HELSA: Hierarchical Reinforcement Learning with Spatiotemporal Abstraction for Large-Scale Multi-Agent Path Finding

Zhaoyi Song[1], Rongqing Zhang[1*], and Xiang Cheng[2]

[1] NaMI Lab, School of Software Engineering, Tongji University, P.R. China

[2] School of Electronics Engineering and Computing Science, Peking University, P.R. China

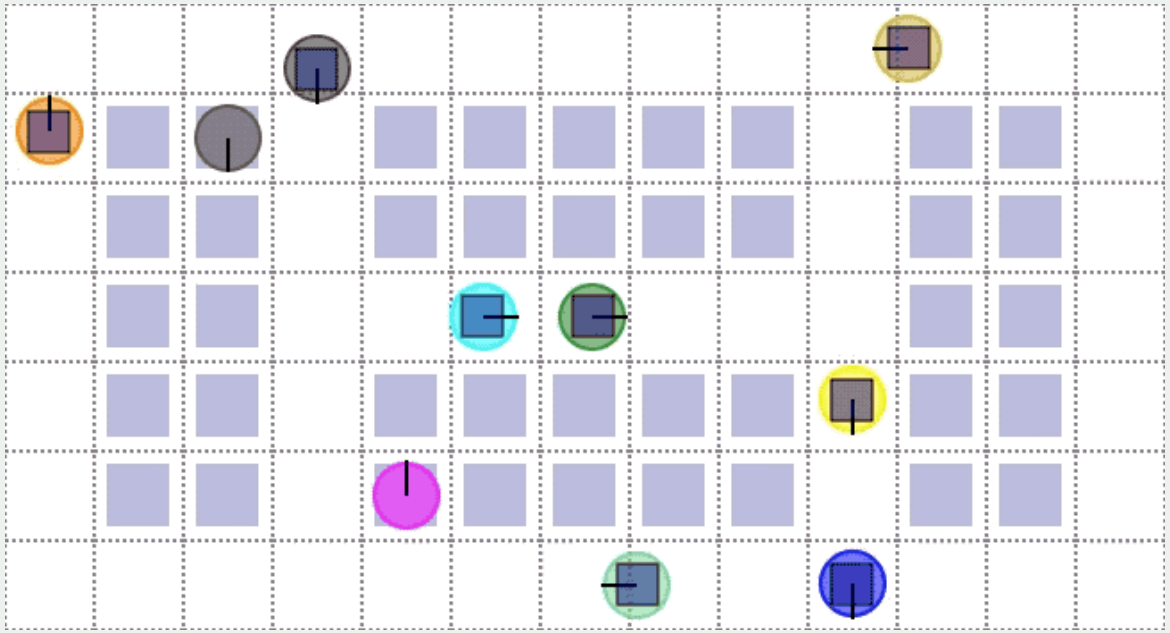
NaMI Network and Machine Intelligence Lab

同濟大學 TONGJI UNIVERSITY

北京大學 PEKING UNIVERSITY

Background

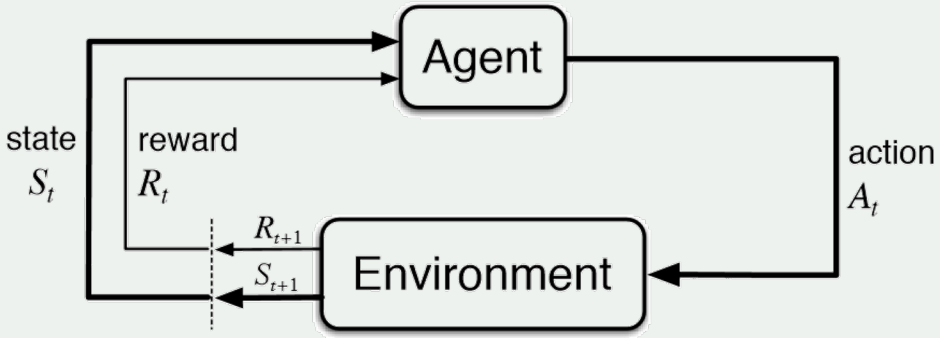
The Multi-Agent Pathfinding (MAPF) Problems

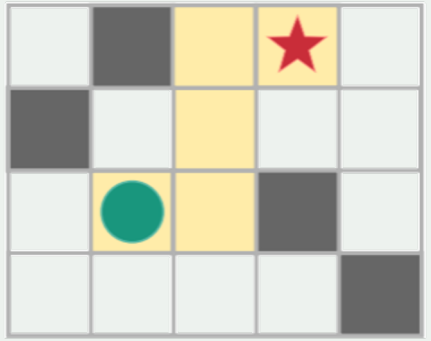


* Images retrieved from https://www.cs.sfu.ca/~hangma/, Amazon Robotics and SuperStock.

Key Challenges

Sparse Rewards in Large-scale Reinforcement Learning



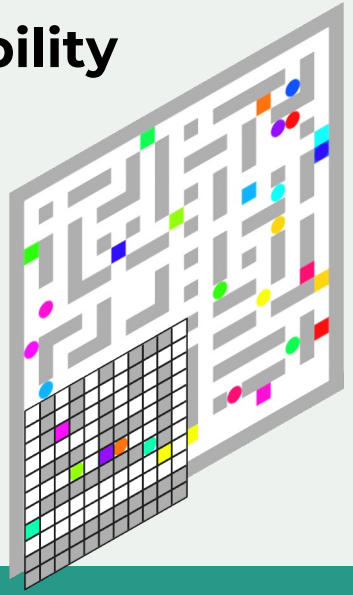


$$R(\boldsymbol{a}) = \mathbb{1}[\text{Execute}(\cdot, \boldsymbol{a}) = \star]$$

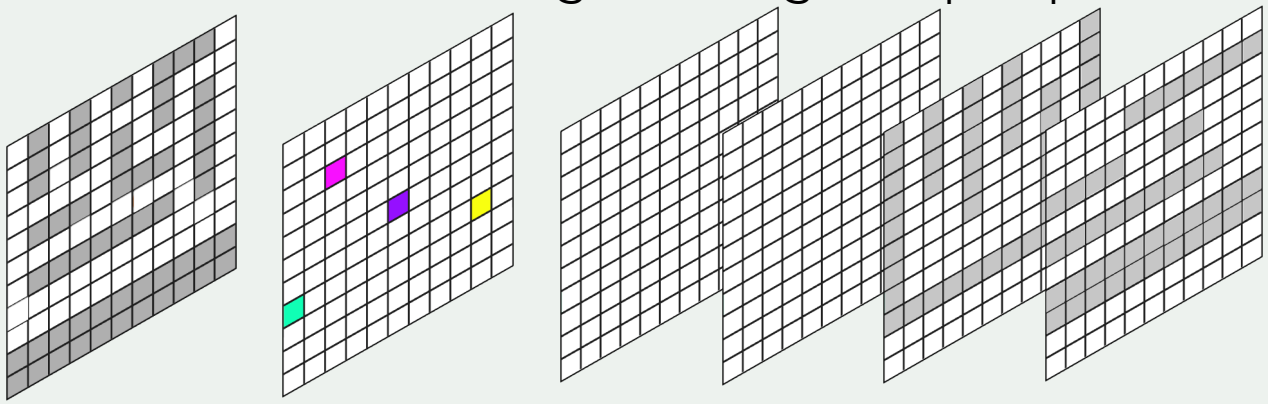
$$\boldsymbol{a}_1 = (\rightarrow, \uparrow, \uparrow, \rightarrow)$$
$$\boldsymbol{a}_2 = (\leftarrow, \rightarrow, \rightarrow, \uparrow, \uparrow, \rightarrow)$$
$$\boldsymbol{a}_3 = (\uparrow, \rightarrow, \rightarrow, \uparrow)$$
$$R(\boldsymbol{a}_1) = R(\boldsymbol{a}_2) = R(\boldsymbol{a}_3) = 1$$

Partial Observability

Grid world



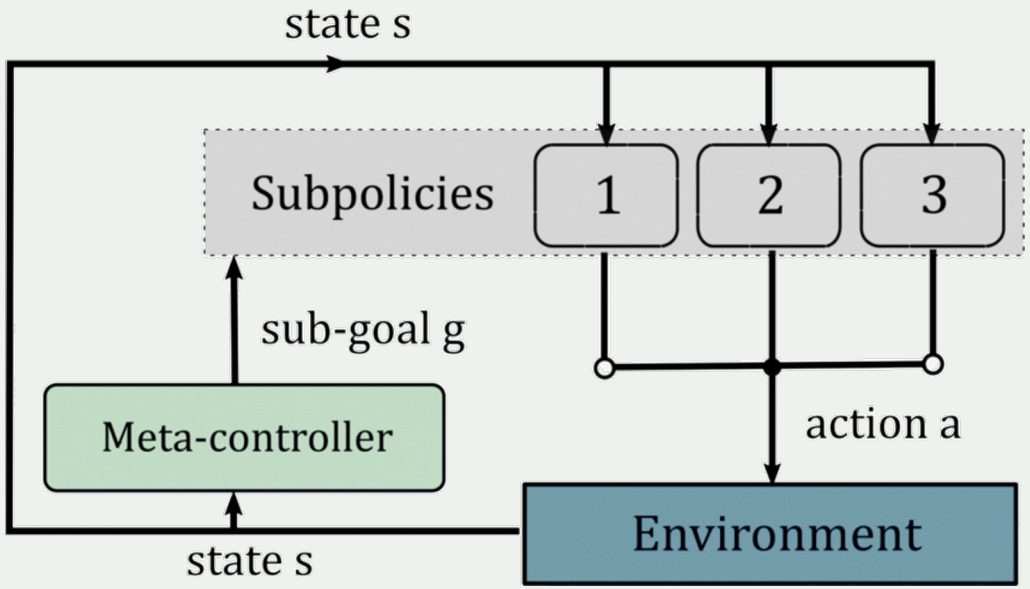
Observation encodings from agent's perspective



Obstacles Agents Heuristic Maps

Overview

Solution: Hierarchical Reinforcement Learning



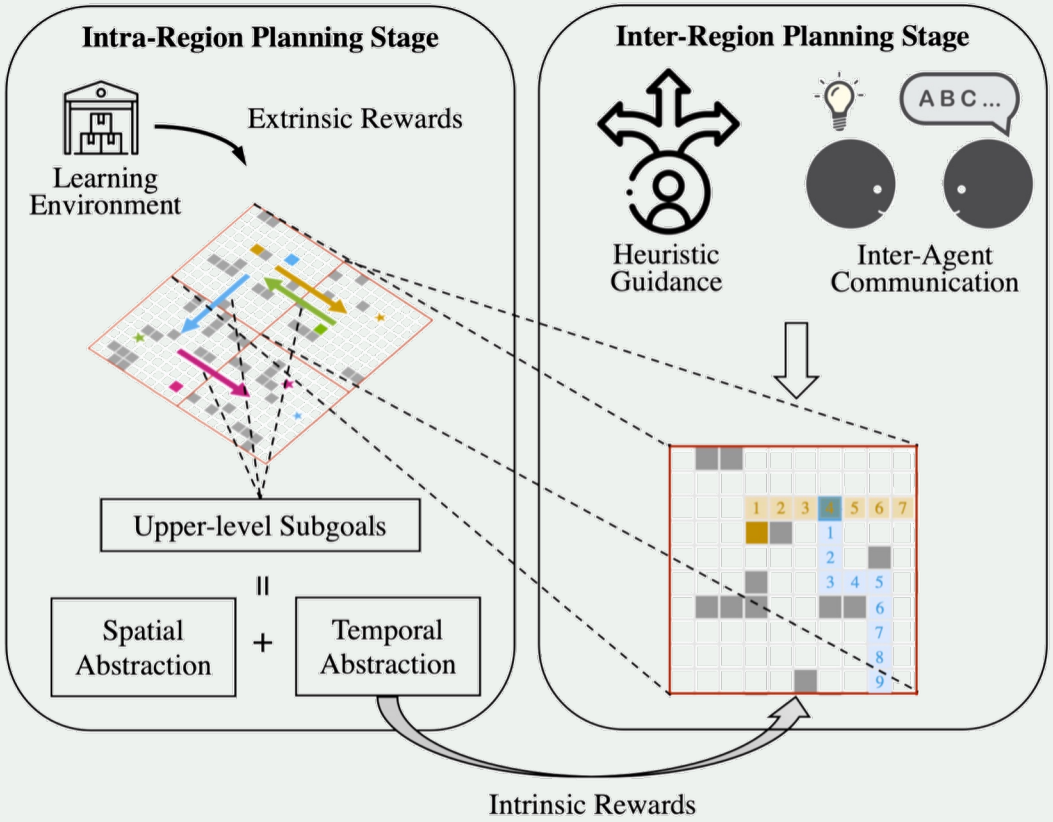
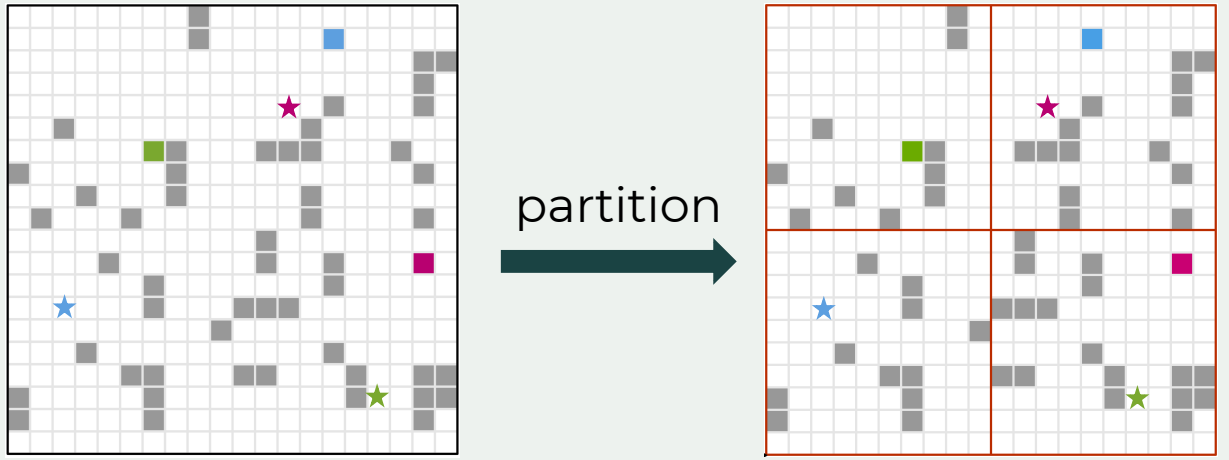


Fig. 1: Illustration of the HELSA framework

The Proposed Framework

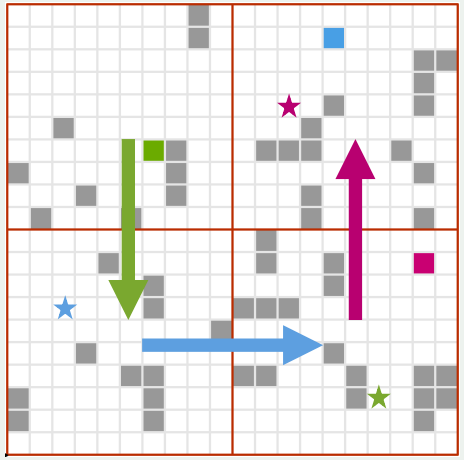
01 Map Partition

Divide the space into a series of **regions** based on hyperparameters.



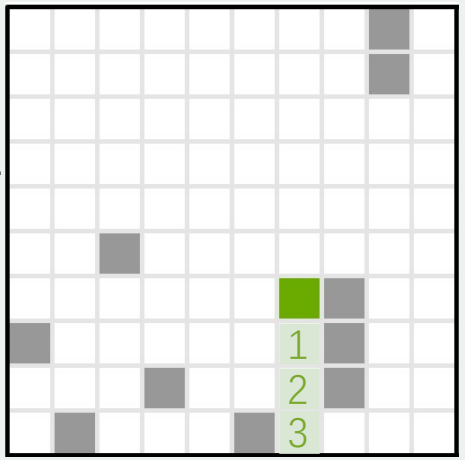
02 Temporal Abstraction

Decompose long-term tasks on a temporal scale into **short-term tasks**



03 Subtask Solver

Solve constrained multi-goal multi-agent pathfinding problems



Lower-Level Controller

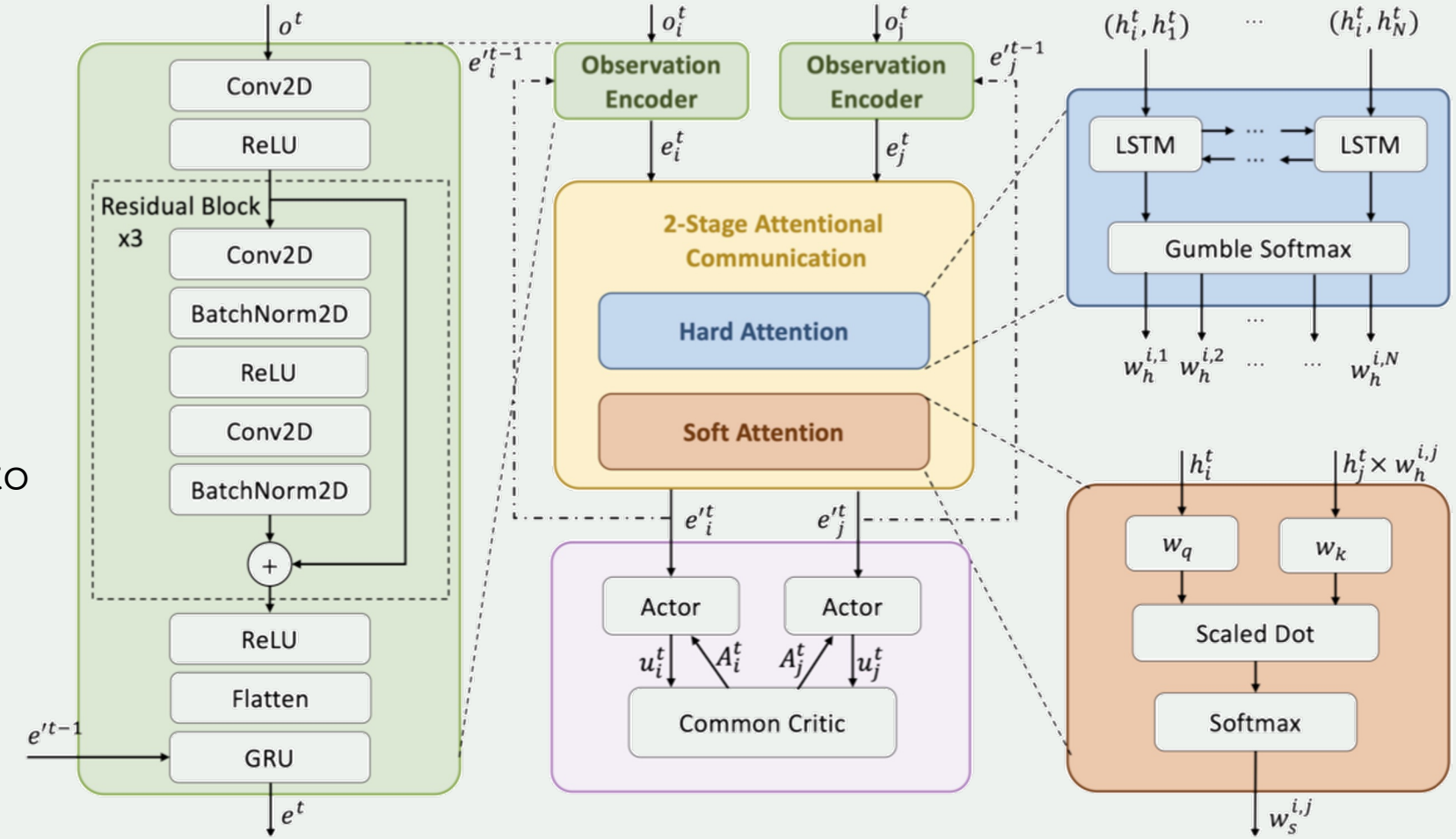
Observation Encoder

- Agents positions
- Obstacle positions
- Heuristic encodings
- Last-step messages

Communication Block

- Hard attention mechanism to filter out irrelevant agents
- Soft attention to calculate relative importance

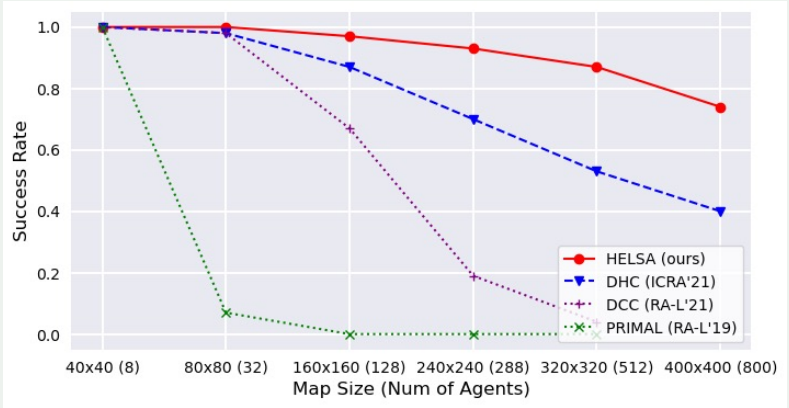
Action Network



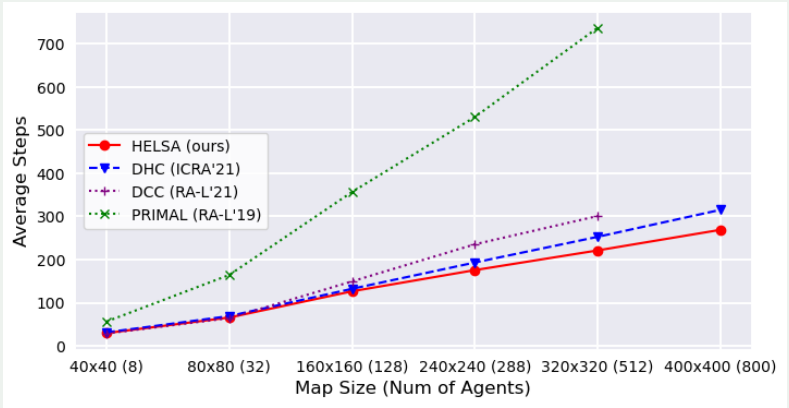
Empirical Analysis on Randomly Generated Datasets

- **Evaluation Metrics**
  - Success Rates
  - Average Steps
  - Makespans
  - Collisions with Agents
  - Collisions with Obstacles

- **Success Rates**



- **Average Steps**



Model	8 agents, 40-sized map, 0.2 density					32 agents, 80-sized map, 0.2 density					128 agents, 160-sized map, 0.2 density				
	SR ↑	AS ↓	MS ↓	CA ↓	CO ↓	SR ↑	AS ↓	MS ↓	CA ↓	CO ↓	SR ↑	AS ↓	MS ↓	CA ↓	CO ↓
PRIMAL [4]	**1.0**	56.49	98.90	0.42	**0.0**	0.88	164.39	305.73	4.12	**0.0**	0.07	356.51	1007.08	113.06	4.27
DHC [6]	**1.0**	31.40	55.77	0.38	**0.0**	0.98	69.18	139.77	3.20	**0.0**	0.87	132.31	399.19	29.38	0.06
DCC [7]	**1.0**	**28.84**	**50.49**	0.40	**0.0**	0.98	**64.47**	**134.34**	5.91	0.01	0.67	149.50	567.41	37.48	**0.0**
HELSA	**1.0**	29.71	52.29	**0.21**	**0.0**	**1.0**	65.85	136.17	**0.54**	**0.0**	**0.97**	**126.51**	**296.14**	**3.69**	**0.0**

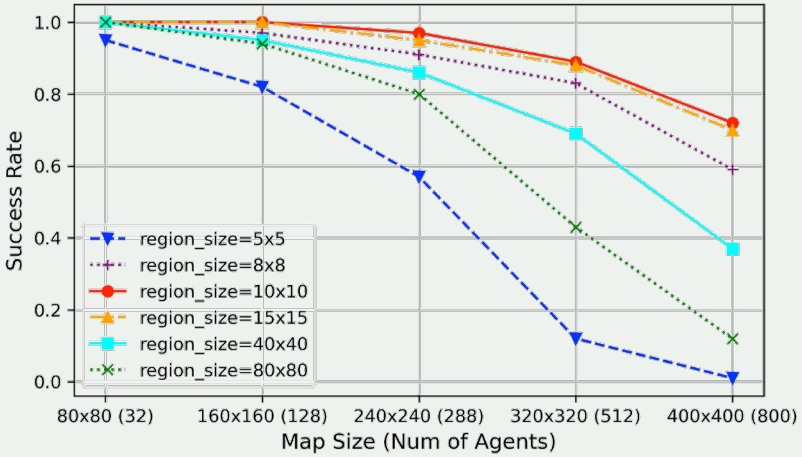
Model	288 agents, 240-sized map, 0.2 density					512 agents, 320-sized map, 0.2 density					800 agents, 320-sized map, 0.2 density				
	SR ↑	AS ↓	MS ↓	CA ↓	CO ↓	SR ↑	AS ↓	MS ↓	CA ↓	CO ↓	SR ↑	AS ↓	MS ↓	CA ↓	CO ↓
PRIMAL [4]	0.0	530.06	1536.0	593.59	34.48	0.0	736.50	2048.0	1498.20	173.49	-	-	-	-	-
DHC [6]	0.70	193.13	804.55	99.52	**0.01**	0.53	252.62	1304.48	236.22	0.30	0.40	315.08	1906.36	468.61	0.71
DCC [7]	0.19	235.32	1375.04	151.88	12.97	0.04	300.78	2020.76	423.40	57.41	-	-	-	-	-
HELSA	**0.93**	**175.56**	**629.58**	**49.41**	0.03	**0.87**	**221.17**	**935.99**	**101.78**	**0.04**	**0.74**	**268.83**	**211.15**	**269.67**	**0.37**

Empirical Analysis

How does the partitioning granularity effect the performance of HELSA?

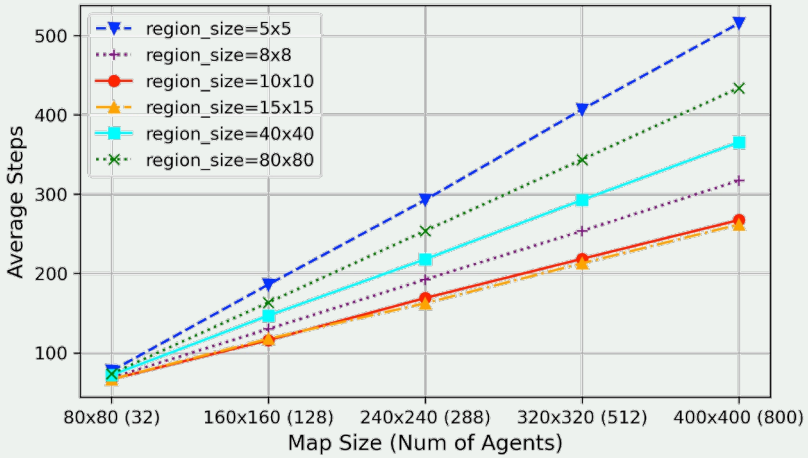
- **Success Rates**

Perform best at approximately 10x10 – 15x15 region size



- **Average Steps**

Perform best at approximately 10x10 – 15x15 region size



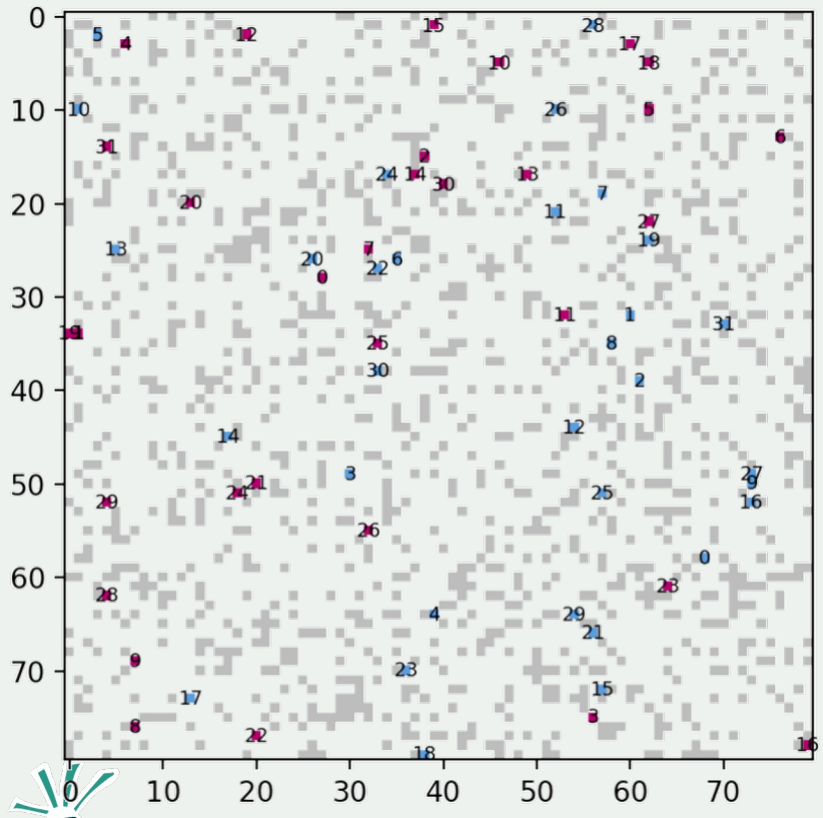
Does the two-stage attention communication lead to better coordination?

TABLE III: Evalutaion of the adopted lower-level controller with other ablations in terms of success rates and average steps.

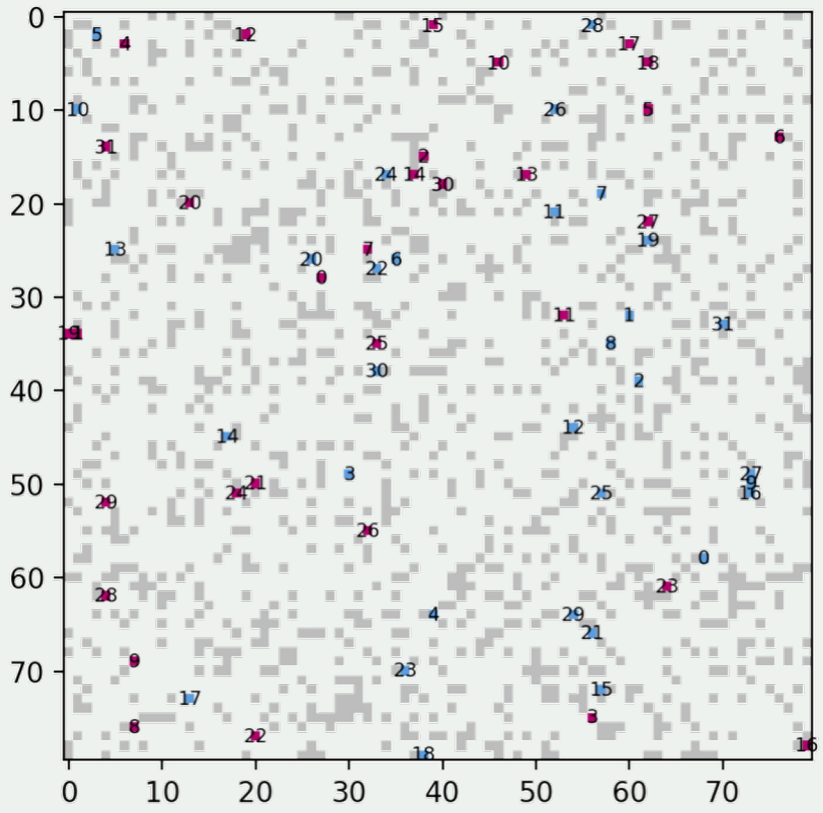
Method	w/ hierarchy?	80-sized map		160-sized map		240-sized map		320-sized map		400-sized map		Avg.	
		SR ↑	AS ↓	SR ↑	AS ↓	SR ↑	AS ↓	SR ↑	AS ↓	SR ↑	AS ↓	SR ↑	AS ↓
COMA+Comm	✓	**1.0**	**65.85**	**0.97**	**126.51**	**0.93**	**175.56**	**0.87**	221.17	0.74	268.83	**0.90**	**171.58**
+Attention		0.98	67.25	0.76	141.95	0.41	219.03	0.07	287.99	0.0	347.63	0.44	212.77
COMA+Comm	✓	**1.0**	66.78	0.95	130.20	0.90	182.13	0.86	**219.75**	**0.77**	**245.00**	0.90	172.97
		0.98	69.89	0.72	147.77	0.35	233.98	0.09	311.93	0.0	387.54	0.43	230.22
COMA	✓	0.95	96.30	0.83	193.39	0.44	323.95	0.04	433.73	0.0	615.19	0.45	332.51
		0.90	139.13	0.43	248.67	0.12	477.53	0.01	633.55	0.0	883.10	0.29	476.40

Simulation Results

HELSA w/ hierarchical controllers



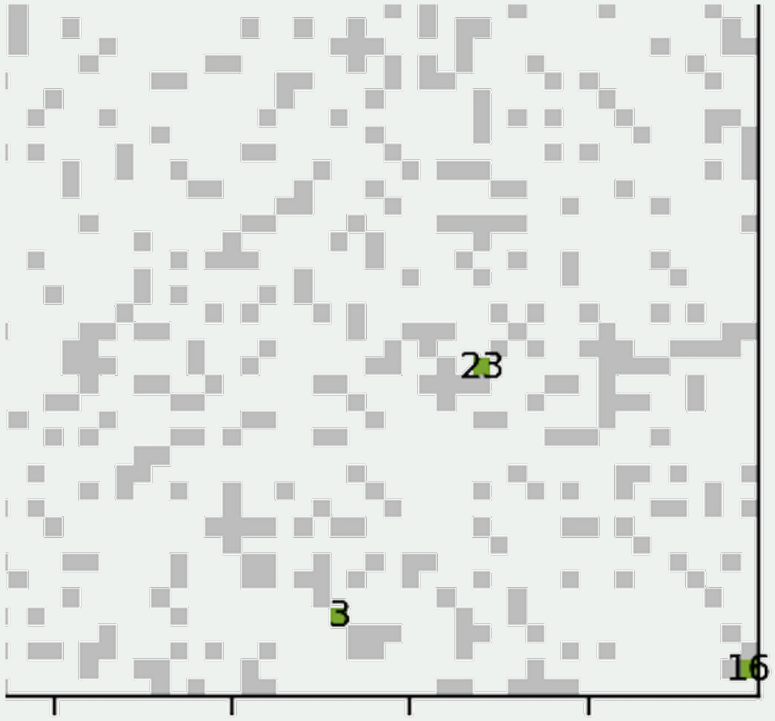
HELSA w/ only low-level controllers



80x80-sized map, 32 agents, 20% obstacle density

Simulation Results

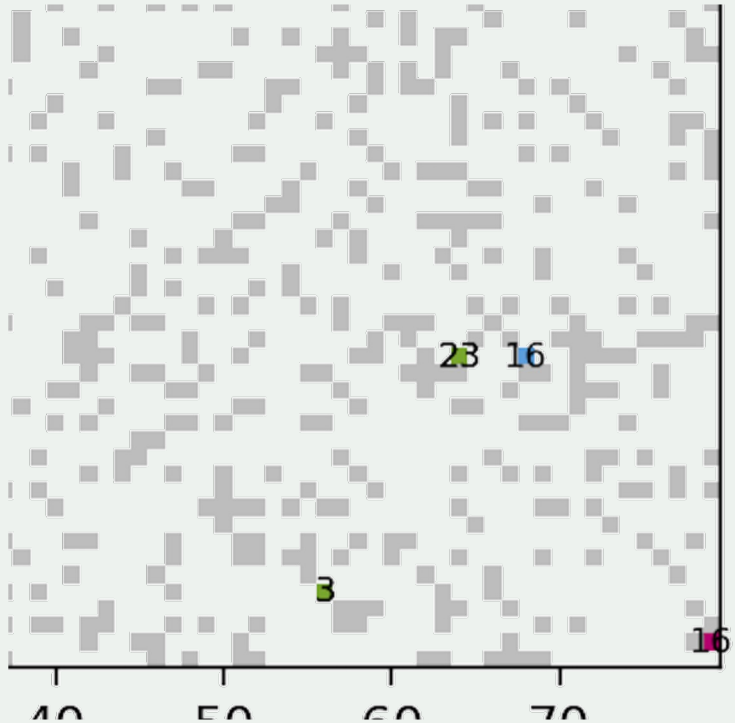
HELSA w/ hierarchical controllers



GOOD Case!

Succeeded at 122th step

HELSA w/ only low-level controllers



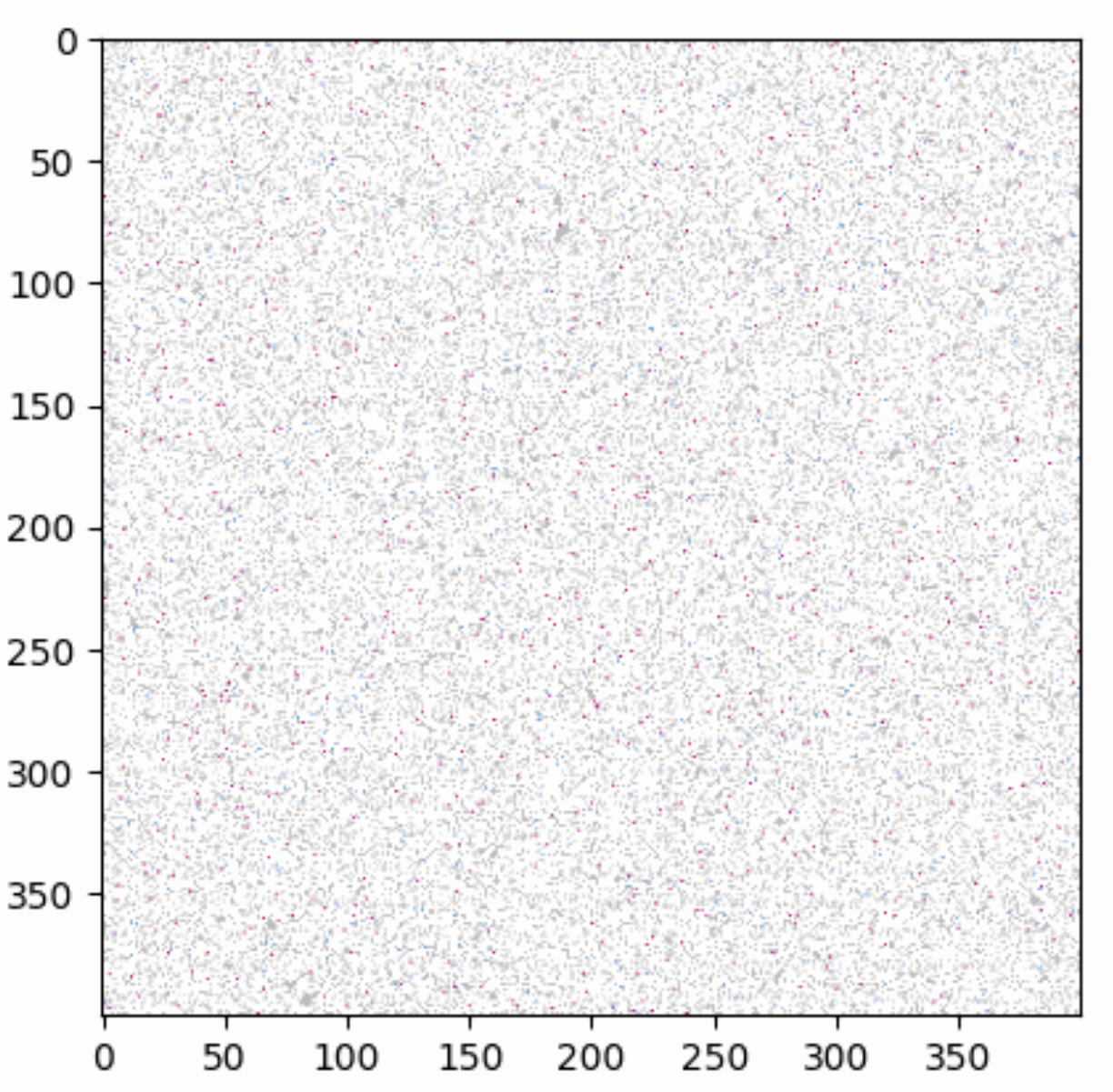
BAD Case!

Agent 16 failed its job

Simulation Results

Showcase of a challenging 400x400-sized scenario, with 800 agents and 32,000 obstacles

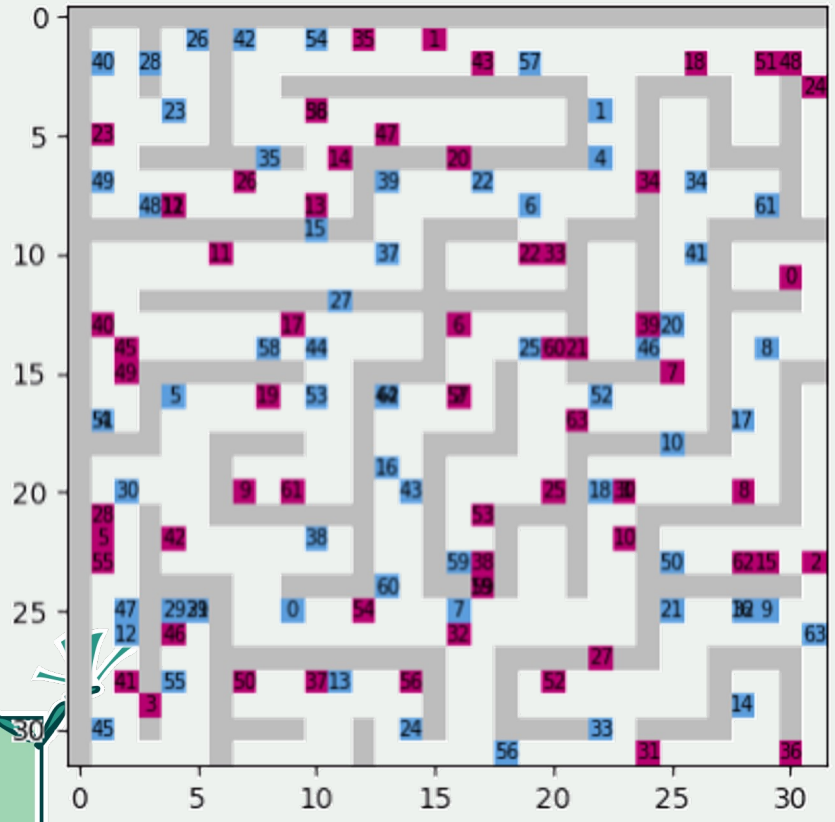
All agents reach their destinations at the 746th timestep



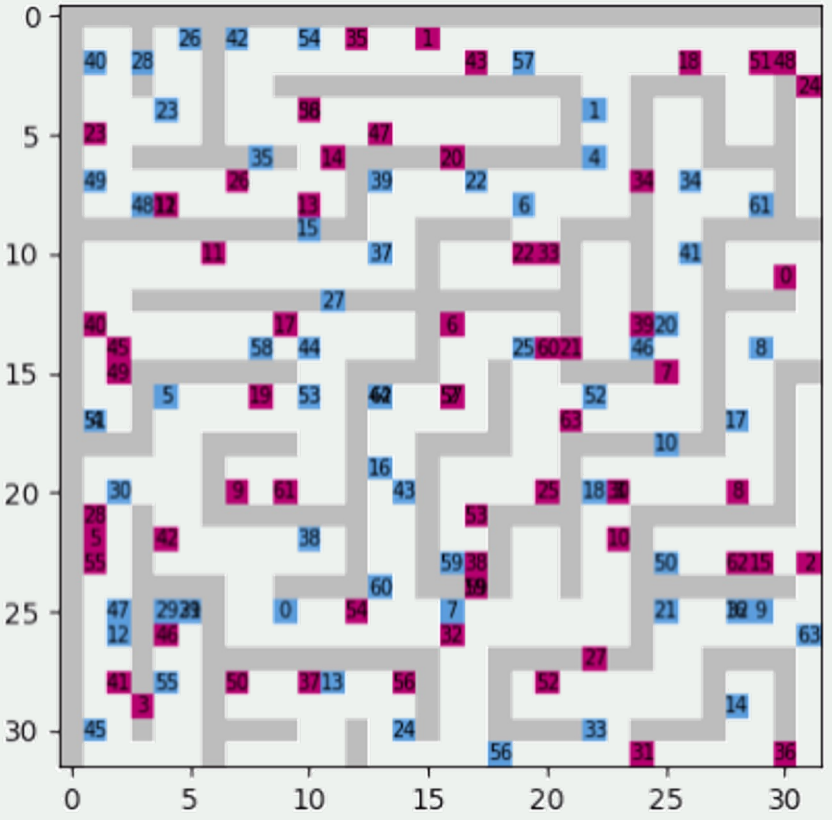
Simulation Results

maze-32-32-2 from mapf.info (SoCS'19, a challenging benchmark), 64 agents

HELSA

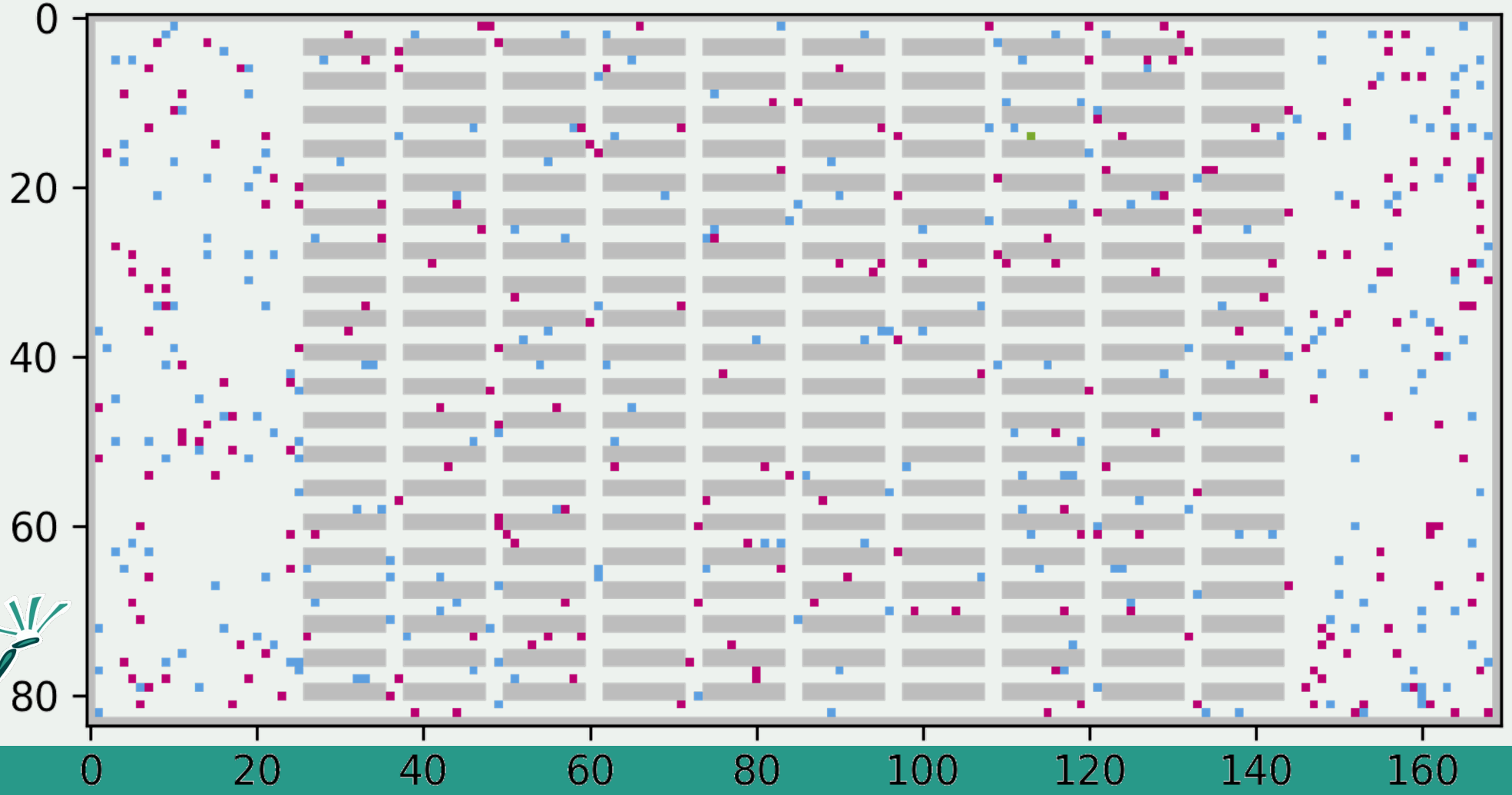


DCC (R-AL' 21)



Simulation Results

warehouse-10-20-10-2-2 from mapf.info (SoCS'19), 256 agents



Thanks for Listening!

HELSA: Hierarchical Reinforcement Learning with Spatiotemporal Abstraction for Large-Scale Multi-Agent Path Finding

Zhaoyi Song[1], Rongqing Zhang[1*], and Xiang Cheng[2]

[1] NaMI Lab, School of Software Engineering, Tongji University, P.R. China

[2] School of Electronics Engineering and Computing Science, Peking University, P.R. China