Deep Curiosity Search: Intra-life exploration can improve performance on challenging deep reinforcement learning problems

Christopher Stanton (U. Wyoming) and Jeff Clune (U. Wyoming, Uber AI Labs)

EVOLVING ARTIFICIAL INTELLIGENCE LABORATORY

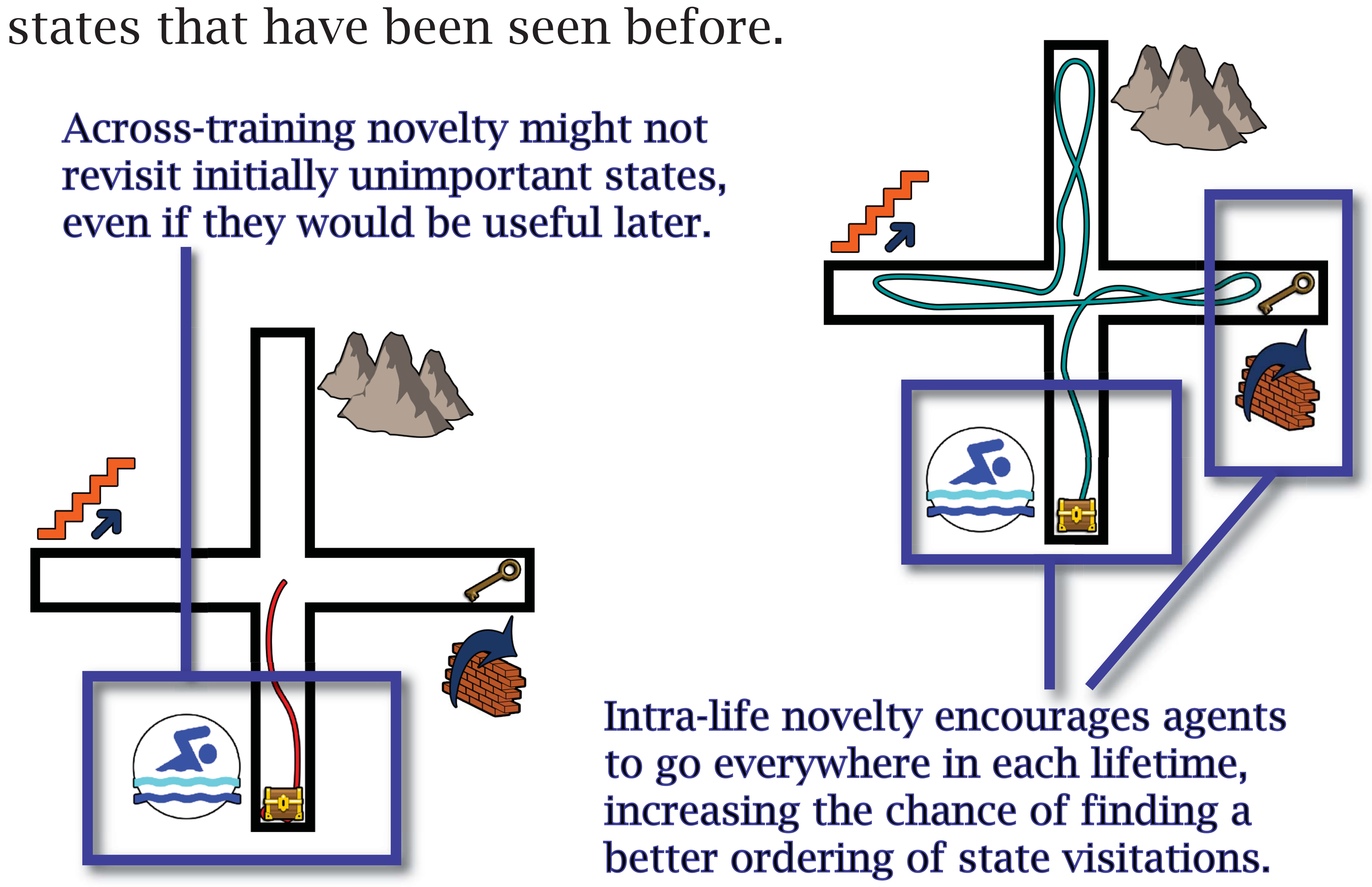
UNIVERSITY of WYOMING

Overview

We introduce Deep Curiosity Search (DeepCS), a directed exploration algorithm inspired by intrinsic motivation in animals, that rewards agents for "doing something new" within their lifetime. Using some domain knowledge, DeepCS matches the performance of other state-of-the-art techniques on hard Atari games like Montezuma's Revenge.

Background

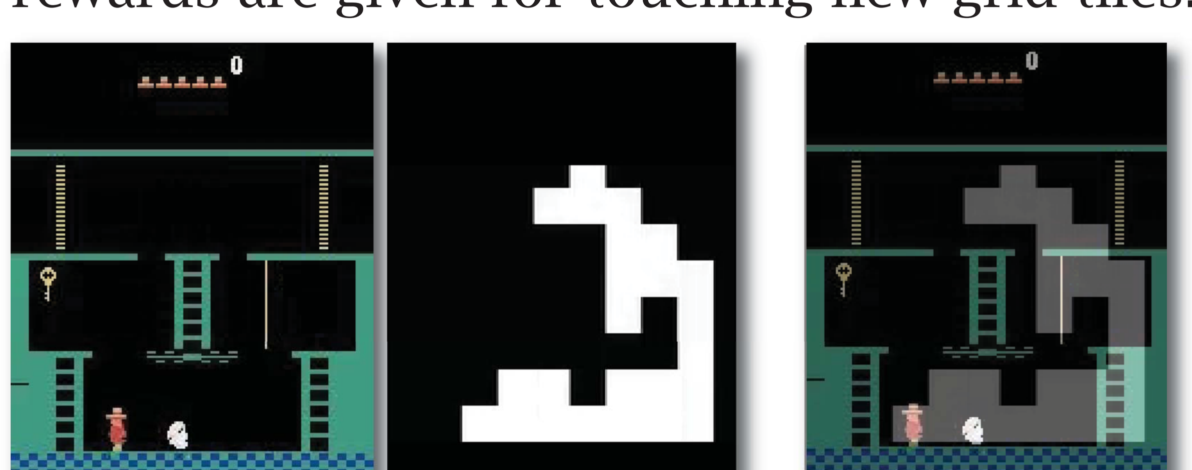
Traditional exploration methods reward ***across-training novelty***: whether a state is new compared to all other states that have been seen before.



Curiosity Search instead encourages ***intra-life novelty***, rewarding agents for visiting new states even if they have been seen in prior training episodes.

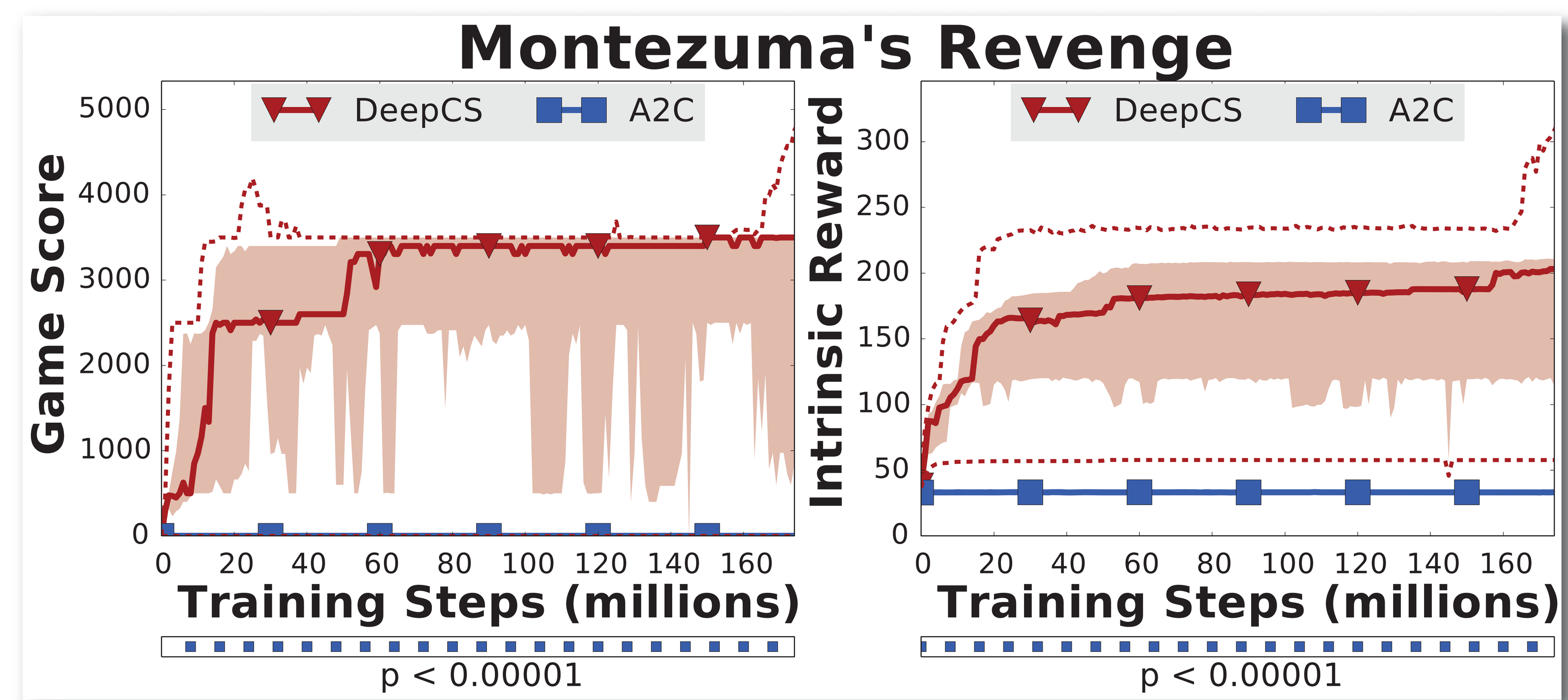
Methods

- Game world is discretized into a uniform grid using RAM.
- Agent receives grid as extra input, as in previous work.[1]
- Intrinsic rewards are given for touching new grid tiles.

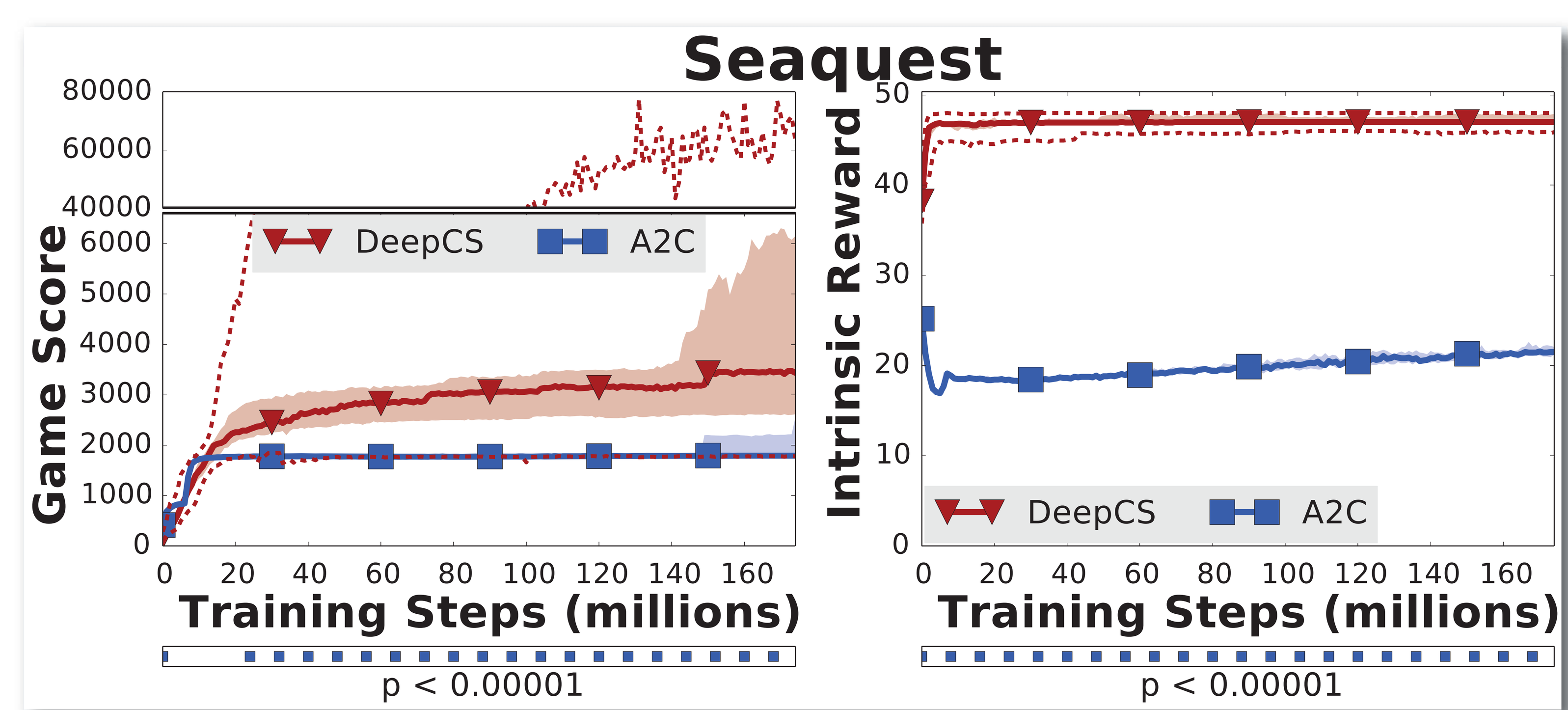


Results

Naïve exploration algorithms usually achieve no score at all on Montezuma's Revenge (MR), while Deep Curiosity Search (DeepCS) matches the performance of other state-of-the-art methods:



DeepCS also improves performance on some games in which directed exploration is seemingly not required, like Seaquest:



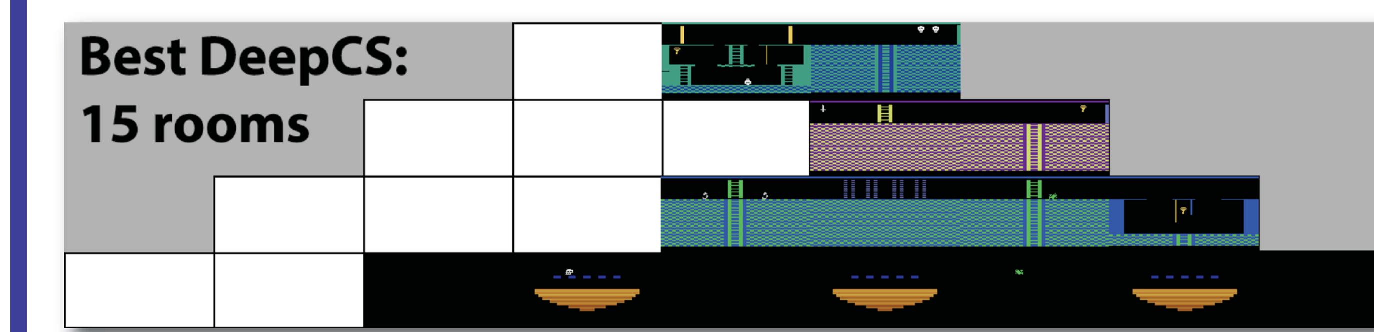
On 5 games, Deep Curiosity Search performs similarly or better than popular naïve- and directed-exploration algorithms:

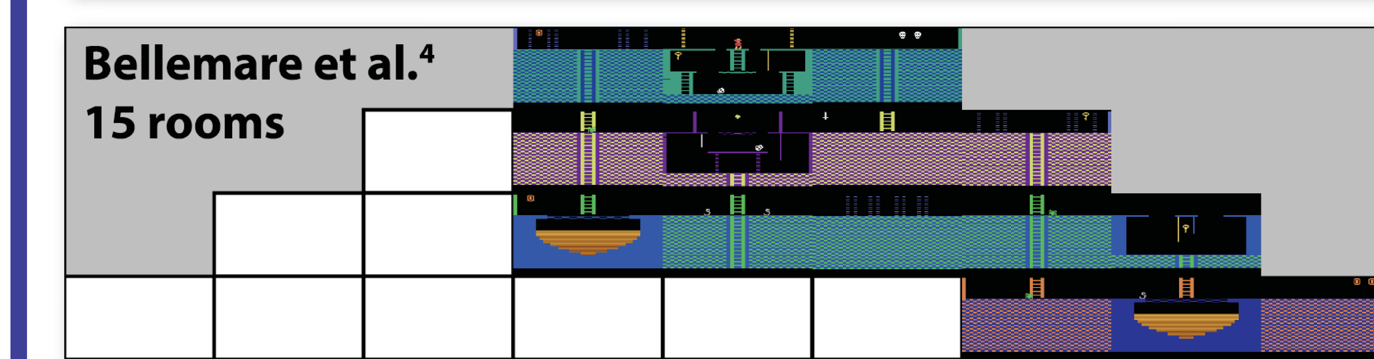
	DQN[2]	A3C[3]	PC[4]	PCn[5]	DeepCS
Amidar	739	283	964	900	**1404**
Freeway	30	0	30	31	**33**
Gravitar	306	269	246	859	**881**
MR	0	53	3439	**3705**	3500
Tutankham	186	156	132	190	**256**
Alien	**3069**	518	1945	1700	2705
Kangaroo	6740	94	5475	**7900**	2837
Pitfall	-	-78	-155	**0**	-186
Private Eye	1788	206	246	**15806**	1105
Seaquest	**5286**	2300	2274	2500	3443
Venture	380	23	0	**1356**	12
Wizard of Wor	3393	**17244**	3657	2500	2134

[1] C. Stanton and J. Clune, "Curiosity Search: Producing generalists by encouraging individuals to continually explore and acquire skills throughout their lifetime," *PLoS ONE*, 11(9): e0162235.
[2] Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529-533, 2015.
[3] Mnih et al., "Asynchronous methods for deep reinforcement learning," in *Proceedings of the 33rd International Conference on Machine Learning*, vol. 48, pp. 1928-1937, 2016.
[4] Bellemare et al., "Unifying count-based exploration and intrinsic motivation," *30th Conference on Neural Information Processing Systems*, 2016.
[5] Ostrovski et al., "Count-based exploration with neural density models," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, pp. 2721-2730, 2017.

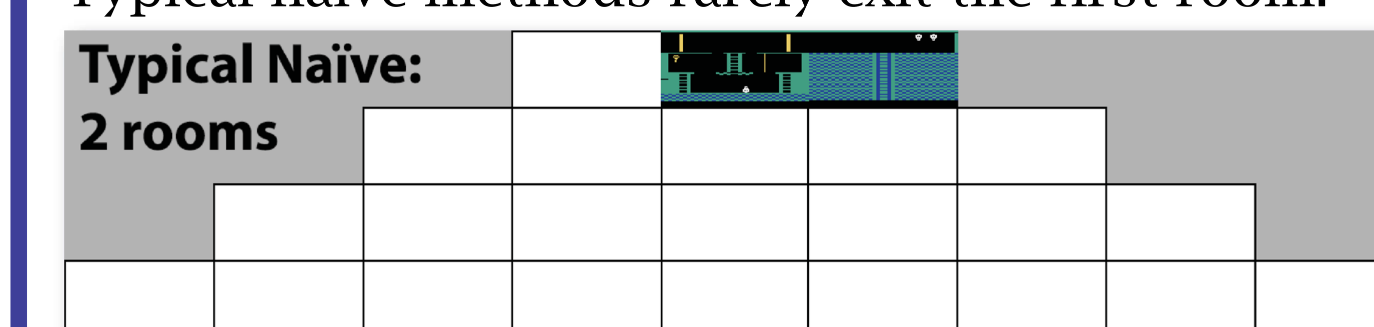
Exploration

The best Deep Curiosity Search agent explores 15 rooms of Montezuma's Revenge, matching other state-of-the-art exploration methods:





Typical naïve methods rarely exit the first room:



Even when the curiosity grid fills up quickly, the brief-lived intrinsic rewards can boost exploration:

Frame 0 (start): the grid has not yet been explored; DeepCS can provide lots of feedback

Frame 700: intrinsic rewards decline; DeepCS can no longer provide feedback

Frame 28,500: even brief presence of intrinsic rewards allows single agents to get 76,000 points

Conclusions

- Encouraging intra-life novelty is an interesting new technique for improving exploration, in both sparse- and dense-reward domains.
- Providing agents with a visual memory of where they have been may help improve exploration.
- More general determination of agent position (e.g. by rewarding novelty in the latent space of an auto-encoder) may help Deep Curiosity Search become a general, even more useful tool.