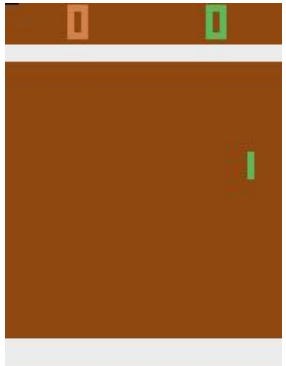
UCLA Samueli School of Engineering

# An Empirical Study of Deep Q-Learning Algorithms



Aditya Golatkar[1], Albert Zhao[1], Cagatay Isil[2], Xiaoran Zhang[2]

[1]Department of Computer Science, [2]Department of Electrical and Computer Engineering

# Deep Q-Networks and Variants

## DQN

$$Y_t^{DQN} = R_{t+1} + \gamma \max_a Q(S_{t+1}, a; (\boldsymbol{\theta_t^-}, \boldsymbol{w_t^-}))$$

## Double DQN

$$Y_t^{DoubleDQN} = R_{t+1} + \gamma Q(S_{t+1}, \text{argmax}_a Q(S_{t+1}, a; (\boldsymbol{\theta_t}, \boldsymbol{w_t})); (\boldsymbol{\theta_t^-}, \boldsymbol{w_t^-}))$$

## Dueling DQN

$$Y_t^{DuelDQN} = R_{t+1} + \gamma \max_a [V(S_{t+1}; (\boldsymbol{\theta_t^-}, \boldsymbol{\beta_t^-}) + (A(S_{t+1}, a; (\boldsymbol{\theta_t^-}, \boldsymbol{\alpha_t^-})) - \frac{1}{|\mathcal{A}|} \sum_a A(S_{t+1}, a; (\boldsymbol{\theta_t^-}, \boldsymbol{\alpha_t^-}))]$$

## Noisy DQN

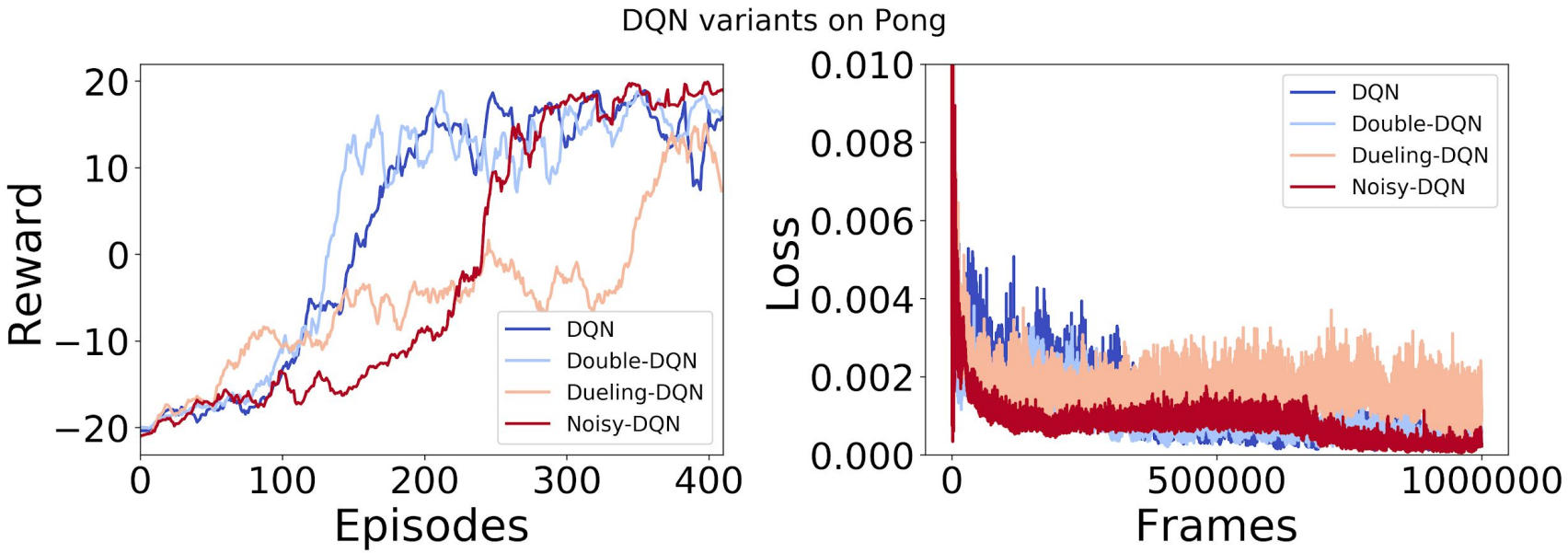
$$Y_t^{NoisyDQN} = R_{t+1} + \gamma \max_a Q(S_{t+1}, a; (\boldsymbol{\theta_t^-}, \boldsymbol{\mu}^{w_t^-} + \boldsymbol{\sigma}^{w_t^-} \odot \boldsymbol{\epsilon})) \quad \boldsymbol{\epsilon} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{0}, \boldsymbol{I})$$

# Default Parameter Settings

| Hyperparameter | Value |
|---|---|
| Number of Frames | 1M |
| Discount Factor $\gamma$ | 0.99 |
| Batch Size | 32 |
| Loss Function | $L_2$ |
| Optimizer | Adam |
| Learning Rate | 1e-4 |
| Target Q Update Frequency | 1K |
| Replay Buffer | Prioritized |
| Replay Buffer Size | 10K |
| Replay Buffer $\alpha$ | 0.6 |
| Replay Buffer Initial, Final $\beta$ | 0.4, 1.0 |
| Replay Buffer $\beta$ Schedule | Linear (100K frames) |
| Initial, Final $\epsilon$ | 1.0, 0.01 |
| $\epsilon$-greedy Schedule | Exponential (Exploration Factor 30K) |

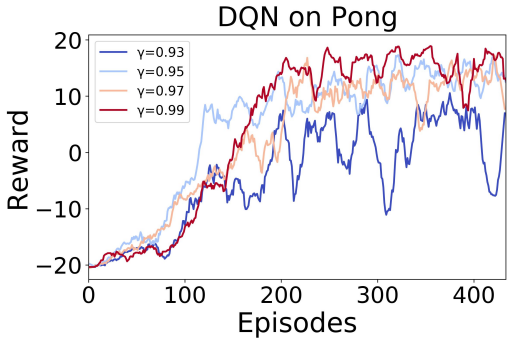
Table 1: Common Hyperparameters for DQN Variants

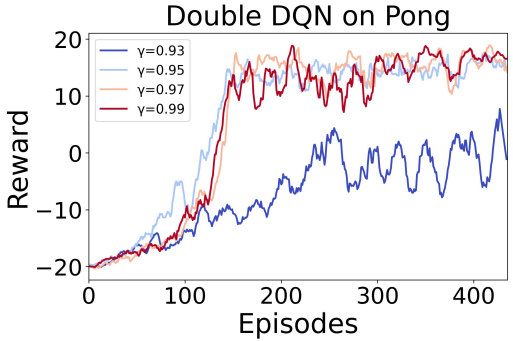
# Deep Q-Networks and Variants

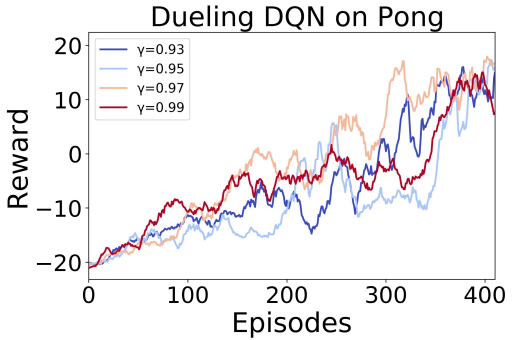


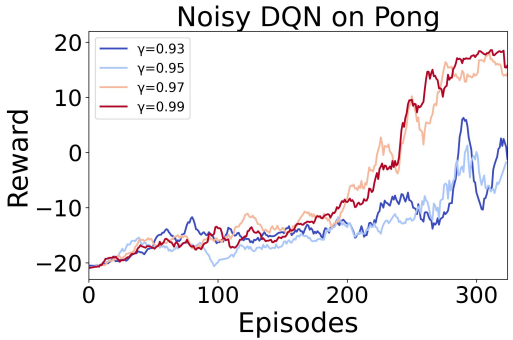
# Effective Horizon

- Agents with wider effective horizon (=1/(1-$\gamma$)) learn better policies.
- Small horizon also results in poor convergence for Double & Noisy-DQN



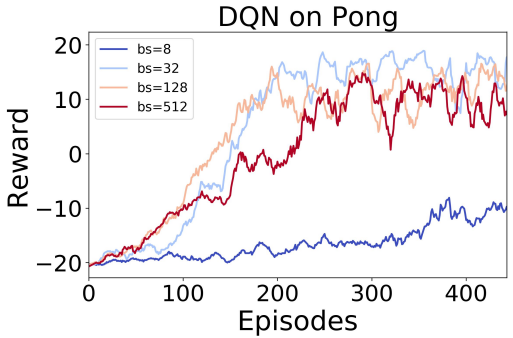



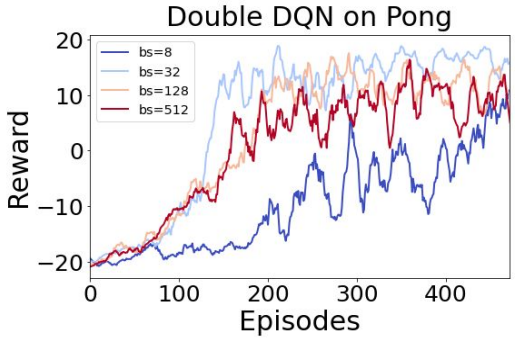



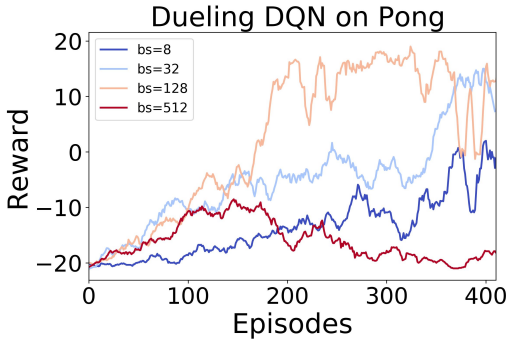



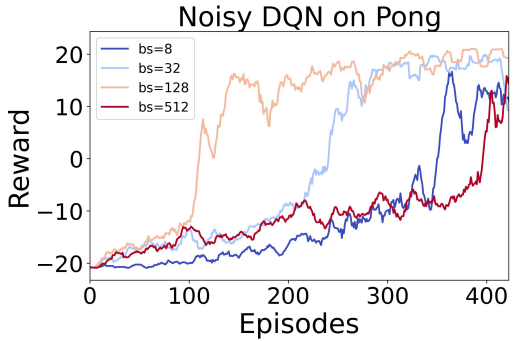
# Batch Size

- Low batch-size (or high gradient noise) results in slower convergence for all the methods.
- Large batch-size (less stochasticity) leads to poor convergence in Deuling & Noisy-DQN



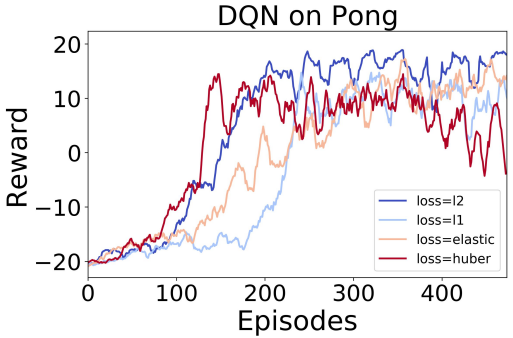
# Loss Function

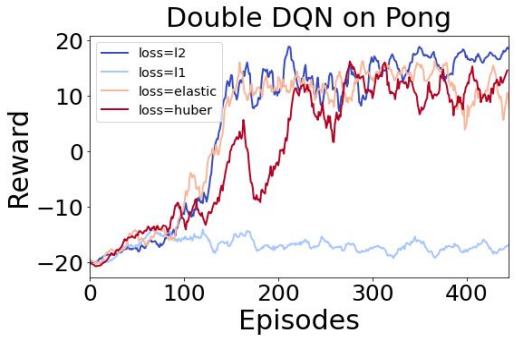
## Elastic-Net Loss

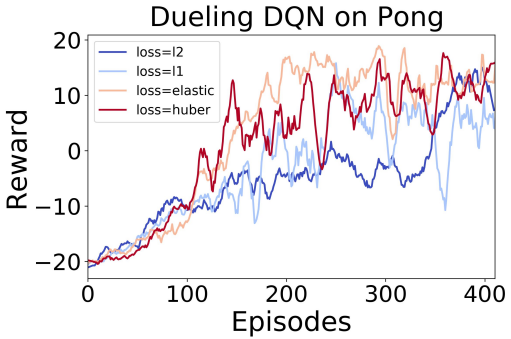
$$l(x, y) = \|x - y\|_2^2 + \lambda \|x - y\|_1$$

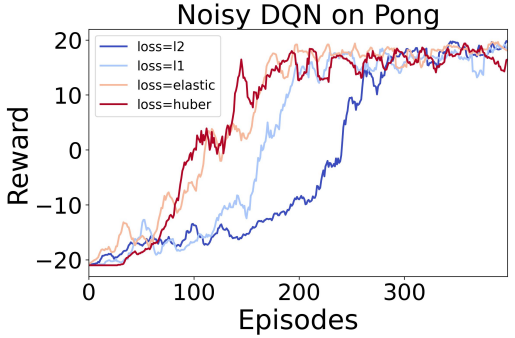
## Huber-Loss

$$l(x, y) = \begin{cases} \frac{1}{2}(x - y)^2 & |x - y| \leq \delta \\ \delta |x - y| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$



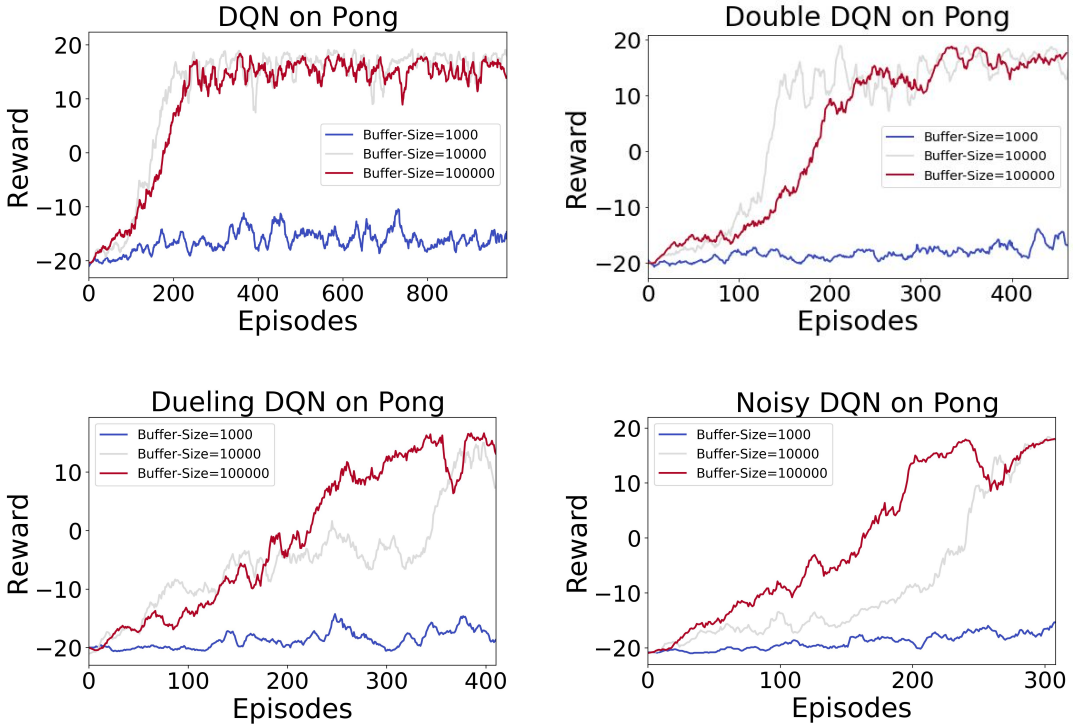



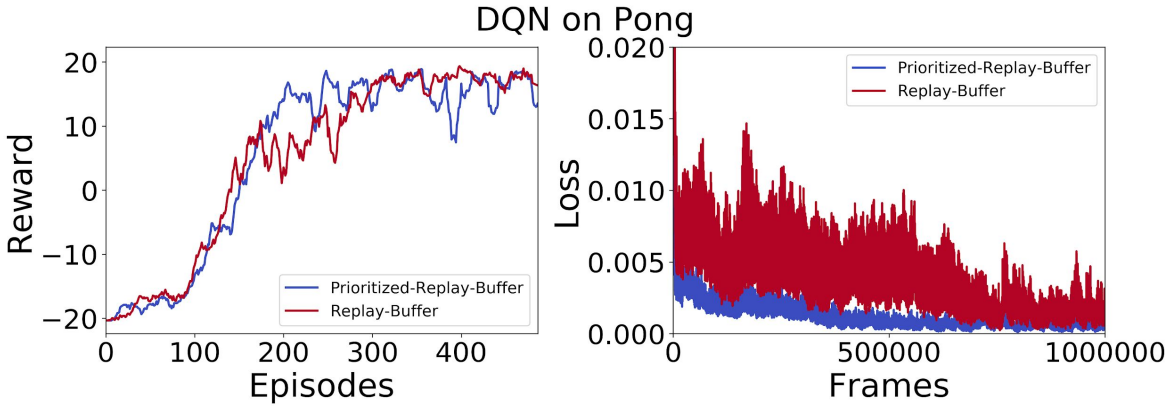
# Replay Buffer Size

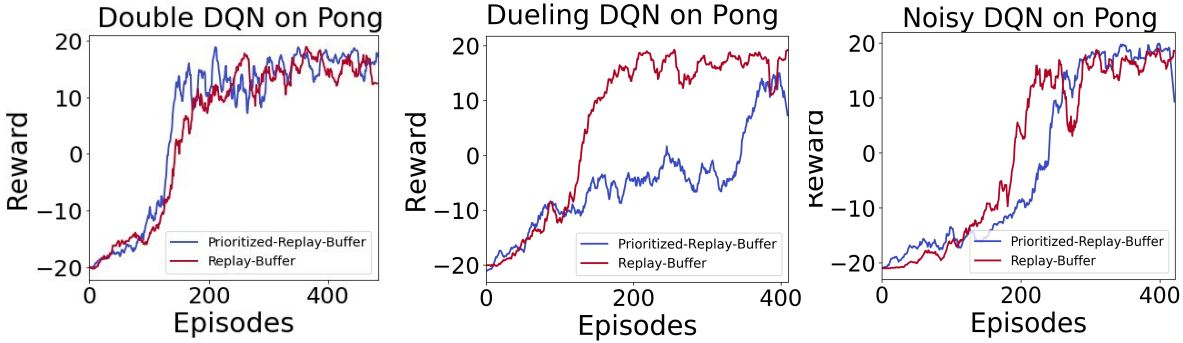
- Here, we use prioritized replay buffer.
- Low replay buffer size leads to slow convergence due to inability to effectively reuse data samples.



# Role of Prioritization

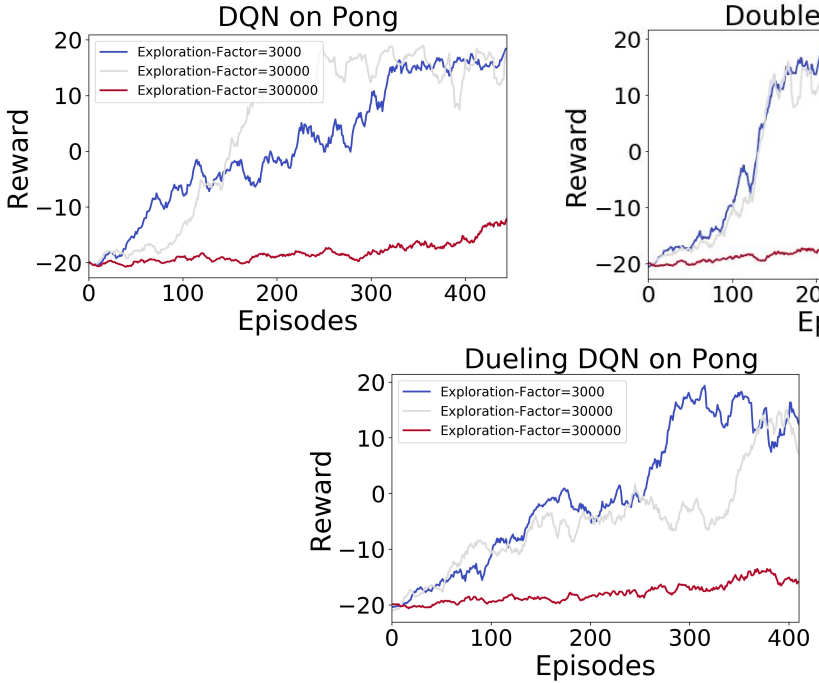
- Prioritization reduces loss variance but does not speed up convergence & hurts for Dueling DQN.
- In the original Dueling DQN paper, prioritization gives only a small benefit (0.74%).



# Exploration

- Over-exploration results in slower convergence.
- Noisy DQN does not use epsilon-greedy exploration method.



# Conclusion

- Empirically analyzed the different variants of DQN.
  - Increasing the horizon length improves the performance.
  - Small batch-size during training leads to suboptimal policies.
  - Optimal choice of loss functions depends on the DQN variant.
  - In general, small buffer size hurts the performance.
  - Prioritization of replay buffer reduces loss variance.
  - Over-exploration leads to slow convergence.

# References

[1] Volodymyr Mnih et al. Human-level control through deep reinforcement learning.Nature, 518(7540):529–533, 2015.
[2] Hado Van Hasselt et al.  Deep reinforcement learning with double q-learning. In Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[3] Ziyu Wang et al. Dueling network architectures for deep reinforcement learning. arXiv preprint arXiv:1511.06581, 2015.
[4] Meire Fortunato et al.  Noisy networks for exploration.  In International Conference on Learning Representations,, 2018.
[5] Tom Schaul et al. Prioritized experience replay. In International Conference on Learning Representations, Puerto Rico, 2016.