Deep Reinforcement Learning Project Presentation

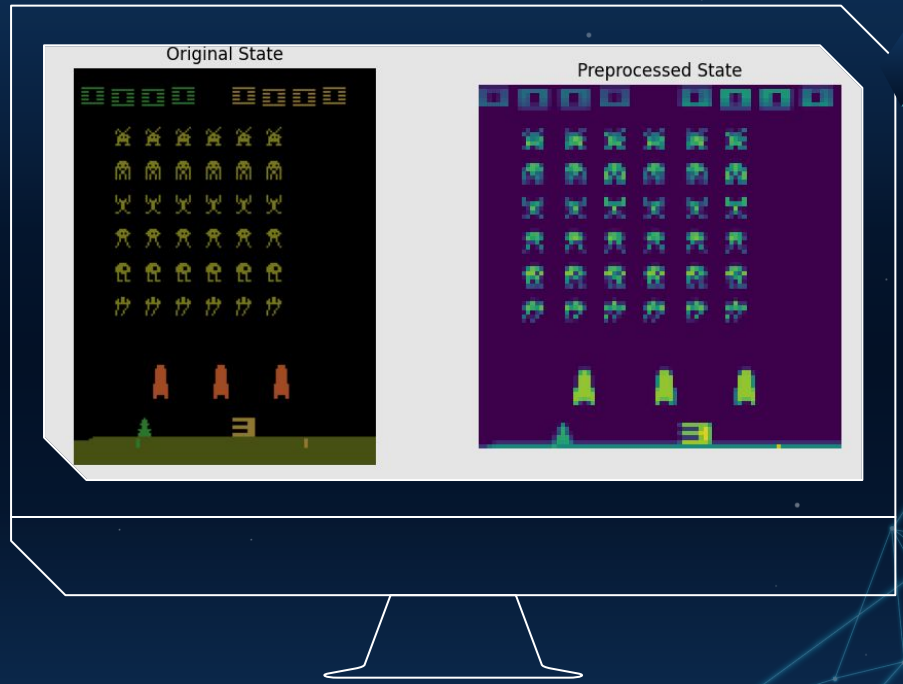
Stavros Nikolaidis
3975

Environment

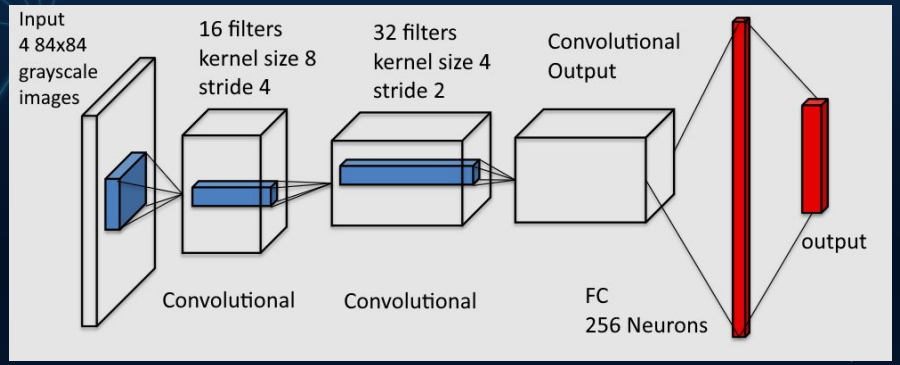
Gymnasium by Farama Foundation [1]

Arcade Learning Environment [2]

Game: Space Invaders



Network Architectures

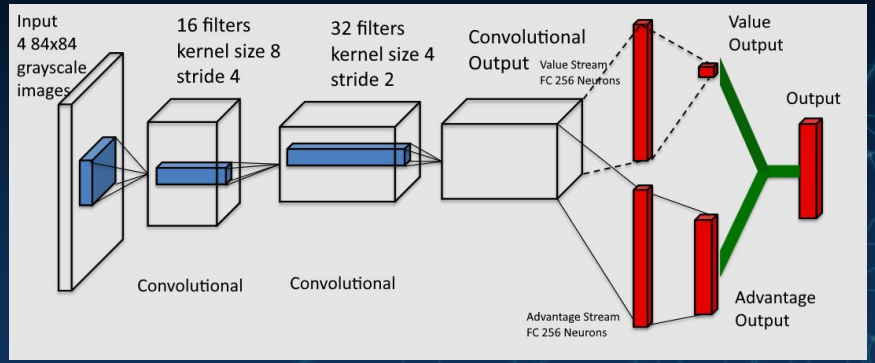


CNN

PyTorch

Dueling Network [3]

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|\mathcal{A}|} \sum_{a'} A(s,a') \right)$$



Deep Q Learning - Key Elements

- Neural network predicts Q-values for each action given a state.
- Epsilon-greedy policy:
  - Exploration (epsilon): Random action selection.
  - Exploitation (1-epsilon): Selects action with the highest predicted Q-value.
- Memory
  - Simple Memory (Random Sampling) or
  - Prioritized Experience Replay
    - Sum Tree structure with a fixed size.
    - Minibatch sampling based on the priority of stored experiences.
    - TD error to compute priority = (|TD error| + ε)^α with experiences having higher TD errors contributing more to the learning process.
- Rewards
  - Default (Raw) Rewards from the environment
  - Normalized Rewards [-1, 1]
  - Simple Rewards (-1 and +1)

Deep Q Learning - Implementation

1. Q-Value Initialization (both main_network and target_network).
2. Action Selection with Epsilon Greedy
3. Action Execution
4. Store Experience (state, action, reward, next_state, done)
5. Mini-Batch Sampling from memory and training
6. Q-Network Update (Training)
    a. Calculate the target network Q-values through Bellman Equation
    b. Calculate the loss between the main network Q-values and the target Q-values
    c. Perform gradient descent to minimize the loss and update the weights of the main Q-network.
7. Target Network Update
    a. After some steps update the target Q-network by copying the weights from the main Q-network to the target Q-network.
8. Epsilon Decay
    a. Gradually decrease the epsilon value to reduce the exploration rate over time.

Experiment 0 - Random

Each step the agent chooses a random move

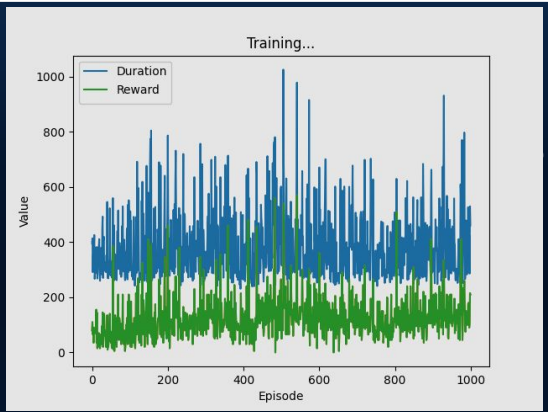
Min Score	Max Score	Median Score	Average Score
0	185	50	62.85

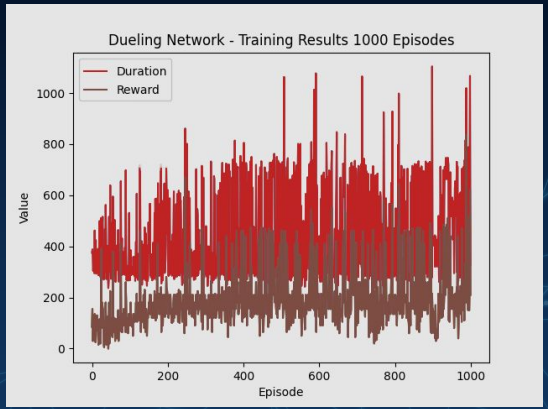
100 Episodes Test on Hard mode

Experiments 1 & 2 – CNN vs Dueling

Hyper-parameters:

- Architecture: CNN vs Dueling Network
- Episodes Trained = 1000
- Memory Size = 10000 (Random Sampling)
- minibatch_size = 8 (train per step)
- γ = 0.99
- ε_start = 1.0
- ε_min = 0.1
- ε_decay_rate = 0.995 (per episode)
- target_network_update_rate = 1000
- Adam Optimizer
  - learning_rate = 0.00025
- Rewards: Default Rewards
- Mode: Hard





Experiments 1 & 2 – CNN vs Dueling

CNN

Table 1: Απόδοση μοντέλων Πειράματος 1 (CNN) για 100 επεισόδια (Hard δυσκολία)

Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	20	495	95	121.05
200	90	430	165	175.65
300	50	415	127.5	141.15
400	60	450	120	143.65
500	30	405	90	105.45
600	5	355	75	103.7
700	65	410	135	166.35
800	30	210	90	94.1
900	50	415	120	131.2
1000	105	550	157.5	186.95

Dueling Network

Table 2: Απόδοση μοντέλων Πειράματος 2 (Dueling Network) για 100 επεισόδια (Hard δυσκολία)

Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	75	210	120	125.2
200	45	470	135	138.15
300	80	430	155	160.8
400	120	490	270	282.75
500	90	590	260	263.45
600	125	645	215	262.35
700	65	745	210	252.4
800	45	745	180	210.9
900	25	370	60	97.45
1000	90	470	210	214.2

Table 3: Σύγκριση CNN και Dueling Network Median Scores

Episodes Trained	Median Scores ανά Μοντέλο									
	100	200	300	400	500	600	700	800	900	1000
CNN	95	165	127.5	120	90	75	135	90	120	157.5
Dueling	120	135	155	270	260	215	210	180	60	210

From Hard to Easy - CNN

Table 1: Απόδοση μοντέλων Πειράματος 1 (CNN) για 100 επεισόδια (Hard δυσκολία)

Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	20	495	95	121.05
200	90	430	165	175.65
300	50	415	127.5	141.15
400	60	450	120	143.65
500	30	405	90	105.45
600	5	355	75	103.7
700	65	410	135	166.35
800	30	210	90	94.1
900	50	415	120	131.2
1000	105	550	157.5	186.95

Hard

Table 4: Απόδοση μοντέλων Πειράματος 1 (CNN) για 100 επεισόδια (Easy δυσκολία)

Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	50	770	180	194.3
200	75	740	180	196.05
300	50	465	155	170.65
400	50	635	185	208.7
500	5	585	125	144.75
600	155	260	155	162.25
700	75	530	180	201.05
800	35	645	135	162.65
900	30	605	140	172.6
1000	15	560	110	154.5

Easy

From Hard to Easy - Dueling

Table 2: Απόδοση μοντέλων Πειράματος 2 (Dueling Network) για 100 επεισόδια (Hard δυσκολία)

Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	75	210	120	125.2
200	45	470	135	138.15
300	80	430	155	160.8
400	120	490	270	282.75
500	90	590	260	263.45
600	125	645	215	262.35
700	65	745	210	252.4
800	45	745	180	210.9
900	25	370	60	97.45
1000	90	470	210	214.2

Hard

Table 5: Απόδοση μοντέλων Πειράματος 2 (Dueling Network) για 100 επεισόδια (Easy δυσκολία)

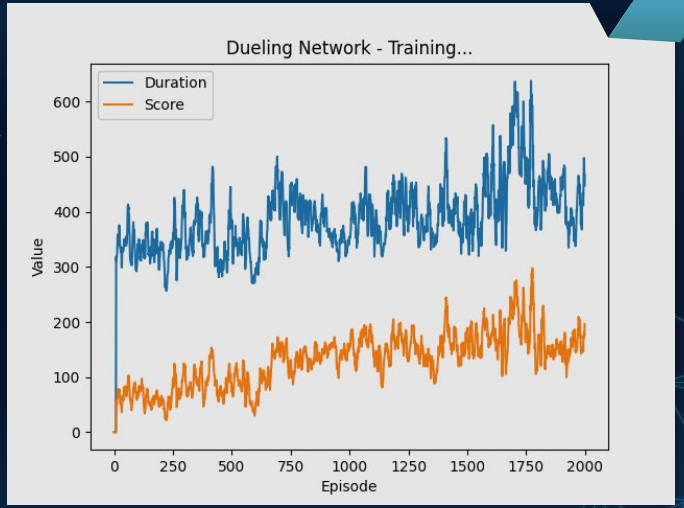
Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	30	280	90	95.05
200	110	525	142.5	167.15
300	20	460	100	122.25
400	65	835	292.5	333.15
500	75	595	260	279.25
600	135	1065	225	282.55
700	120	755	155	237.45
800	35	695	95	156.7
900	30	710	210	228.7
1000	80	845	315	347.8

Easy

Experiment 3

Hyper-parameters [4]:

- Architecture: Dueling Network
- Episodes Trained = 2000
- Memory Size = 20000 (Prioritized Experience Replay)
- minibatch_size = 32 (train per 4 steps)
- γ = 0.99
- ε_start = 1.0
- ε_min = 0.1
- ε_decay_rate = 0.99 (per episode)
- target_network_update_rate = 10000
- Adam Optimizer
  - learning_rate = 0.00025
- Rewards: Normalized Rewards
- Mode: Hard



Experiment 3

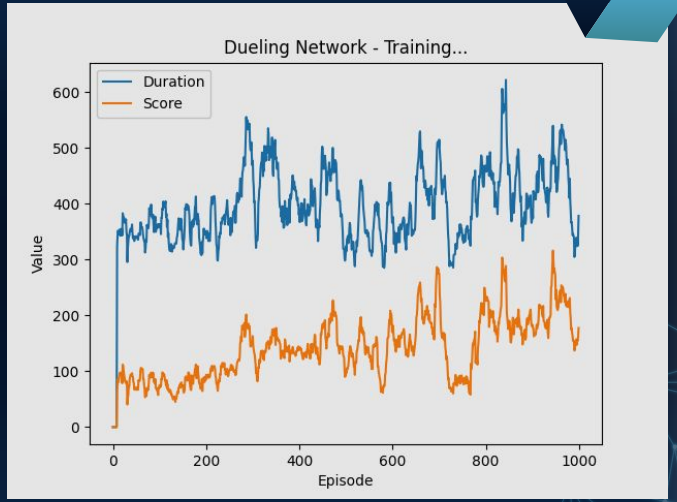
Table 6: Απόδοση μοντέλων Πειράματος 3 για 100 επεισόδια (Hard δυσκολία)

Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	0	160	57.5	61.8
200	0	310	35	42.6
300	5	160	20	34.7
400	20	415	82.5	92.95
500	55	255	120	122
600	5	210	75	75.2
700	105	515	197.5	209.8
800	20	215	135	127.45
900	90	410	155	155.25
1000	105	240	180	173.05
1100	75	410	155	157.15
1200	35	425	155	150.6
1300	30	415	180	191
1400	35	425	190	177.9
1500	40	610	152.5	167.25
1600	35	460	180	184.5
1700	30	535	210	220.1
1800	40	545	155	207.55
1900	10	430	132.5	131.6
2000	90	410	210	204.5

Experiment 4

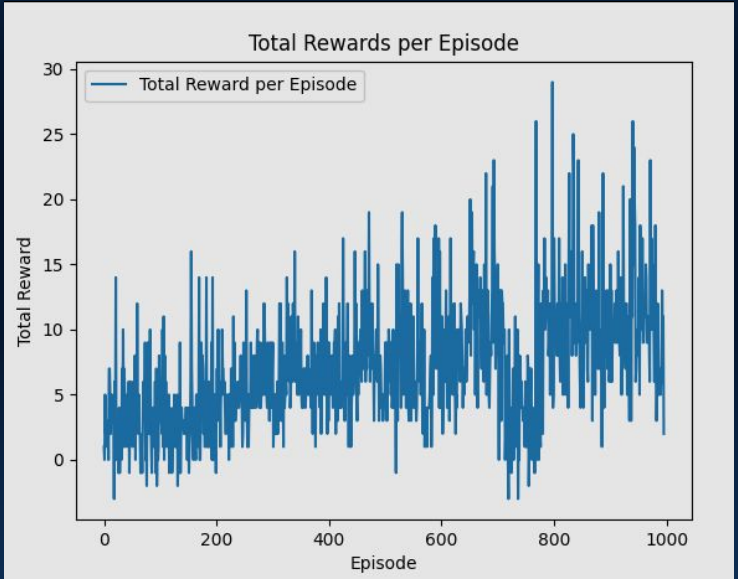
Hyper-parameters:

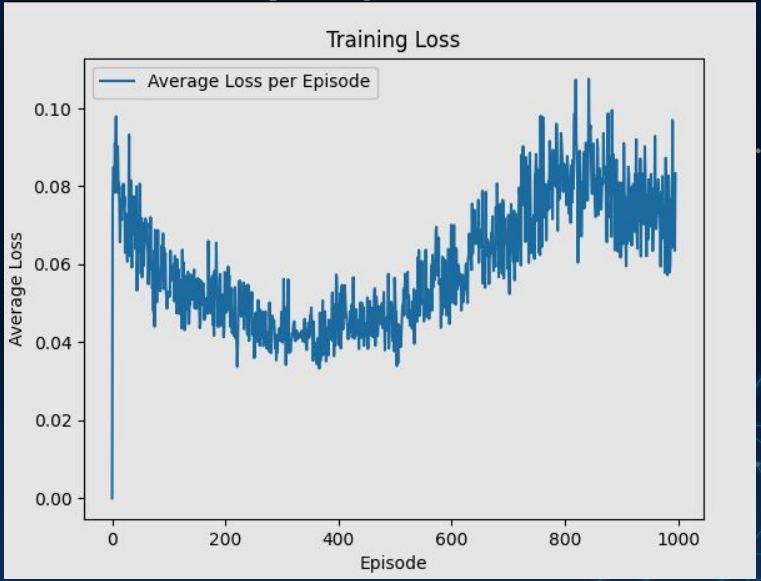
- Architecture: Boosted Dueling Network*
- Episodes Trained = 1000
- Memory Size = 100000 (Prioritized Experience Replay)
- minibatch_size = 8 (train per 4 steps)
- γ = 0.99
- ε_start = 1.0
- ε_min = 0.1
- ε_decay_rate = 0.99999 (per step)
- target_network_update_rate = 10000
- Adam Optimizer
  - learning_rate = 0.0001
- Rewards: Simple Rewards
- Mode: Hard



* Boosted Dueling Network is the same architecture but with double filters and neurons per layer and one extra convolutional layer.

Experiment 4





Experiment 4

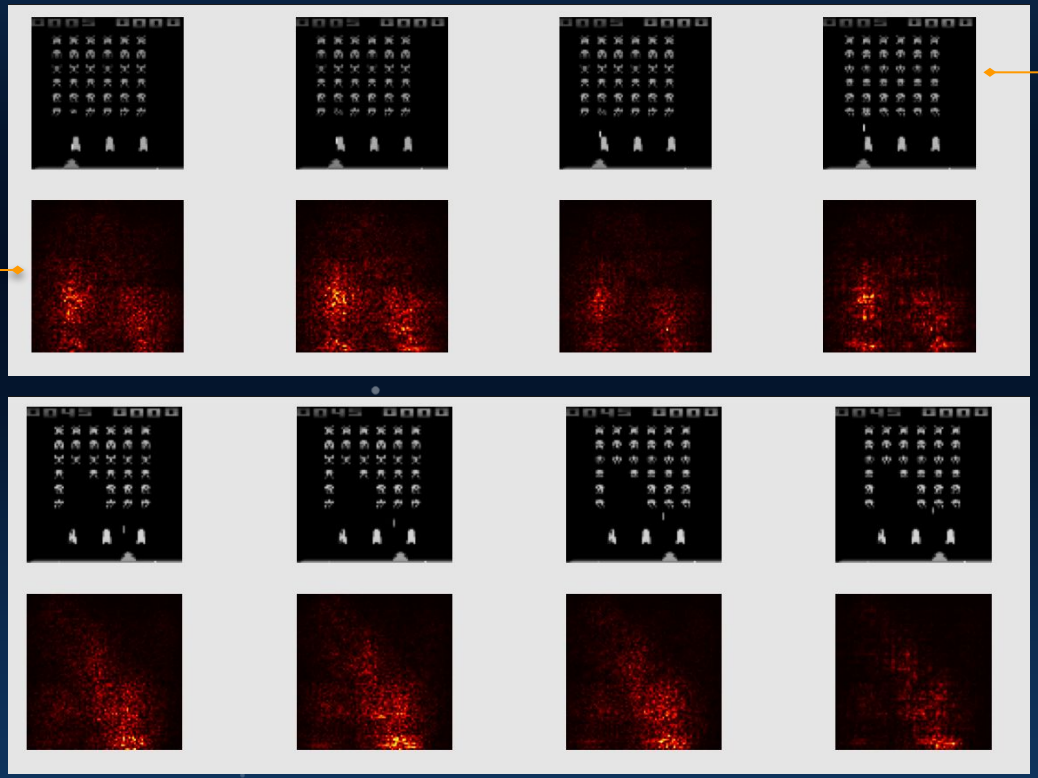
Table 7: Απόδοση μοντέλων Πειράματος 4 για 100 επεισόδια (Hard δυσκολία)

Episodes Trained	Min Score	Max Score	Median Score	Average Score
100	5	345	55	79.5
200	60	410	127.5	137.15
300	240	265	240	247.95
400	30	420	135	141.15
500	60	165	60	81.75
600	45	395	60	69.25
700	95	790	245	254.1
800	80	565	230	235.25
900	75	490	150	161.55
1000	90	510	180	191.65

Saliency Maps

Saliency Maps

Which part of the input is more important to the decision making



Input

The Input image for the network

Results for DRL Project

Result 1

Dueling Network is a better architecture that evaluates the value of a state and the advantage of each action and helps the agent make better decisions.

Result 2

A Prioritized Experience Replay Memory helps Deep Q Learning retrieve more useful experiences for the learning process (network's training).

Result 3

Manually fine-tuning the hyperparameters of both the Deep Q Learning Algorithm and the network is of critical sense.

Result 4

Saliency maps visualize the parts of an input image that are most important to an agent's decision-making process.

Results for DRL Project

Result 5

Different types of rewards can significantly influence the way agents play. Making the agent focus only on score makes the agent play only for quick high scores. On the other hand, a standard positive reward for executing an enemy and a standard negative reward for losing a life encourage the agent to focus on defeating as many enemies as possible, which indirectly leads to higher scores.

Created with DALL-E

Bibliographic references

[1] Mark Towers et al. “Gymnasium”. In: (Mar. 2023).

[2] M. G. Bellemare et al. “The Arcade Learning Environment: An Evaluation Platform for General Agents”. In: Journal of Artificial Intelligence Research 47 (June 2013), pp. 253–279.

[3] Ziyun Wang et al. “Dueling Network Architectures for Deep Reinforcement Learning”. In: International Conference on Machine Learning. 2015. Url: https://api.semanticscholar.org/CorpusID:5389801

[4] H. V. Hasselt, Arthur Guez, and David Silver. “Deep Reinforcement Learning with Double Q-Learning”. In: AAAI Conference on Artificial Intelligence. 2015. url: https://api.semanticscholar.org/CorpusID:5389801