PHENIKAA UNIVERSITY

DEEP REINFORCEMENT LEARNING

Topic: Tetris Game Solving using Deep Reinforcement Learning

Presented by: Anh Tuan Nguyen
ID: 22014593
Class: K16 AI&RB

Lectures: ThS. Hoang Dieu Vu
Faculty: Electrical & Electronics Engineering

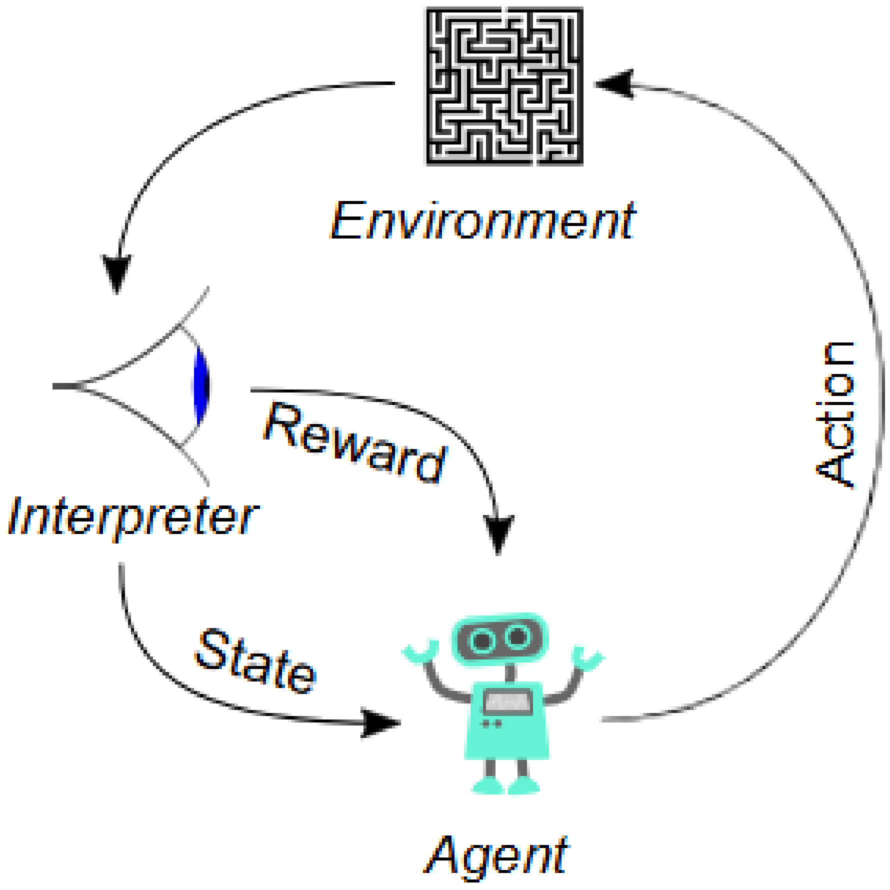
Table of Contents

01

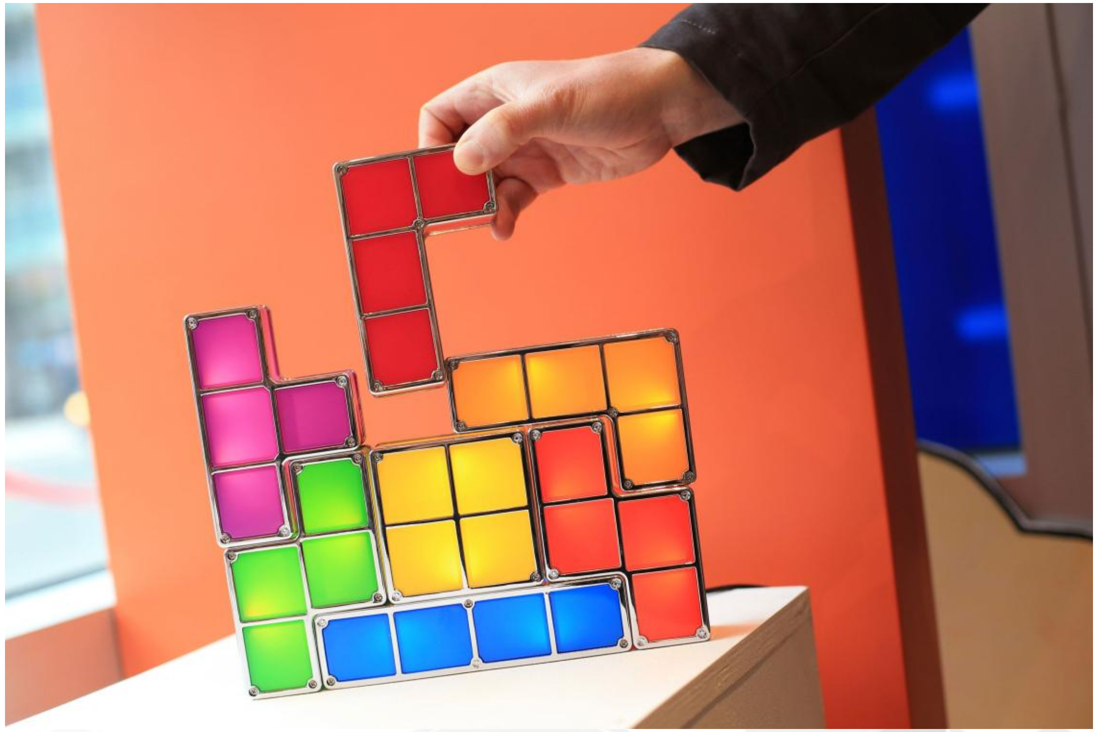
OVERVIEW

OVERVIEW

Deep reinforcement learning blends trial-and-error decision making with deep neural networks: an agent processes high-dimensional inputs (like images), acts within an environment, and learns to maximize cumulative rewards.



WHY TETRIS?



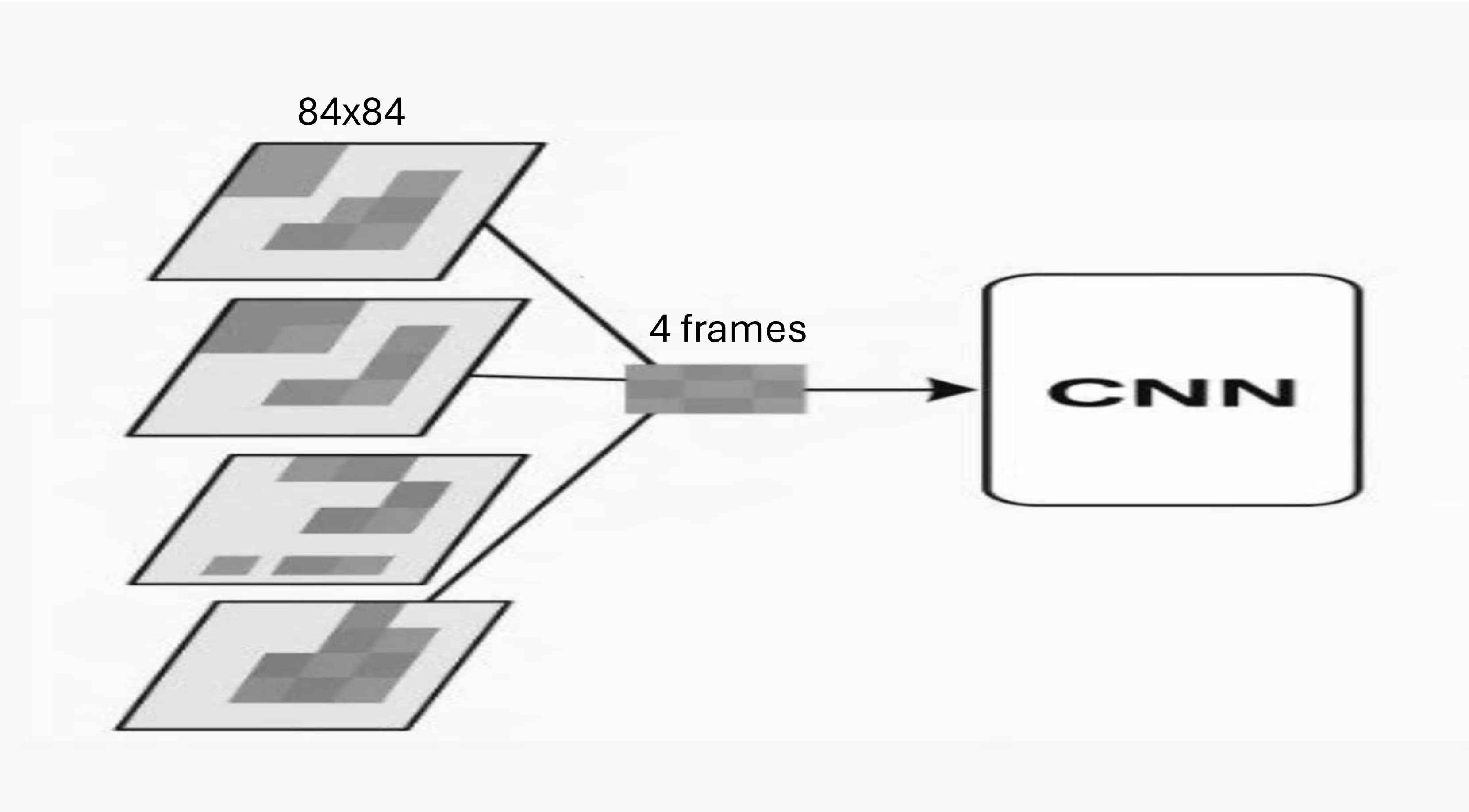
Tetris is a classic block-stacking puzzle with simple rules but deep strategic depth: rotate and place falling pieces to clear full rows. Tetris demands both short-term decision making and long-horizon planning, perfectly illustrating the exploration–exploitation tradeoff.

PHENIKAA UNIVERSITY

02

STATE, ACTION, REWARD OF TETRIS GAME

STATE



ACTION

ACTION	ENCODED
MOVE LEFT	“0”
MOVE RIGHT	“1”
ROTATE	“2”
SOFT DROP	“3”
HARD DROP	“4”
HOLD	“5”

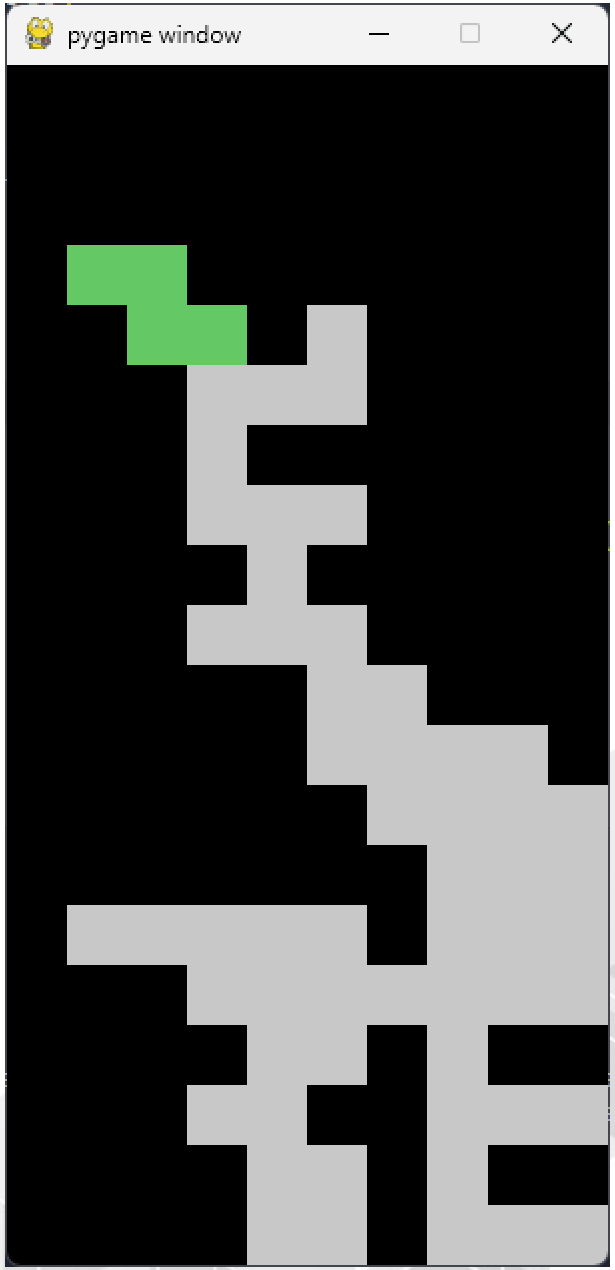
REWARD

EVENT	REWARD
Clear n line (n = 1, 2, 3,...)	LINE_SCORE[n]
Piece "lock" (each placement)	+ SURVIVAL_BONUS
Light hole (above mid-line)	– ALPHA × (# light holes)
Dead hole (below mid-line)	– BETA × (# dead holes)
Game over	– GAME_OVER_PENALTY

03

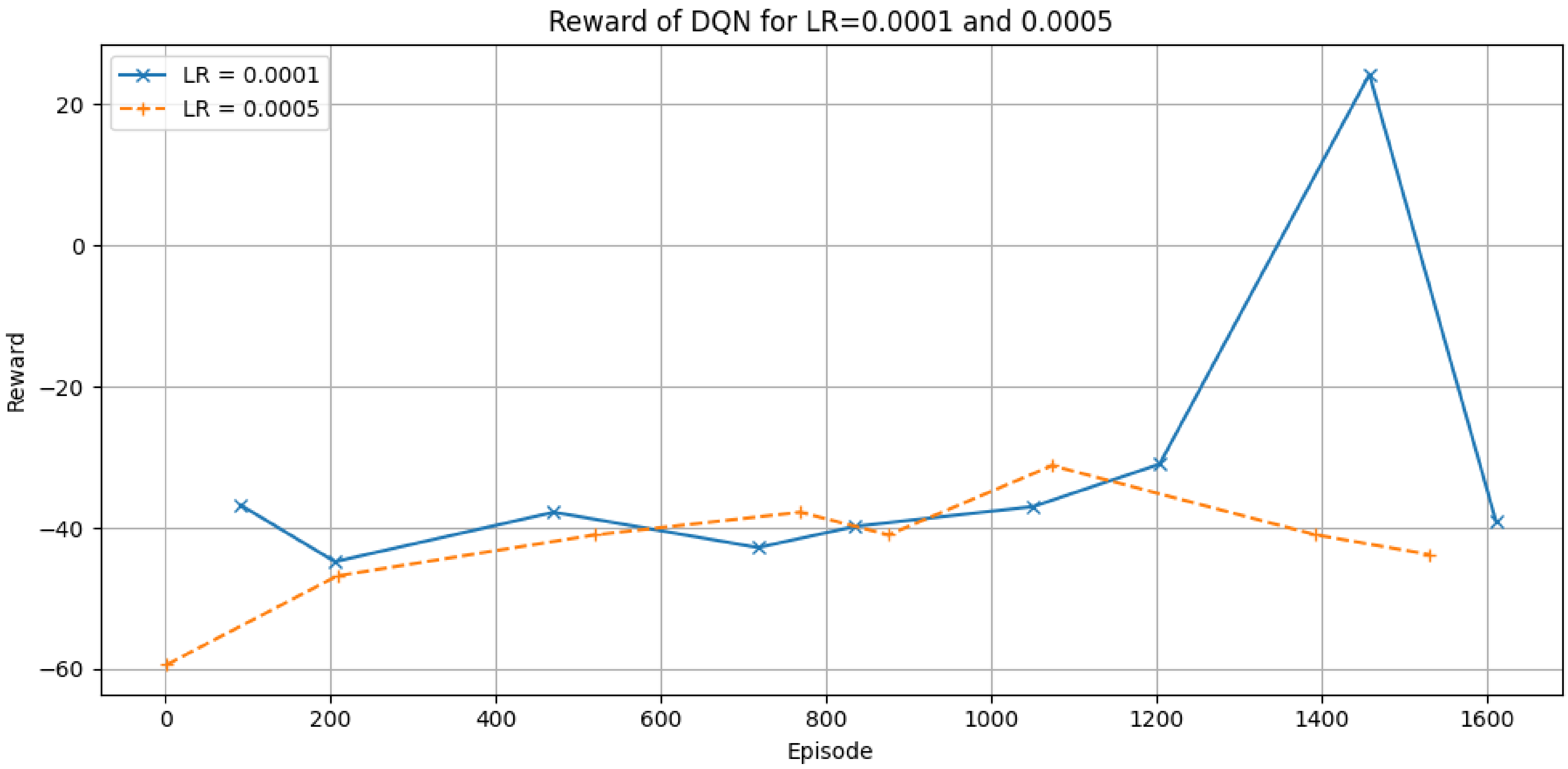
RESULTS AND CONCLUSION

USER INTERFACE (UI)



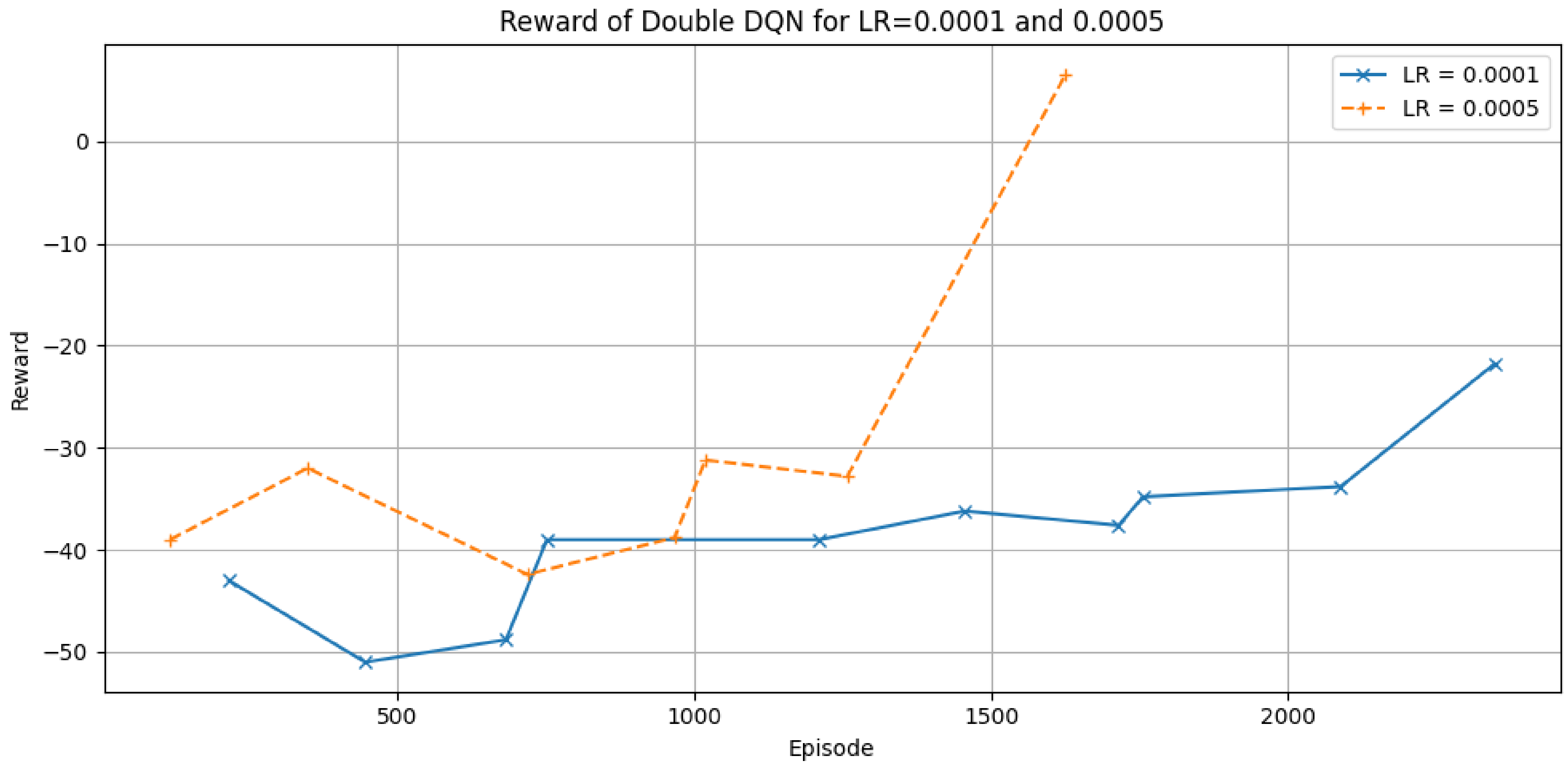
PHENIKAA UNIVERSITY

DQN



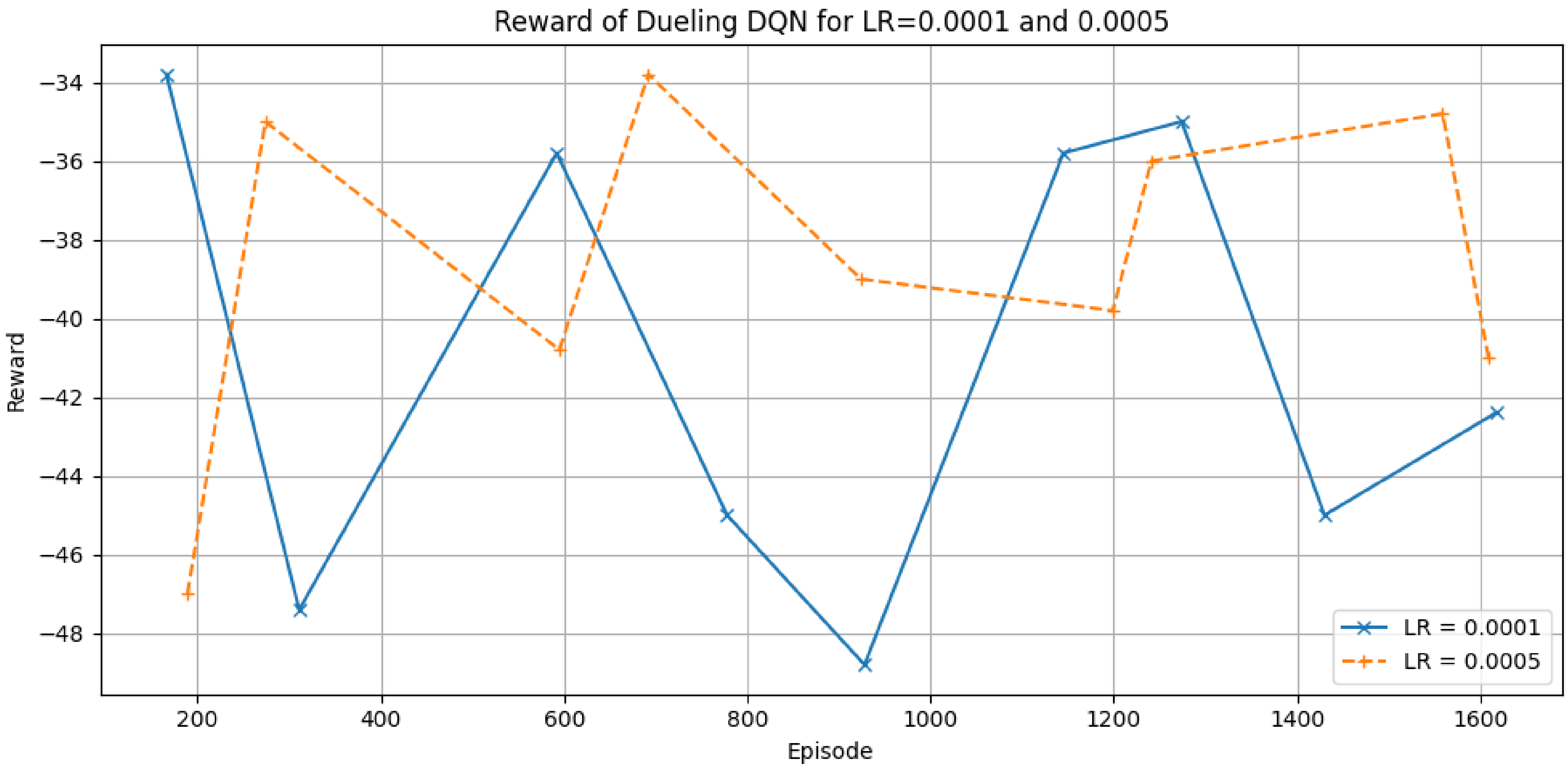
Reward of DQN for LR=0.0001 and 0.0005
LR = 0.0001
LR = 0.0005
Reward
Episode
20
0
−20
−40
−60
0
200
400
600
800
1000
1200
1400
1600

Double DQN

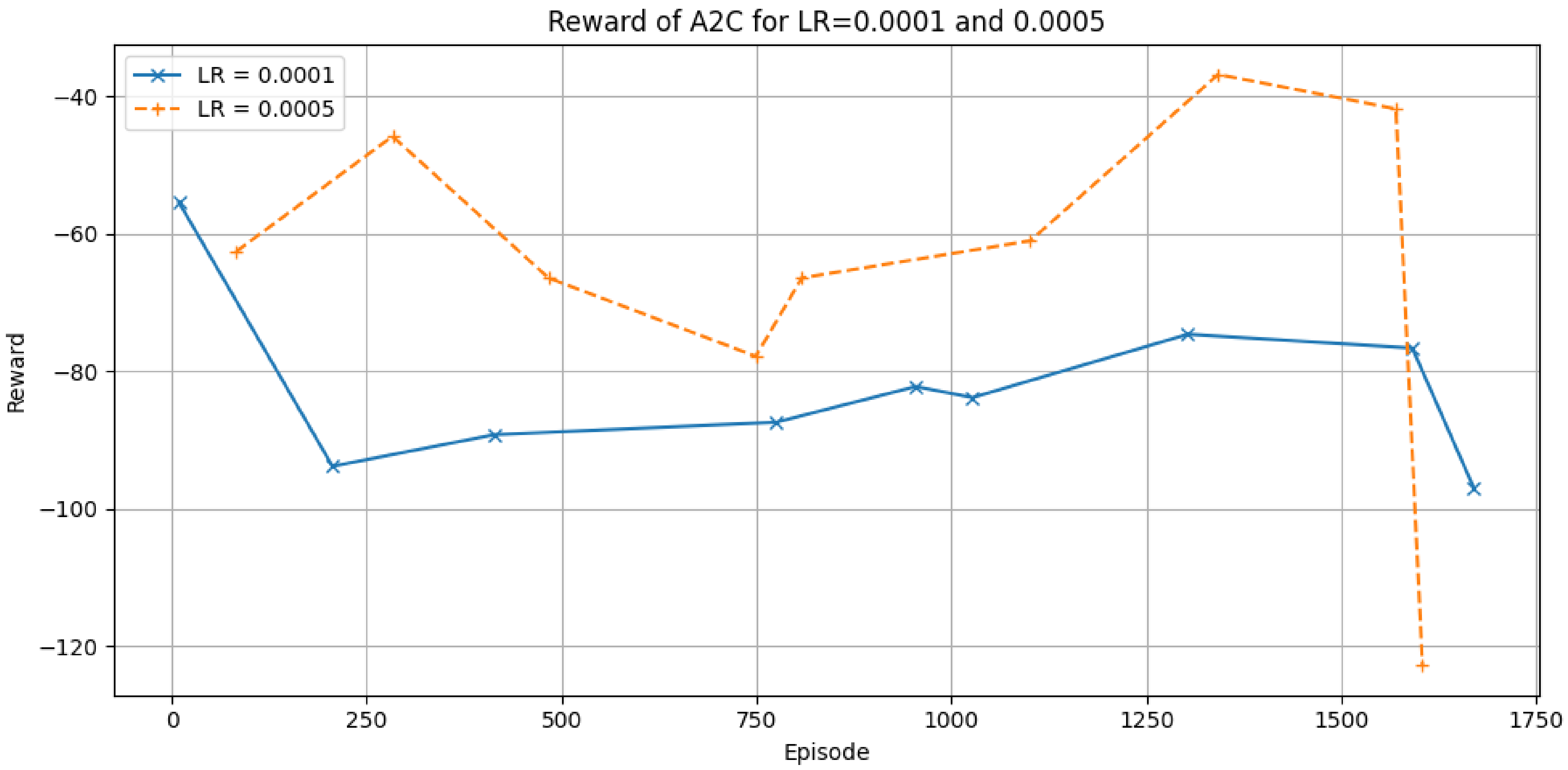


Reward of Double DQN for LR=0.0001 and 0.0005
LR = 0.0001
LR = 0.0005
Reward
Episode
0
−10
−20
−30
−40
−50
500
1000
1500
2000

Dueling DQN


Reward of Dueling DQN for LR=0.0001 and 0.0005
Reward
−34
−36
−38
−40
−42
−44
−46
−48
Episode
200
400
600
800
1000
1200
1400
1600
LR = 0.0001
LR = 0.0005

Advantage Actor–Critic (A2C)



Conclusion

Dueling DQN proved to be the most effective approach in our Tetris experiments, delivering the fastest convergence, highest average scores, and the smoothest learning curves. Double DQN was a close runner-up, significantly reducing overestimation bias compared to vanilla DQN and yielding more stable performance. The A2C agent showed promise as an actor-critic alternative—its policy gradients learned meaningful strategies but converged more slowly and with higher variance. Plain DQN, while capable of learning, suffered from value overestimation and erratic rewards, making it the least reliable of the four.

THANK YOU

For Your Attention