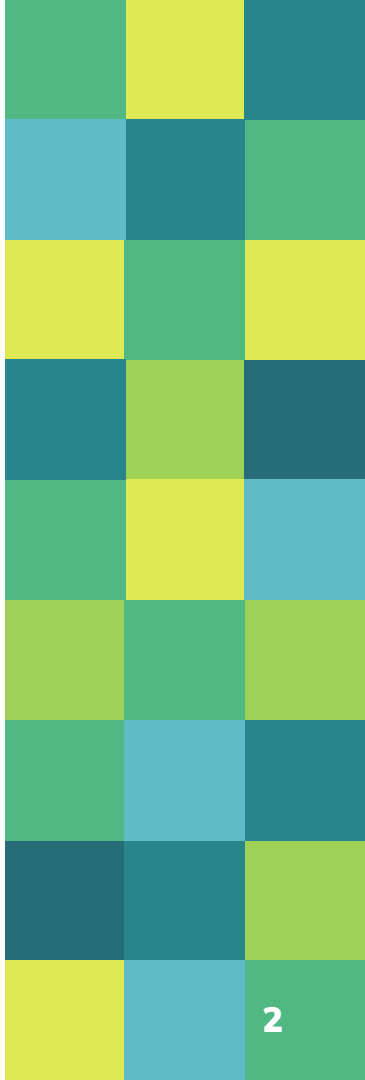
An RL Agent That Can Actually Play 2048

Jyoti Prakash Maheshwari
Prakhar Agrawal

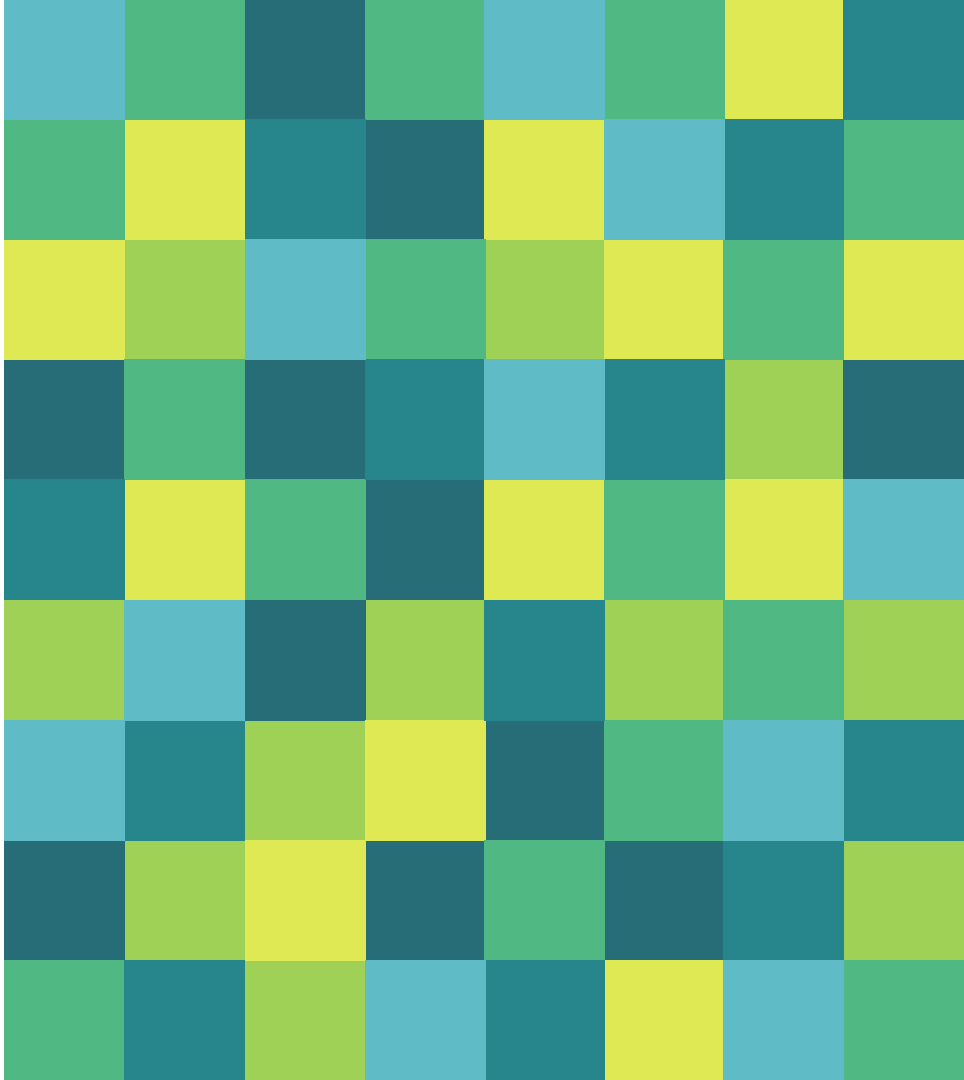
AGENDA

- How 2048 Works
- RL Environment Setup
- Learning Approaches That Sucked
- Learning Approach(es) That Didn't
- Live Demo

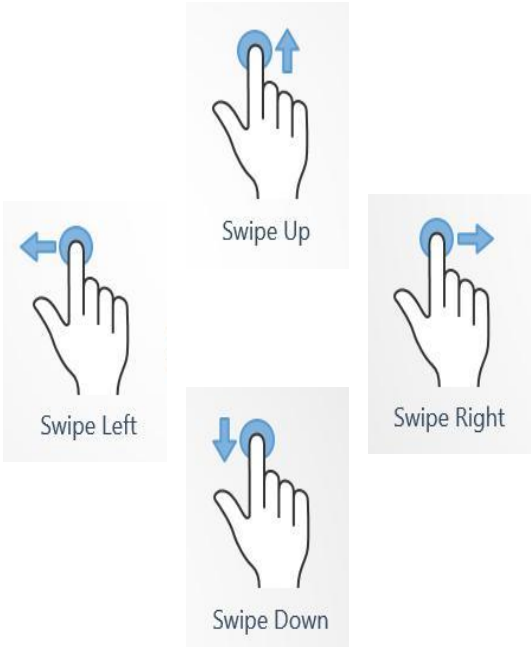


1.

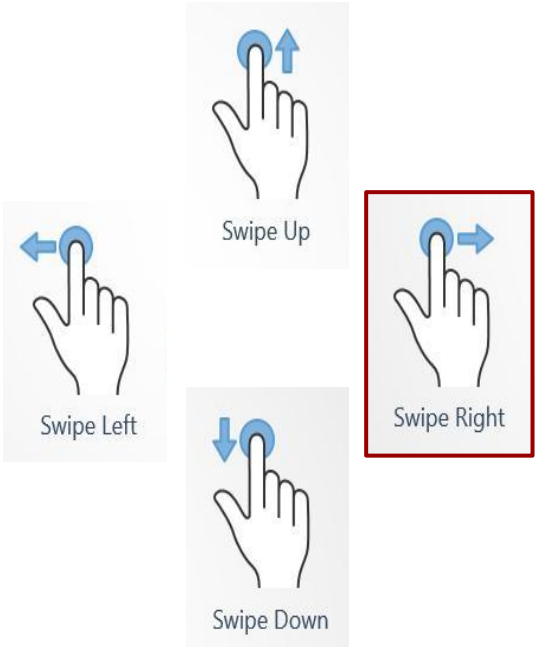
How 2048 Works



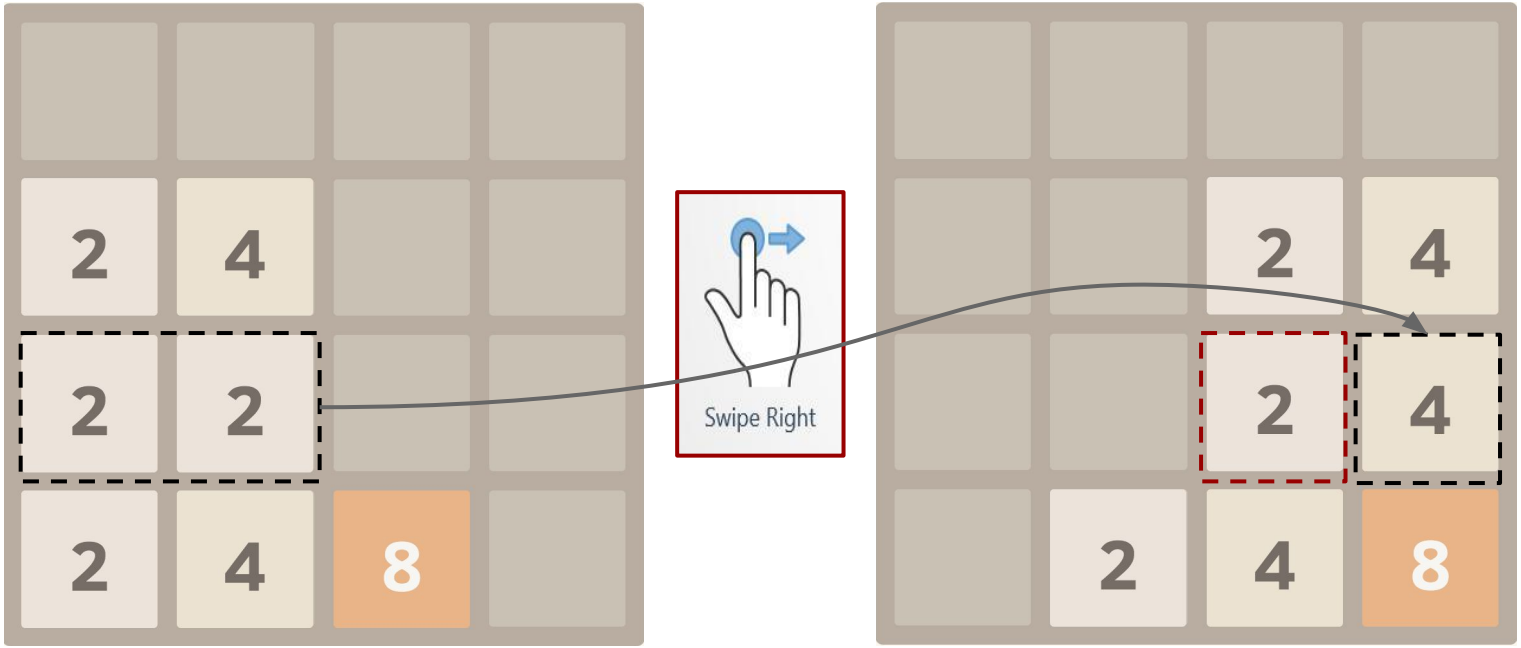
Game Board and Possible Actions



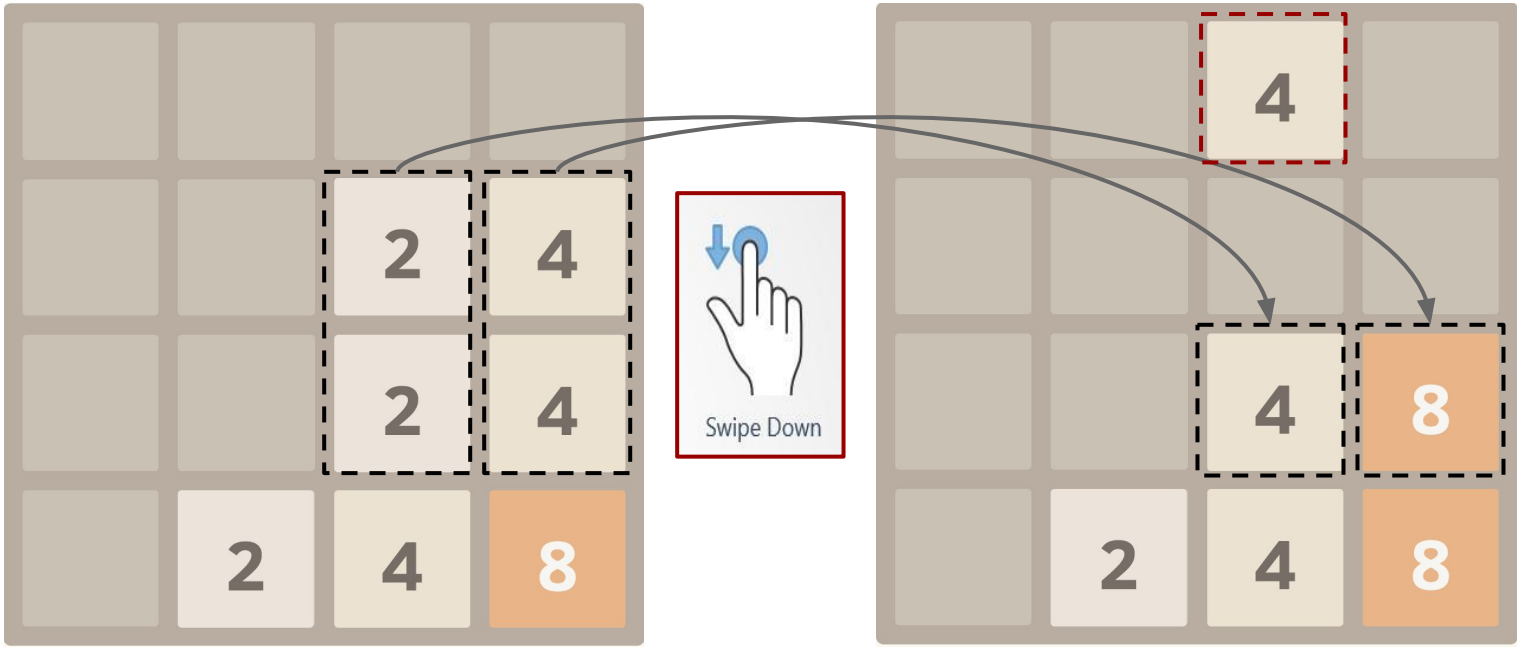
Game Board and Possible Actions



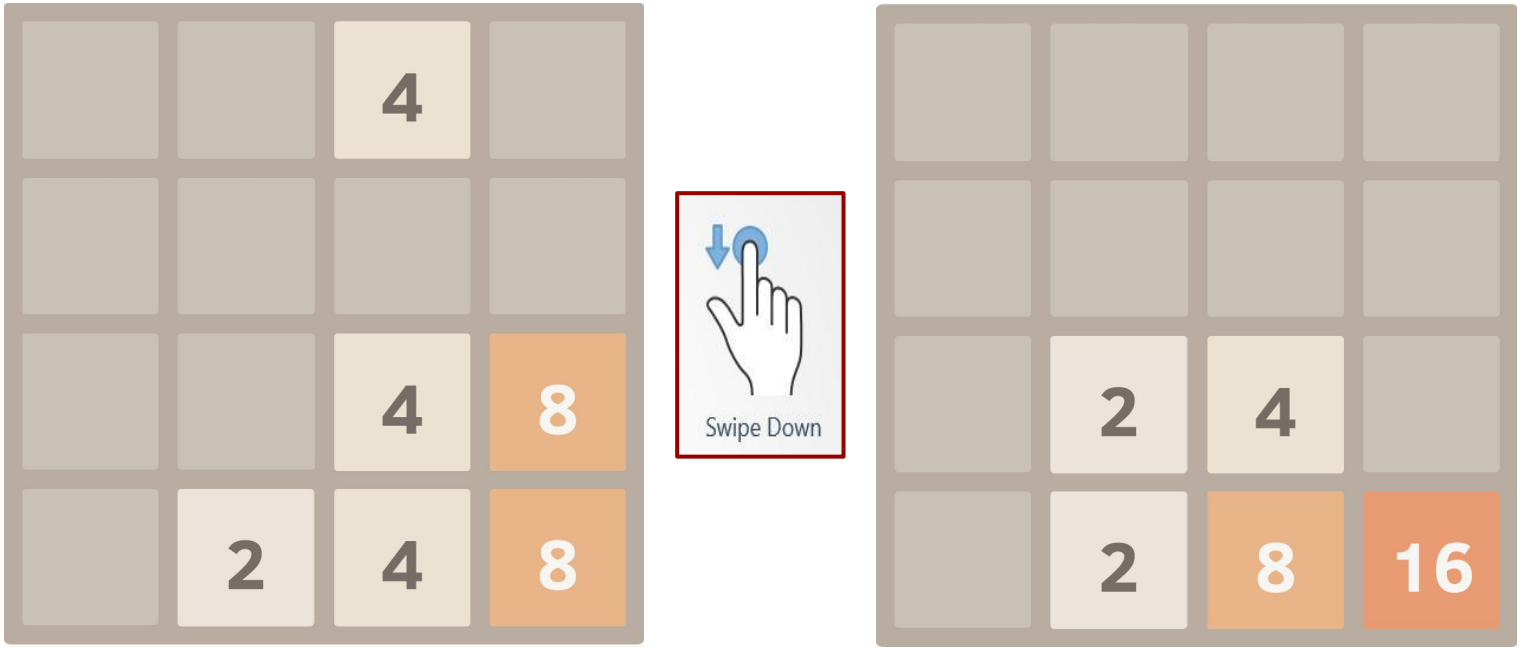
Game Board and Possible Actions



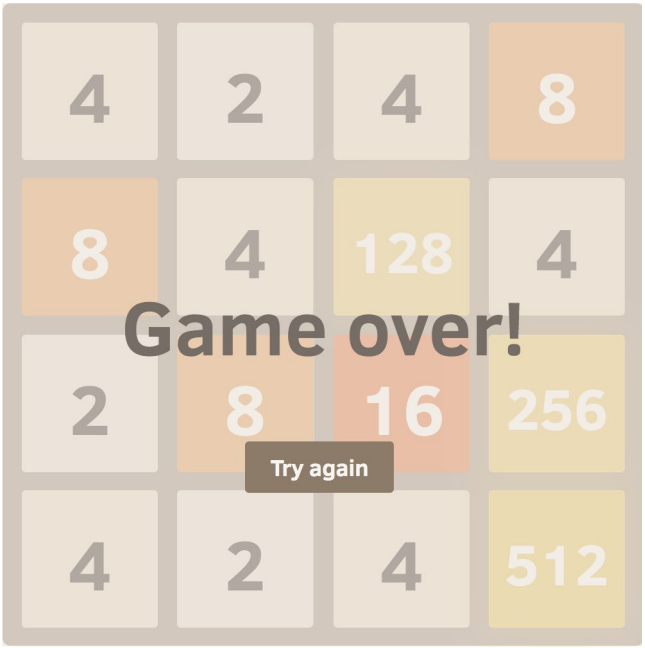
Game Board and Possible Actions



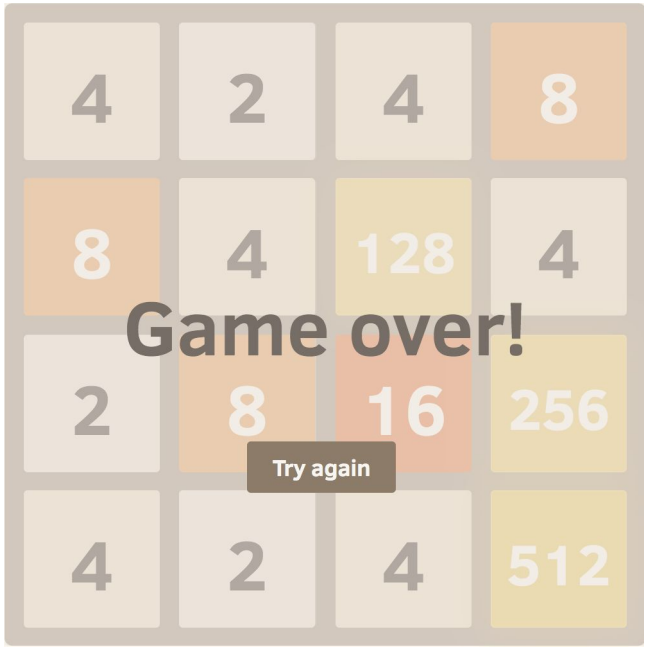
Game Board and Possible Actions



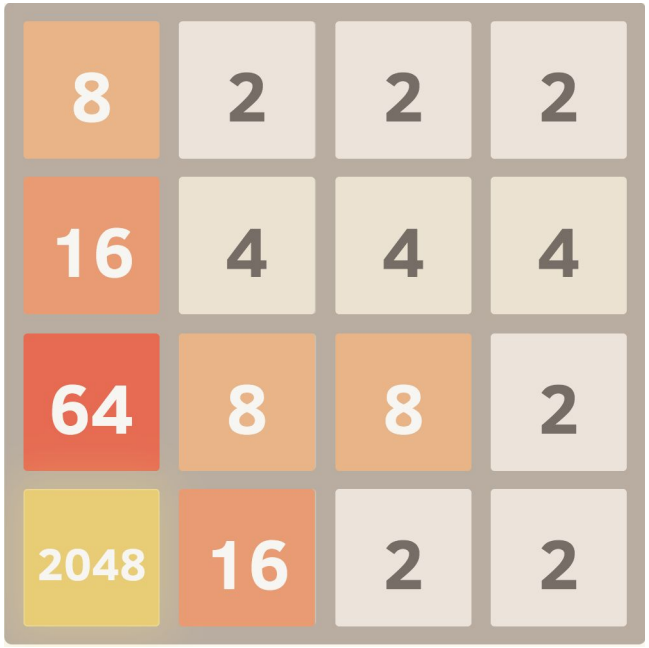
Play Until You Run Out of Moves



Play Until You Run Out of Moves

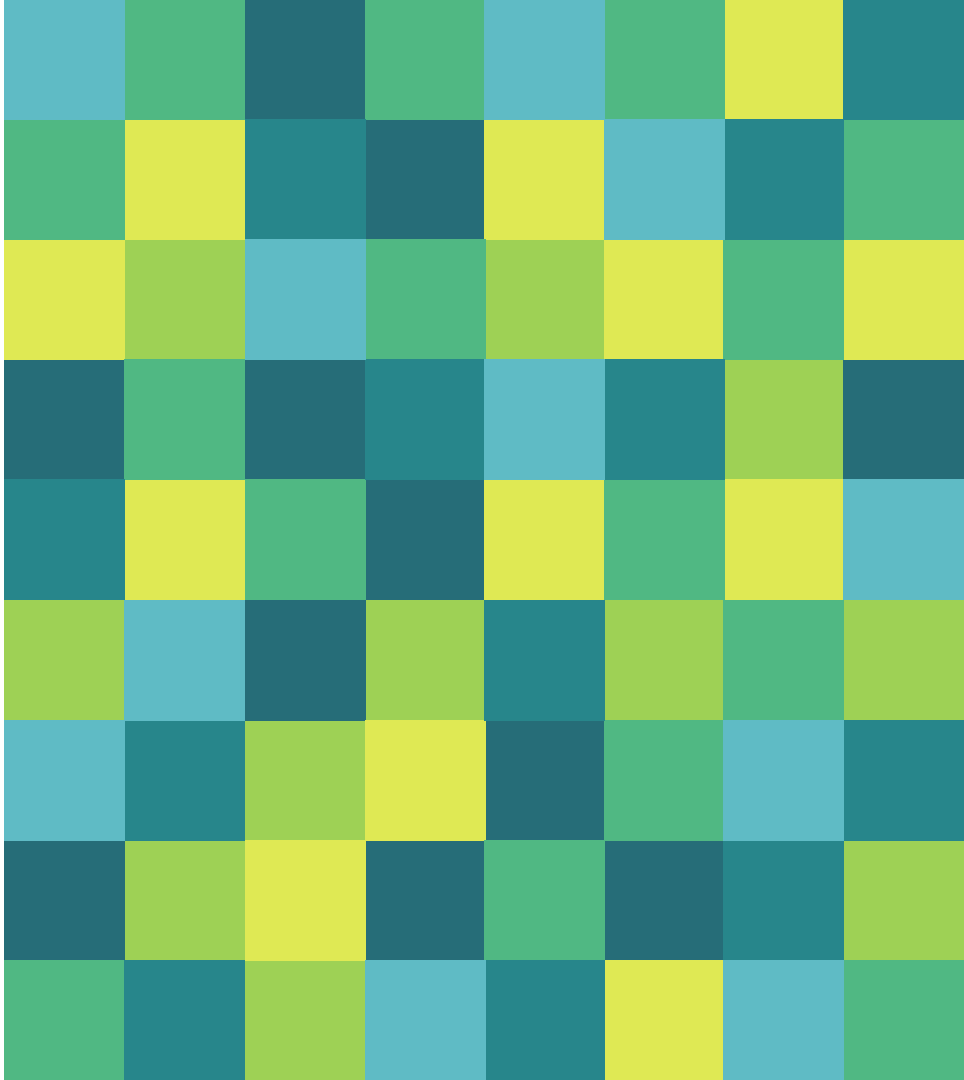


OR



2.

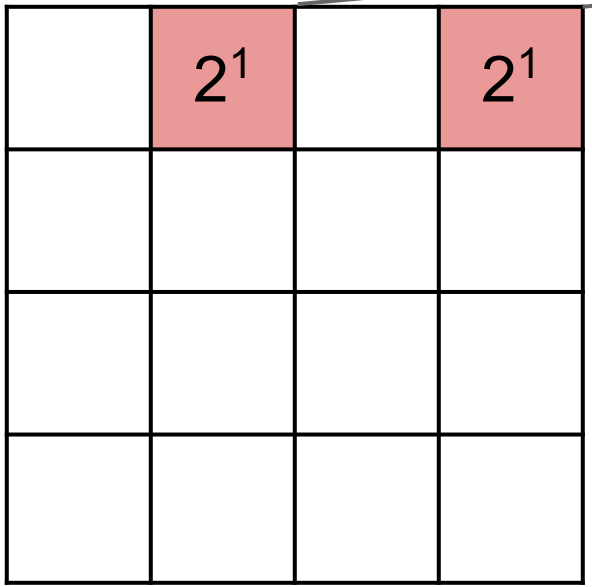
RL Environment Setup



ENVIRONMENT SUMMARY

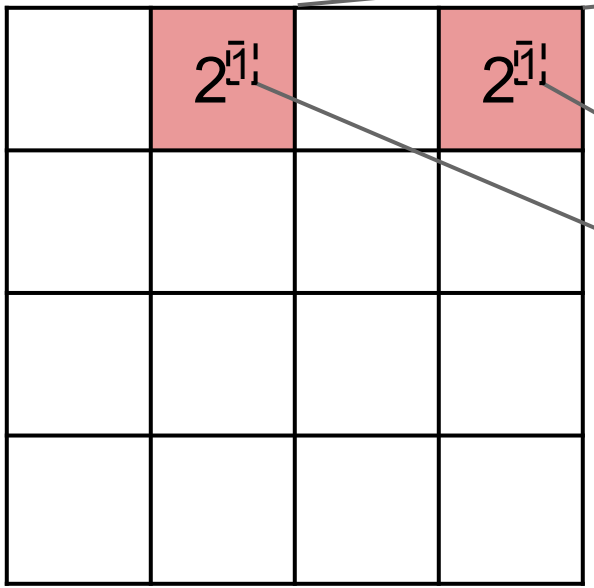
- ENVIRONMENT = Game Board
- AGENT = Trained RL Agent
- STATES = Numbers on Game Board
- ACTIONS = Left, Up, Right, Down
- REWARDS = Total Value of Merged Numbers

Initialization



Two boxes chosen at random

Initialization

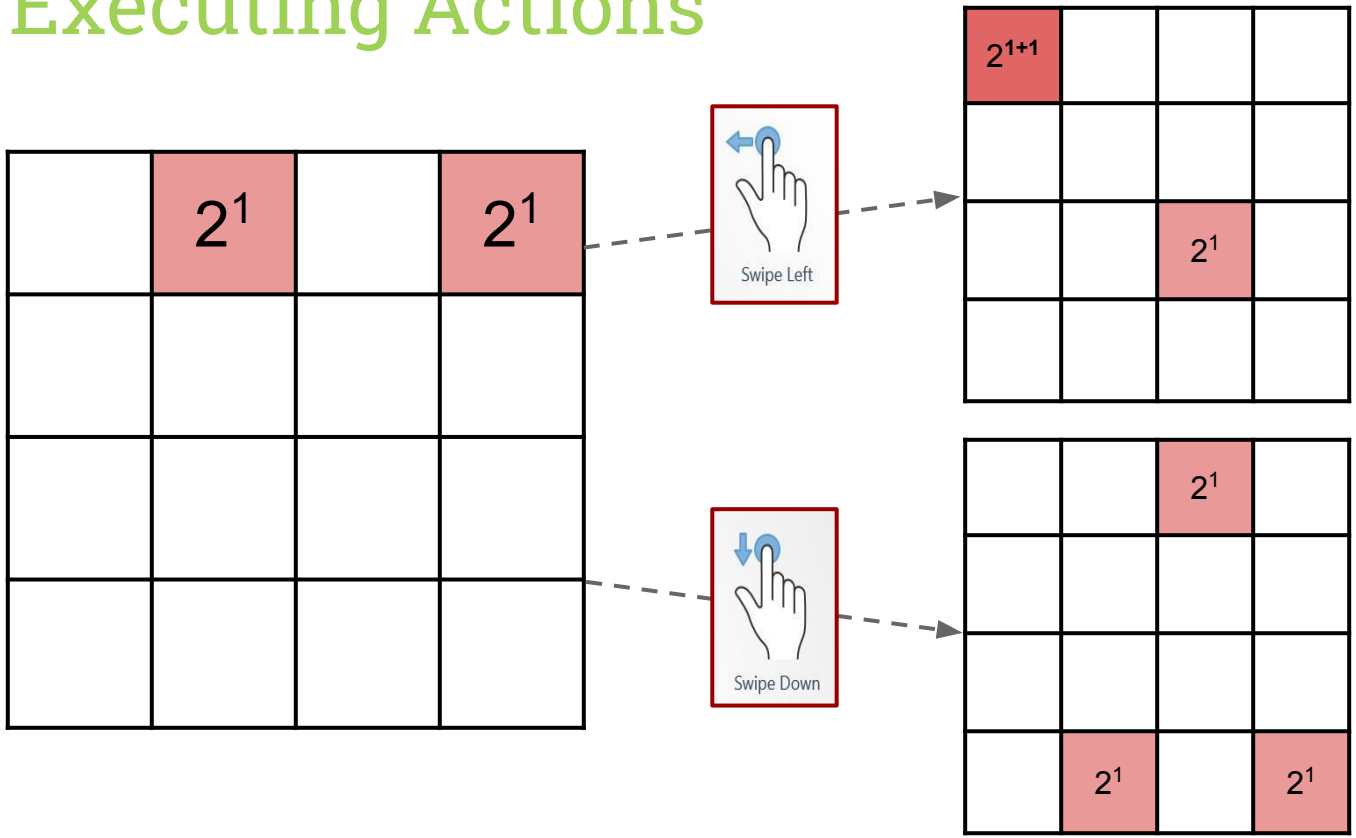


Two boxes chosen at random

Exponent for 2 chosen with prob. distn.

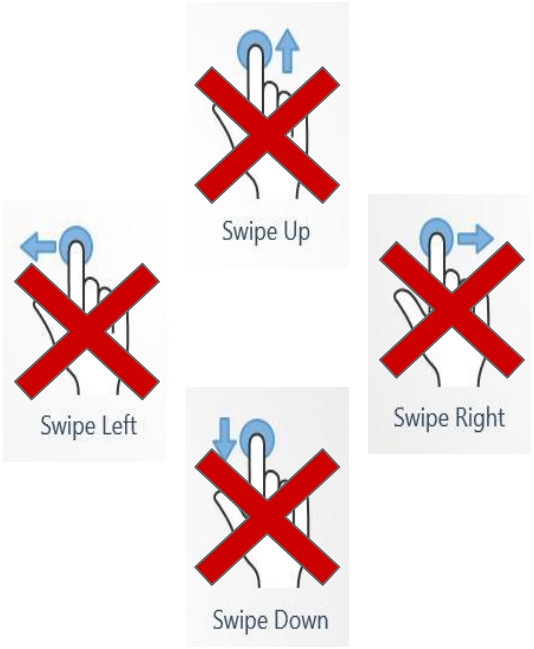
X = {1 [P=0.9]}
{2 [P=0.1]}

Executing Actions



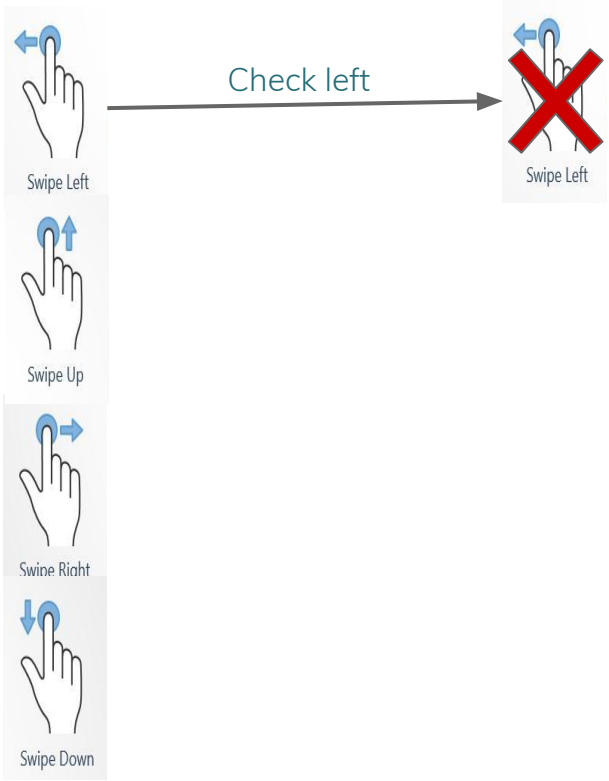
Detecting Termination

2^2	2^1	2^2	2^3
2^3	2^2	2^7	2^2
2^1	2^3	2^4	2^8
2^2	2^1	2^2	2^9



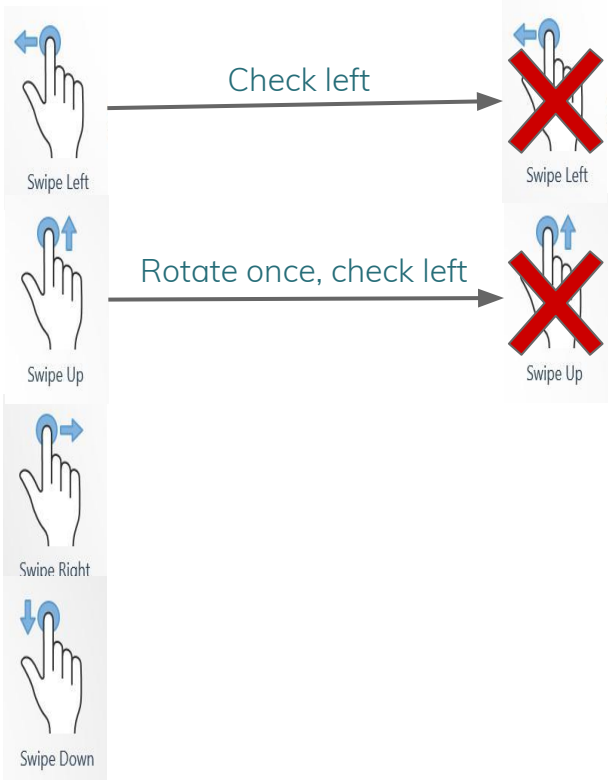
Optimizing Search Space

2^2	2^1	2^2	2^3
2^3	2^2	2^7	2^2
2^1	2^3	2^4	2^8
2^2	2^1	2^2	2^9



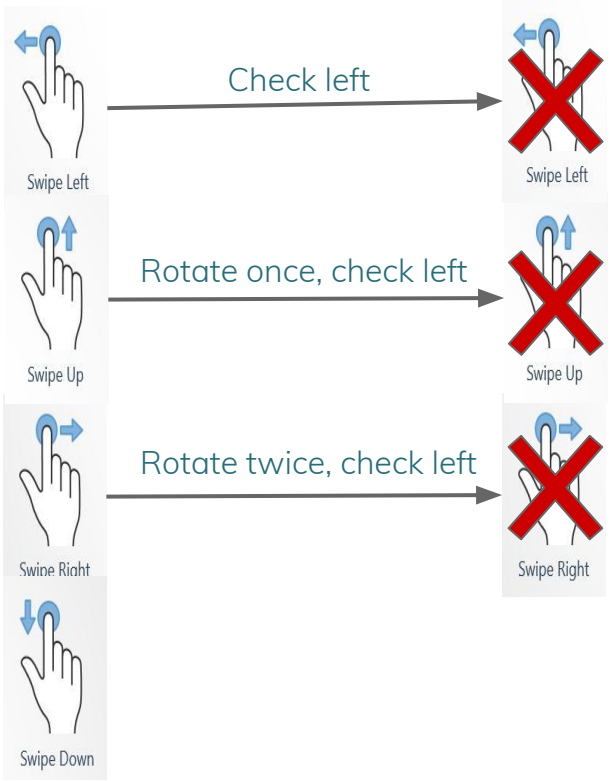
Optimizing Search Space

2^3	2^2	2^8	2^9
2^2	2^7	2^4	2^2
2^1	2^2	2^3	2^1
2^2	2^3	2^1	2^2



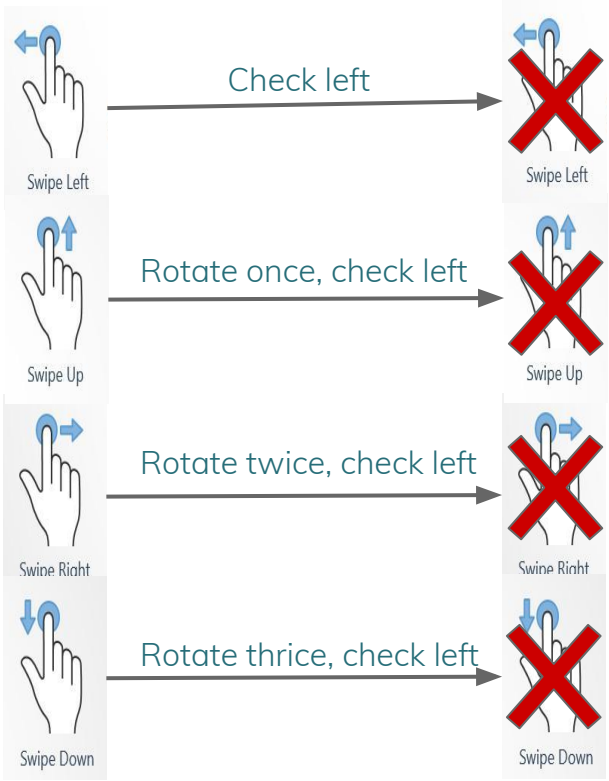
Optimizing Search Space

2^9	2^2	2^1	2^2
2^8	2^4	2^3	2^1
2^2	2^7	2^2	2^3
2^3	2^2	2^1	2^2

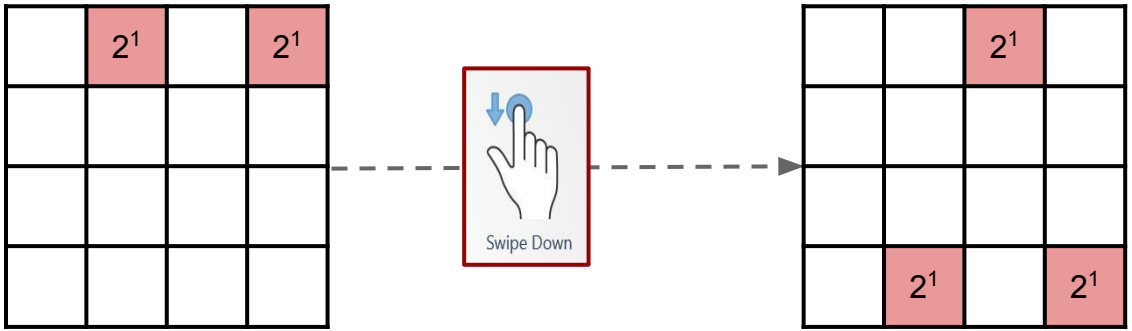


Optimizing Search Space

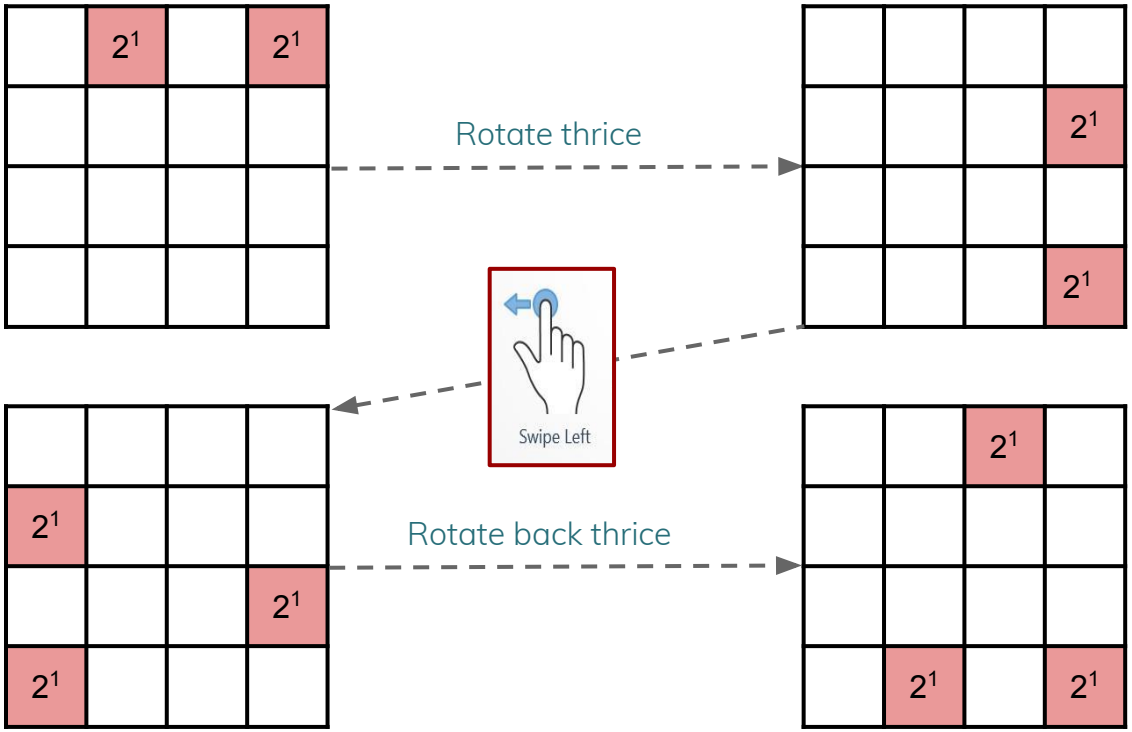
2^2	2^1	2^3	2^2
2^1	2^3	2^2	2^1
2^2	2^4	2^7	2^2
2^9	2^8	2^2	2^3



Optimizing Actions

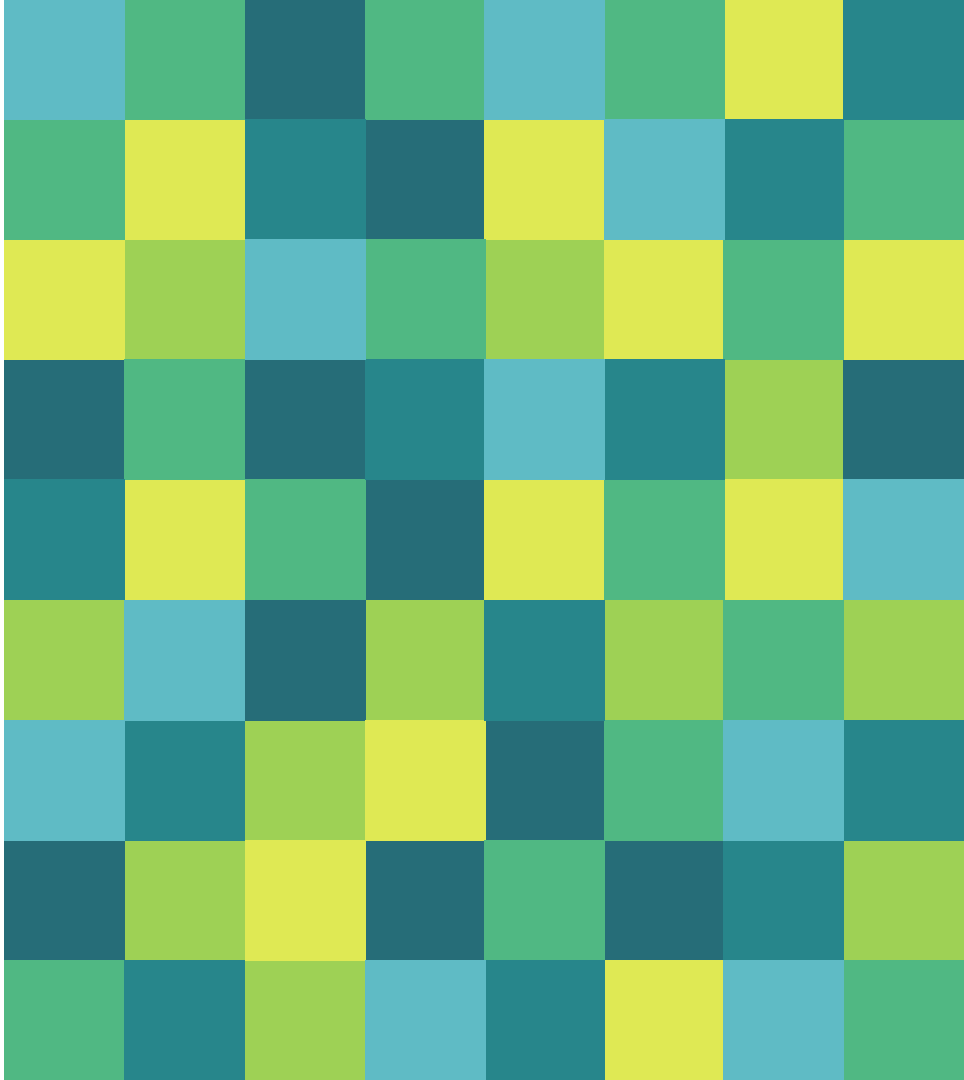


Optimizing Actions



3.

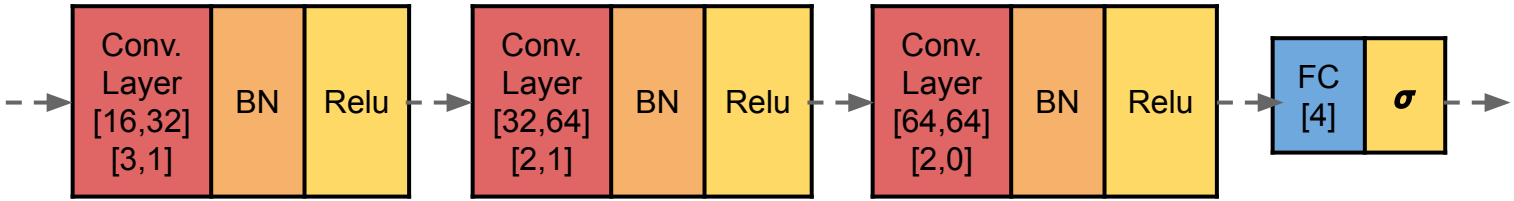
Learning Approaches That Sucked



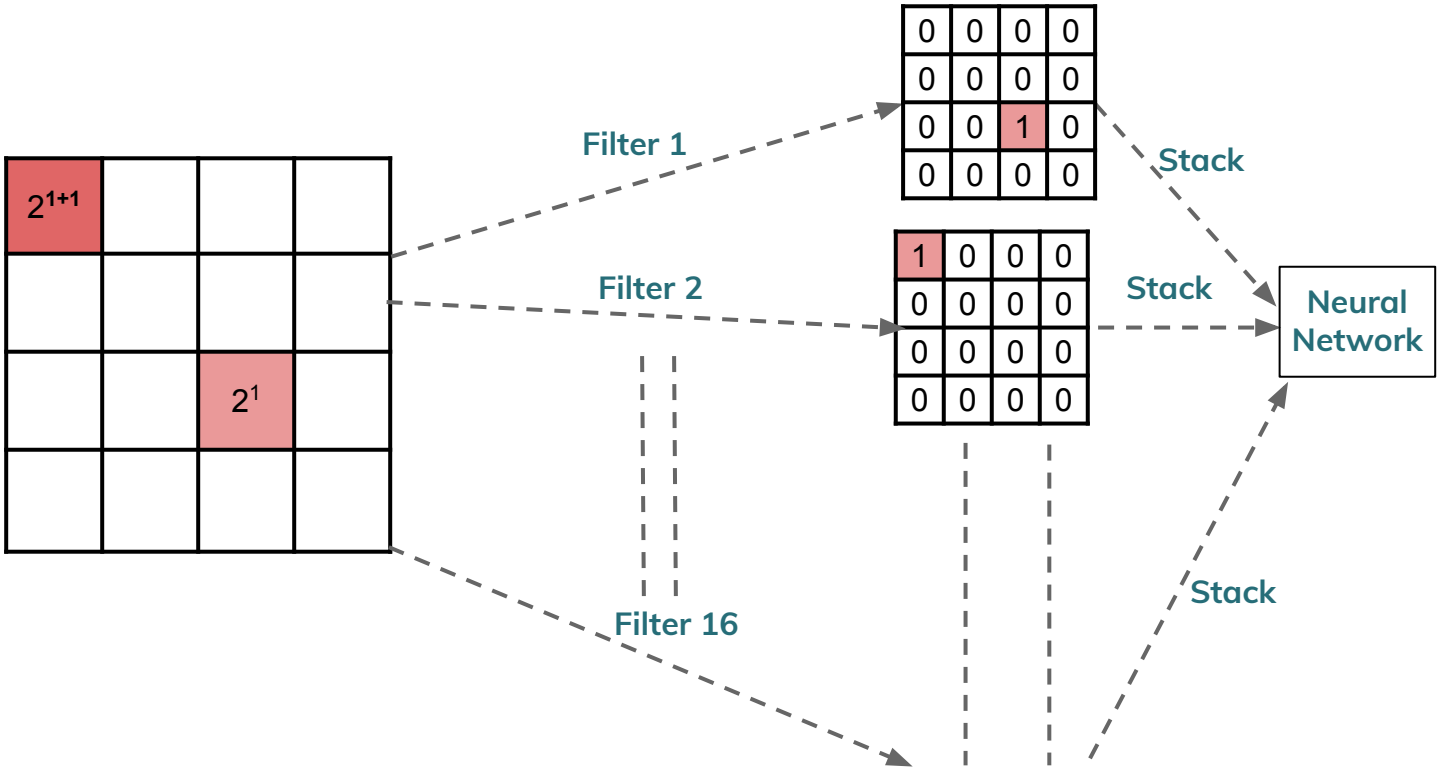
Learning Approaches That Sucked

Approach	Mean Score	Max Score
Random	1093	2736
Q-Learning	1181	3324
DDQN	1205	3530
Human Level	14321	20214

Network used for DDQN

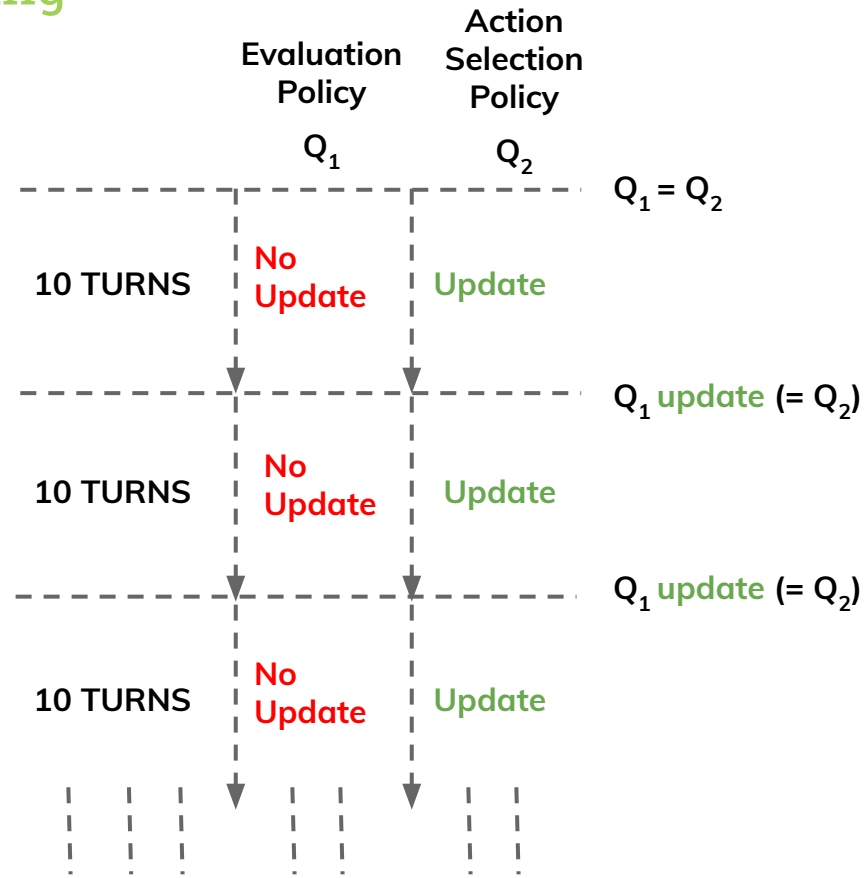


Network used for DDQN



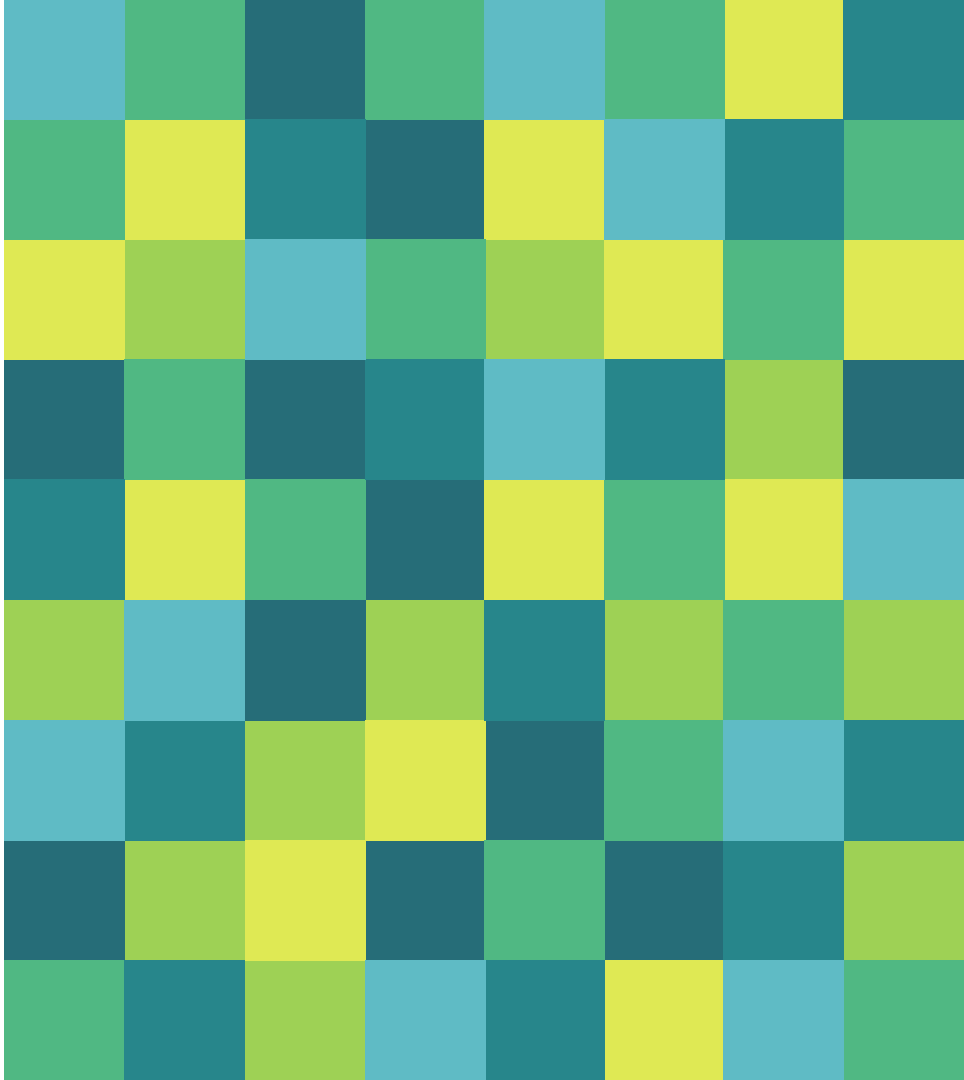
Learning Policy for DDQN

Fixed Q Learning



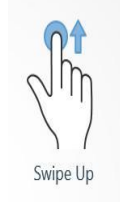
4.

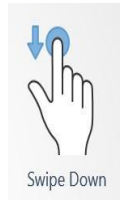
Learning Approach(es) That Didn’t



Monte Carlo Tree Search (One Step)

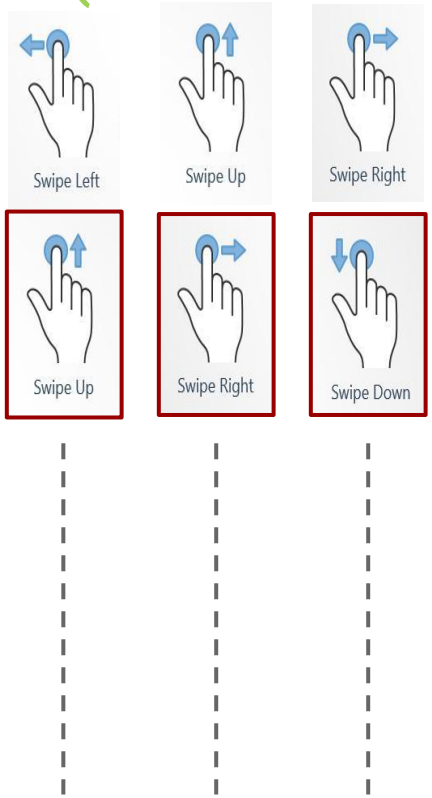
	2^1		2^1





Monte Carlo Tree Search (Multi-Step)

	2^1		2^1



And This Really Works!!

Approach	Mean Score	Max Score
Random	1093	2736
Q-Learning	1181	3324
DDQN	1205	3530
MC (1-step)	1811	6192
MC (2-step)	7648	16132
MC (3-step)	8609	16248
Human Level	14321	20214

5.

Live Demo

Thanks!