ARTIFICIAL INTELLIGENCE
National Laboratory

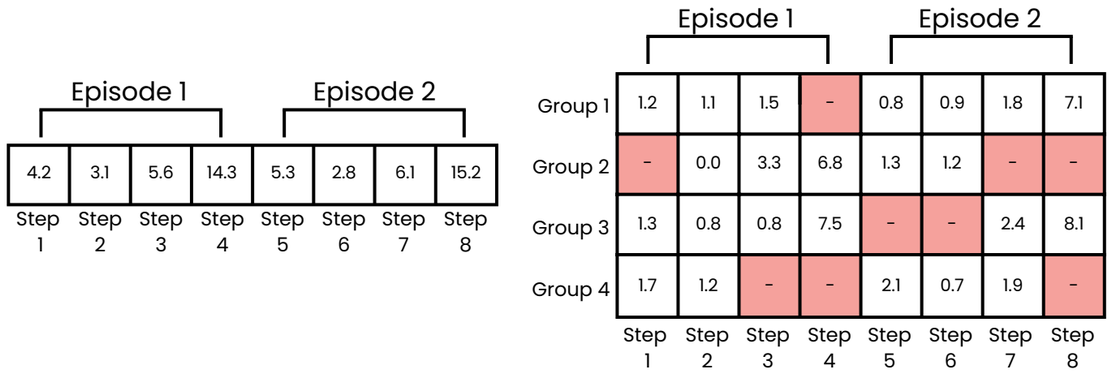
Scalable Distributed Reinforcement Learning in Multi-Agent Environments

Takács Tamás, Magyar Gergely, Gulyás László

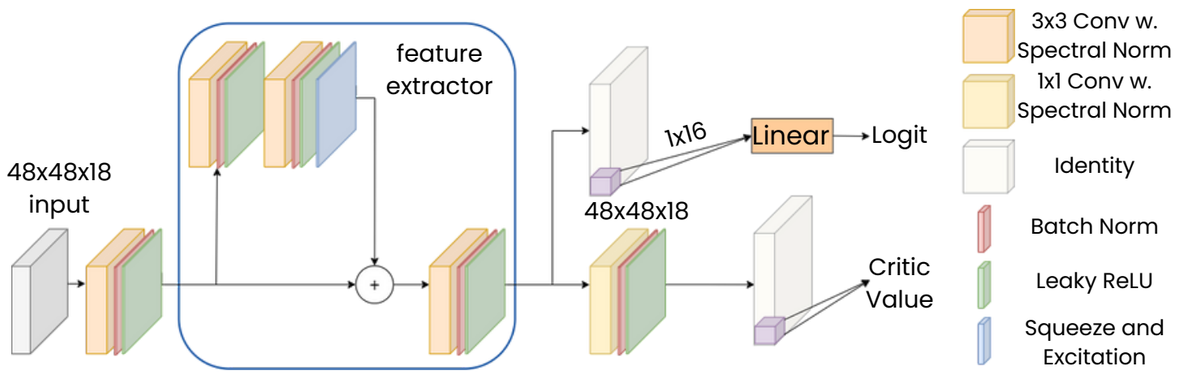
Eötvös Loránd University, Faculy of Informatics, Department of Artificial Intelligence

Novel reinforcement learning algorithms often face **scalability and compatibility issues in multi-agent environments** due to their optimization for single-agent settings. The lack of standardized methods for adaptation restricts their broader applicability, especially when dealing with rapidly changing numbers of controllable entities and massive scaling. Challenges include credit assignment, extensive memory usage, and increased computational time, leading to slow, destabilized training and suboptimal resource utilization. We propose a hybrid architecture, combining monolithic and distributed approaches, resulting in a **30-times reduction in model size** and learning basic skills **24 times faster with 600-times fewer training examples** compared to related works in the same environment. We also introduce **trajectory separation**, achieving a **3-times speed increase in training** convergence. Our **M-PPO-based hybrid** model achieves:

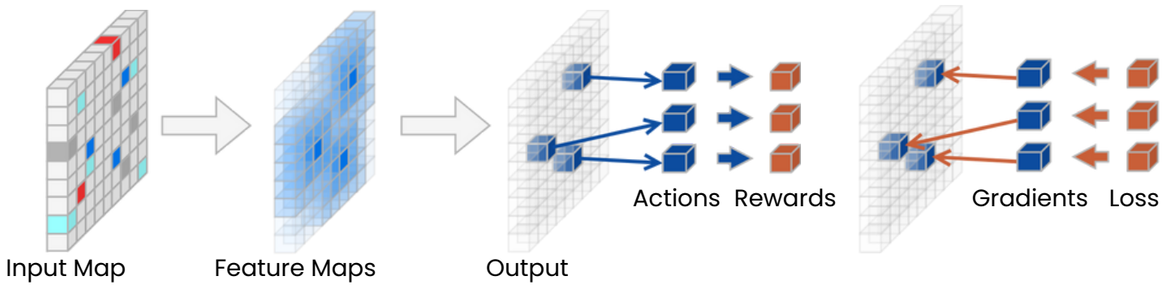
- a performance-based environment
- **separated agent trajectories**
- separated reward, advantage, probability and entropy calculation
- **reinforced positive behavior in early training**



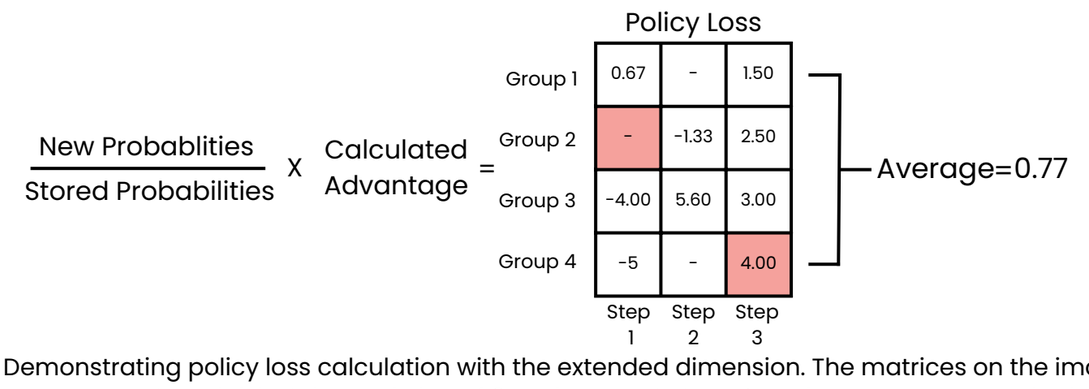
Showing the difference between **global rewards** (left) and **rewards distributed into groups** using trajectory separation (right)



The structure of the **final feature extractor** and the **critic** and **actor** heads utilized on top of M-PPO.



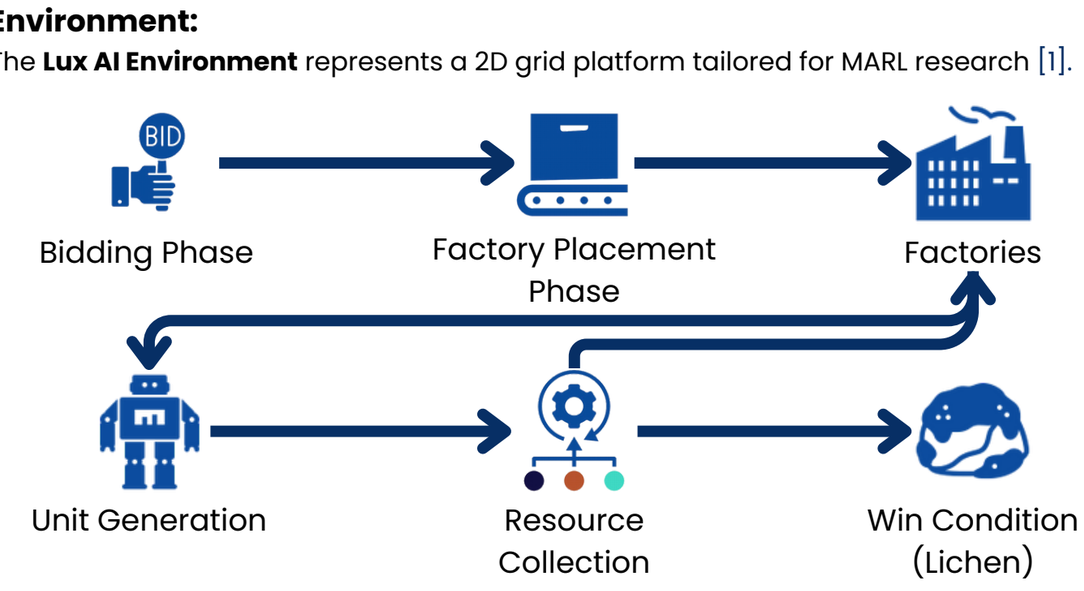
Our **novel trajectory separation method** achieves distributed rewards by producing one trajectory per entity or cluster and calculating loss based on individual performance.



Demonstrating policy loss calculation with the extended dimension. The matrices on the image represent a mini-batch consisting of 3 environment steps. Inactive groups are marked.

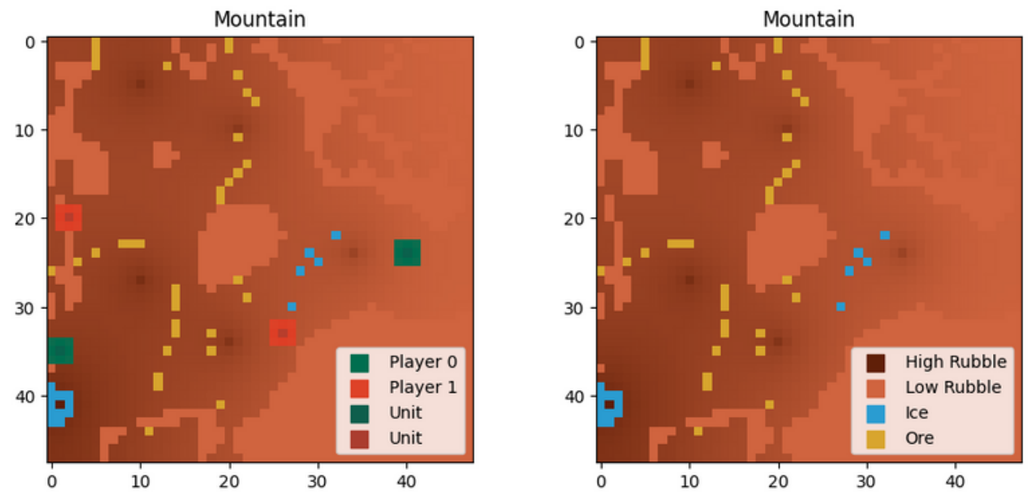
Environment:

The **Lux AI Environment** represents a 2D grid platform tailored for MARL research [1].



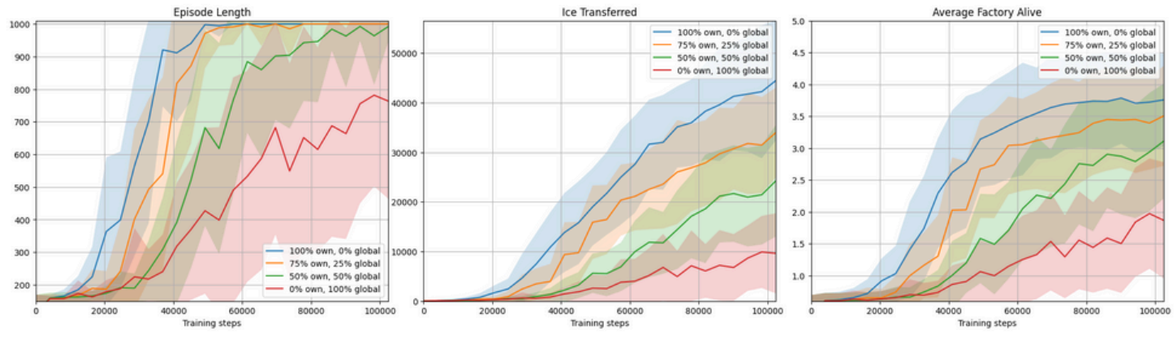
Simplified environment loop of the **Lux AI Environment**. An episode starts with the **Bidding Phase** and ends after **1000** steps. The player with the most lichen collected wins.

The **Lux AI Competition** has already hosted two seasons on Kaggle, one set in a forest environment and the other on Mars.



A visual representation of the grid environment, generated from a specific seed, showcasing a **Gaussian filter-based factory placement** policy on two identical maps.

In our study, **we compared our results with other competition submissions, conducted experiments on various credit assignment setups, and performed ablation studies.**



Results demonstrate the effectiveness of trajectory separation, revealing that higher agent-specific reward weights heavily improve performance. **Episode Length** indicates team survival duration, while **Ice Transferred** and **Average Factory Alive** reflect resource collection efficiency.

References:

[1] Chen, H. et al. Emergent collective intelligence from massive-agent cooperation and competition 2023.
[2] Tao, S., Pan, I., et al. Lux AI Season 2 2023. https://kaggle.com/competitions/luxai-season-2.

Acknowledgements:

This work was supported by the Institute for **Business Cooperation Scholarship** under grants #E111SIVE. We extend our gratitude to **Eötvös Loránd University and the Department of Artificial Intelligence** for providing the resources essential for this project's development.

ELTE | IK INFORMATIKAI KAR
MESTERSÉGES INTELLIGENCIA TANSZÉK
Széchenyi Terv Plusz Építsük együtt Magyarországot!
MAGYARORSZÁG KORMÁNYA
Az Európai Unió finanszírozásával NextGeneration EU