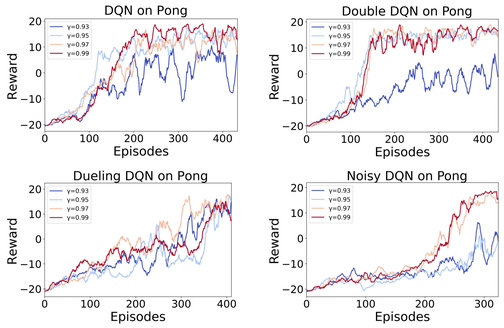
An Empirical Study of Deep Q-Learning Algorithms

Aditya Golatkar[1] (UID: 505221372), Albert Zhao[1] (UID: 605231634), Cagatay Isil[2] (UID: 705430690), Xiaoran Zhang[2] (UID: 605429606)
[1] Department of Computer Science, [2] Department of Electrical and Computer Engineering

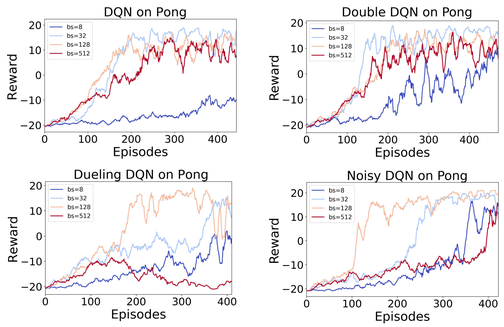
Effective Horizon

Agents with wider effective horizon (=1/(1-γ)) learn better policies.

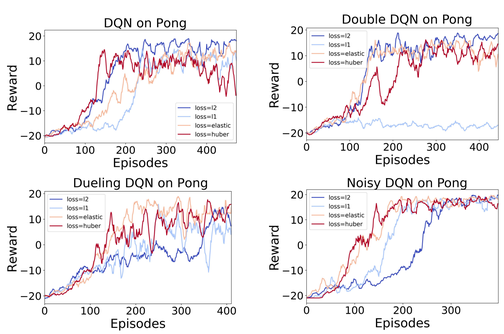


Batch-Size

Low batch-size (or high gradient noise) results in slower convergence.



Loss Function



Deep Q-Networks and Variants

DQN[1]

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

Convolutional Layers; Fully Connected Layers

Double DQN[2]

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

Current Network; Target Network

Dueling DQN[3]

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

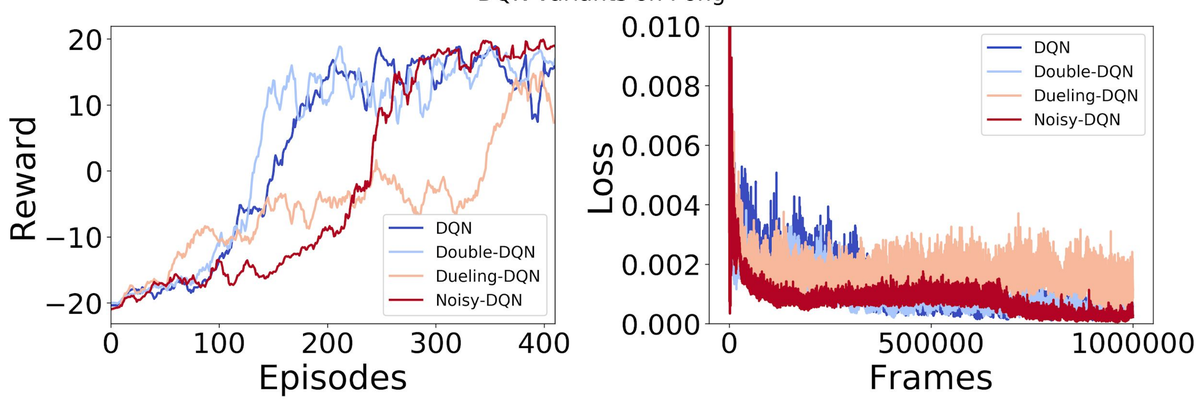
FC Layer for Value Function Stream; FC Layer for Advantage Function Stream

Noisy DQN[4]

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

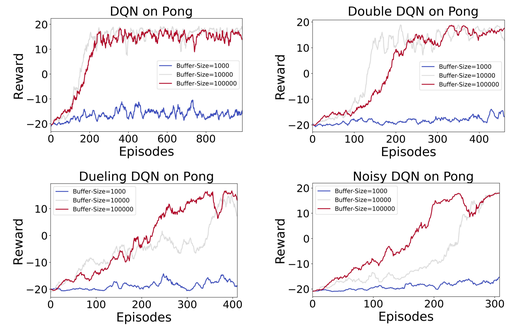
Mean of FC layers; Standard deviation of FC layers; Noise with normal distribution

DQN variants on Pong



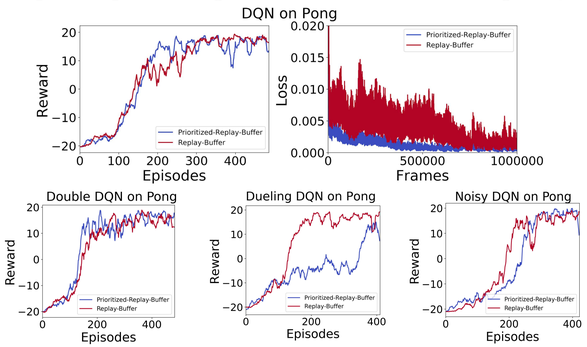
Replay Buffer Size

Low replay buffer size leads to slow convergence due to inability to effectively reuse data samples.



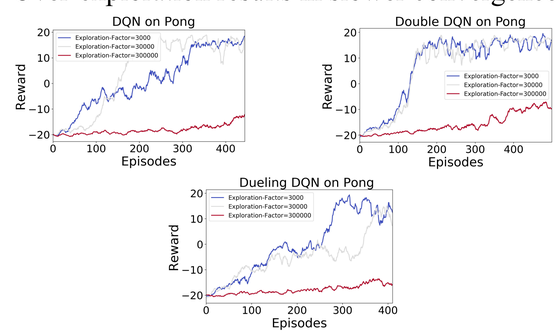
Role of Priority

Prioritization[5] reduces loss variance but does not speed up convergence & hurts for Dueling DQN.



Exploration

Over-exploration results in slower 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.