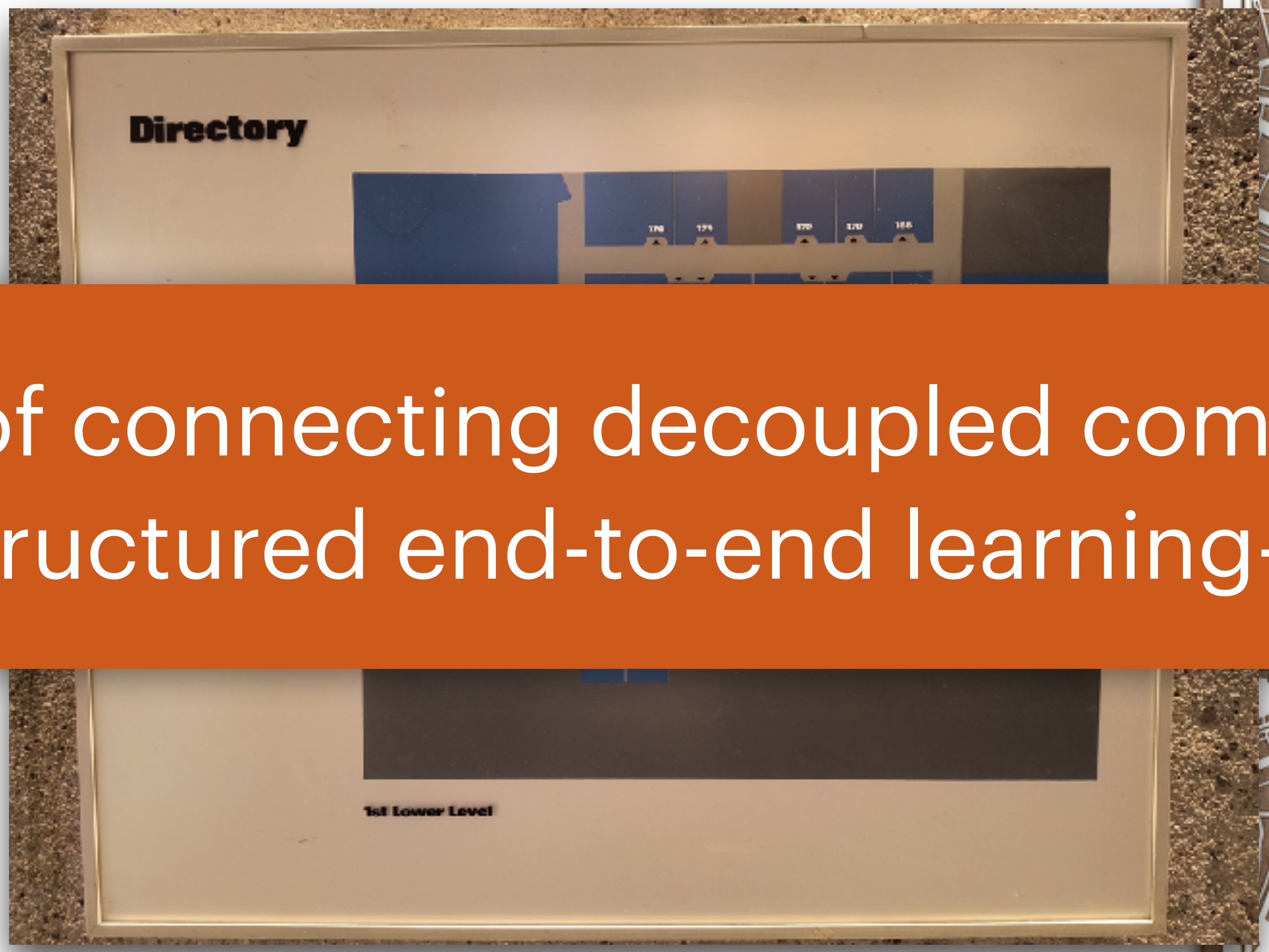
Navigation in a Novel Environment

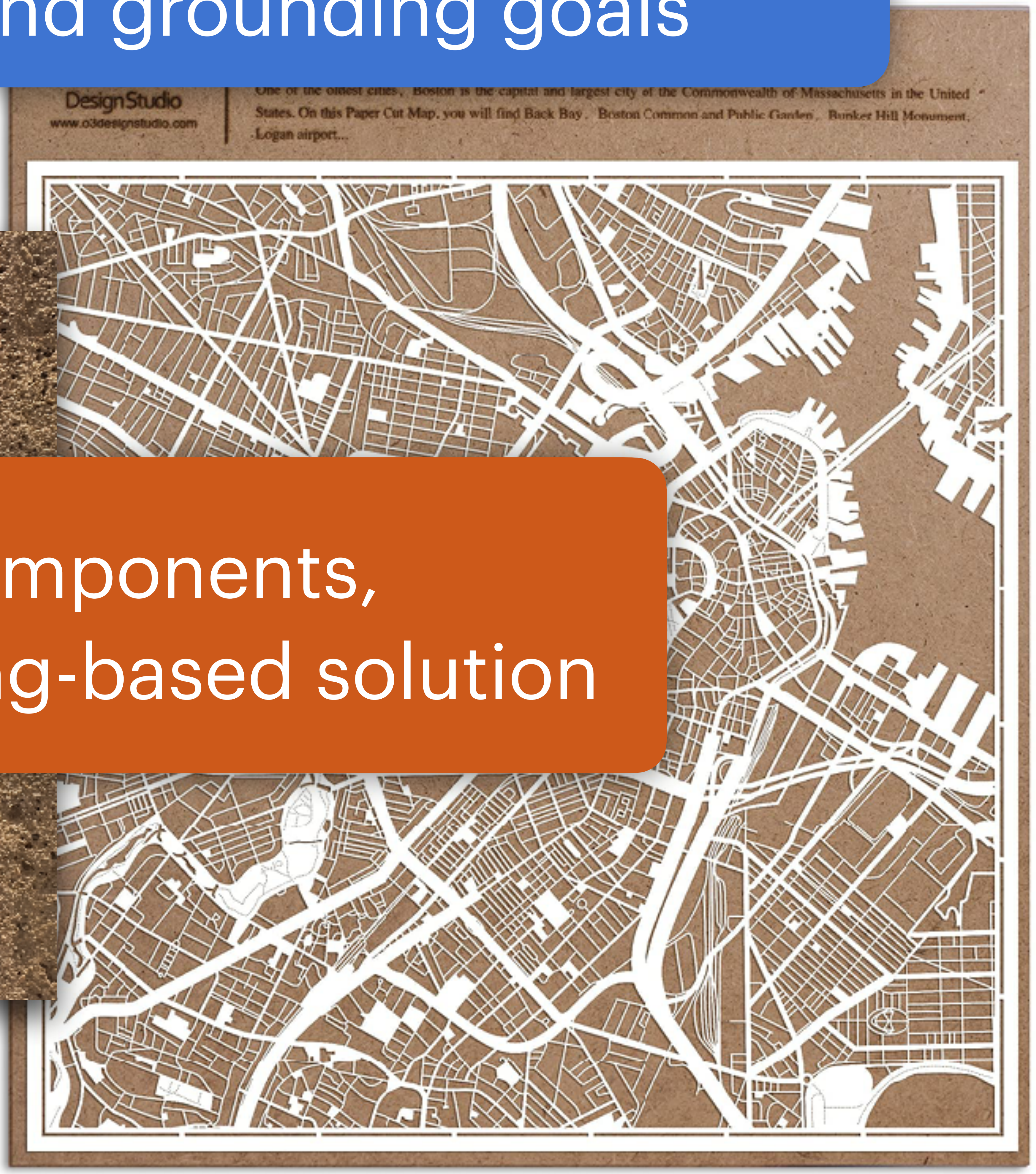
How do we navigate in environments we've never explored before?

Navigating Novel Environment using Maps

A map is typically required for navigation and grounding goals





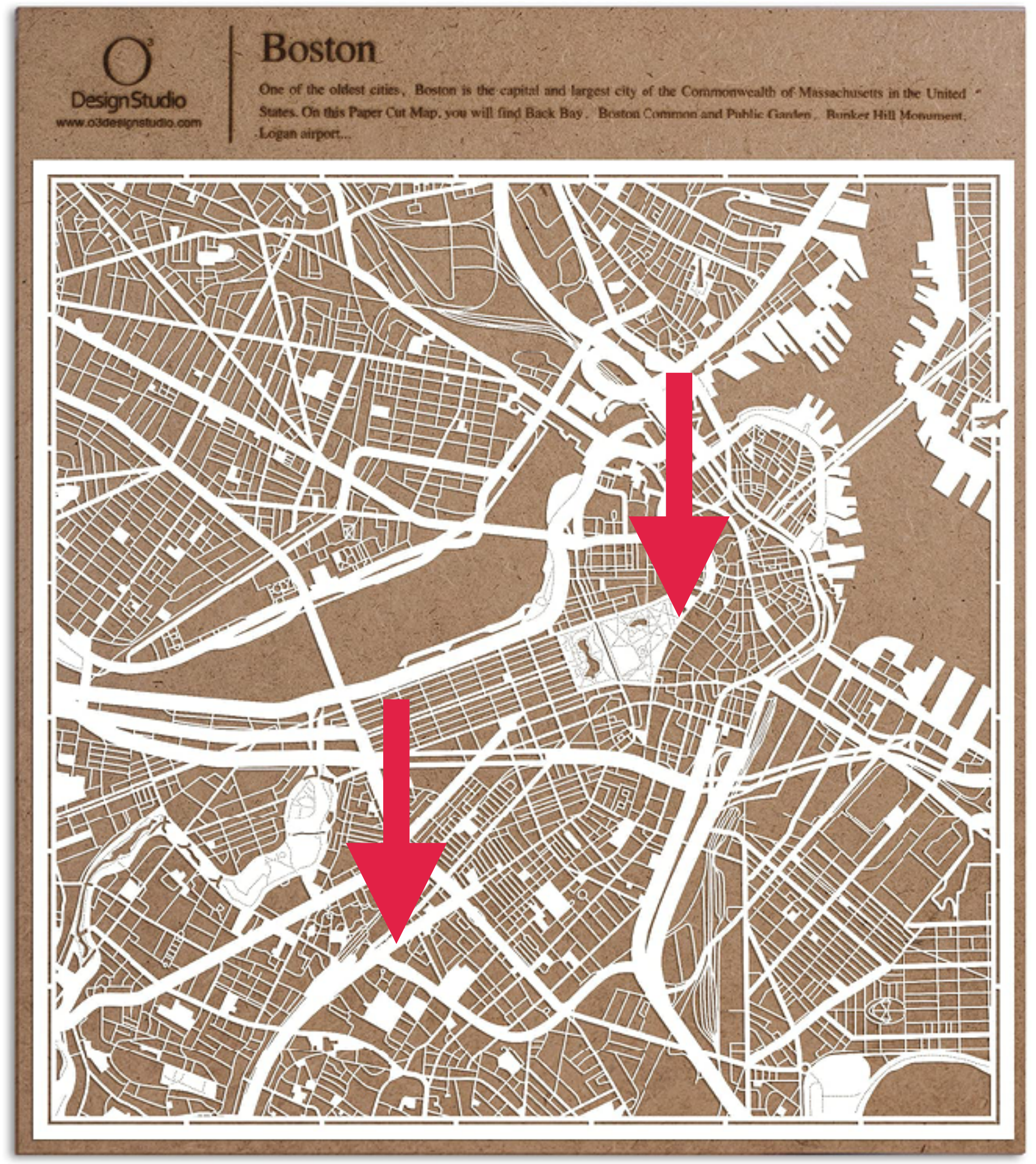


Instead of connecting decoupled components,
we explore a structured end-to-end learning-based solution

Learning Map-based Navigation

Single Task Example

Instead of connecting decoupled components,
we explore a structured end-to-end learning-based solution



Task Input m
Abstract Top-down Map

Step Input o_t
Environment Observation

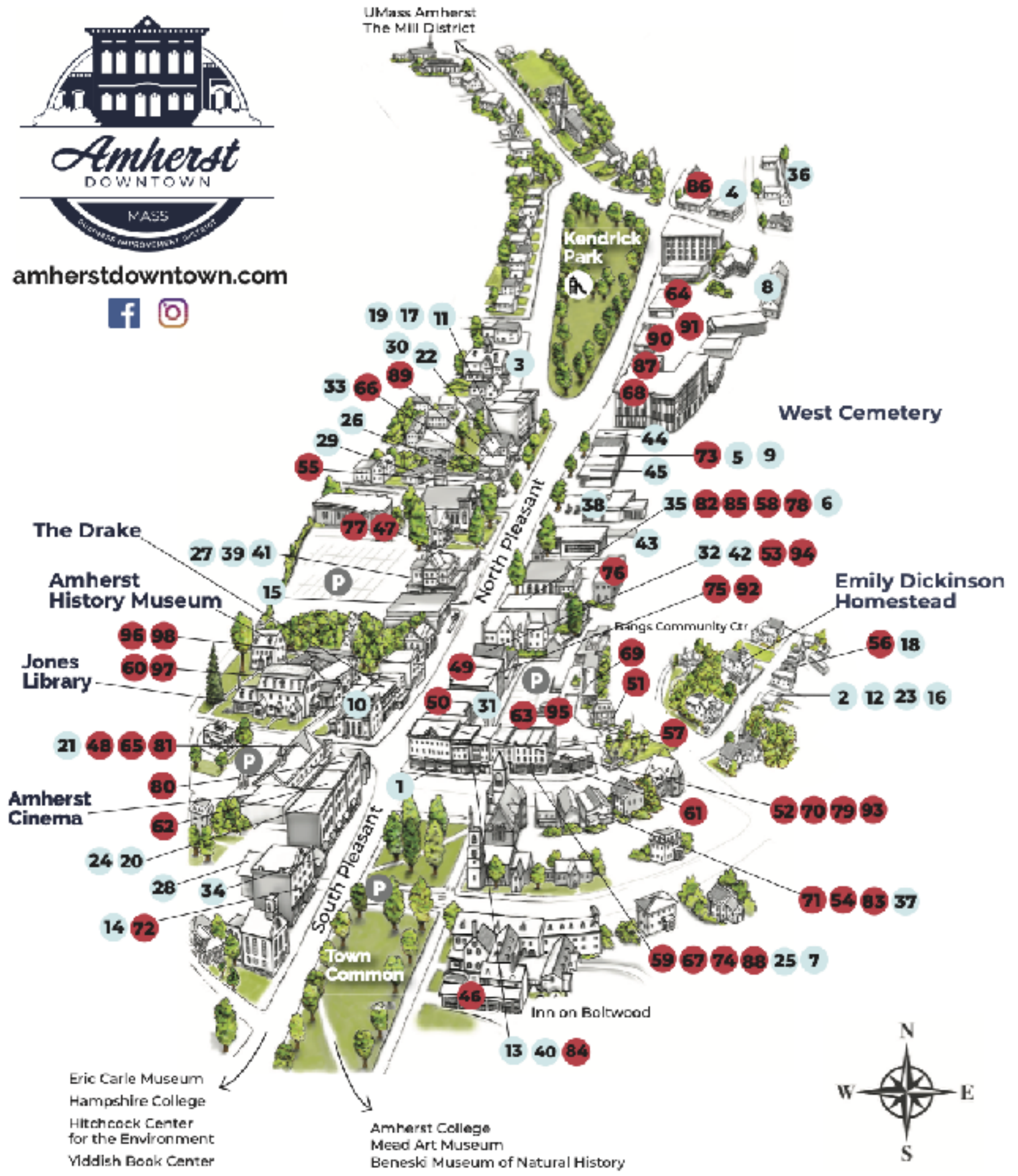
Typical Steps:

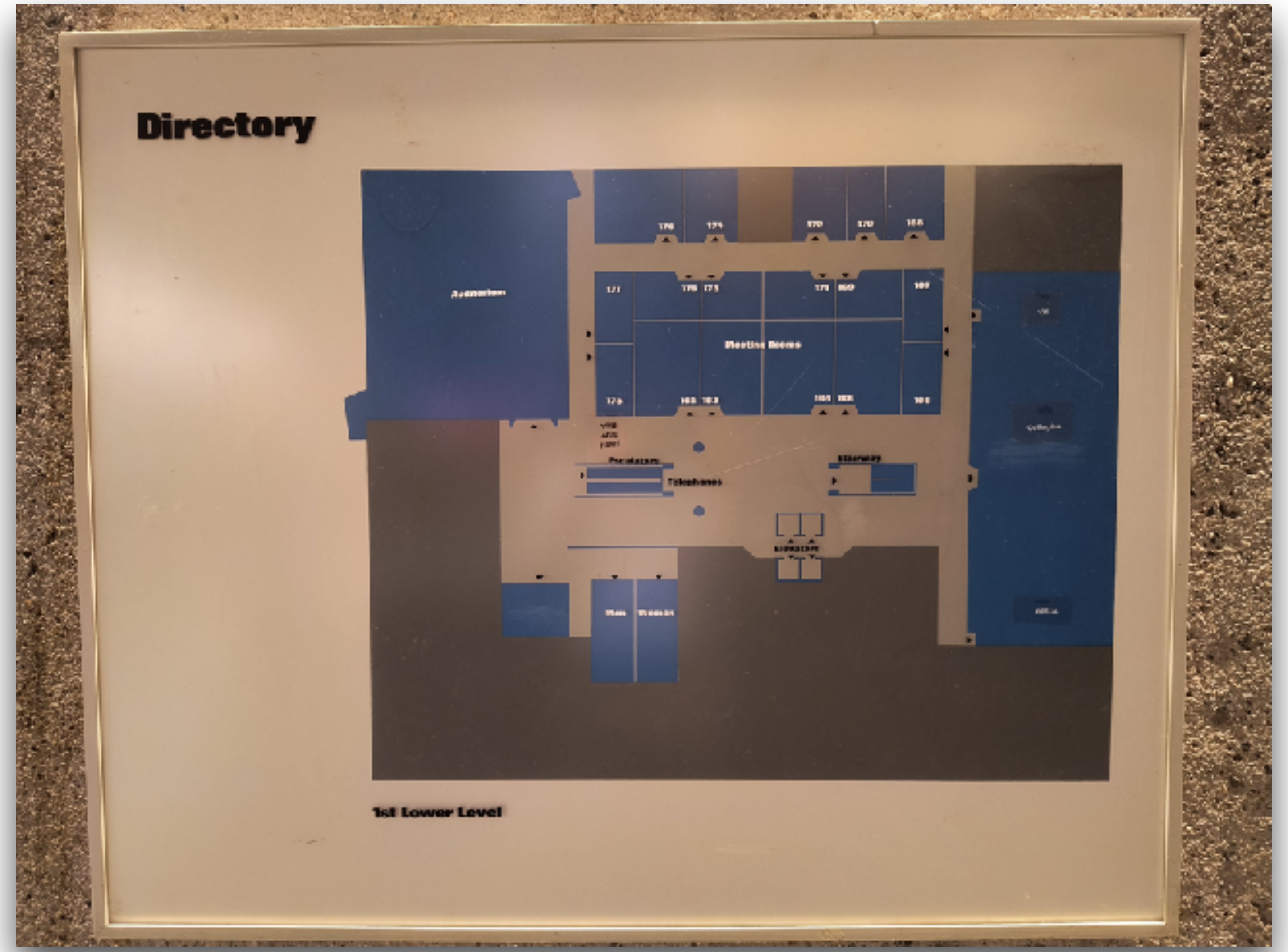
- Given Goal on Map
- Localizing it on Map
- Navigating on Map Image
- Ground to Actual Actions
- Output an Action

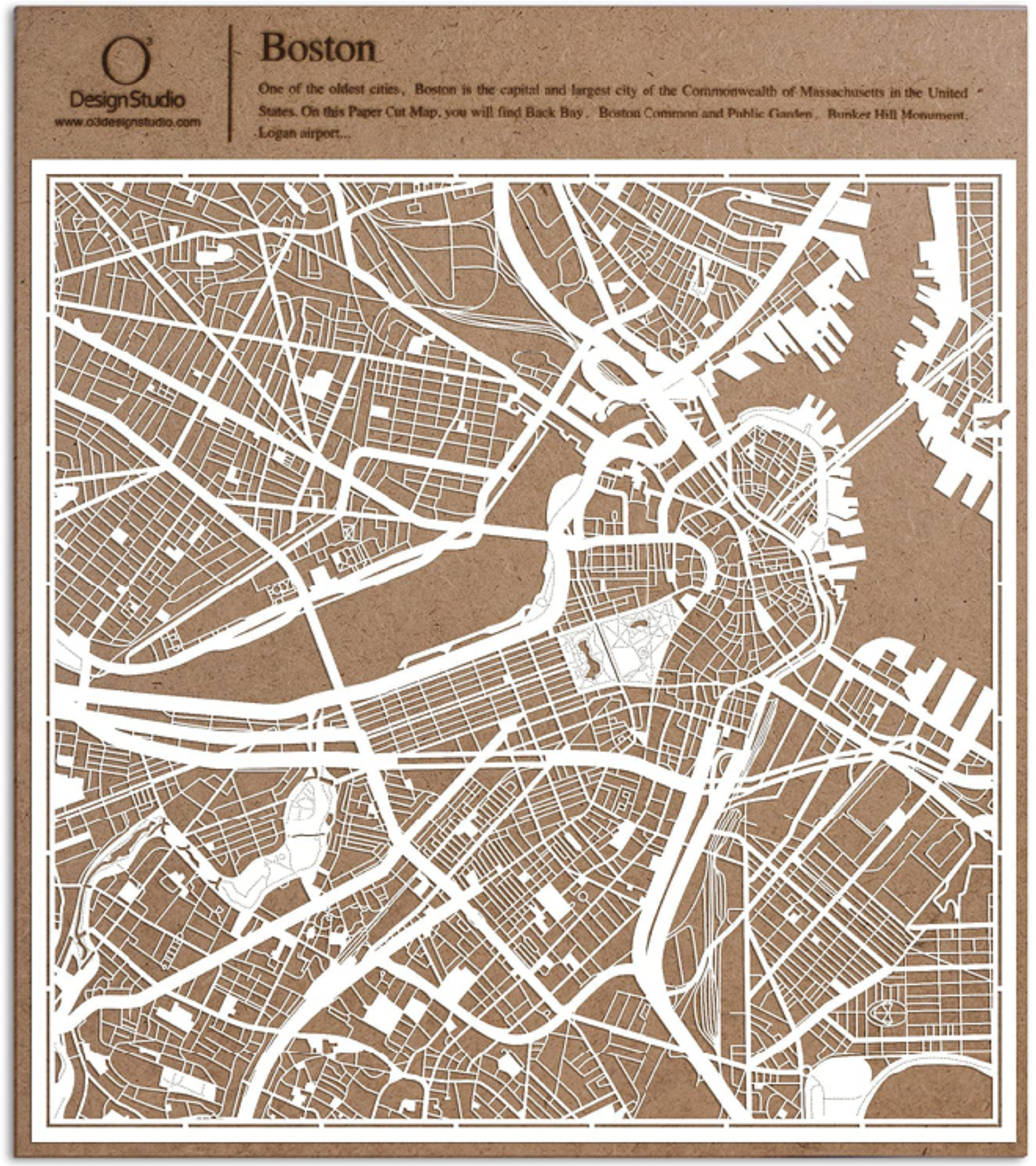
Learning Map-based Navigation

Multi-task Training and Generalization

Objective: Train on a distribution of maps to enable generalization to novel maps

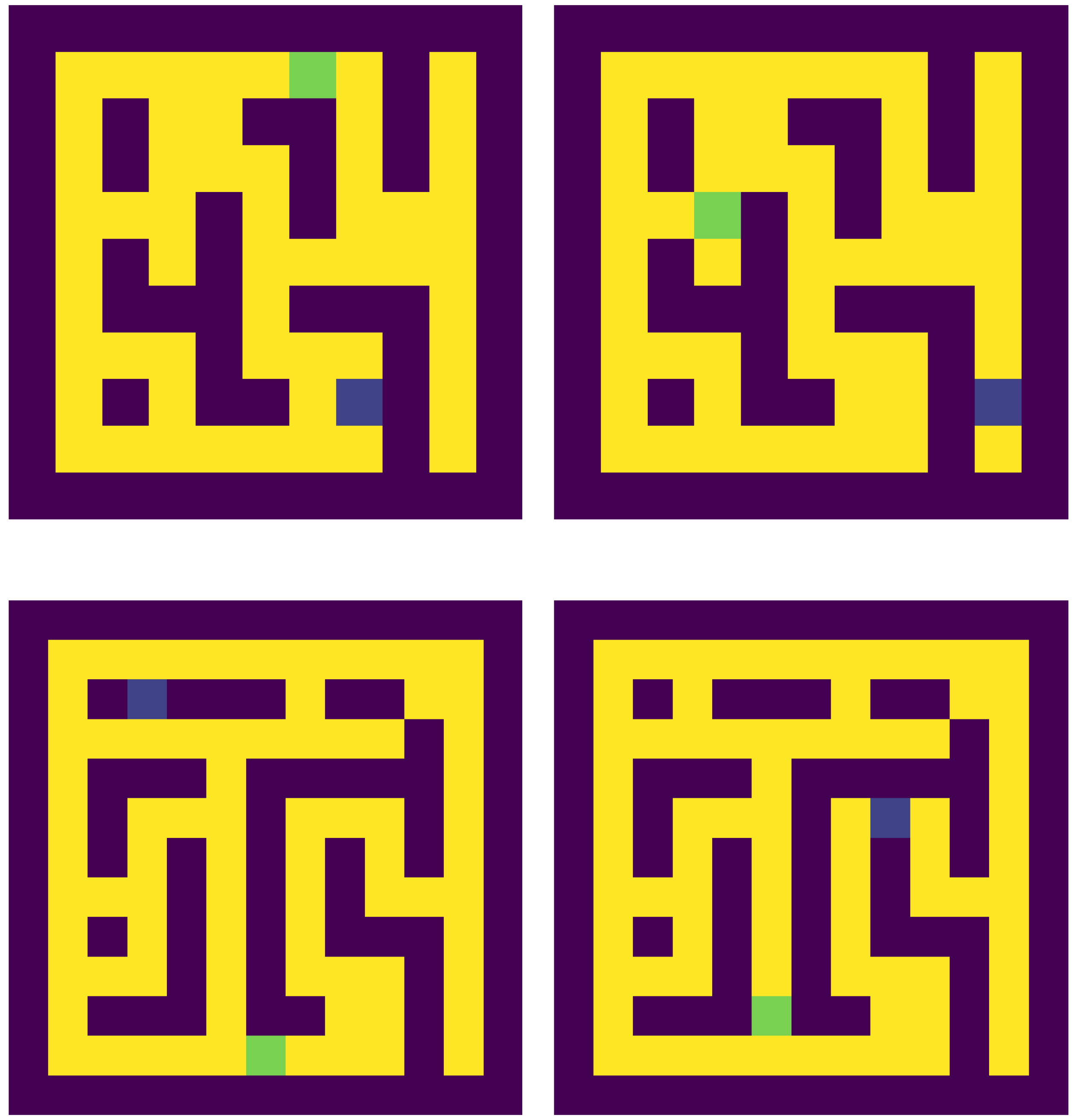






Formulation: Map-based *Maze* Navigation

As Multitask & Goal-conditioned RL Problem



Training

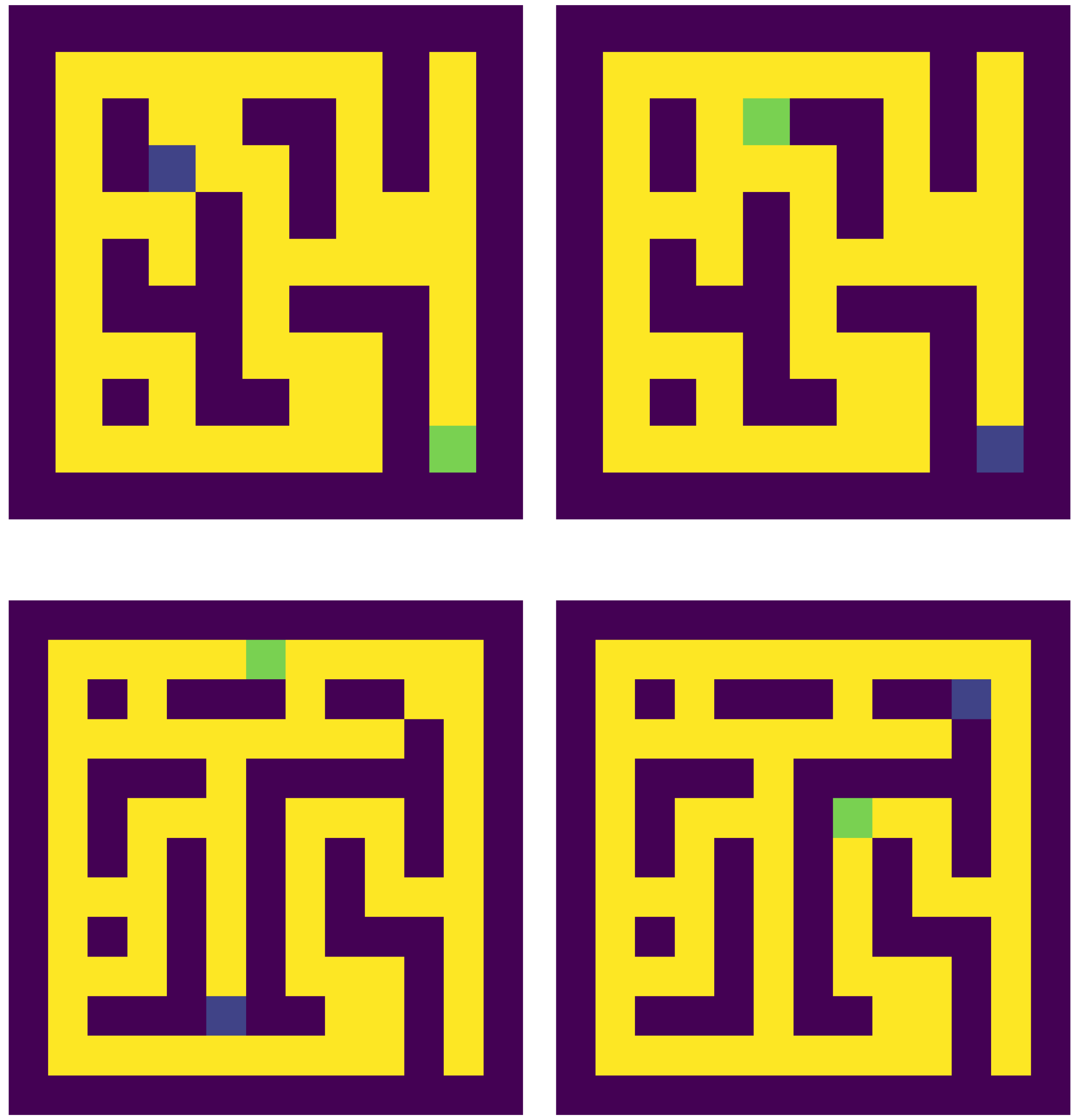
Learning Navigation Behaviors

Given State, Map and Goal
Plan the actions:

$$\text{Plan}\left(s, m, g, f_\phi\right)$$

Outcome: Map-conditioned Planner

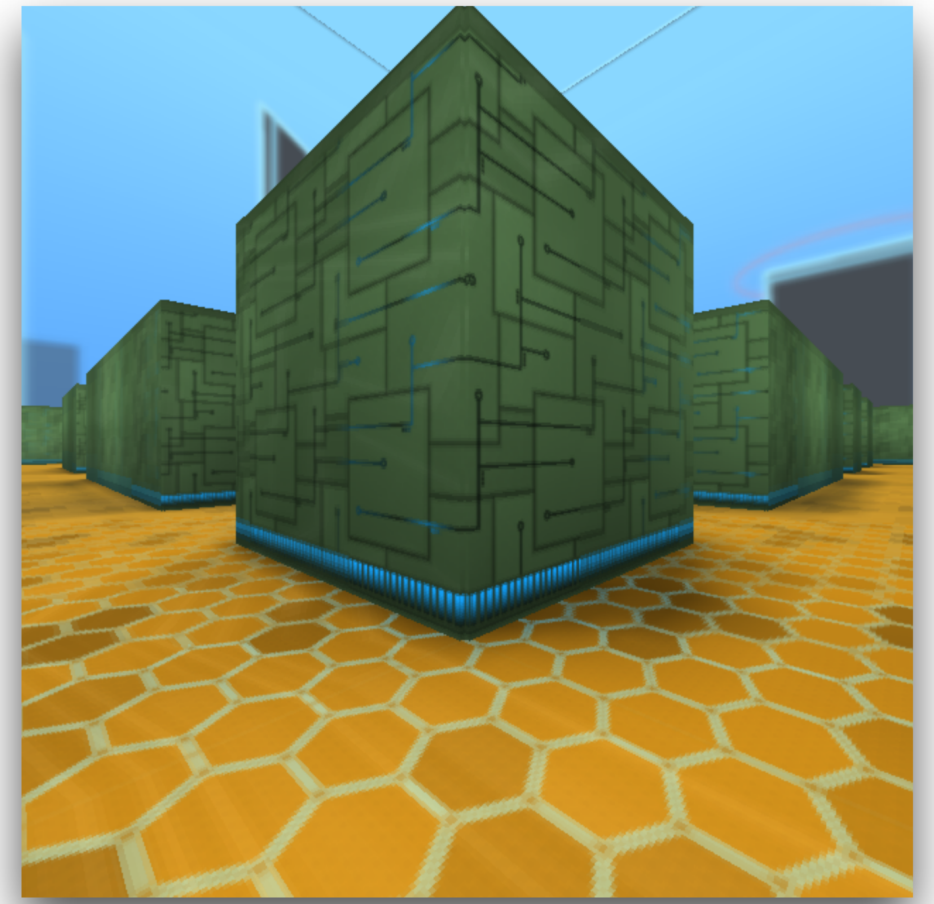
$$\pi(s, [m, g]; \theta)$$



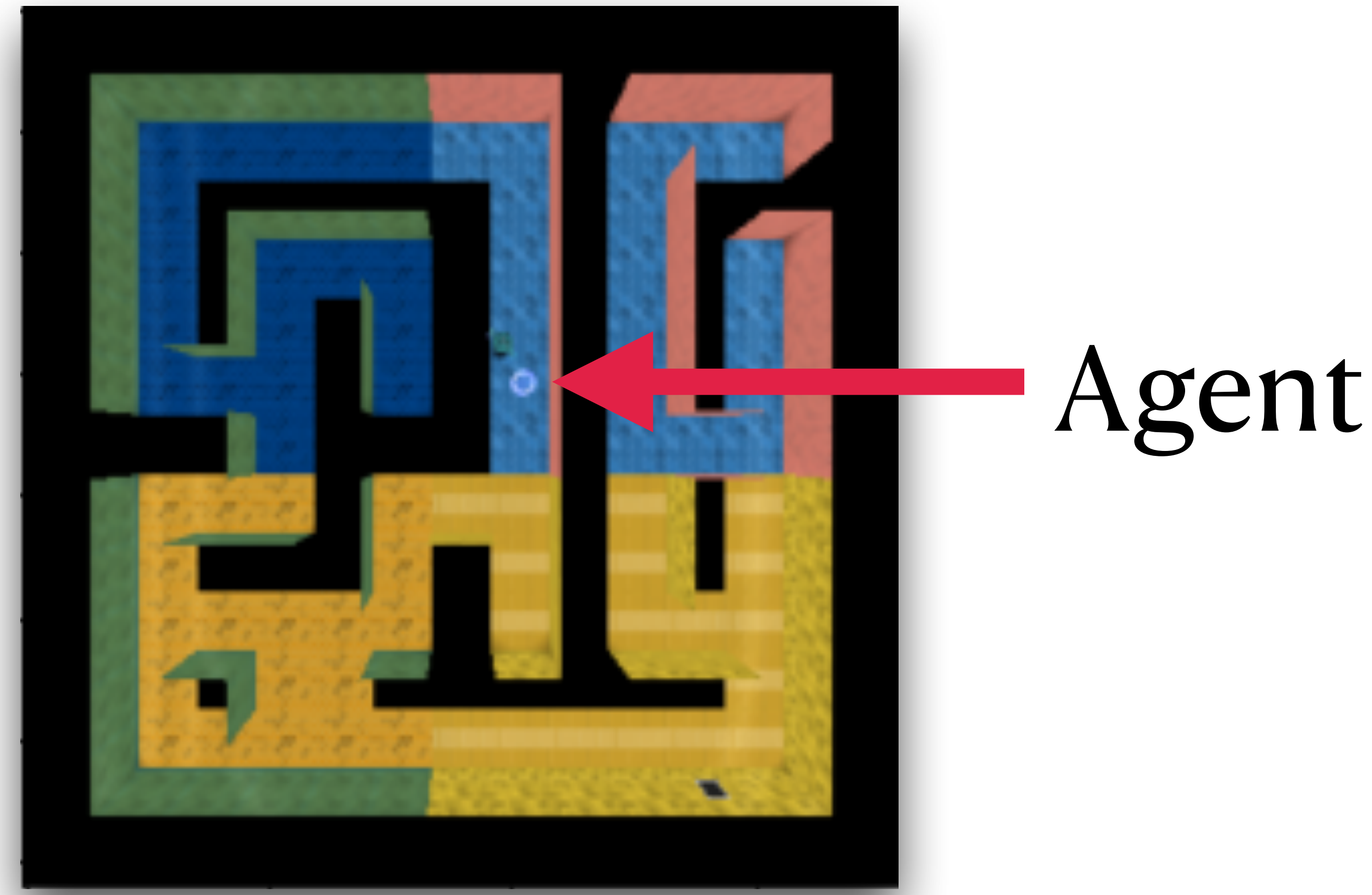
Zero-shot Navigation

Simulation Environment

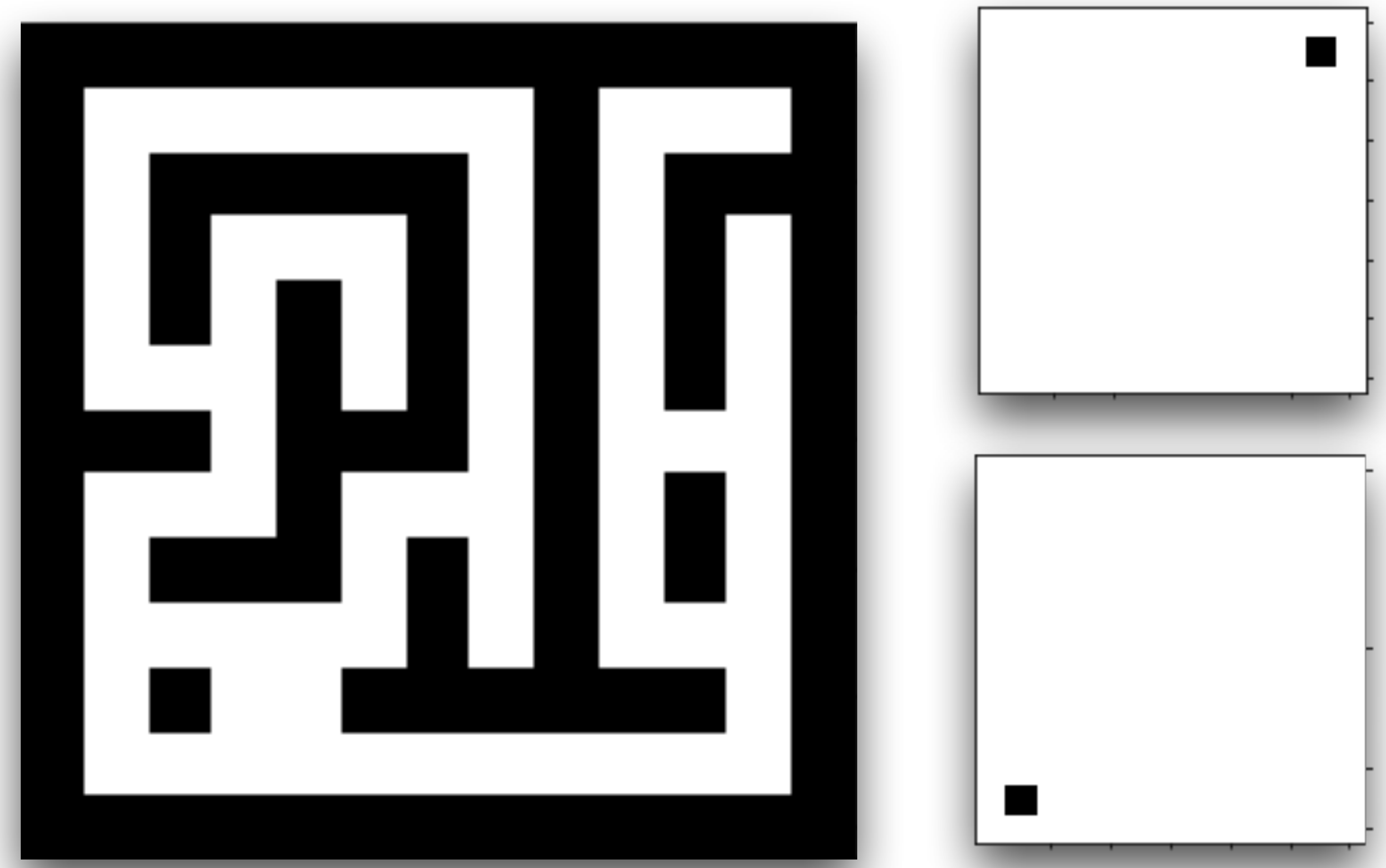
DeepMind Lab



Agent World
Not Input



Top-down View
Not Input

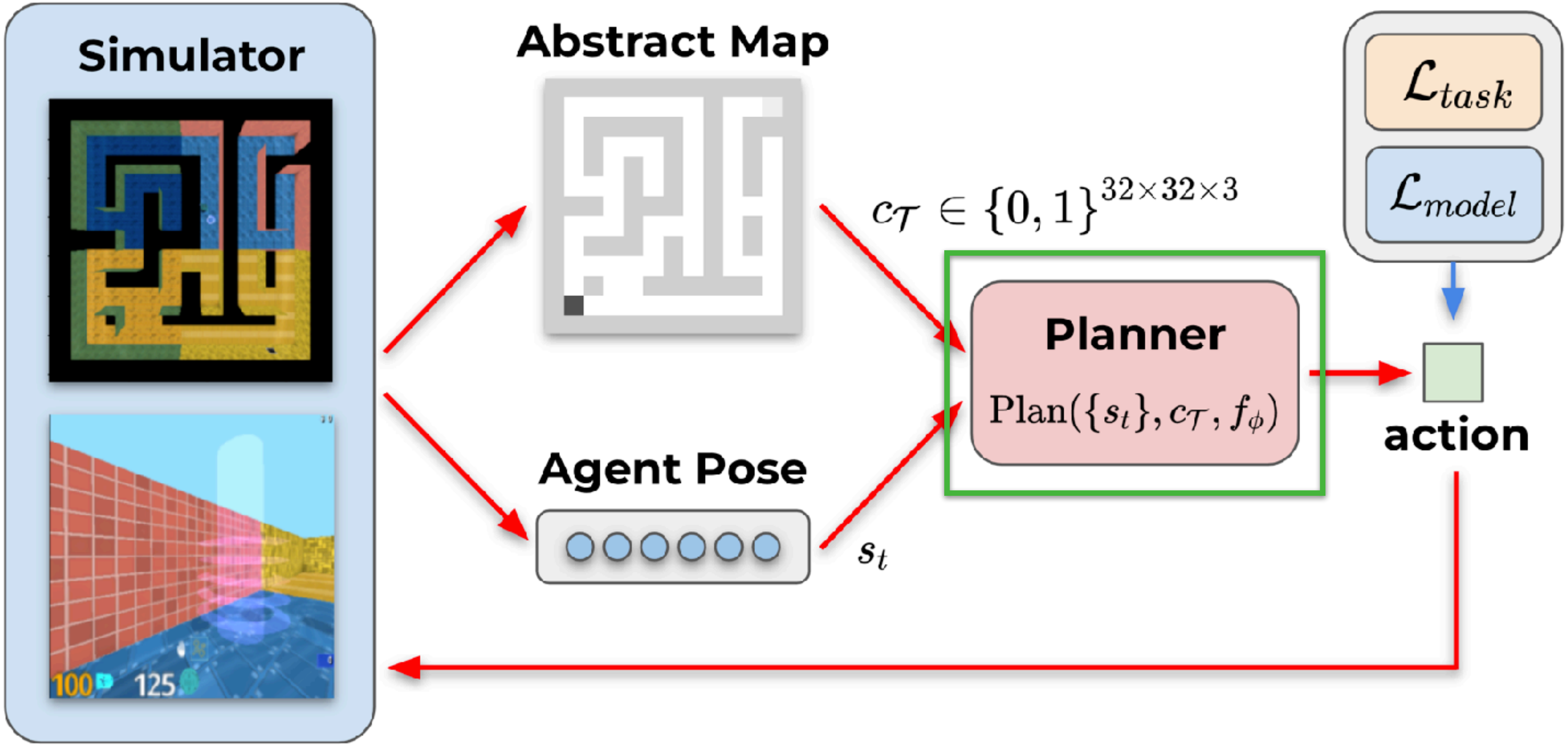


Abstract Map, Goal and Start
Task Input (context)

- State space = position $\mathbb{R}^3$ × orientation $\mathbb{R}^3$ × translational & rotational velocity $\mathbb{R}^6$
- Action space = {forward, backward, strafe left, strafe right, look left, look right}
- Reward $R_{\mathcal{G}}(s, a) = -\mathbb{I}[l(s) \neq g], g \in \mathcal{S}_{\mathcal{G}}$

Overview: Learned Map-conditioned Agent

"Map-conditioned Model-based Navigator"



Overview: Key Design

"Map-conditioned Model-based Navigator"

The algorithm and training procedure are similar to MuZero (Schrittwieser et al. 2020)

- However, MuZero is designed for single-task

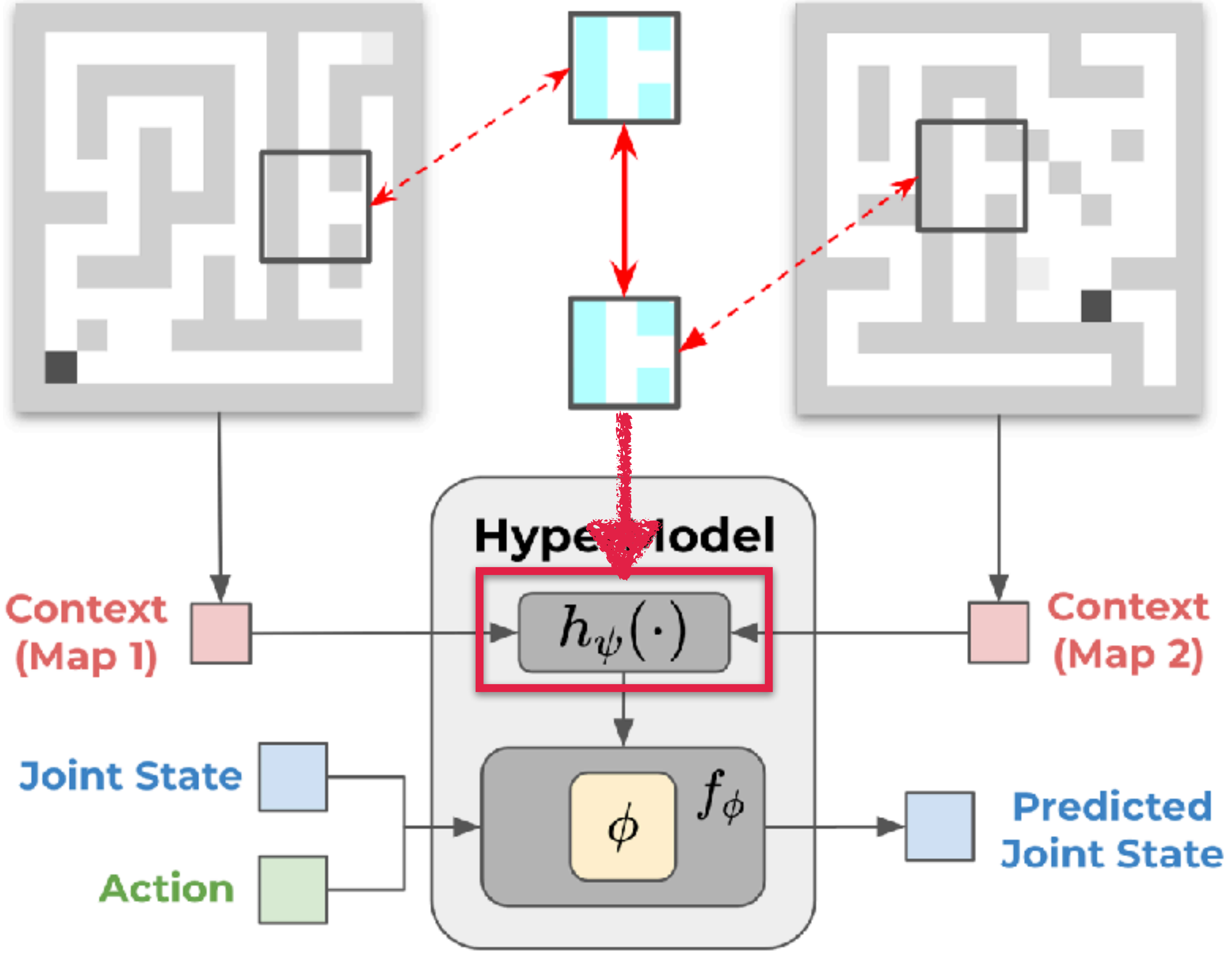
We design the agent for our (1) sparse-reward goal-conditioned (2) multi-task map-based navigation setup:

- Use model-based planning approach for longer-horizon planning (MCTS)
- Modeling dynamics via HyperNetworks (Ha et al., 2017; von Oswald et al., 2020)
- Multi-step Hindsight Experience Replay (HER) for sequence relabelling

Task-conditioned Hypermodel

Forward Pass

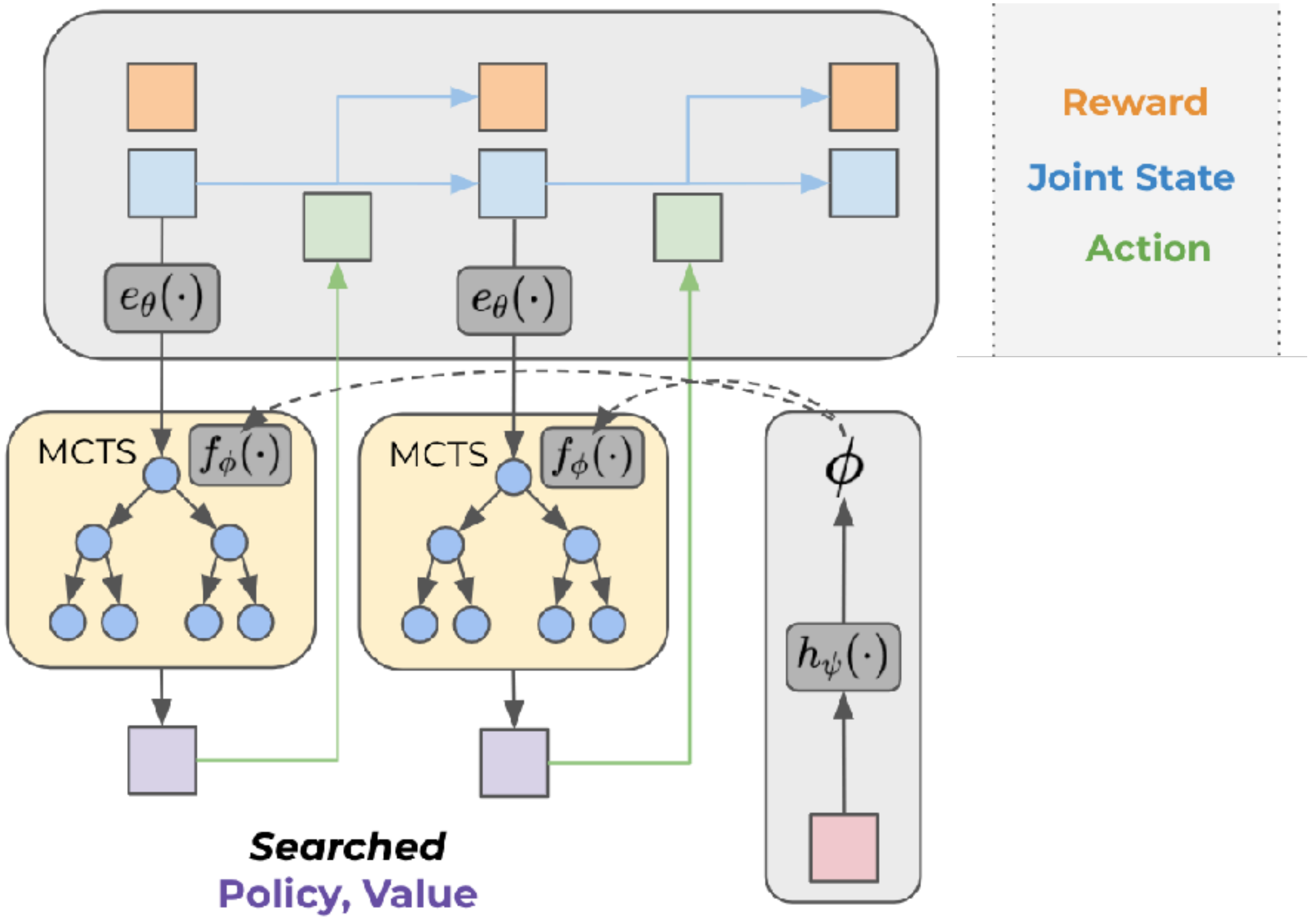
- Hypermodel $h_\psi : c \mapsto \phi, \quad f_\phi : s, a \mapsto s'$
- A hypernetwork h_ψ outputs weights of each transition network f_ϕ
    - The transition "computation" is thus *shared* between *tasks*
- *HyperNetworks* (Ha et al., 2016; von Oswald et al., 2019)



Planning using Learned Hypermodel

Forward Pass / Navigation Computation

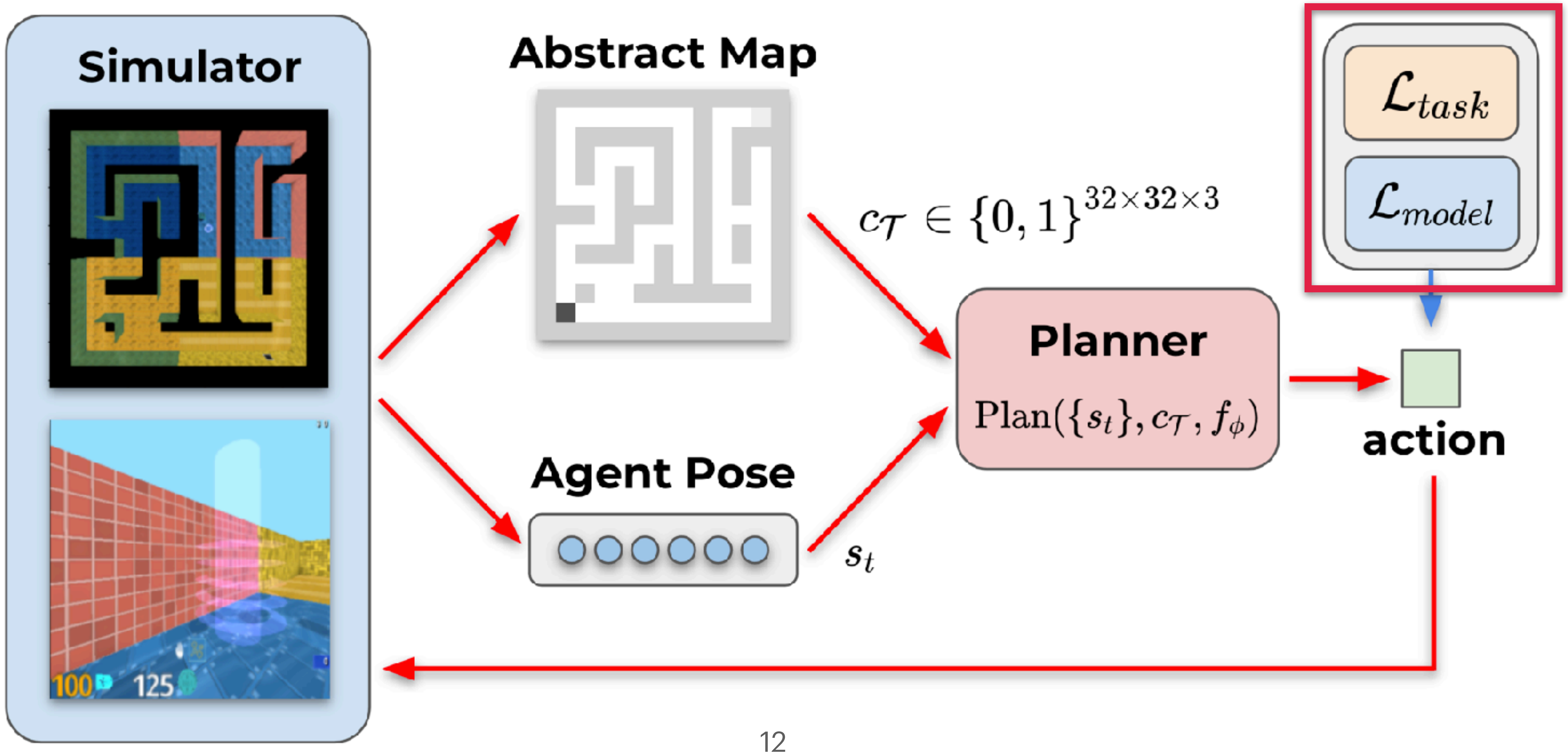
- Planning using *Monte-Carlo tree search*
  - 1. Use hypermodel to predict next states
  - 2. Search policy and value using MCTS
  - 3. Take action sampled from searched policy
  - 4. Repeat, Store $\left(c_{\mathcal{T}}, \left\{s_t, a_t, r_t, s_{t+1}\right\}_t\right)$



Overview

Training Objectives

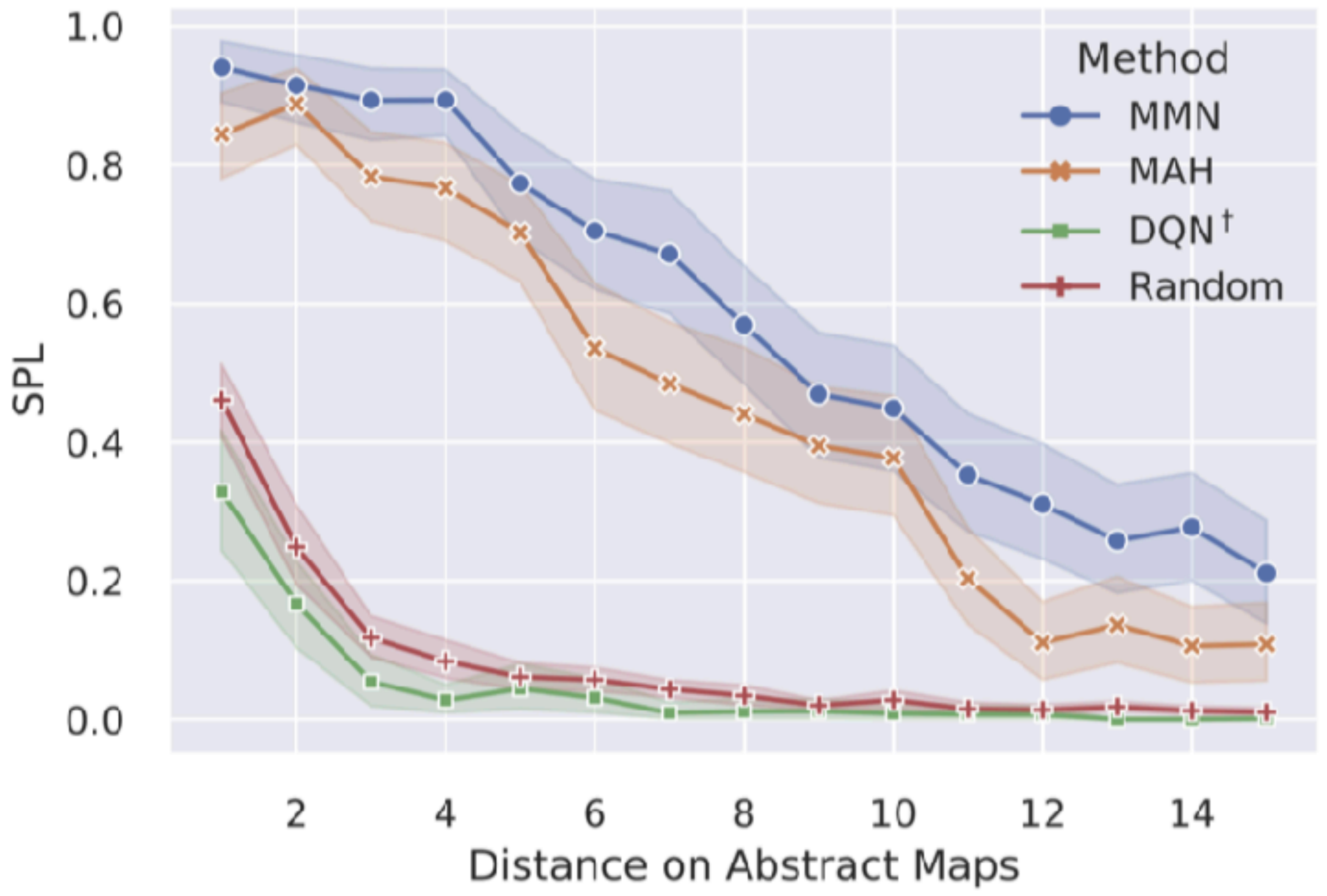
- (1) Task Loss + n-step Goal Relabelling MuZero (Schrittwieser et al., 2019), *HER* (Andrychowicz et al., 2017)
- (2) Auxiliary Model Loss: minimizing hypermodel prediction loss

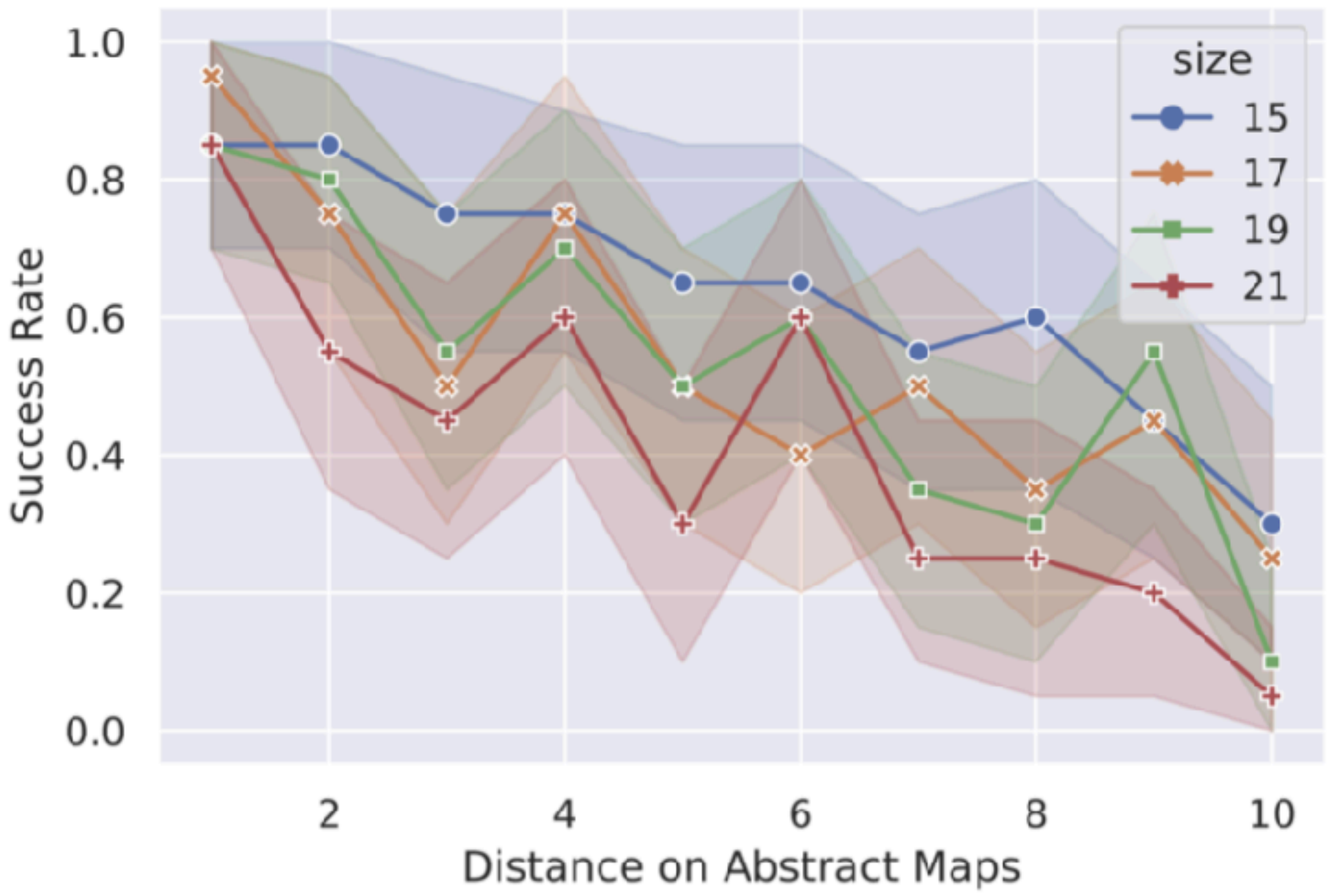


Zero-shot Navigation on Novel Maps

Key Results

- Evaluating on 20 unseen 13×13 maps with 5 goals in distance $[1,15]$ for each map
- **MMN** = Map-conditioned Model-based Navigator, ***model-based*** method; **MAH** = ***model-free*** variant

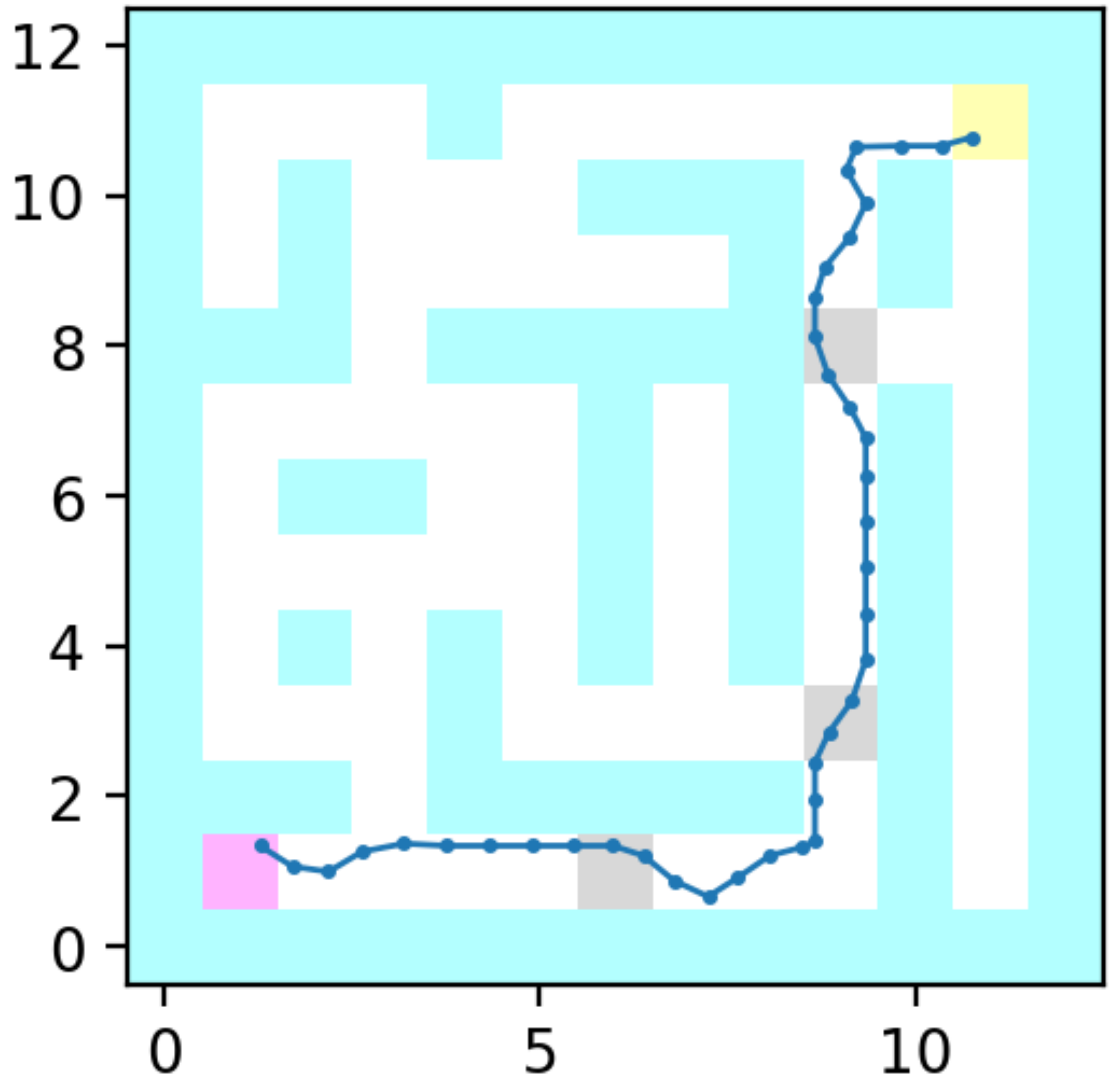




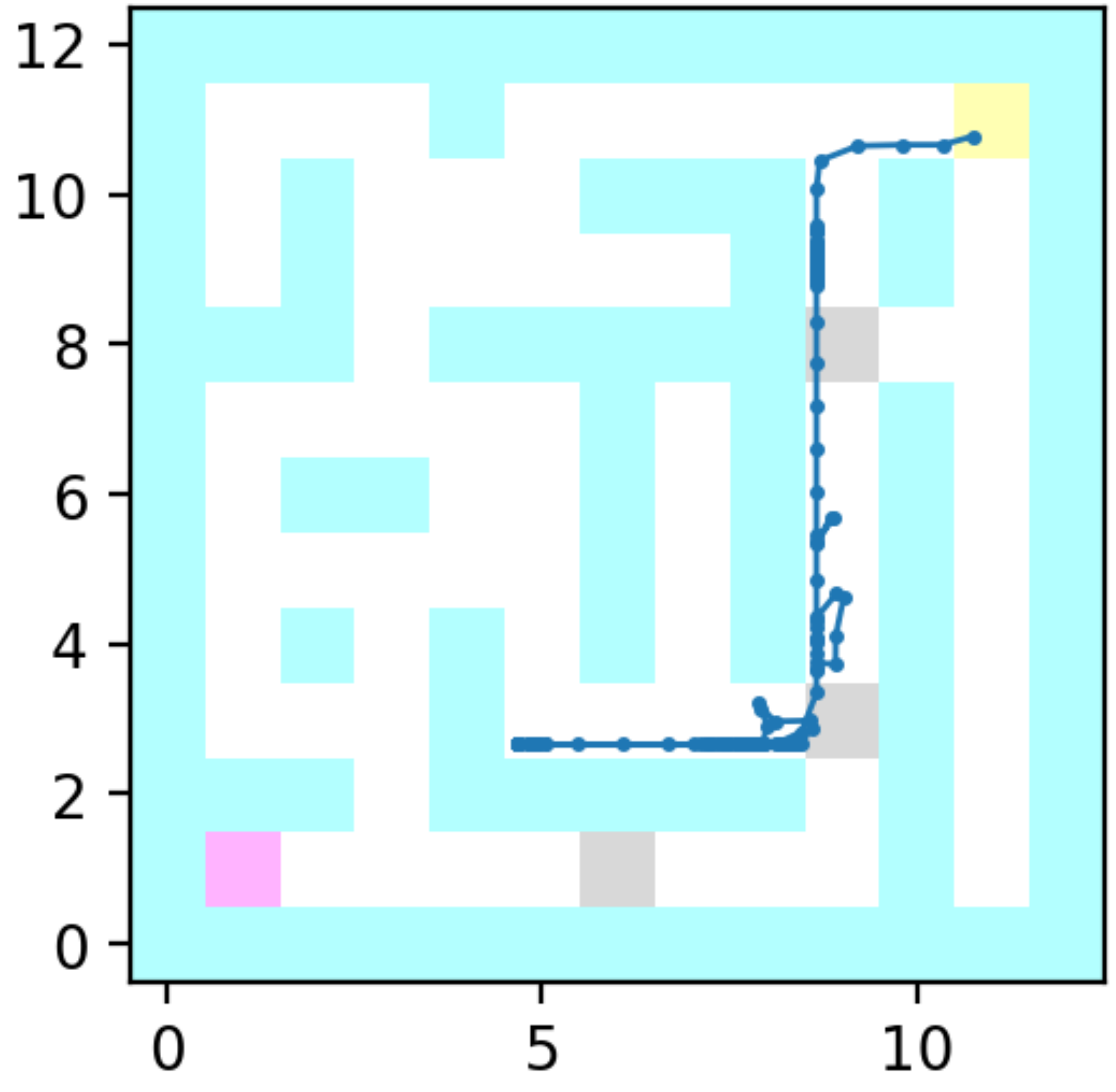
Zero-shot Hierarchical Navigation

Key Results

- ***Zero-shot*** evaluation of trained agents on an ***unseen*** map
- Agents use maps as *images*
- Landmarks (grey) are provided by a *oracle*
  - It is ***not*** required elsewhere
- Our model-based agent generalizes *better* and needs *less exploration* on *unseen* maps



MMN (Model-based)



MAH (Model-free)

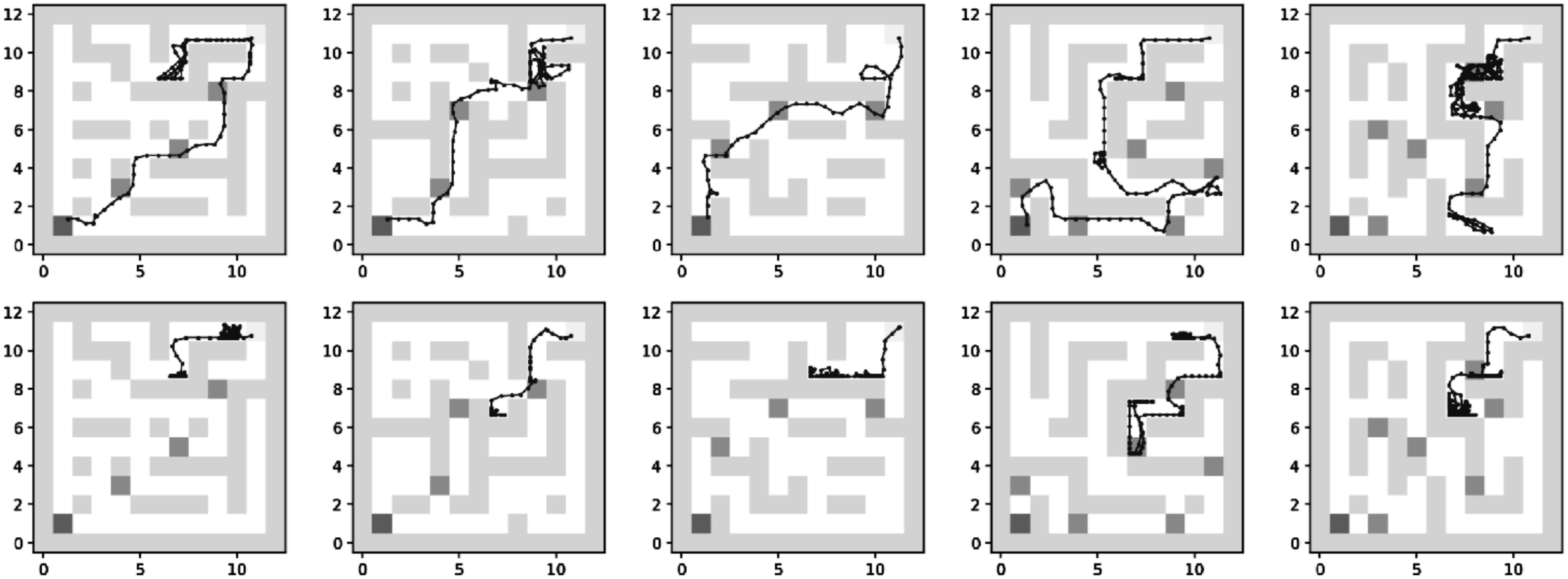
Zero-shot Hierarchical Navigation

Visualization

Map 1 | Map 2 | Map 3 | Map 4 | Map 5

MMN (Model-based)

MAH (Model-free)



Hierarchical Navigation Task with Landmarks (from top-right to bottom-left)

Zero-shot Hierarchical Navigation

Quantitative: various landmark distances

Landmark Distance	1	2	3	4	5	5 (SR)
MMN	**0.61**	**0.59**	**0.68**	**0.45**	**0.63**	**0.80**
MAH	0.24	0.42	0.45	0.41	0.28	0.45
DQN†	0.00	0.00	0.00	0.00	0.00	0.00
Random	0.00	0.00	0.00	0.00	0.00	0.00

- Hierarchical navigation performance for various distances between the landmarks
- Showing the SPL metric for landmark distances 1~5 and SR for distance 5 only
- Planning-based MMN greatly outperforms MAH (model-free) one and other baselines

Thank you!

Please come to our poster session: 08/11 5-7pm

Applied reinforcement learning: Room 174/176

Website:

Please reach out for more questions:

zhao.linf@northeastern.edu