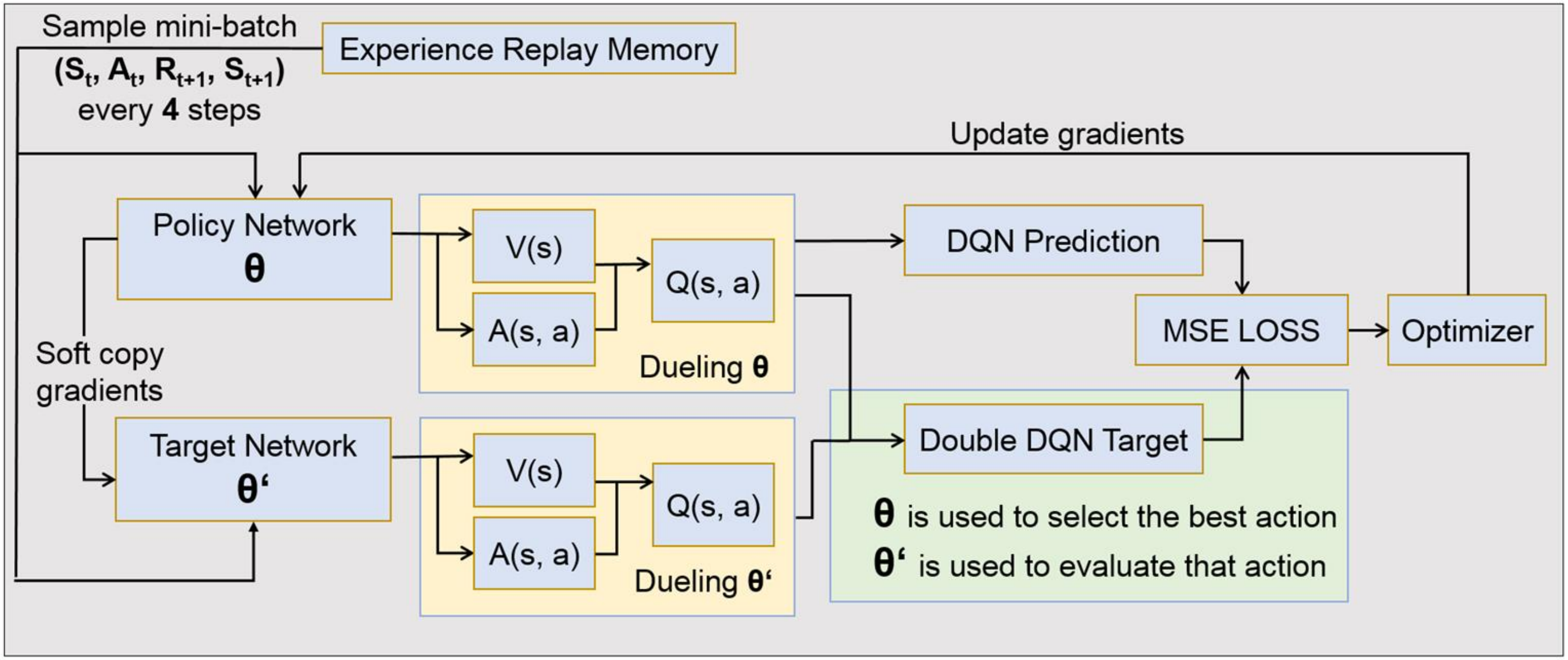
Choice of Implemented Algorithm

- Choice: DQN[1] (discrete action space)
  - Popular algorithm, can solve Atari games
  - Lots of possible improvements[3][4][5]
  - Agents with discretized actions can solve given problem



DQN Target: $Y_t^{\text{DoubleQ}} \equiv R_{t+1} + \gamma Q(S_{t+1}, \operatorname{argmax} Q(S_{t+1}, a; \boldsymbol{\theta}_t); \boldsymbol{\theta}'_t)$

Figure 1: Schematic diagram of DQN Algorithm[1] with Double Q Learning[3] and Dueling Networks[4]

Performance Analysis

Experiment	K steps	Episodes	Evaluation (1000 ep)	Train score (last 100 ep)	Success metric	Converged / Total
Baseline	688.07±161.59	2851.13±714.40	303.42±177.98	191.71±126.34	Eval ≥ 400 (every 20K steps)	10/15
Best (eval @5K)	**189.02±28.41**	**879.73±80.56**	414.96±18.02	59.81±45.44	Eval ≥ 400 (every 5K steps)	15/15
Best (eval @20K)	197.41±35.29	893.93±110.19	414.85±12.16	81.16±57.91	Eval ≥ 400 (every 20K steps)	15/15
Best (train score)	512.08±162.60	1521.27±210.33	**420.25±3.10**	**400.42±0.75**	Score of last 100 episodes ≥ 400	14/15
Best (shaped reward)	256.03±54.99	1027.00±163.02	409.688±40.87	13.23K±71.80K	Eval ≥ 400 (every 5K steps)	15/15

Table 1: Quantitative Analysis of 5 selected experiments

- Careful HP optimization improved the results
- Different success metric have their advantages
- High variance, repeated experiments needed
- Highly time efficient
  - **~5min** GPU runtime for experiment *Best (eval @5K)*

References

- [1] Mnih, V. et al (2015), Human Level Control Through Deep Reinforcement Learning
https://web.stanford.edu/class/psych209/Readings/MnihEtAlHassibis15NatureControlDeepRL.pdf
- [2] Lillicrap, T. et al (2015), Continuous control with deep reinforcement learning
https://arxiv.org/abs/1509.02971
- [3] van Hasselt, H. et al (2016), Deep Reinforcement Learning with Double Q-learning
https://arxiv.org/abs/1509.06461
- [4] Wang, Z. et al (2015), Dueling Network Architectures for Deep Reinforcement Learning
https://arxiv.org/abs/1511.06581
- [5] Hessel, M. et al (2017), Rainbow: Combining Improvements in Deep Reinforcement Learning
https://arxiv.org/abs/1710.02298
- [6] Maroti, A., RBED: Reward Based Epsilon Decay
https://arxiv.org/abs/1910.13701