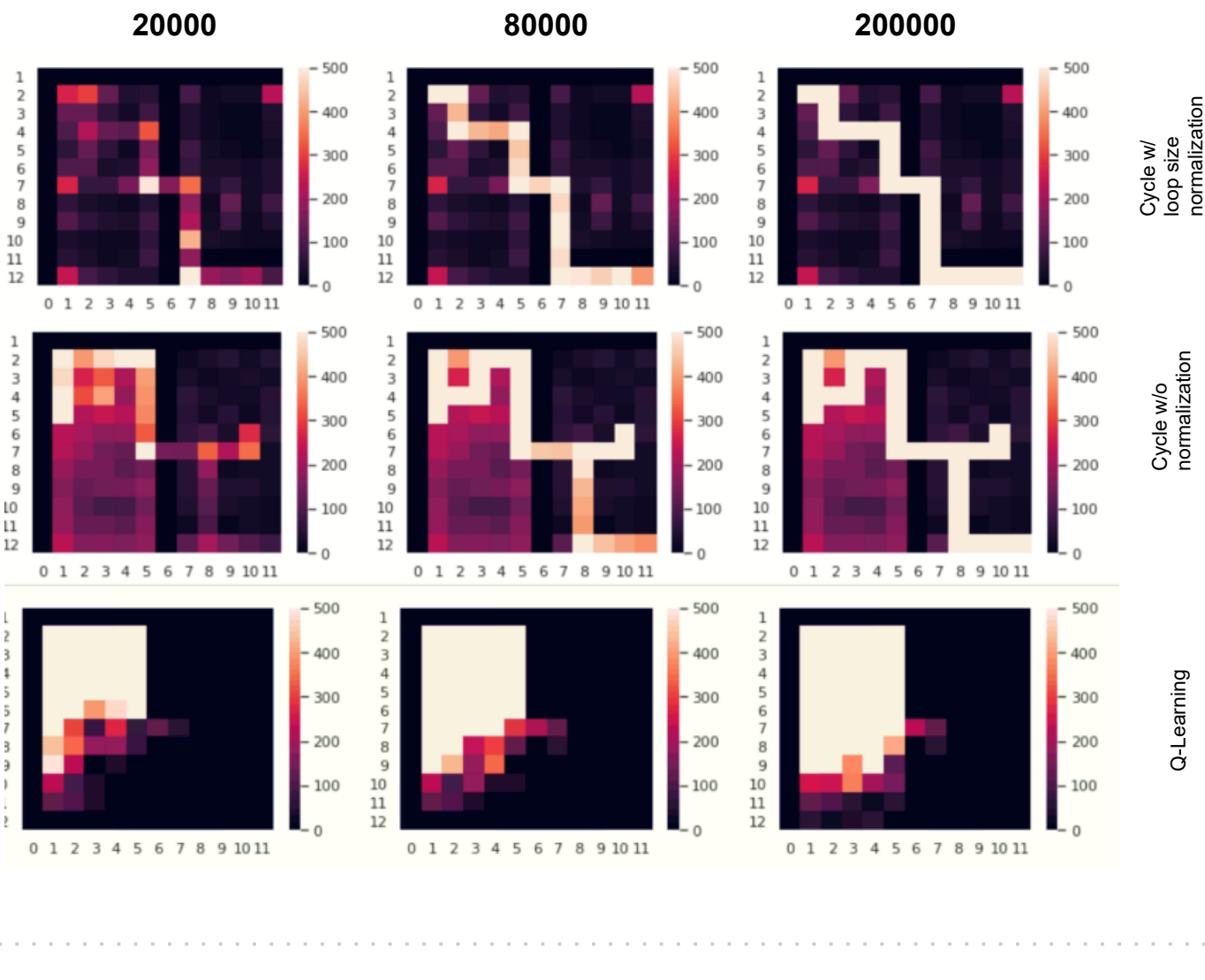
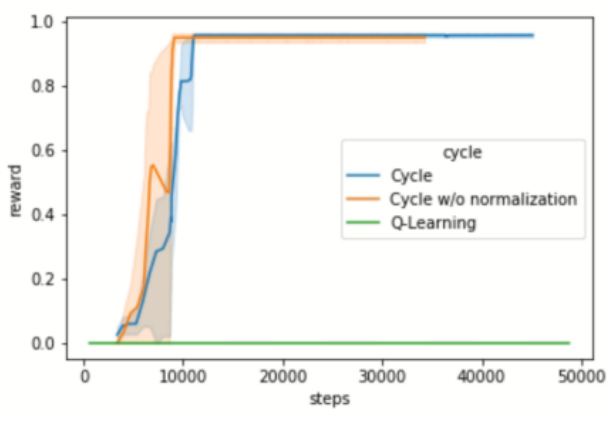
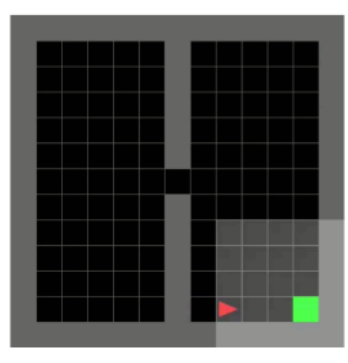
Results on SimpleCrossing 8x8 with one crossing

The cyclophobic agent is able to solve the environment while the Q-learning agent is unable to. Moreover, dividing the fixed cycle penalty -1, by the size of the loop greatly improves exploration.

Note that reward is only given at the green goal state!

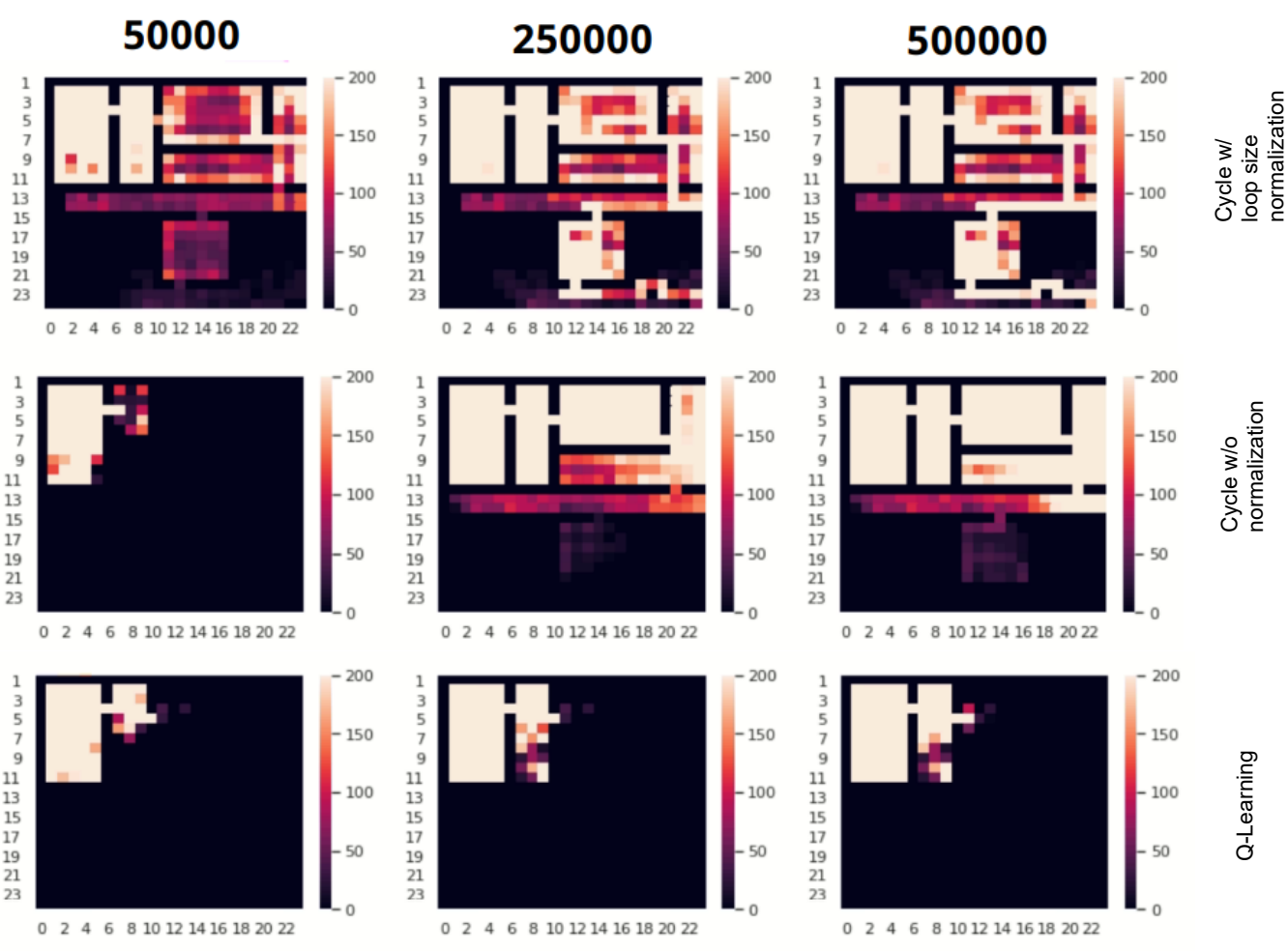


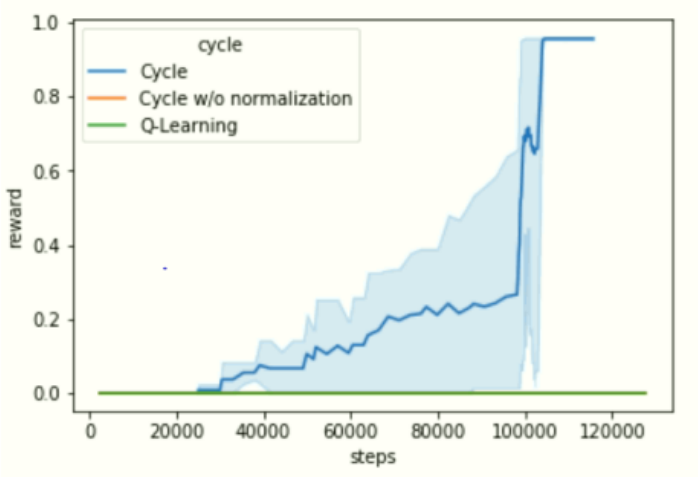


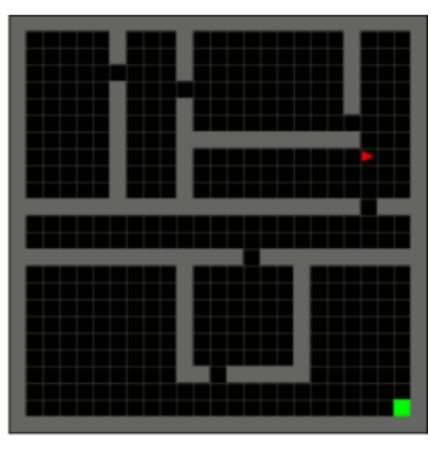


Results on SimpleCrossing 24x24 with multiple crossings and rooms.

Again, the cyclophobic agent is able to solve the environment while the Q-learning agent barely able to get out of the second room. Therefore, cycles as inductive bias are extremely efficient for exploration.





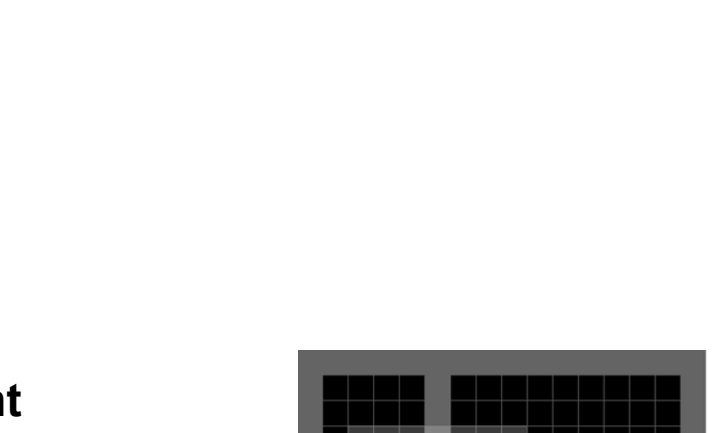


Results on DoorKey environment compared to state of the art.

In the DoorKey environment the agent must pickup a key to open a door. The agent only receives a reward when reaching the green goal state. Position of key, door and wall are shuffled randomly after every episode. By learning a representation of the objects via cycles as inductive bias and the action-vector as representation we are able to be up to par with state of the art method while using considerably less steps.

DoorKey-8x8

DoorKey-11x11



Results on UnlockPickup-8x8 environment.

The UnlockPickup environment requires the agent to open a door with a key and to pickup an object. This is especially difficult because the box can be destroyed erronously by toggling it. It thus requires the agent to not perform this action and to drop the key it is carrying in order to pickup the box.