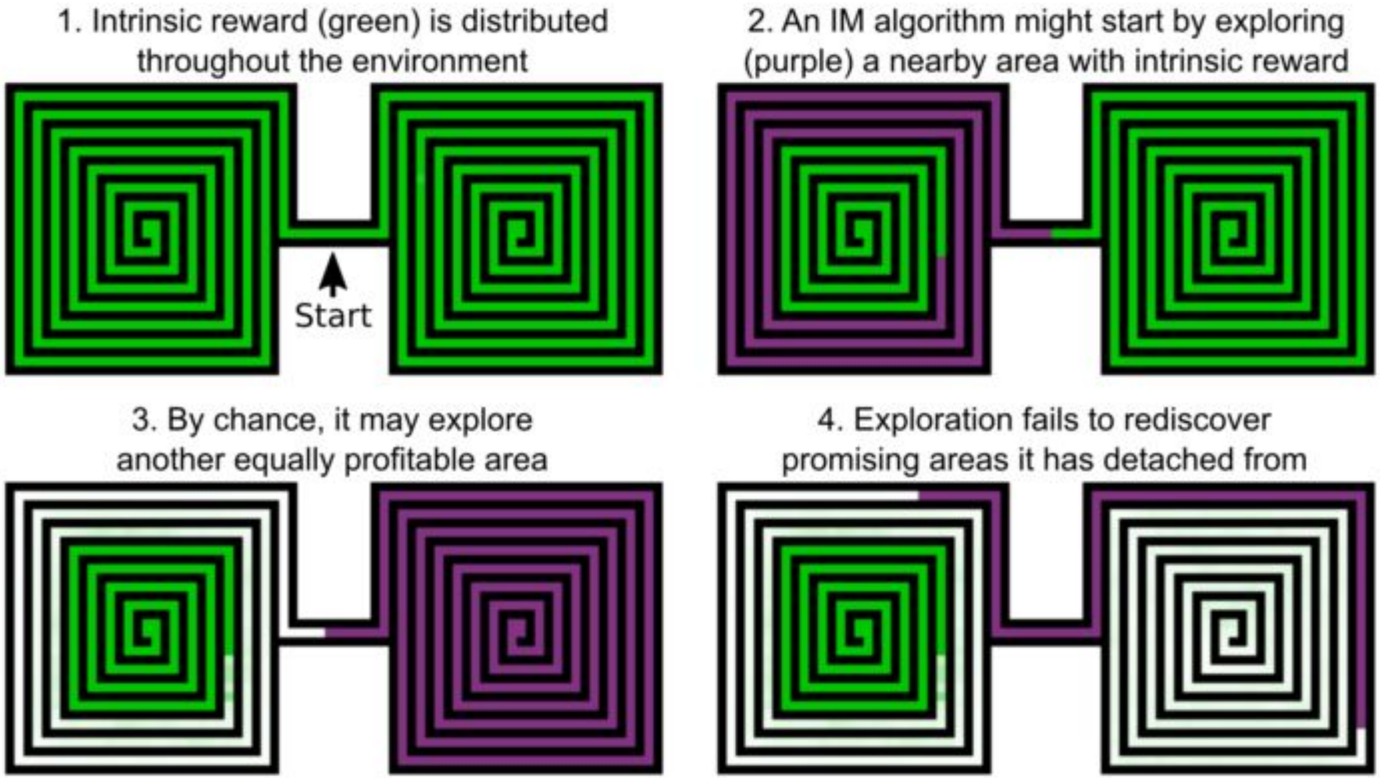
Go-Explore: a New Approach for Hard-Exploration Problems*

Kapranov Ivan

Confusing points

- Is it RL?!
- Wait, wait. It is a brute force!
- They using domain knowledge
- It is a heuristic for ATARI

Intrinsic reward



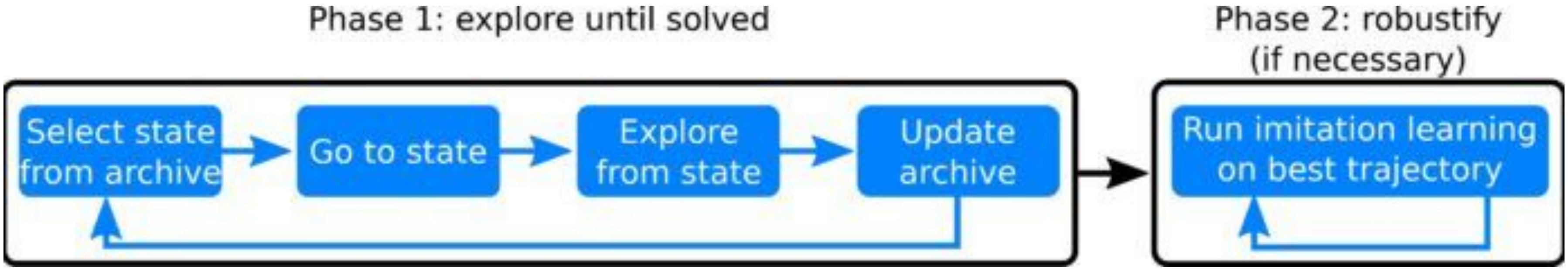
1. Intrinsic reward (green) is distributed throughout the environment

2. An IM algorithm might start by exploring (purple) a nearby area with intrinsic reward

3. By chance, it may explore another equally profitable area

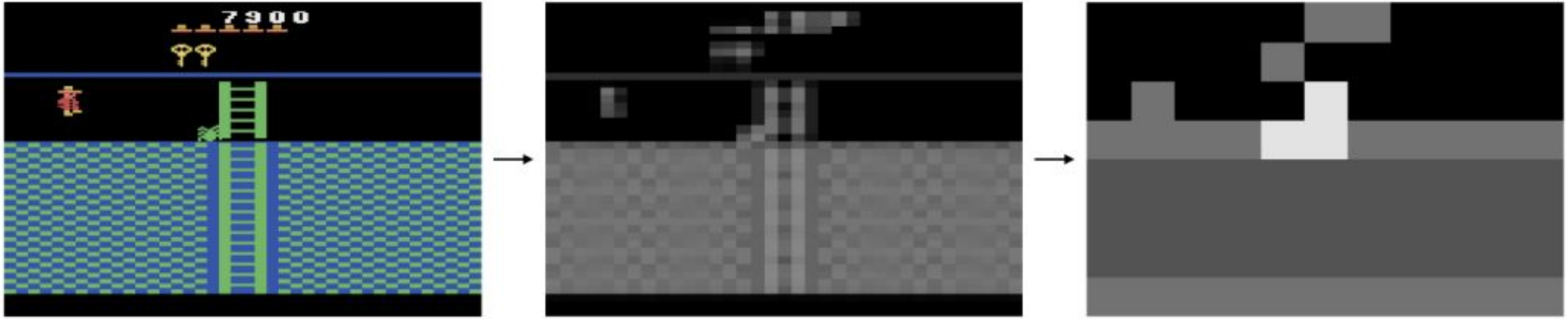
4. Exploration fails to rediscover promising areas it has detached from

Go Explore


Phase 1: explore until solved
Select state from archive
Go to state
Explore from state
Update archive
Phase 2: robustify (if necessary)
Run imitation learning on best trajectory

State -> Cell

11 * 8 pixels, 8 colors



+ heuristics

Exploration

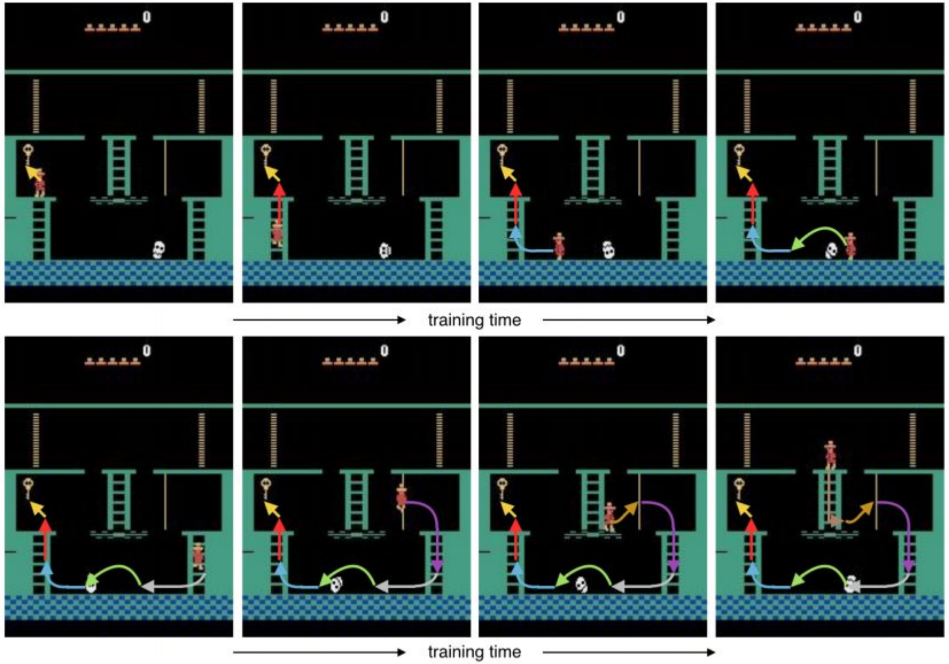
- 100 steps
- With 95% probability repeat the previous action

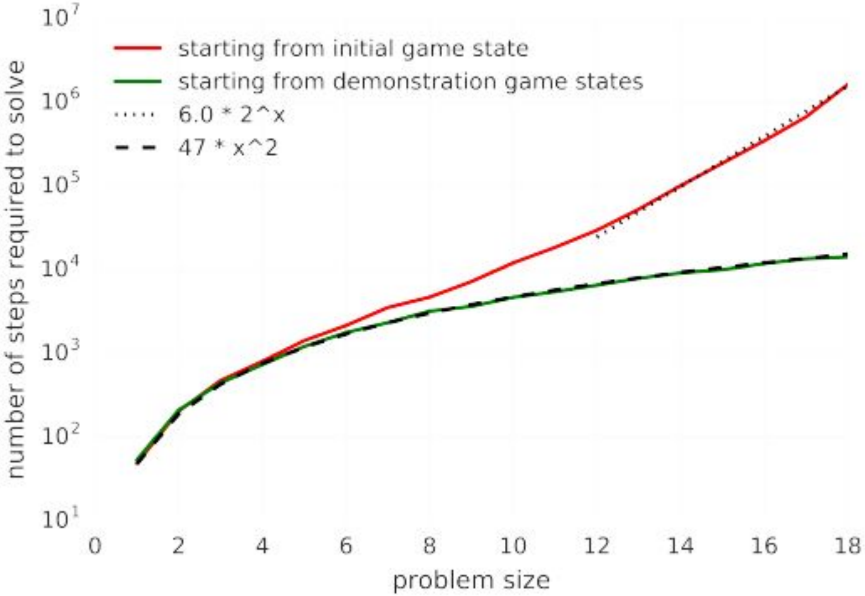
Archive

- If current Cell is new, archive it
- Else, compare rewards
- Else, compare trajectory length

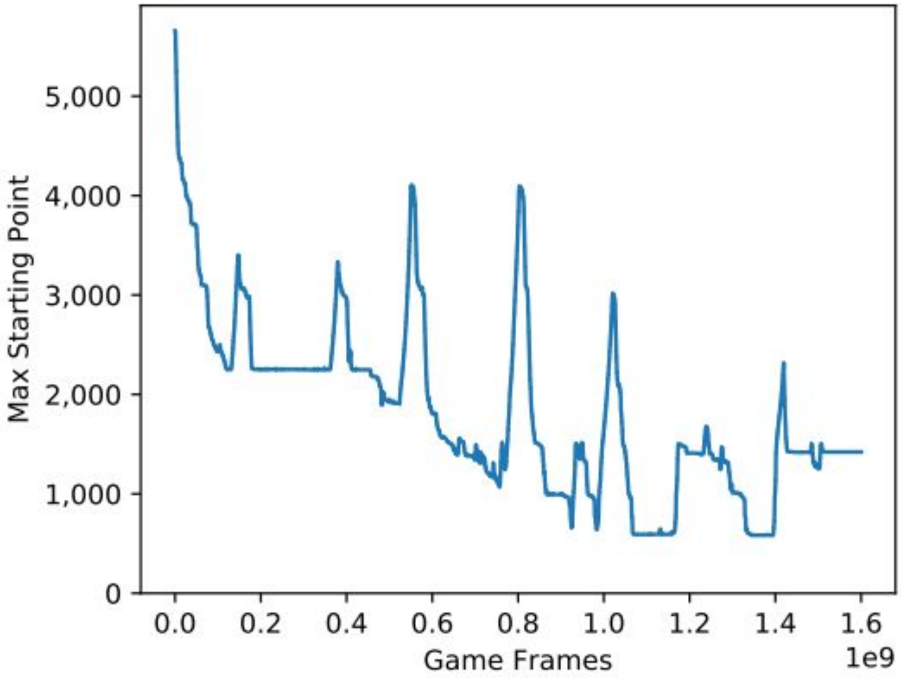
Robustification - imitation learning

- Learning Montezuma's Revenge from a Single Demonstration

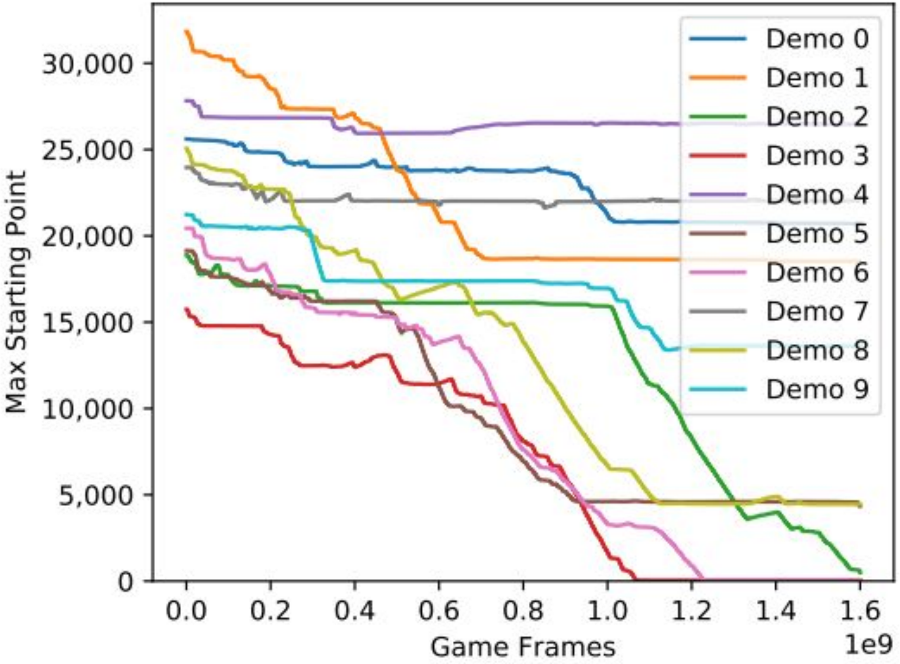




Experiments

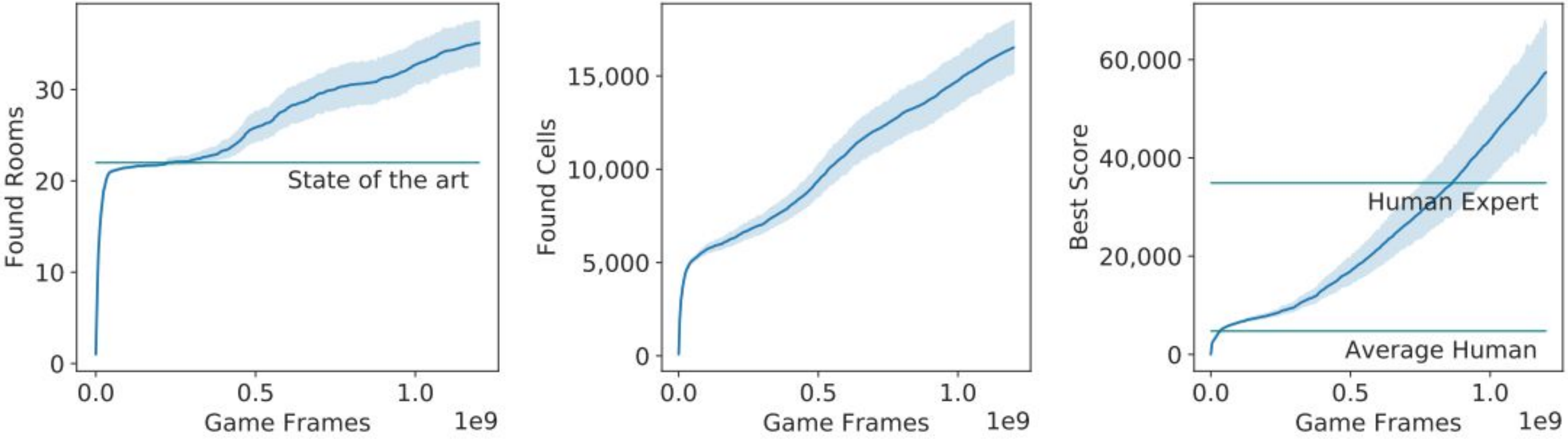


(a) Failed robustification with 1 demonstration



(b) Successful robustification with 10 demonstrations

Result “without” domain knowledge

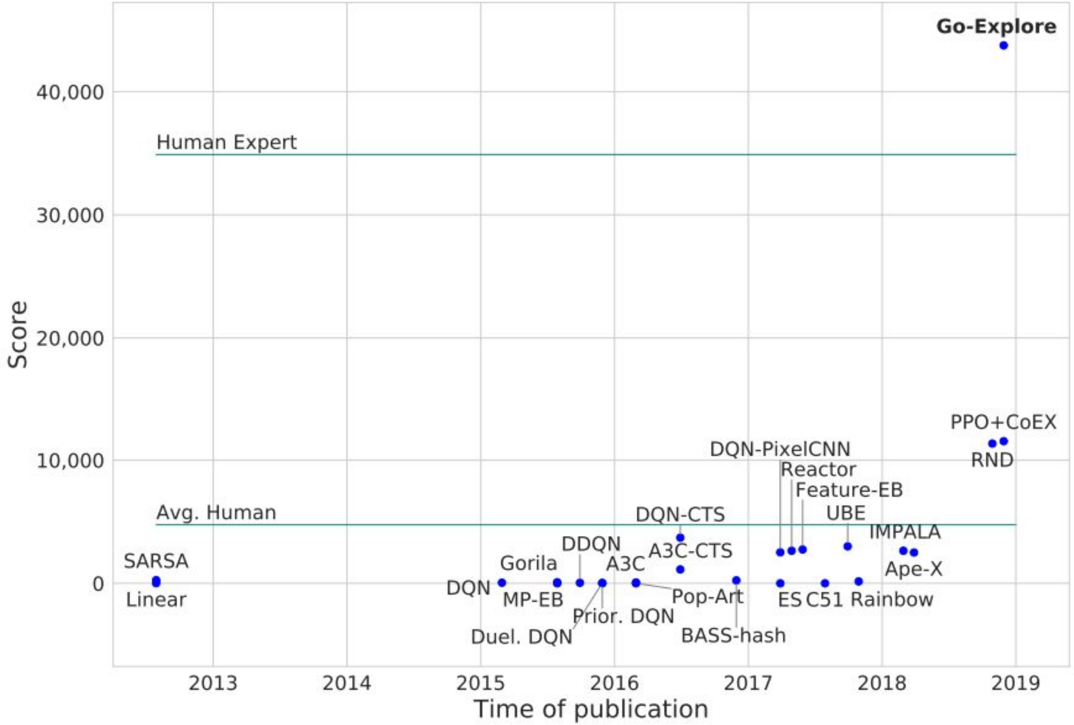


(a) Number of rooms found

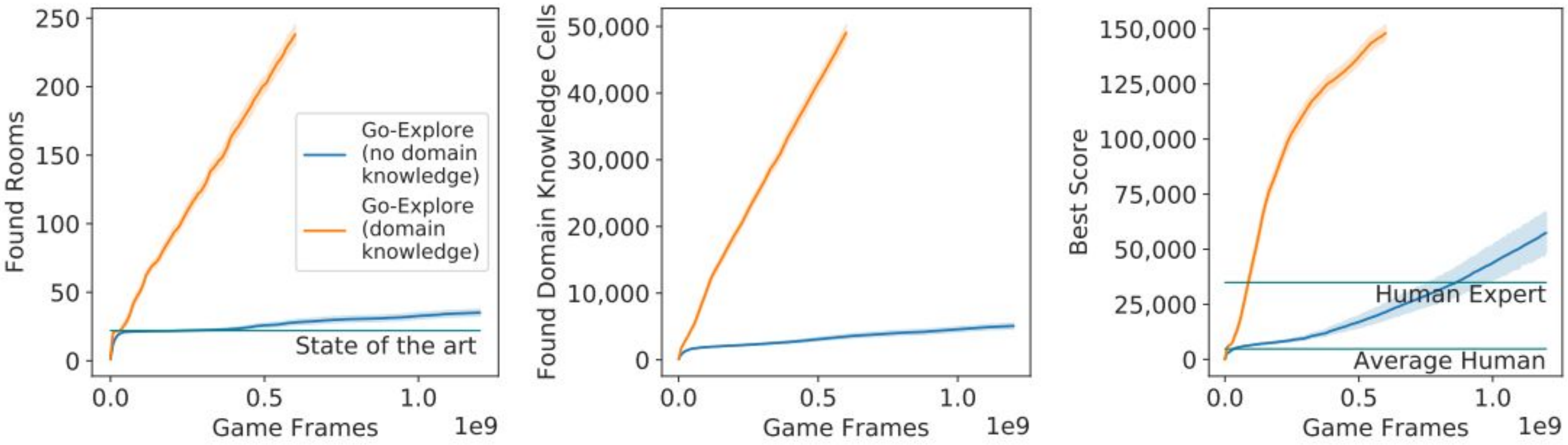
(b) Number of cells found

(c) Maximum score in archive

Result “without” domain knowledge



Result with domain knowledge

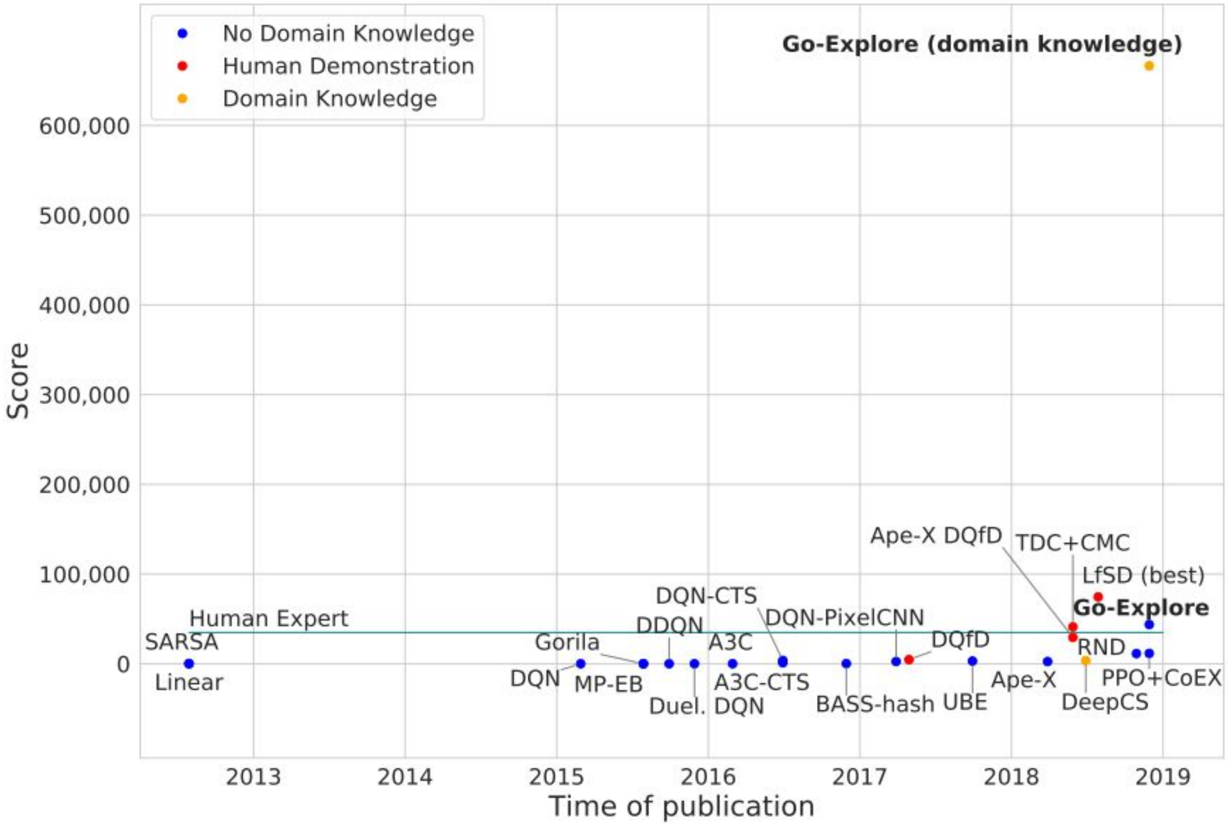


(a) Number of rooms found

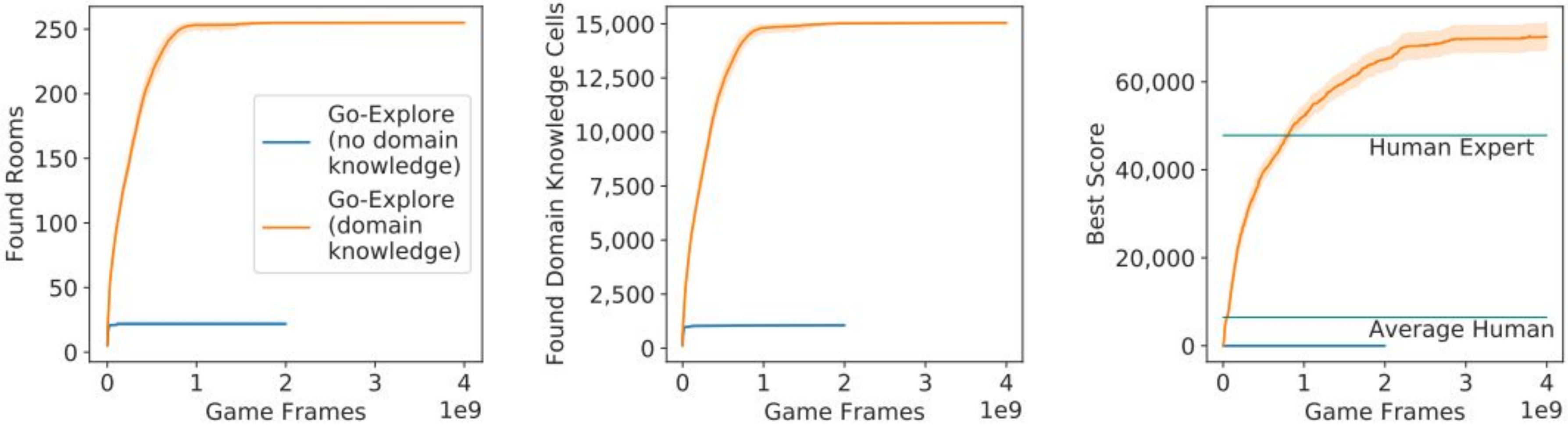
(b) Number of cells found

(c) Maximum score in archive

Result with domain knowledge



Pitfall



(a) Number of rooms found

(b) Number of cells found

(c) Maximum score in archive