Scaling All-Goals Updates in Reinforcement Learning Using Convolutional Neural Networks

Fabio Pardo, Vitaly Levdik, Petar Kormushev

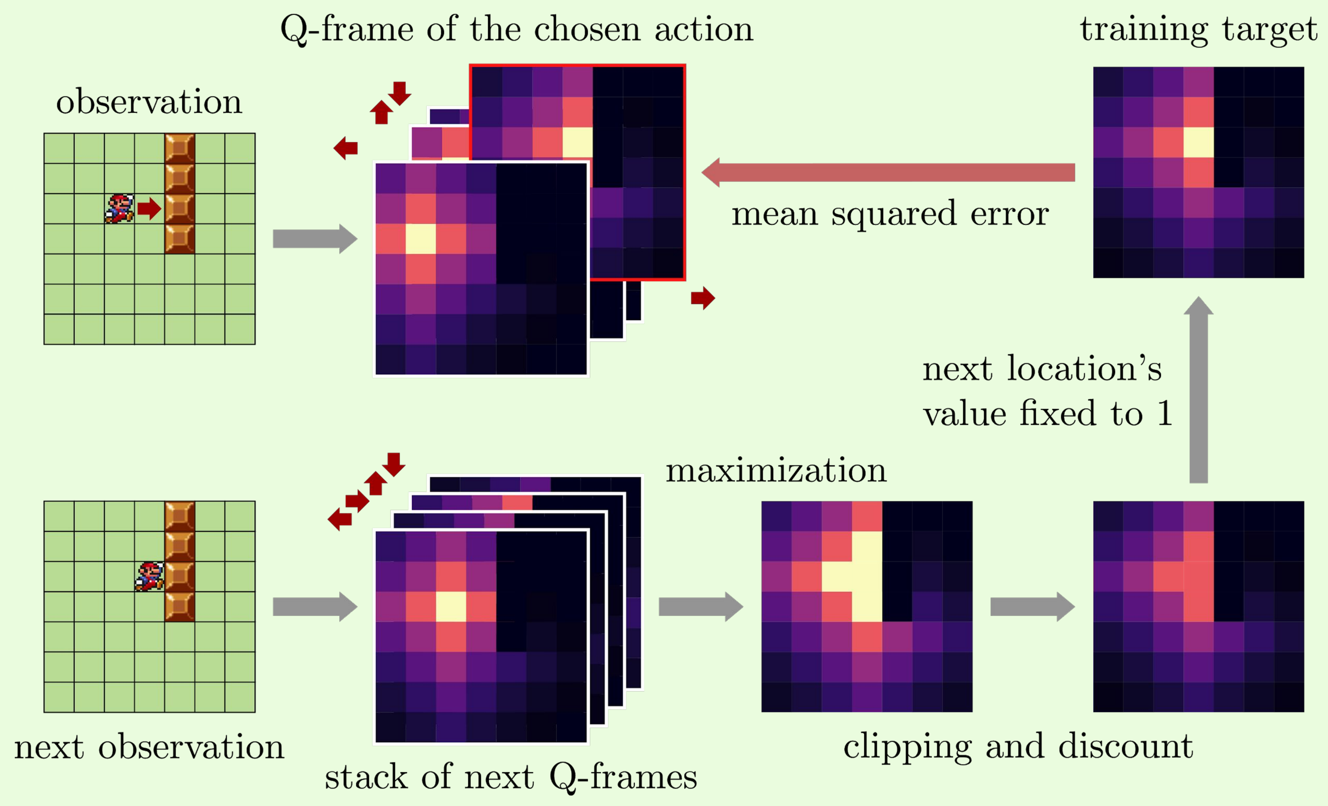
Imperial College London

1. Learn to reach goals

Naive approach: learns via RL one goal at a time
Goal relabeling: replays past transitions with new goals (which ones?)
All-goals update: replays past transitions with all possible goals (how to scale?)

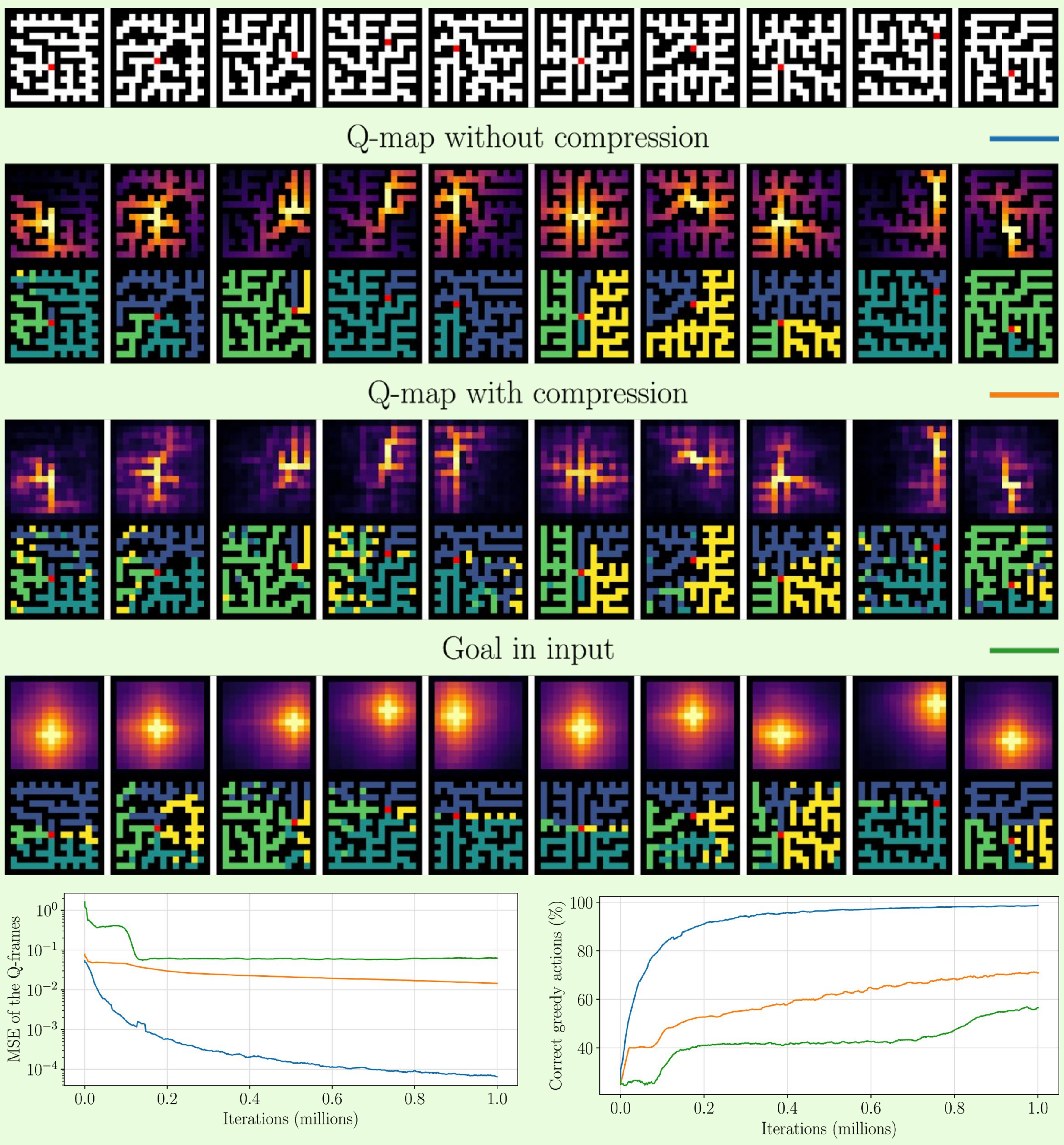
Proposed Q-map:

- Uses Q-learning towards all goals simultaneously
- Produces hundreds of Q-values towards all the goals in output
- Uses convolutions to exploit correlations between visual features and reachability



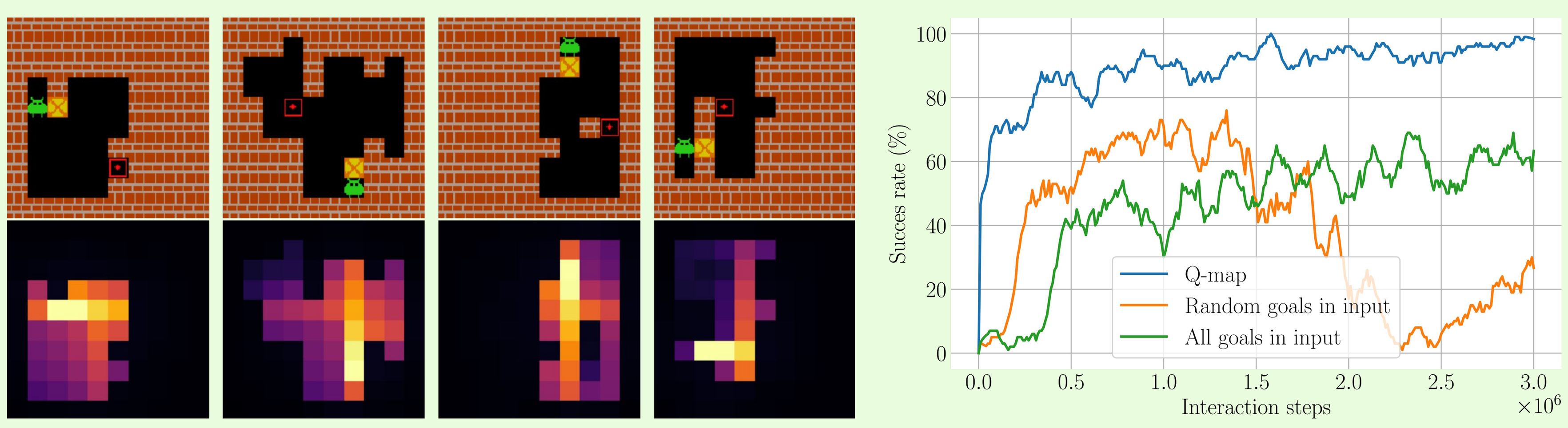
Random mazes:

- All-goals in output is better than all-goals in input
- Compression architecture (conv → dense → deconv) is worse



Random Sokoban:

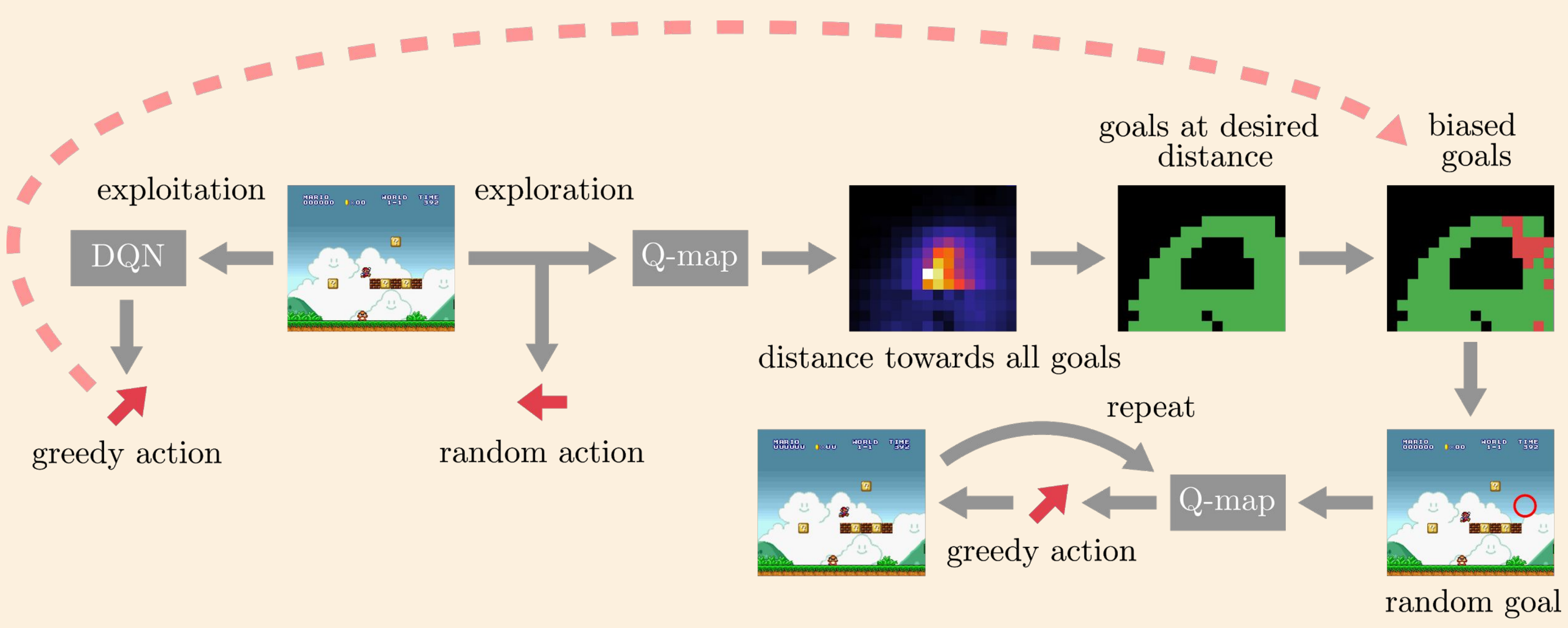
- All-goals in output is better than all-goals in input
- Random goals in input is unstable



2. Explore by reaching goals

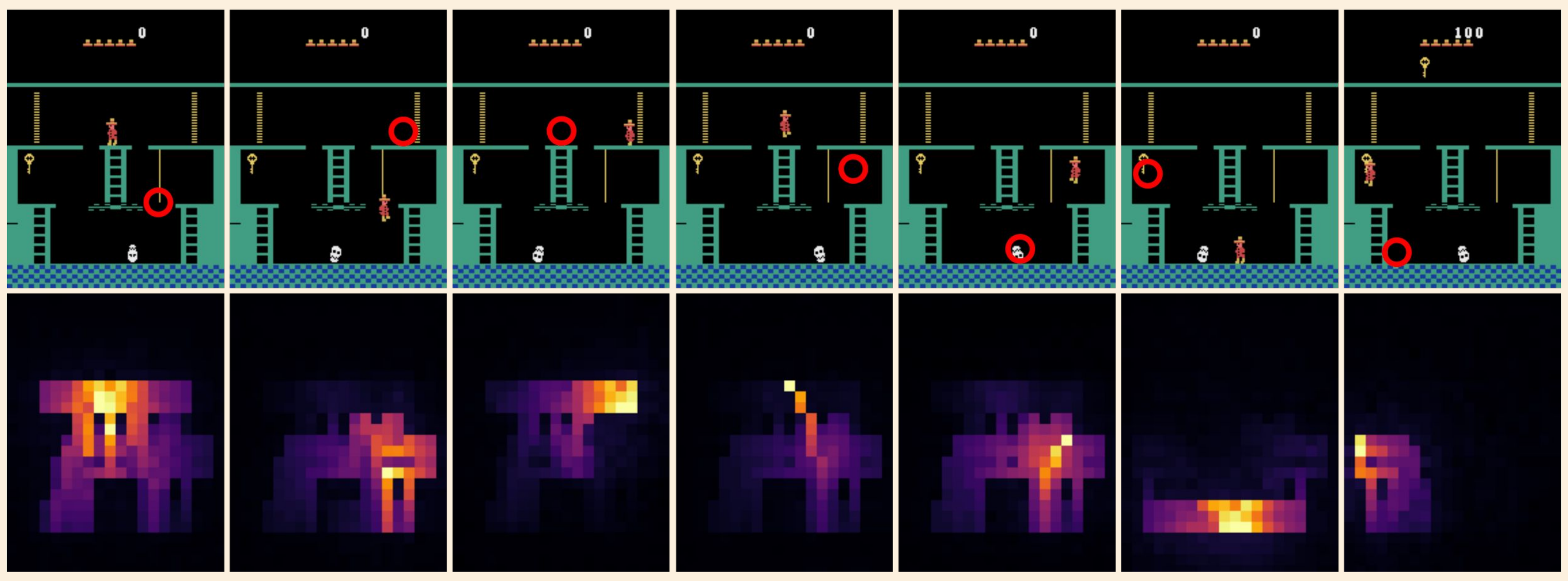
Proposed exploration:

- Trains a Q-map to reach goals
- Selects a goal in reasonable range and reaches it using the Q-map
- Optionally selects goals in the direction of task-directed actions



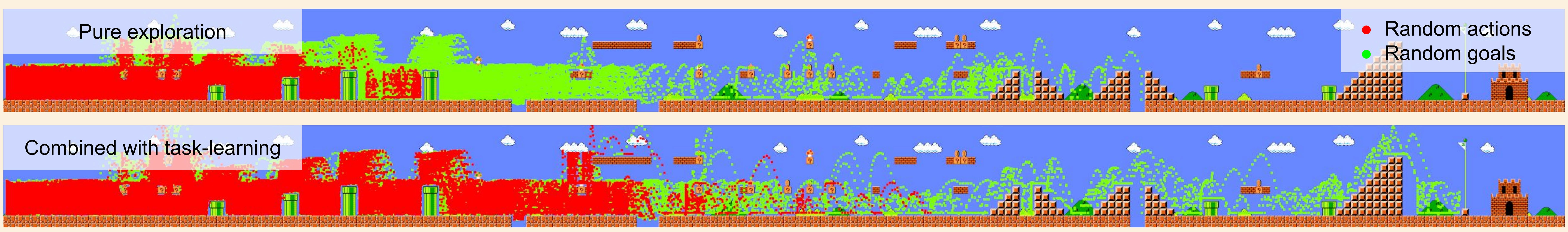
Montezuma's Revenge:

- Random actions lead to the key once in 5 million steps
- Training Q-map + random goals selection lead to the key 398 times



Super Mario Bros. (All-Stars):

- Random goals doubles the exploration range
- Partially transfers mapping to other levels



Website: sites.google.com/view/q-map-rl **Contacts:** f.pardo@imperial.ac.uk, v.levdik@imperial.ac.uk, p.kormushev@imperial.ac.uk

Acknowledgments: This research has been supported by Dyson Technology Ltd. and computation resources were provided by Microsoft Azure