Deep Reinforcement Learning using Memory-based Approaches

Dai Shen
Stanford University

Apurva Pancholi
Omnisenz Inc

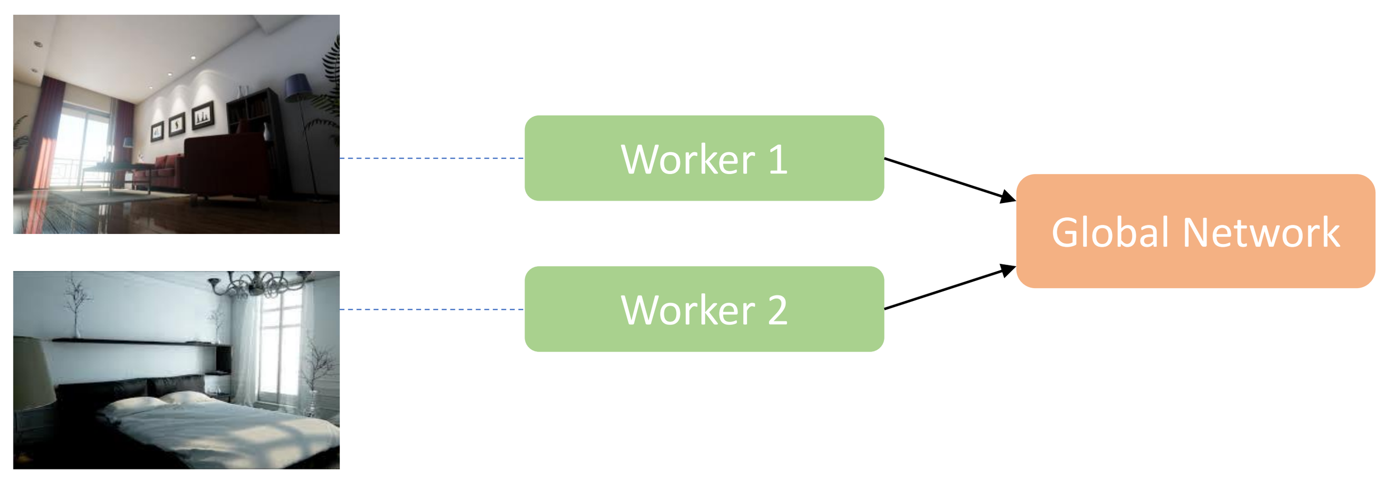
Manish Pandey
Synopsys Inc

Problem of Visual Navigation

- Navigate towards a target with visual input
- Applications in Robotics, autonomous ground vehicles, UAVs
- Learn relationships between action and environment changes
- Well suited for deep learning

Related Work

- Map-less navigation - Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning (Zhu)
- Dynamic Reinforcement Learning - Game playing, obstacle avoidance using monocular vision
- Physics Engines – Realistic interactive Newtonian world simulation (Mottaghi)
- Asynchronous methods for deep reinforcement learning



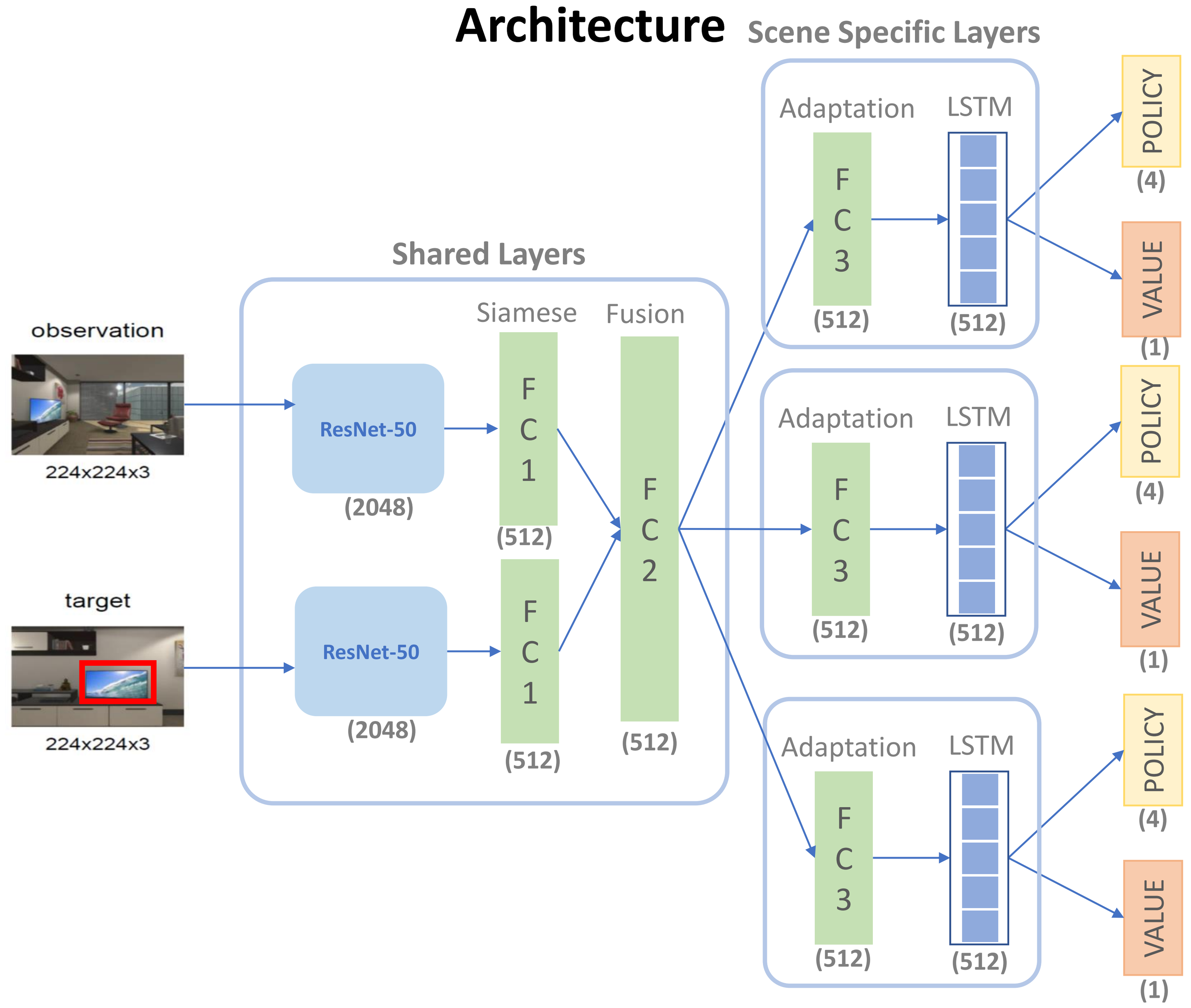
Problem Statement

- Can we add state to deep reinforcement learning to improve Quality of Navigation (QoN) ?
- Measures of QoN: Path length, Number of Collisions, Training Speed
- How well can we learn policy for targets not trained on?
- What memory architectures work well?

Thor Framework and Dataset

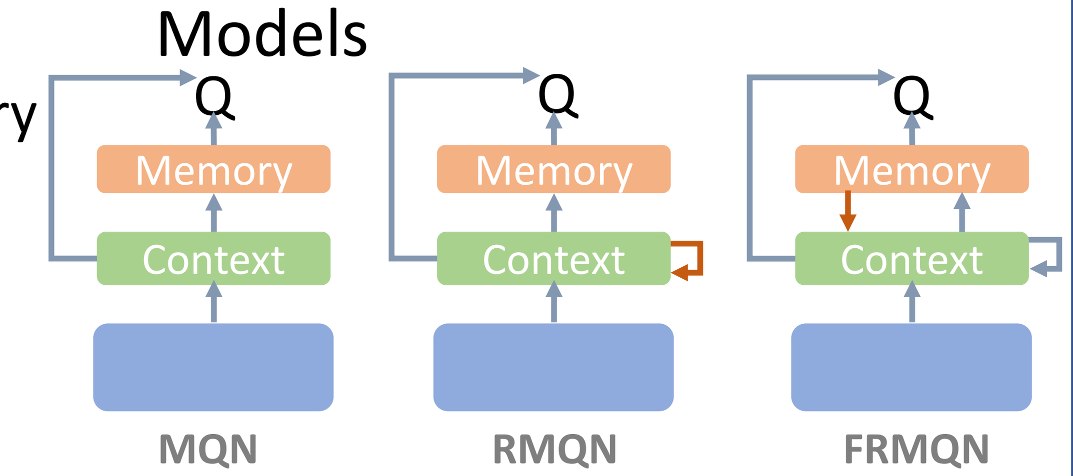
- 4 scene types with ~68 objects
- 224x224x3 RGB images
- Agent navigates to target object specified with a RGB image
- Python API to interact with scene
- (F,B,L,R) -> new image/collision
- Pre-trained ResNet-50 layers

Architecture



Models

- Add context memory to DRL
- Multiple memory architectures



Experimental Results

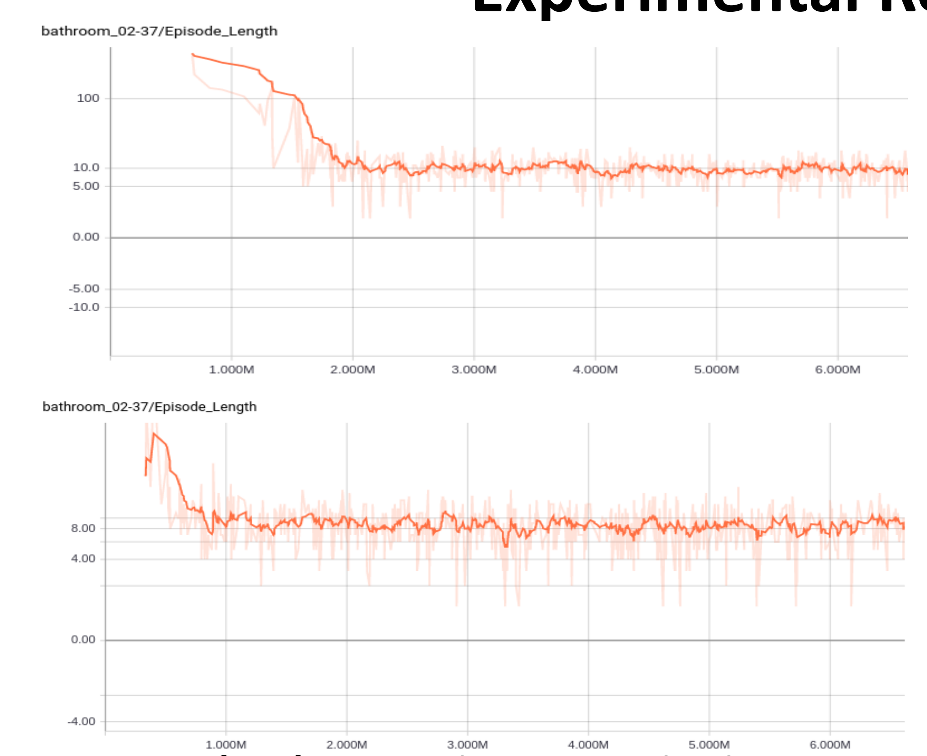


Fig 1: LSTM (top) vs Baseline - train bathroom02_26

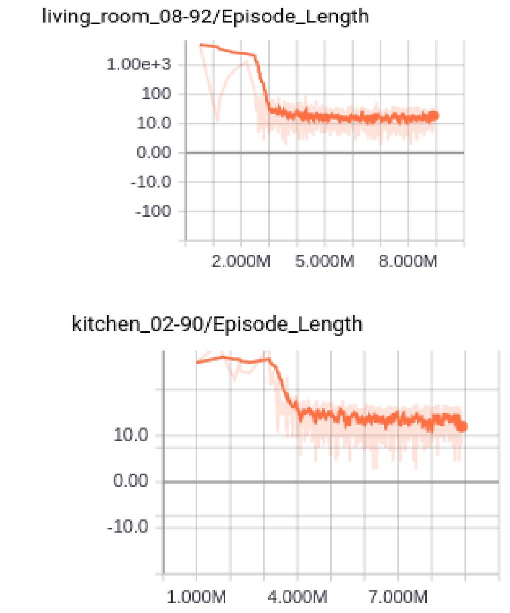


Fig 2: Longer convergence in complex scenes

	Untrained Network	Training Without Target	Training With Target
Episode Length	425.07	312.93	9.55
Reward	1.28	5.71	9.90
Collisions	49.82	13.04	0.16

Table 1: Path lengths, rewards and collision with and without training

Scene_target	Episode Length	Steps (M)	Episode Length	Steps (M)	Episode Length Improvement	Steps Increase Percentage
	Baseline	Baseline	LSTM	Baseline	LSTM/Base	LSTM/Base
bathroom_02_26	9.10	0.96	8.20	2.63	11.0%	174%
bathroom_02_37	9.00	1.00	8.70	2.14	3.4%	114%
bathroom_02_43	8.40	1.34	8.70	2.61	-3.4%	94.8%
bathroom_02_53	8.5	1.15	8.1	2.18	4.9%	89.6%
bathroom_02_69	7.8	1.34	7.6	2.8	2.6%	109%

Table 2: Baseline vs LSTM Episode Lengths

Initial Conclusions and Ongoing Work

- Adding memory context results in small improvements in episode path lengths. The training time (number of images in training episode increased by 80-170%), suggesting a tradeoff.
- Continue tuning and evaluation of memory models
- Determining approaches to better debug DRL networks and speeding up training convergence

Acknowledgements

The team would like to thank Yuke Zhu for his help with numerous questions on Thor, DRL and related topics.