Model-based Navigation in Environments with Novel Layouts using Abstract 2-D Maps

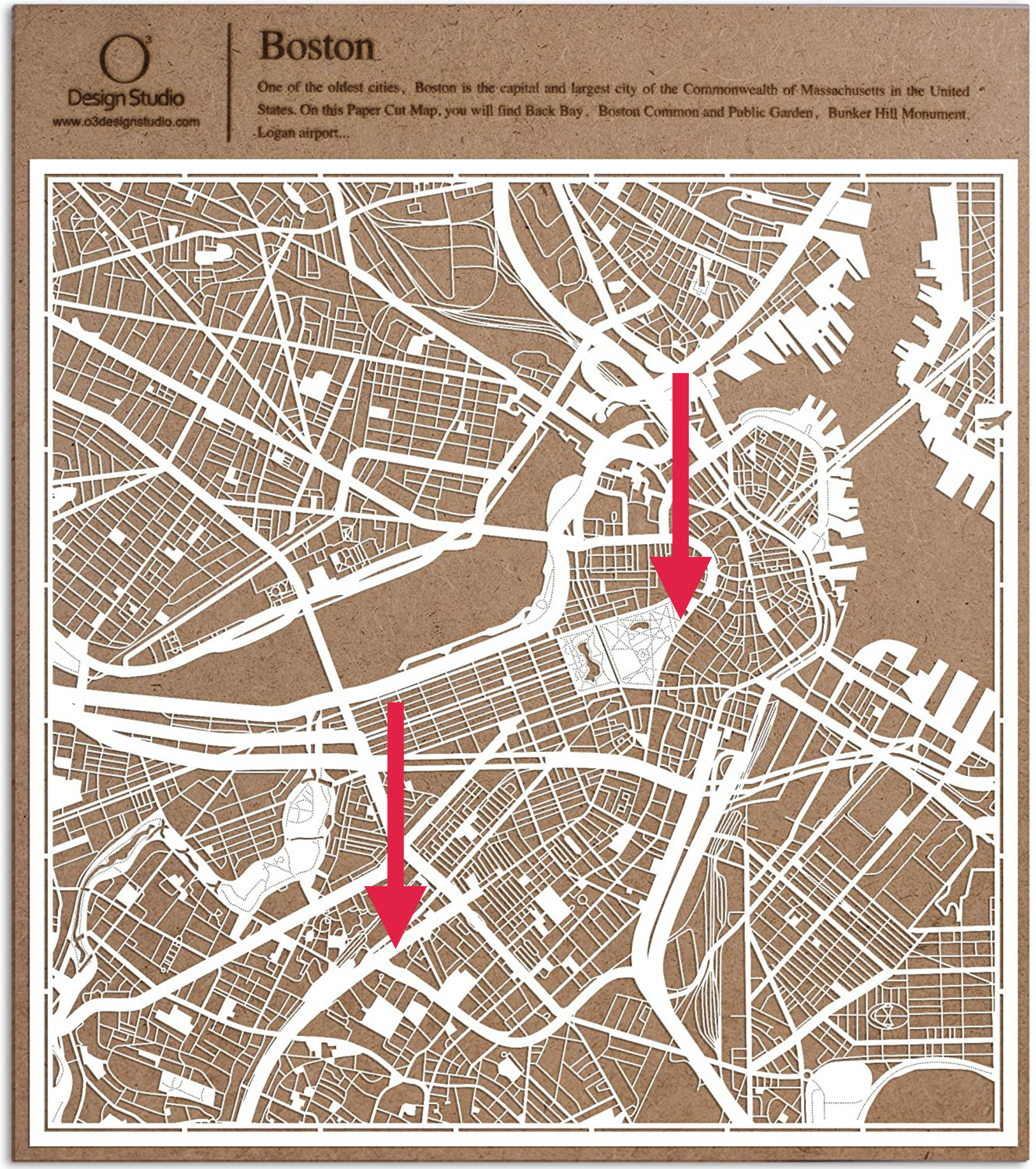
Linfeng Zhao, Lawson L.S. Wong

NeurIPS Deep Reinforcement Learning Workshop 2020

Khoury College of Computer Sciences, Northeastern University, Nov 2020

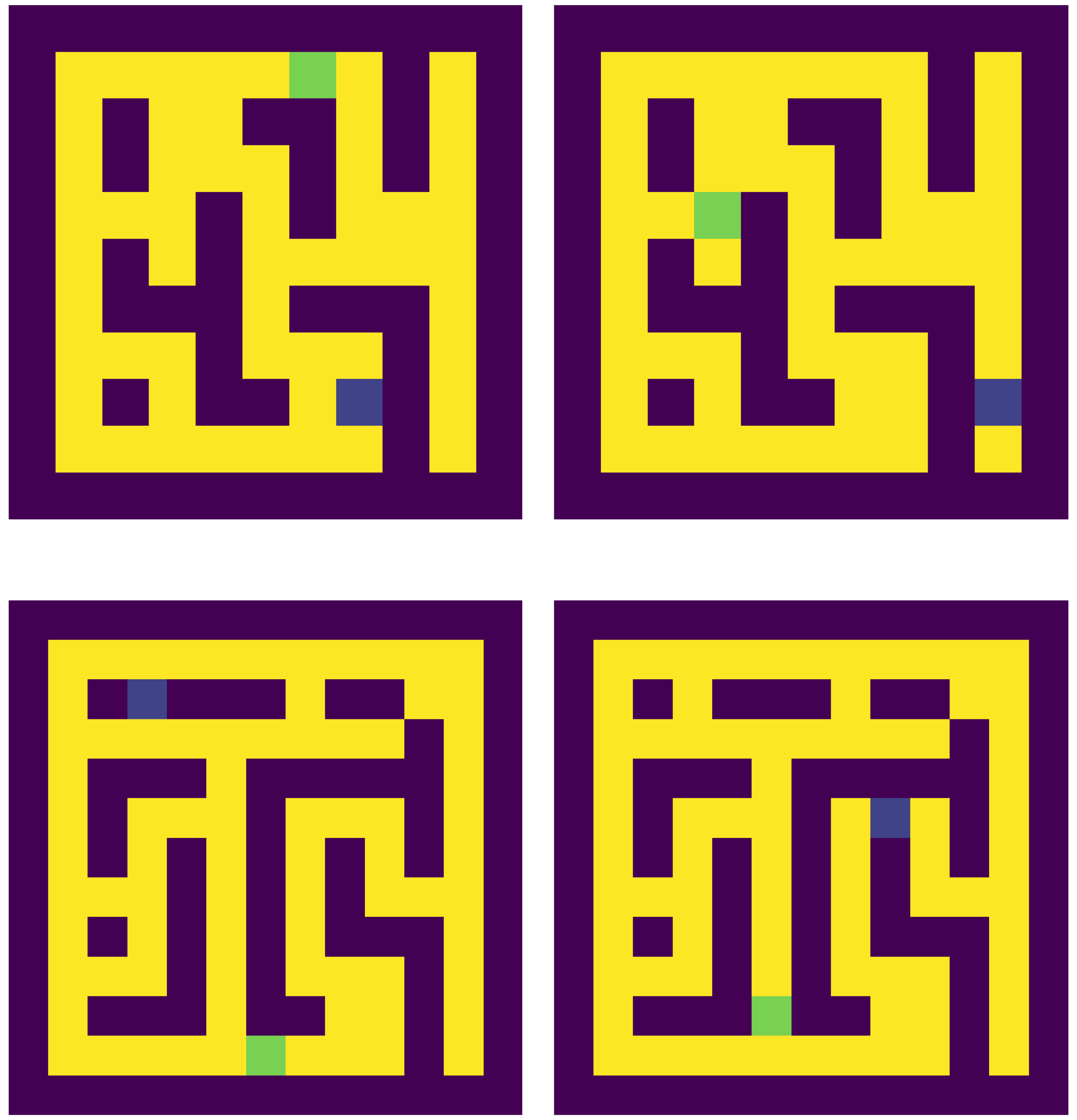
Map-based Navigation

Motivating Example



Map-based *Maze* Navigation

Abstract Map and Goal



Training

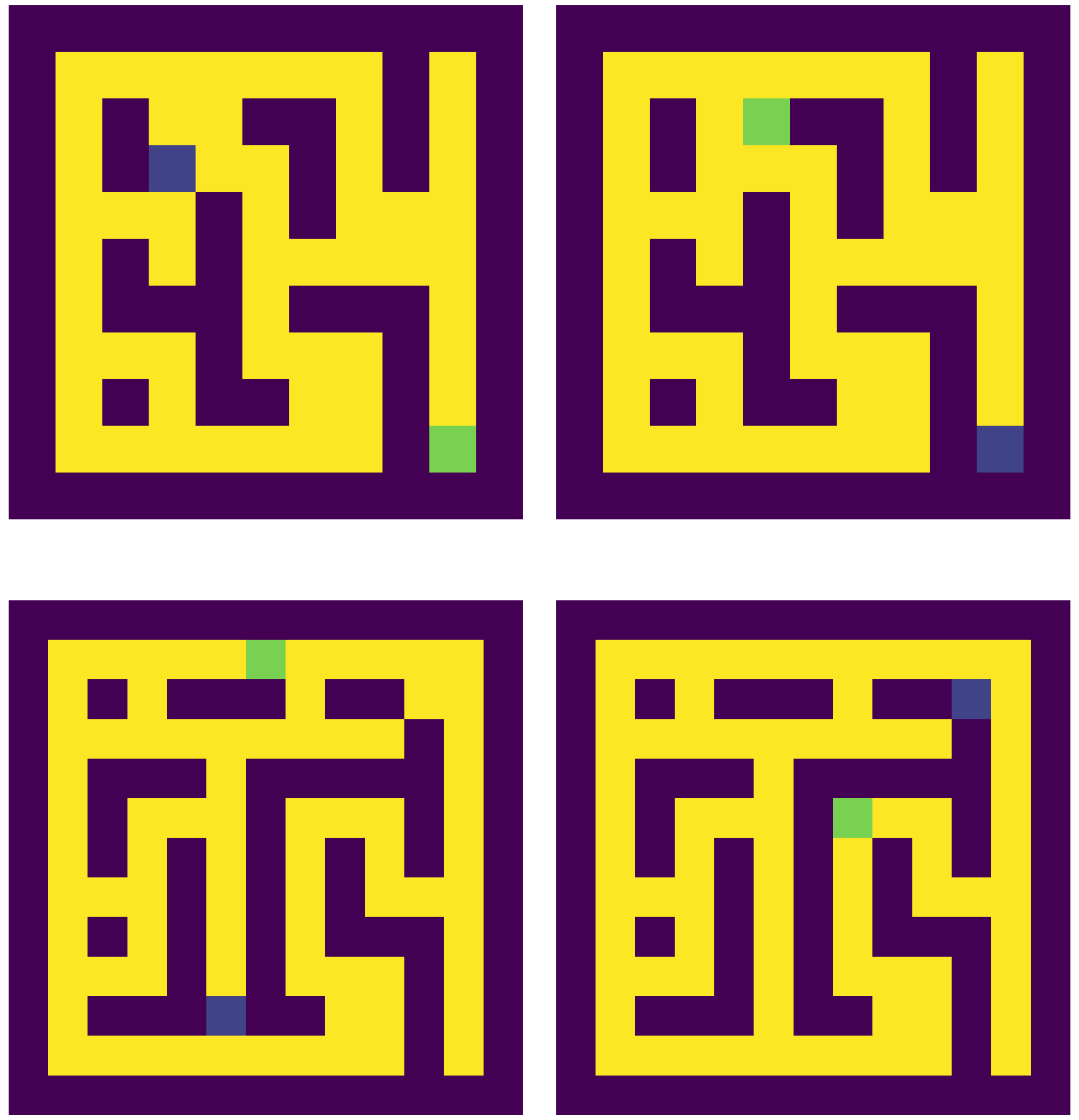
Learning a Planner

Given State, Map and Goal

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

Generalize: Goal-conditioned Policy

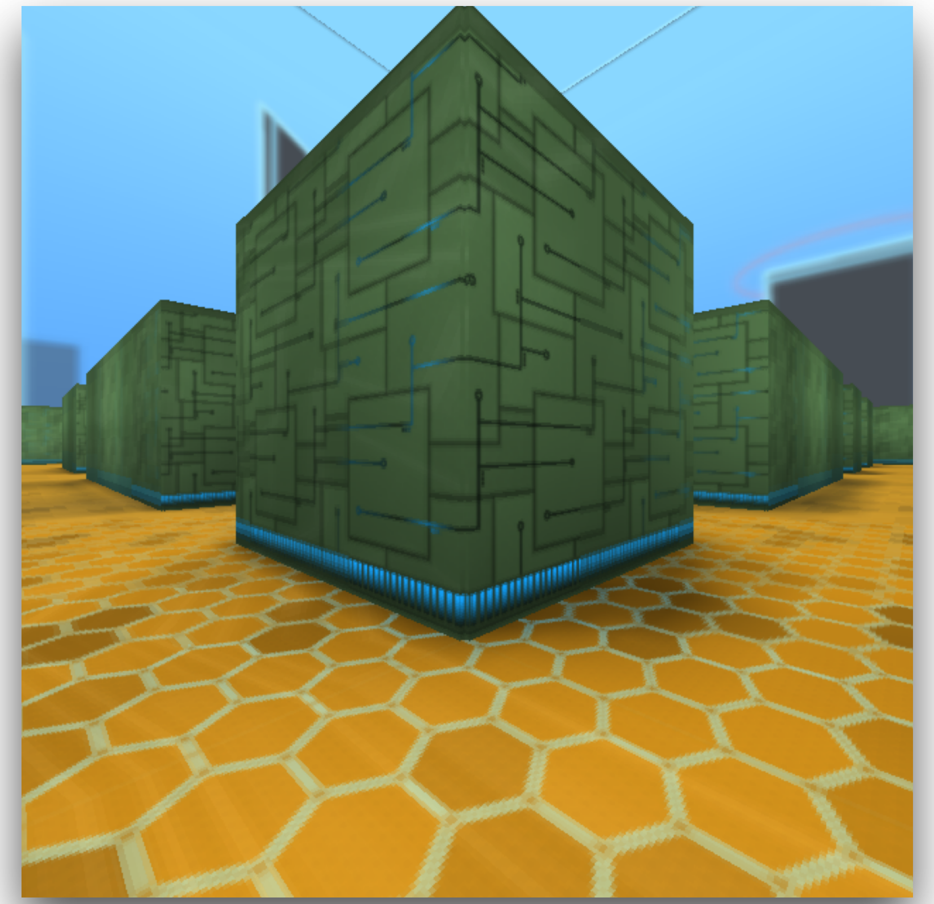
$$\pi(s, [m, g]; \theta)$$



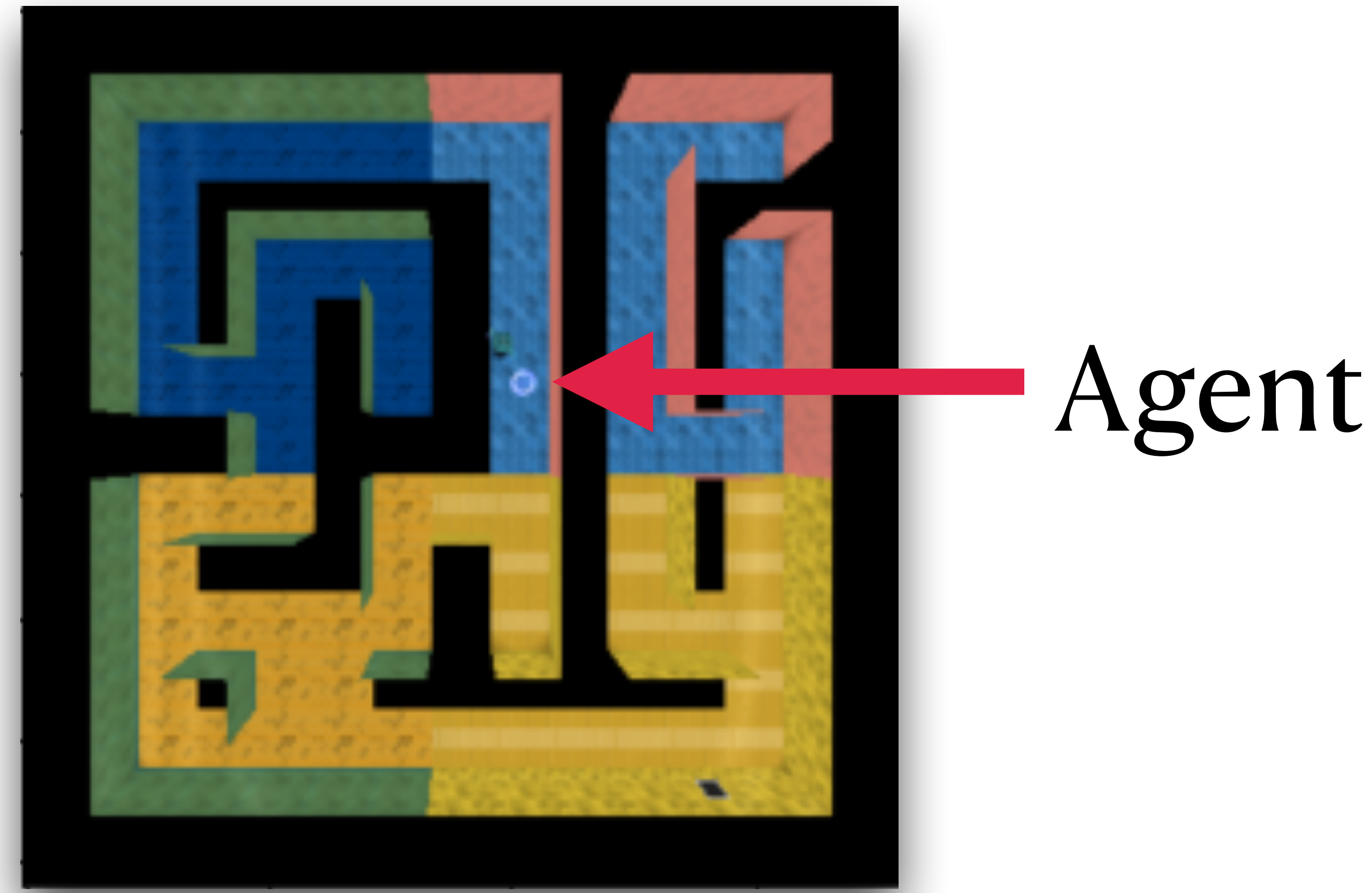
Zero-shot Navigation

Environment

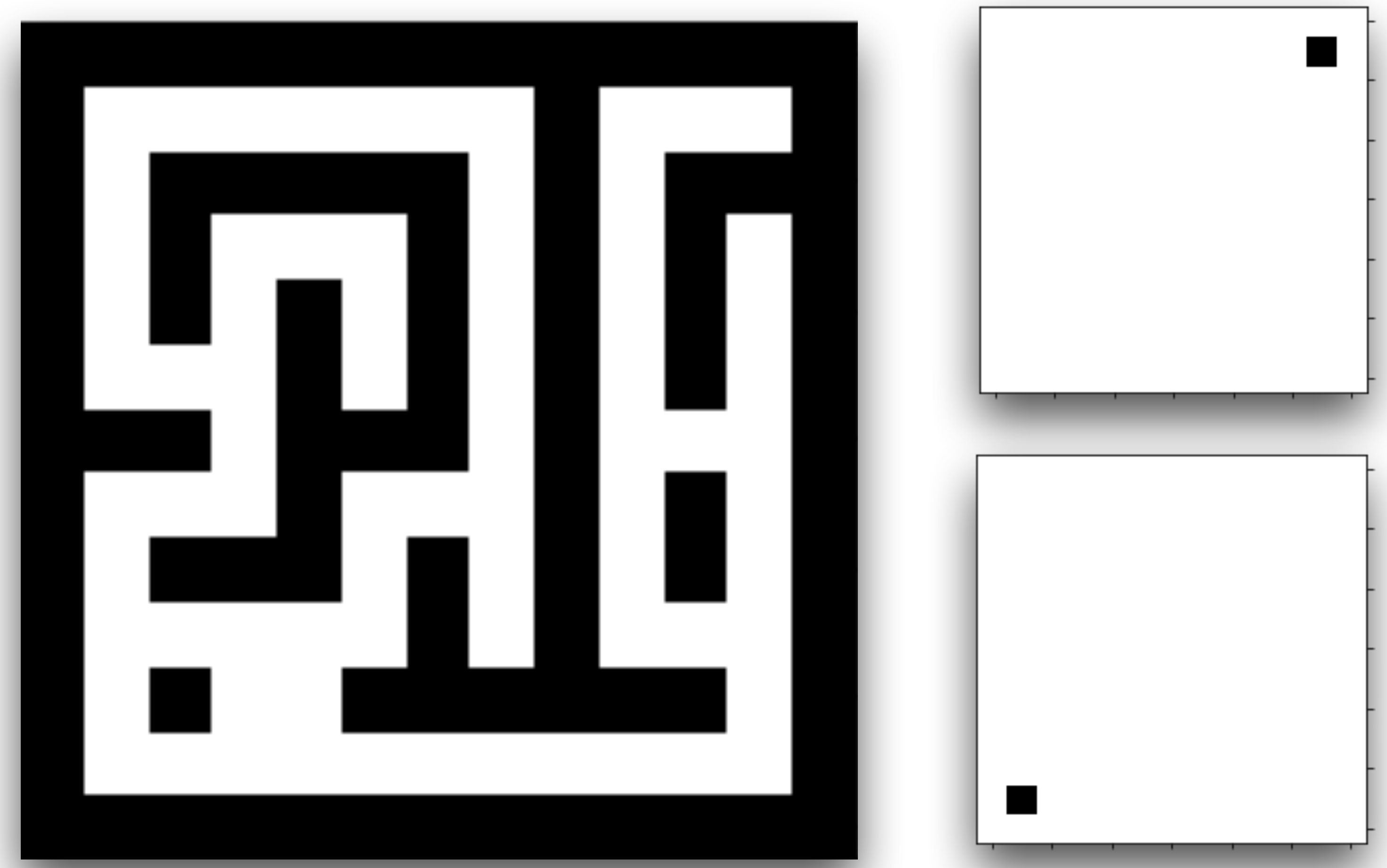
DeepMind Lab



Agent World
Not Input



Top-down View
Not Input

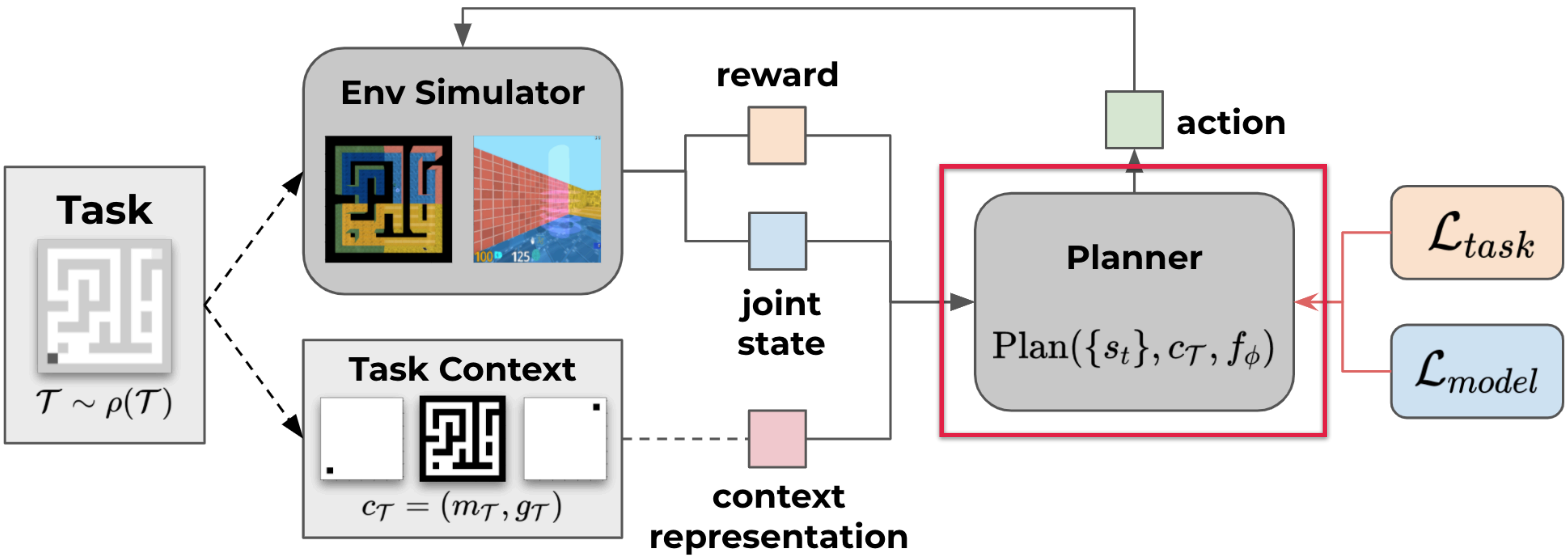


Abstract Map, Goal and Start
Task Input (context)

- 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

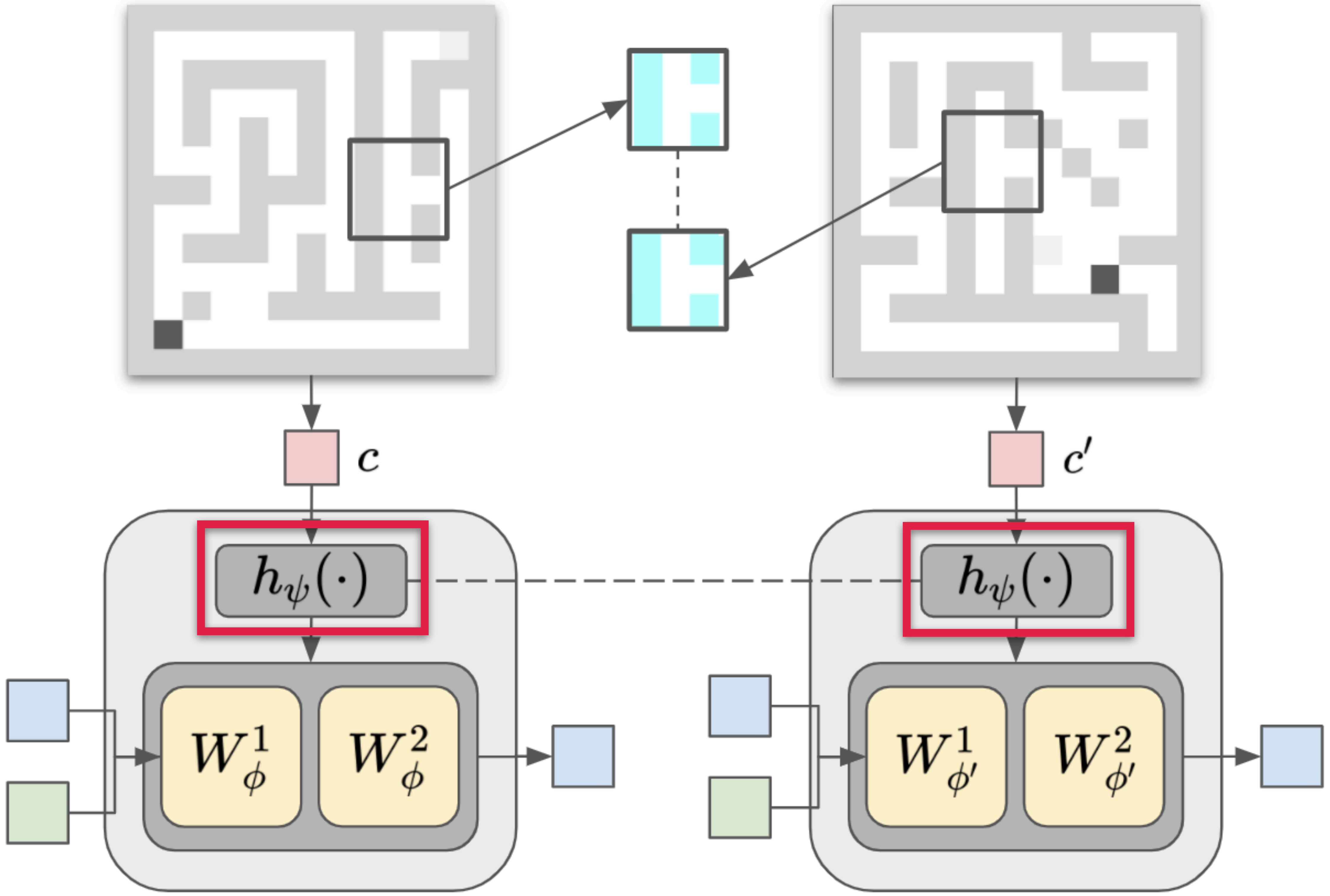
Map-conditioned Model-based Navigator



Task-conditioned Hypermodel

Forward Pass

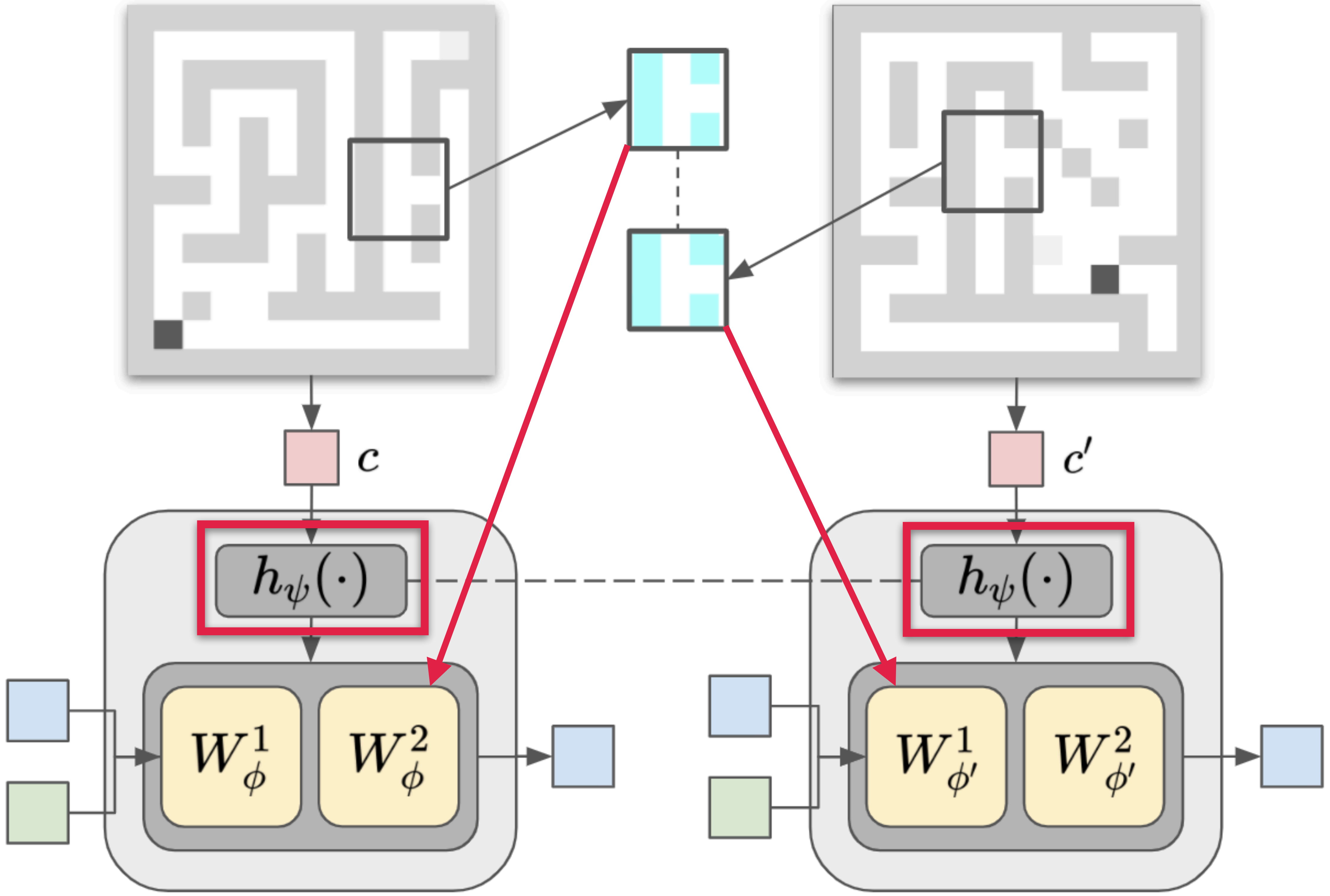
- Hypermodel $h_\psi : c \mapsto \phi, \quad f_\phi : s, a \mapsto s'$
- A hypernetwork h_ψ outputs weights of each transition network f_ϕ
  - The transition "computation" is thus *shared* between *tasks*
- *HyperNetworks* (Ha et al., 2016; von Oswald et al., 2019)



Task-conditioned Hypermodel

Forward Pass

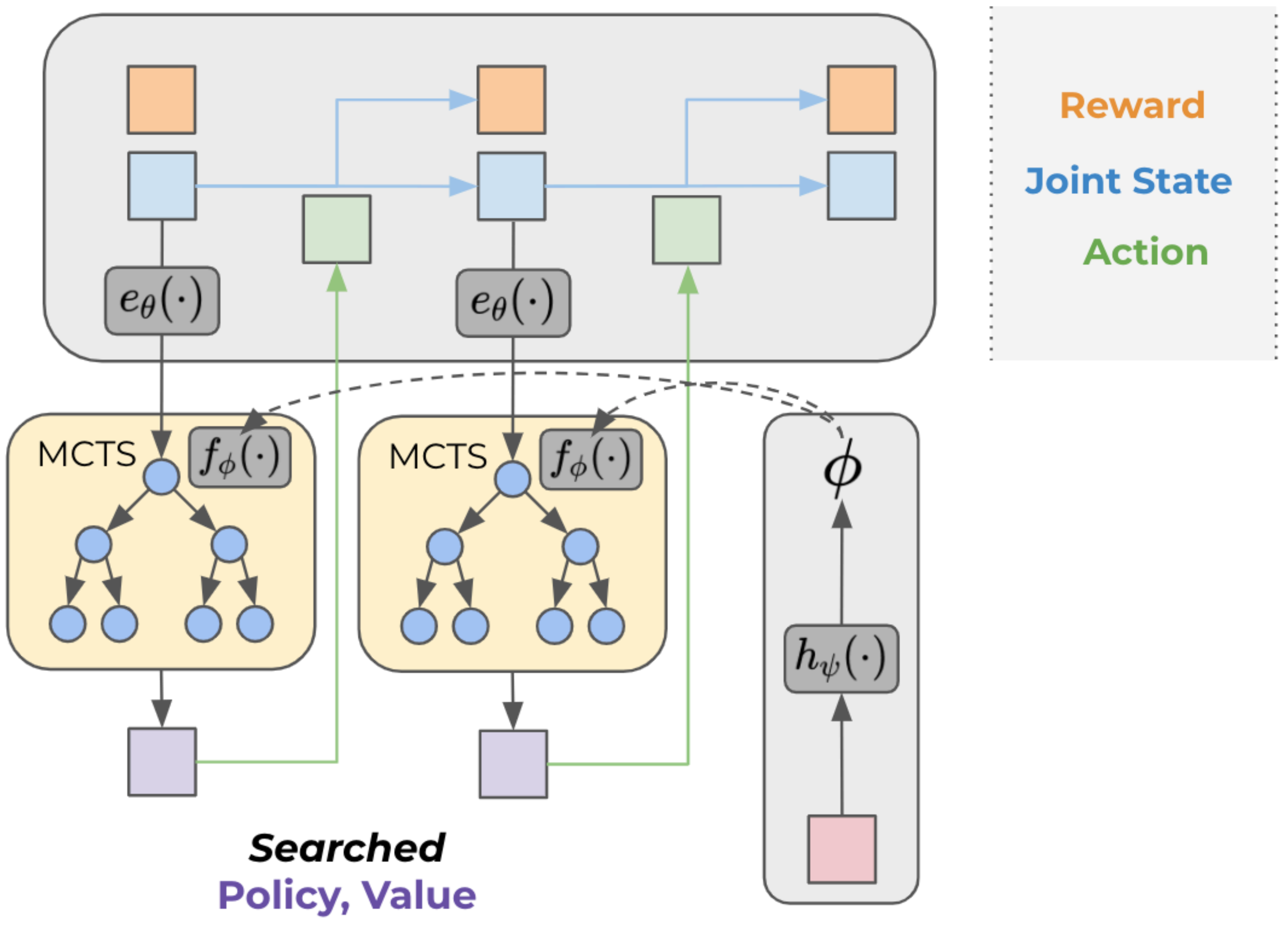
- Hypermodel $h_\psi : c \mapsto \phi, \quad f_\phi : s, a \mapsto s'$
- A hypernetwork h_ψ outputs weights of each transition network f_ϕ
  - The transition “computation” is thus *shared* between *tasks*
- *HyperNetworks* (Ha et al., 2016; von Oswald et al., 2019)



Planning using Learned Hypermodel

Forward Pass / Navigation Computation

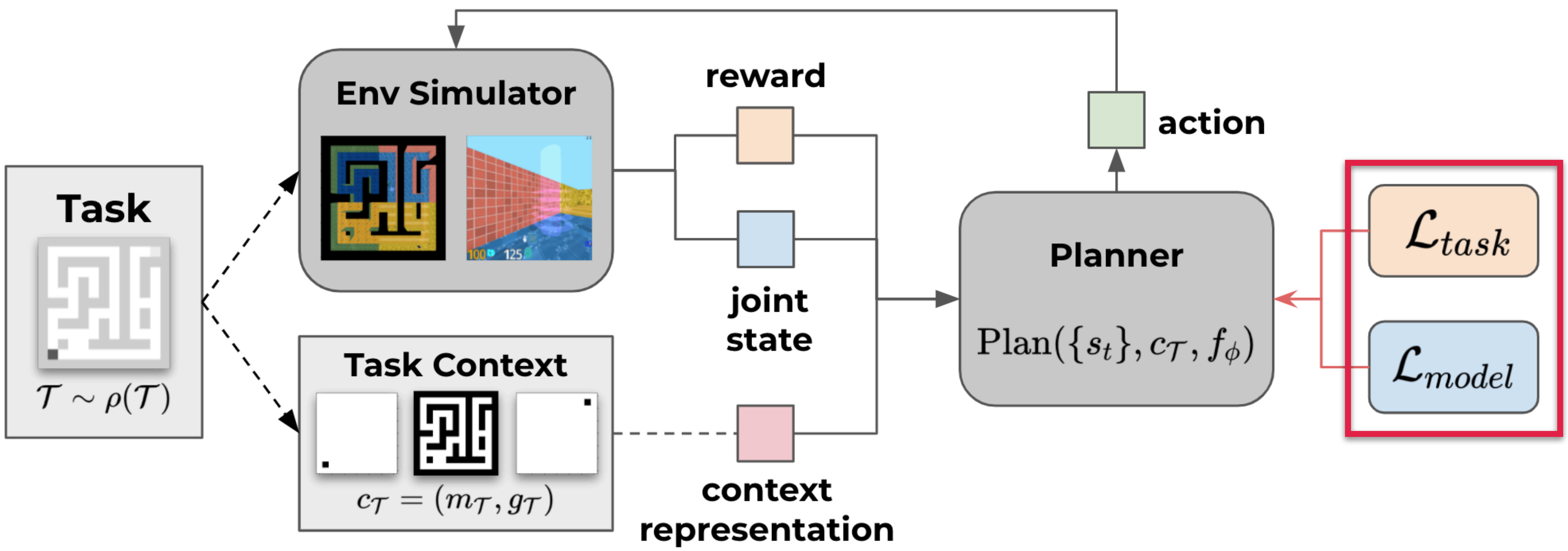
- Planning using *Monte-Carlo tree search*
  - 1. Use hypermodel to predict next states
  - 2. Search policy and value using MCTS
  - 3. Take action sampled from searched policy
  - 4. Repeat, Store $\left(c_{\mathcal{T}}, \left\{s_t, a_t, r_t, s_{t+1}\right\}_t\right)$



Overview

Training Objectives

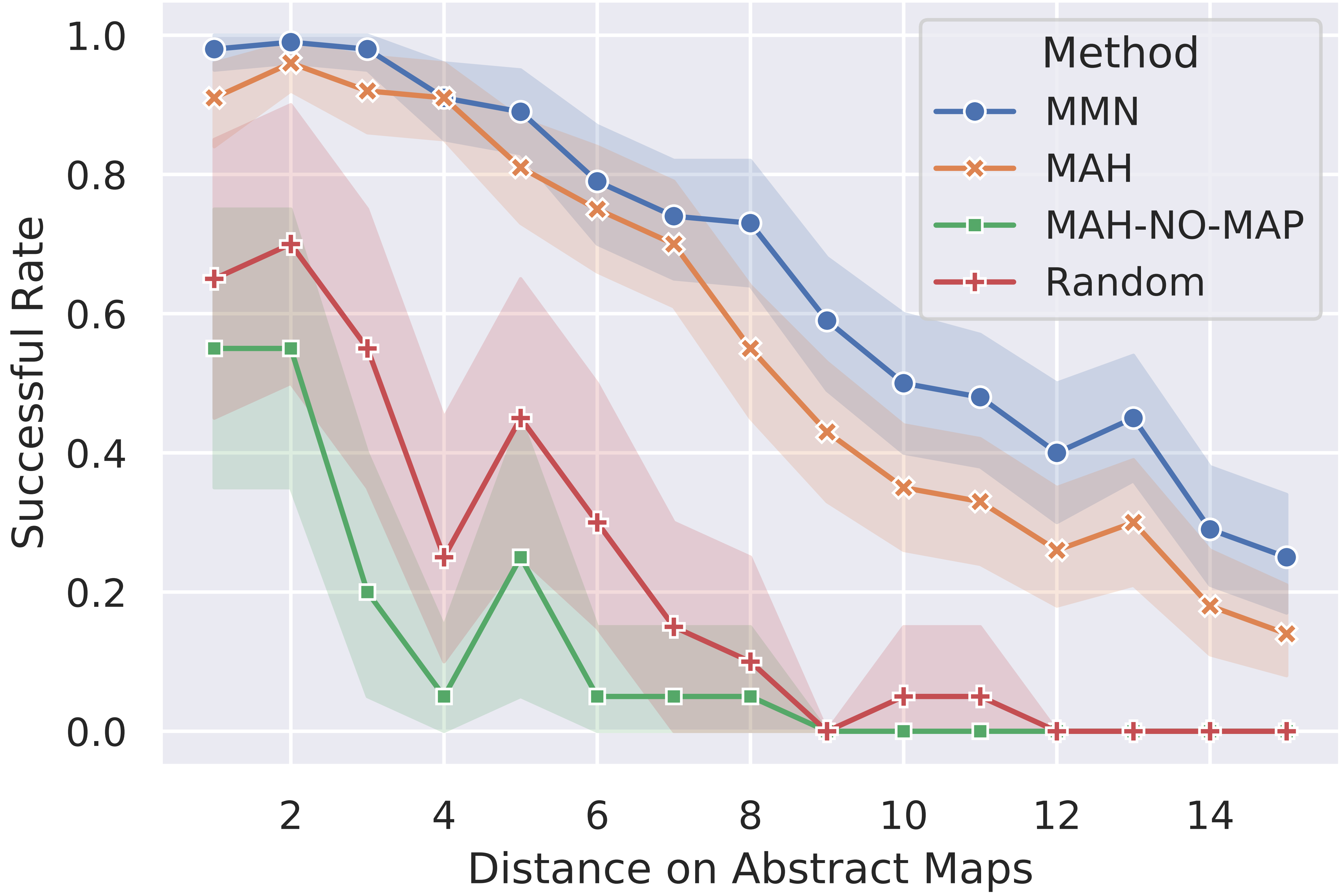
- (1) Task Loss + n-step Goal Relabelling MuZero (Schrittwieser et al., 2019), *HER* (Andrychowicz et al., 2017)
- (2) Auxiliary Model Loss: minimizing hypermodel prediction loss

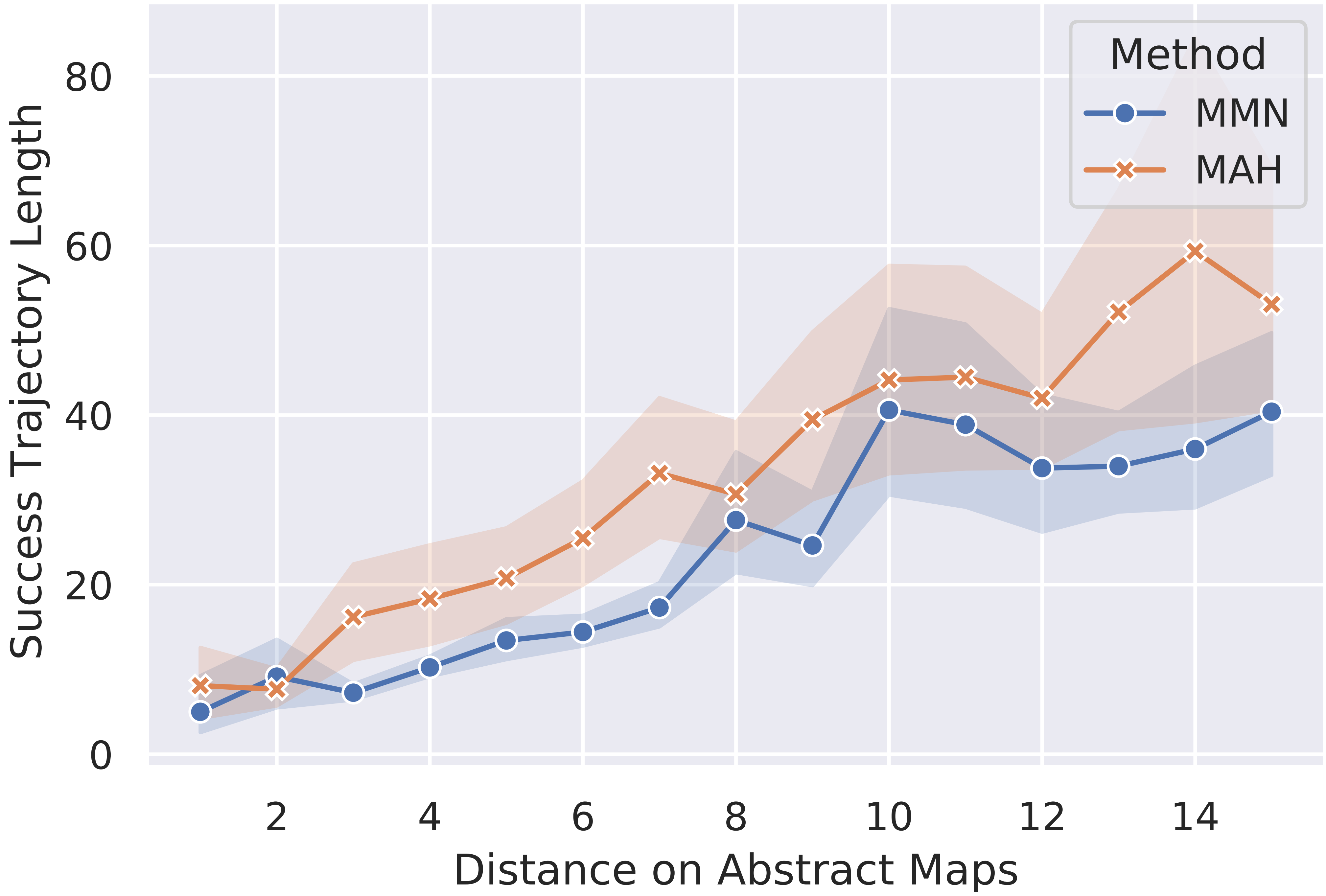


Zero-shot Navigation on Novel Maps

Key Results

- Evaluation on 20 unseen 13×13 maps with 5 goals in distance $[1,15]$ for each map
- MMN = Map-conditioned Model-based Navigator, MAH = Map-conditioned Ape-X DQN with HER





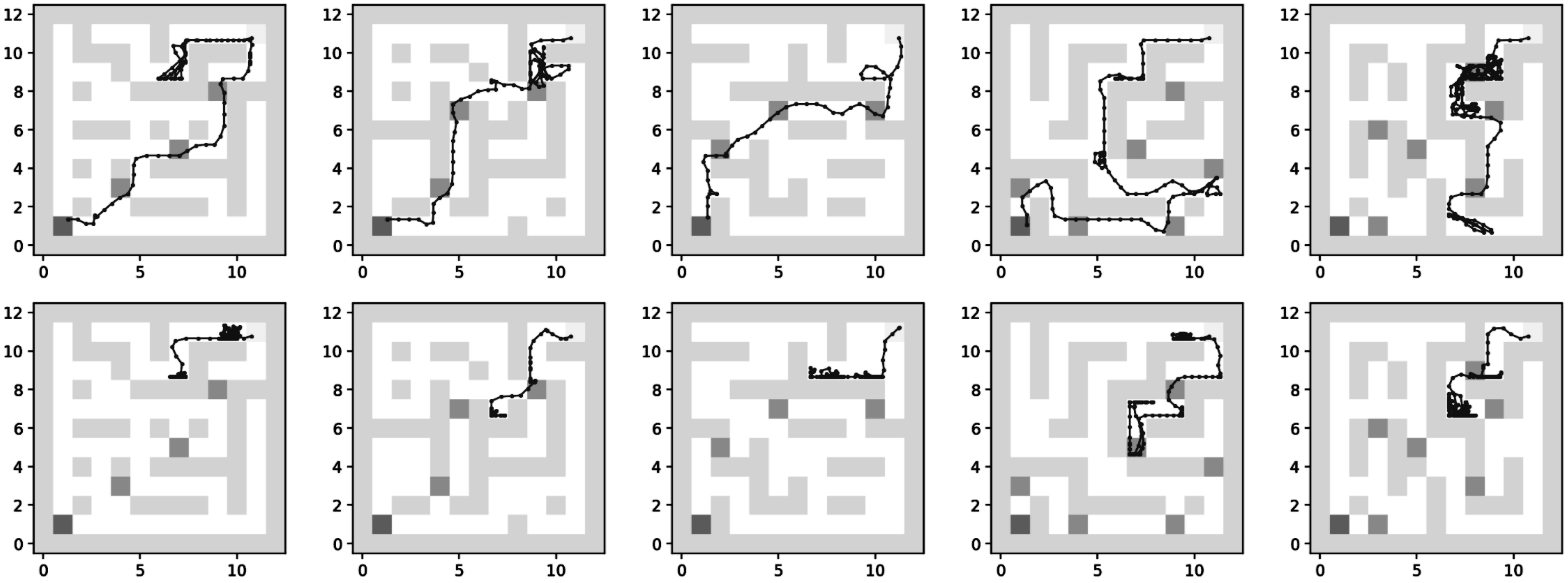
Zero-shot Hierarchical Navigation

Key Results

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)

Thank You