Reinforcement Learning Agent for Grid World Room Cleaning

To develop an RL agent capable of cleaning dirt patches in a 2D grid world while navigating obstacles and optimizing its path. The agent is implemented using:

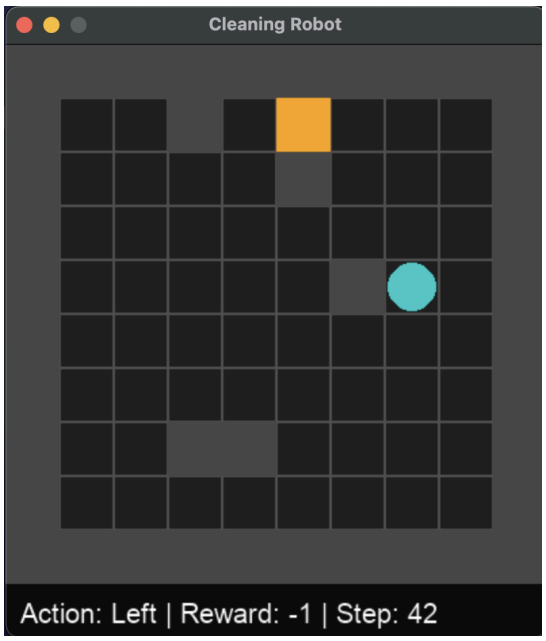
- Q-Learning
- Deep Q-Network (DQN)

Bonus: Support for multiple dirt patches in larger environments.

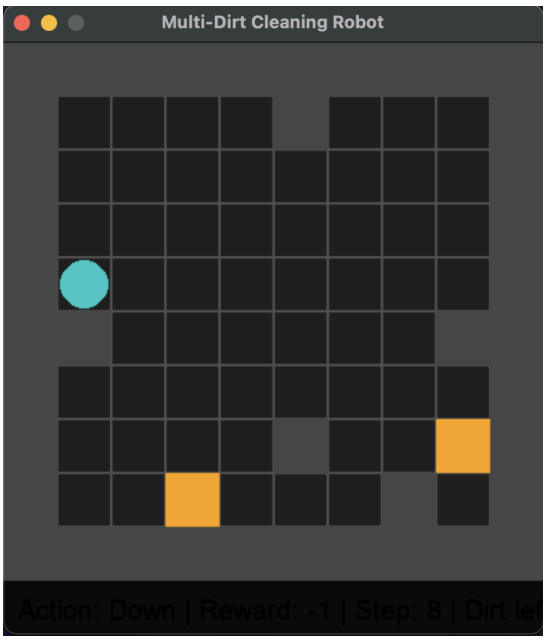
Environment Design

Component	Description
Grid Size	NxN (scalable up to 1000x1000)
Agent Actions	Move up, down, left, right
Obstacles	Impassable grid blocks
Dirt Patches	Cleaning goal (terminal or intermediate)

Interface	reset(), step(action), render() compatible with OpenAI Gym
Visualization	PyGame-based
Reward Structure	+100 (clean dirt), -1 (move), -10 (hit wall)



(Figure1)



(Figure2)

Note:

- **Figure 1** illustrates the running state of the Q-Learning-based Reinforcement Learning agent in a standard grid environment.

- **Figure 2** demonstrates the Q-Learning agent operating in an extended environment with **multiple goal states** (multiple dirt patches), showcasing its ability to generalize and handle more complex scenarios.

Algorithmic Approaches

Q-Learning

- Q-values stored as a dictionary (state-action pairs)
- ε-greedy exploration
- Efficient for small/medium grids

Deep Q-Learning (DQN)

- Neural network approximation of Q-values
- Includes experience replay, target network
- Suitable for larger/high-dimensional environments

Hyperparameter Tuning

Parameter	Q-Learning Range	DQN Value
Learning Rate (α)	0.1 – 0.2	0.2

Discount Factor (γ)	0.9 – 0.99	0.99
Exploration Rate (ε)	0.05 – 0.2	0.5
Batch Size	-	128
Target Update Freq	-	1000
Max Steps/Episode	1000	10

Q-Learning Results by Grid Size

Grid Size	Best Params (α, γ, ε)	Avg Reward	Training Time (s)	Iterations/sec
10x10	(0.2, 0.9, 0.05)	59.06	0.01	3,176.83
100x100	(0.2, 0.95, 0.05)	54.26	0.01	2,805.55

1000x1000	(0.1, 0.99, 0.2)	44.50	0.04	2,510.03

Final Q-Learning Performance (1000 Episodes)

Metric	Value
Avg Reward	92.16
Final Episode Reward	98
Min/Max Reward	-246 / 99
Training Speed	11,936.29 it/s

DQN Results

Metric	Value
Avg Reward	-10.66

Final Episode Reward	-10.00
Min/Max Reward	-22 / 95
Training Speed	159.92 it/s
Best Batch Size	128
Target Update Freq	1000

Visual Analysis

Observation	Description
Q-Learning	Rapid reward increase, good convergence

DQN	High variance, unstable learning, low final reward
Generalization	Q-Learning is more robust for discrete state space
DQN Needs Improvements	Potential gains via CNNs and reward normalization

Q-Learning vs DQN Comparison

Criteria	**Q-Learning**	**DQN**
Avg Reward	92.16	-10.66
Final Reward	98	-10
Min-Max Range	-246 to 99	-22 to 95
Training Speed	11,936.29 it/s	159.92 it/s
Scalability	Excellent	Slower

Stability	High	Low
Tuning Requirement	Low-Moderate	High
Architecture	Tabular	Neural Network

Bonus Task: Multiple Dirt Patches

Approach

- Modified state to include a binary dirt-patch mask
- Extended Q-table to track multiple dirt states

Challenges

Challenge	Description
State Explosion	2^n states for N dirt patches
Memory Usage	Large Q-table due to exponential growth

Efficient Exploration	Required reward shaping and heuristics

Outcome

- Cleaned up to 5 dirt patches in a 20x20 grid using Q-Learning
- Achieved good performance via compact state encoding and episodic resets

Conclusion

Q-Learning significantly outperforms DQN in this discrete, grid-based environment. Its simplicity, stability, and speed make it ideal for problems with manageable state spaces. While DQN is designed for larger or continuous spaces, it failed to converge consistently in this setup without architectural refinements. Future work could explore:

- CNN-based DQN for spatial feature extraction
- Curriculum learning for multiple dirt patches
- Transfer learning for large-scale cleaning tasks