Scaling Multi-Agent Reinforcement Learning via State Upsampling

Luis Pimentel*, Rohan Paleja*, Zheyuan Wang, Esmaeil Seraj, James E. G. Pagan & Matthew Gombolay
{lpimentel3, rohan.paleja, pjohnwang, eseraj3}@gatech.edu, jepagan@sandia.gov, matthew.gombolay@cc.gatech.edu

Georgia Tech CORE ROBOTICS LAB
ROBOTICS SCIENCE AND SYSTEMS

Introduction

In Multi-Agent Reinforcement Learning (MARL), agents must learn to individually make decentralized decisions to achieve high-performance collaboration. Collaboration can greatly improve performance and task efficiency and is often a required step to accomplish an overarching mission successfully [1].

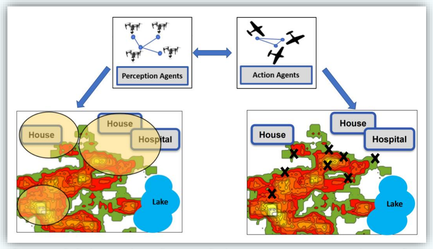


Figure 1. Real-world inspired, multi-agent scenario, where perception agents must spot wildfire locations for action agents to put out [2].

- Learning in large environments with large numbers of agents is challenging due to the exponential increase in the joint state-action space.
- This requires increased training times and high-performance compute, making it difficult to scale to large multi-agent systems for real-world applications.

Contribution

We present a novel and efficient transfer learning method to accelerate training in MARL via an upsampling procedure to upscale the domain size, and a tensor representation to upscale number of agents.

Background

Increasing the joint state-action space in MARL causes the following challenges:

- Increased non-stationarity from an agent's perspective.
- Increasingly complex credit-assignment to any agent [3].
- Large variance in the policy gradient during training [4].

Transfer learning methods offer a way to speed up training in difficult tasks by leveraging prior knowledge learned in less challenging tasks.

Our work augments MARL frameworks to enable knowledge transfer from pre-training in tasks with small environments and number of agents, to more complex environments with larger environment size, and numbers of agents.

References

[1] Esmaeil Seraj, Zheyuan Wang, Rohan Paleja, Daniel Martin, Matthew Sklar, Anirudh Patel, and Matthew Gombolay. Learning efficient diverse communication for cooperative heterogeneous teaming. In Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, pages 1173–1182, 2022.
[2] Esmaeil Seraj, Xiyang Wu, and Matthew Gombolay. Firecommander: An interactive, probabilistic multi-agent environment for joint perception-action tasks. arXiv preprint arXiv:2011.00165, 2020.
[3] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.
[4] Amanpreet Singh, Tushar Jain, and Sainbayar Sukhbaatar. Learning when to communicate at scale in multiagent cooperative and competitive tasks. arXiv preprint arXiv:1812.09755, 2018.
[5] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. Advances in neural information processing systems, 30, 2017
[6] Yaru Niu, Rohan Paleja, and Matthew Gombolay. Multiagent graph-attention communication and teaming. In Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, pages 964– 973, 2021.

Overview

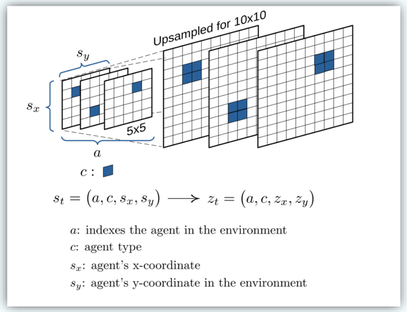


Figure 2. Visual of our tensor-based state representation and upsampling transformation.

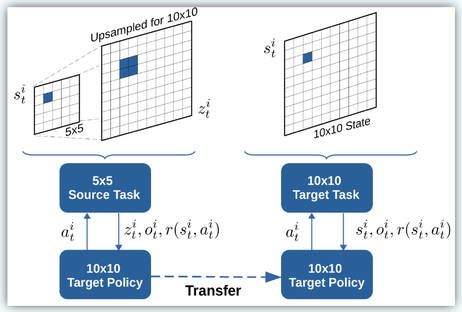


Figure 3. Visual of our transfer method.

- States & observations are represented as tensors.
- This allows scaling to larger team sizes without additional parameterization.
- We upsample an agent's state, s_t^i, to generate z_t^i, the input to the target policy in the source task.

Our upsampling method allows the same policy to be used in the source and target tasks and helps transfer spatial knowledge to the target task to achieve greater performance than training from scratch.

Experiments

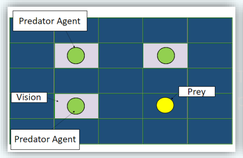


Figure 4. Homogenous PP [4] grid-world domain.

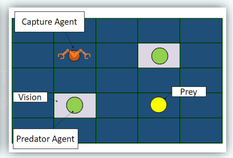


Figure 5. Heterogenous PCP [1] grid-world domain.

MARL Frameworks Deployed

- IC3Net [4], MAGIC [6], and HetNet [1].

Source Tasks

- 5x5 World
- 3 total agents

Simple

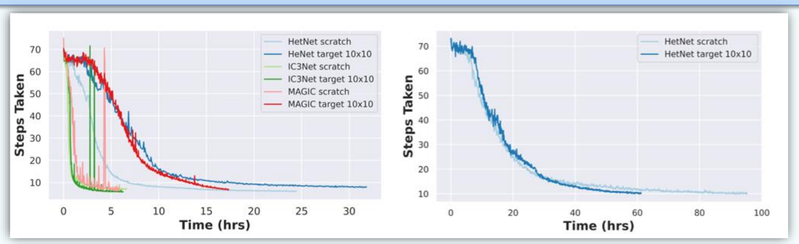
Target Tasks

- 10x10 World
- 5 total agents

Challenging

Pre-training Results

Q1) Does pre-training under our upsampling method affect performance within the source task?



a) 5x5 PP Source task. **b) 5x5 PCP Source task.**

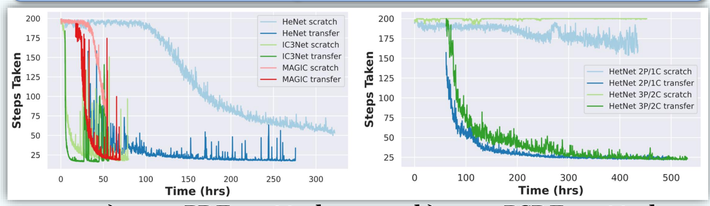
Figure 6. Learning curves for pre-training in the source task using our upsampling method.

- Pre-training with our method converges and achieves similar performance as training from scratch.

Knowledge learned in the source task can be leveraged when transferred the more difficult target task to accelerate learning performance.

Transfer Results

Q2) What are the benefits obtained via our policy transfer procedure?



a) 10x10 PP Target task. **b) 10x10 PCP Target task.**

Figure 7. Learning curves for training in the target task. Transferred policies are shifted to account for time spent pre-training.

Experiment			Time (hrs)	Performance (Avg. Steps-Taken)	
				Scratch	Transfer
HetNet	10×10	5P	100.00	189.70	**27.13**
IC3Net	10×10	5P	10.00	55.59	**20.72**
MAGIC	10×10	5P	30.00	192.61	**69.67**

(a) Experiments in the homogeneous PP [4] domain.

Experiment			Time (hrs)	Performance (Avg. Steps-Taken)	
				Scratch	Transfer
HetNet	10×10	2P/1C	200.00	184.62	**30.19**
HetNet	10×10	3P/2C	350.00	199.85	**25.34**

(b) Experiments in the homogeneous PCP [1] domain.

Table 1. Performance in a fixed amount of aggregate training time.

- In a fixed amount of training time, **transferred policies** achieve up to a **6.11× - 7.88×** performance improvement

Trained from scratch
Unable to learn high performance behavior after ≥ 400 hours in PCP.

Trained from transfer
Learns relatively high performance behavior in ≤ 40 hours after transfer in PCP

- Transferred policies achieve convergence and much better performance with much **less aggregate training time.**

Conclusion

- **We present a transfer learning method that allows us to leverage knowledge learned in simple tasks to accelerate learning in difficult tasks.**
- **Empirically, we show our method can achieve greater performance with significantly less training experience and time, across multiple MARL algorithms and domains.**

Authors

Luis Pimentel · Rohan Paleja · Zheyuan Wang · Esmaeil Seraj · James E. G. Pagan · Matthew Gombolay