Mitigation of Policy Manipulation Attacks on Deep Q-Networks with Parameter-Space Noise

Vahid Behzadan and Arslan Munir

Department of Computer Science

Kansas State University

First International Workshop on Artificial Intelligence Safety Engineering (WAISE), 2018

Email: behzadan@ksu.edu

http://www.vbehzadan.com

Outline

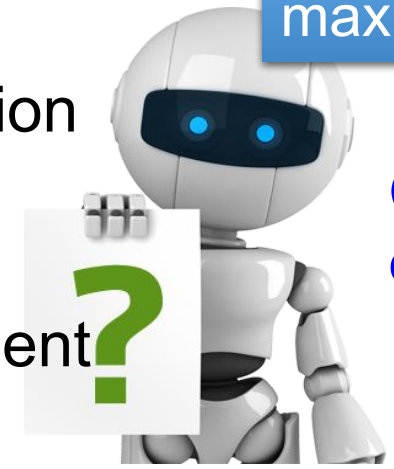
- Deep Q-Networks
- Adversarial examples
- Vulnerability of Reinforcement Learning (RL) to adversarial examples
- Mitigation of RL attacks via parameter-space noise
- Conclusion

Reinforcement Learning

Agent learns to take actions maximizing expected reward.

Observation

State



Agent

Action

Change the environment

Thank you.

Reward

Environment

Q-Learning

Learn optimal policy through optimization of Action-Value function (a.k.a., Q-function)

- Definitions:
  - **Action-Value Function**

$$Q^{\pi}(s,a) = E[R_t | s_t = s, a_t = a]$$

  - **Deterministic Policy** $\pi: S \rightarrow A$: Mapping of states to corresponding actions
  - **Stochastic Policy** $\pi(a|s)$: Probability distribution of taking action a at state s.

Q-Learning

Objective: Derive optimal policy π^* based on optimal Q

$$Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a)$$

Iterative formulation:

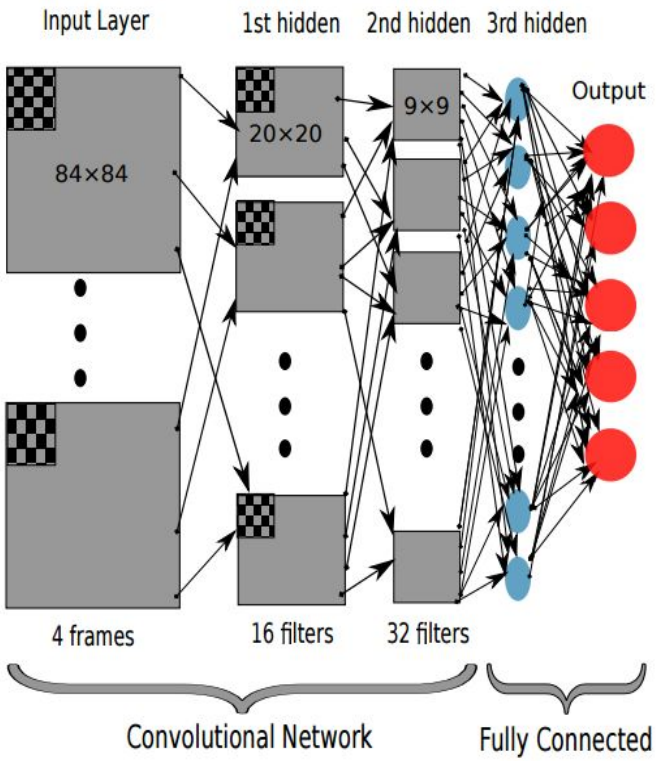
$$Q(s,a) = R_{s,a} + \gamma \max_{a'} Q(s',a')$$

Solution:

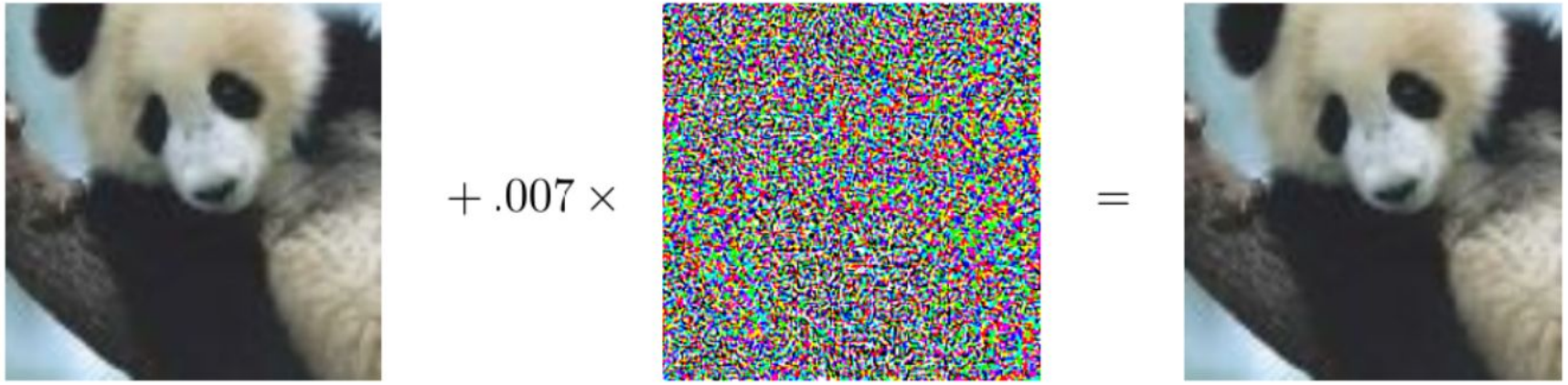
- Bellman / Dynamic programming
- Parametrization as $Q(s,a,\theta)$
- Solve via neural nets, where θ corresponds to weights, hence ***Q-Network***

Deep Q-Networks (DQNs)

- **Deep**: Feature Learning
- **Non-iid Data**: Experience replay
- **Oscillation**: Fix parameters
- **Unbounded**: Normalize rewards to [-1, 1]



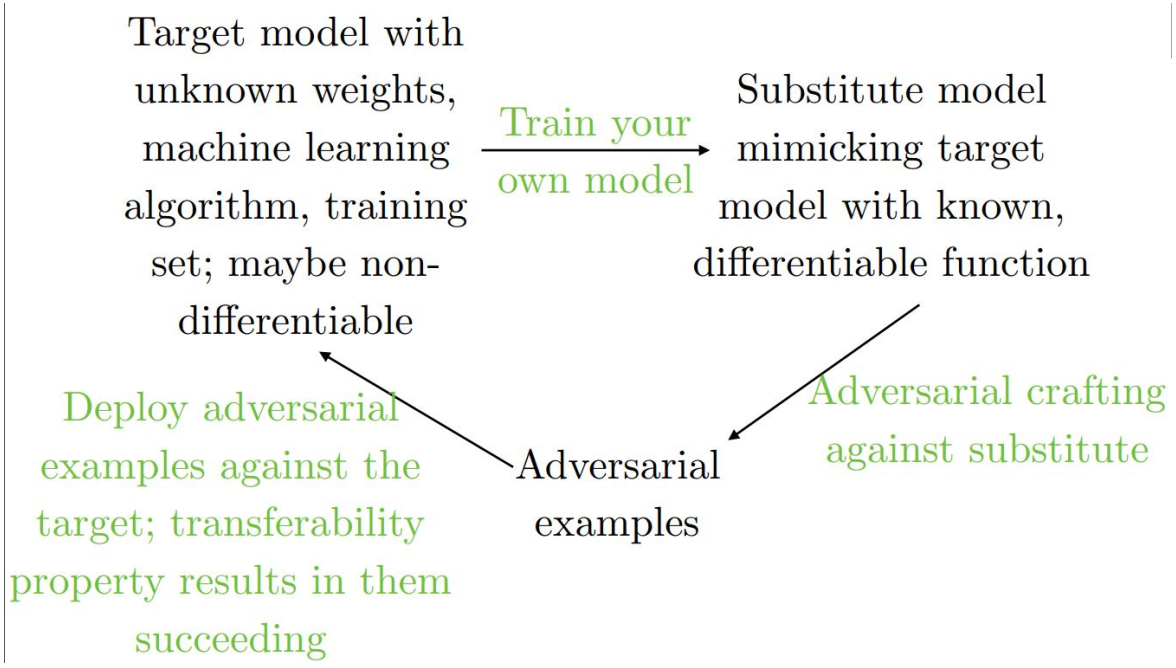
Adversarial Examples



Papernot, Nicolas, et al. "Practical black-box attacks against deep learning systems using adversarial examples." *arXiv preprint* (2016).

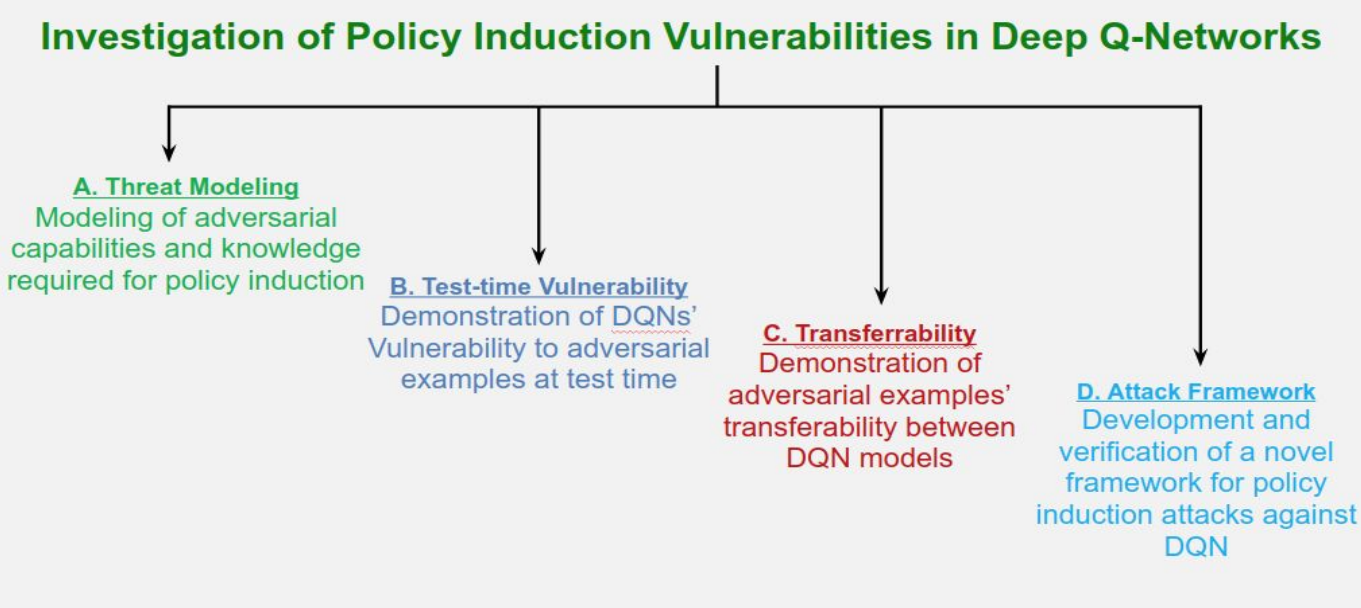
Transferability of Adversarial Examples

Adversarial example for one model can manipulate another model trained on similar datasets (Papernot 2016)



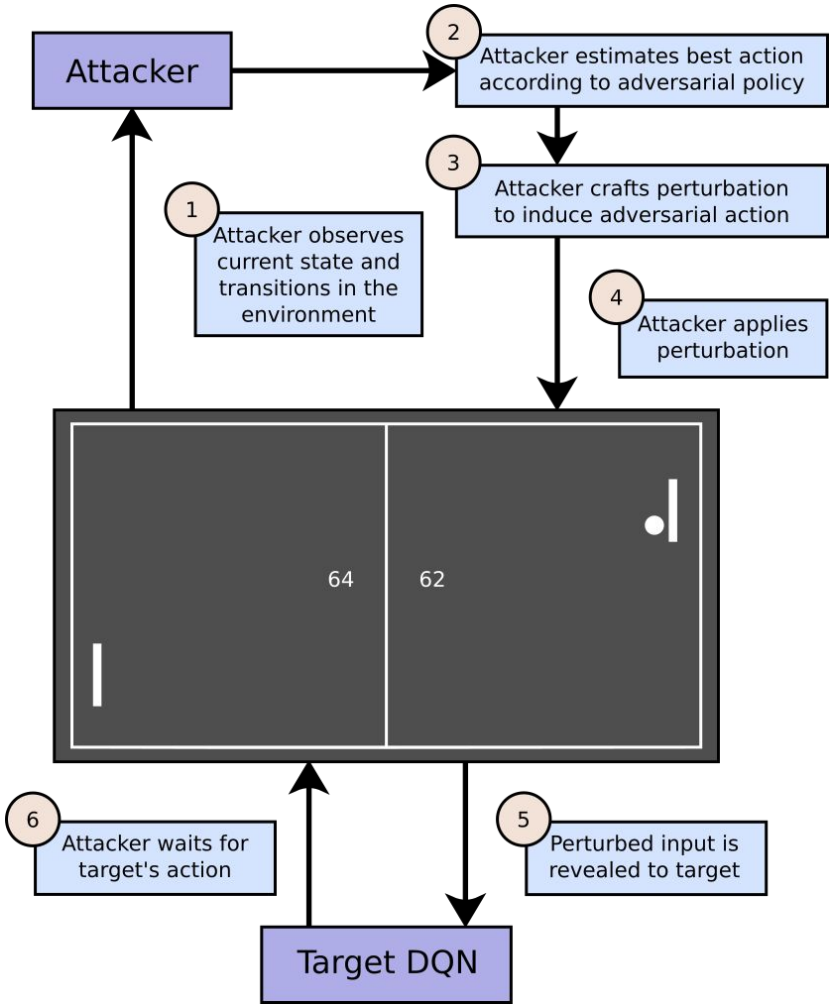
Adversarial Examples vs RL

DNNs used in DQN are no different from those of classifiers, hence Adversarial Examples



Behzadan and Munir. "Vulnerability of deep reinforcement learning to policy induction attacks." *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer, Cham, 2017.

Exploitation Methodology



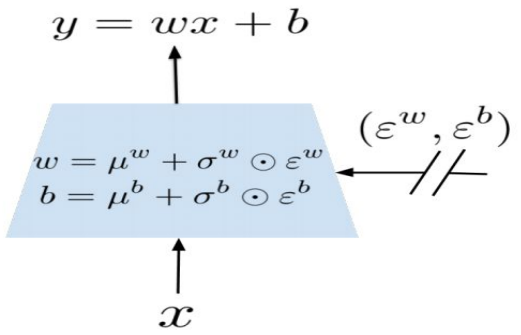
Mitigation via Parameter-Space Noise

- Typical exploration mechanism in RL : ε-greedy
  - Start by taking random actions with probability ε = 100
  - Monotonically decrease ε, increase chances of taking learned actions

$$a_t = \begin{cases} a_t^* & \text{with probability } 1 - \epsilon \\ \text{random action} & \text{with probability } \epsilon \end{cases}$$

Parameter-Space Noise

- Novel exploration mechanism
  - Plappert et al. (2017) - Parameter Space Noise for Exploration
  - Fortunato et al. (2017) - Noisy Networks for Exploration
- Method (NoisyNet): Introduce zero-mean random noise to the learnable parameters of neural network in deep RL
- Shown to enhance exploration and convergence in deep RL benchmarks



Fortunato, Meire, et al. "Noisy networks for exploration." *arXiv preprint arXiv:1706.10295* (2017).

Hypothesis

The enhanced generalization and increased randomization of NoisyNet exploration can alleviate the impact of policy manipulation attacks at both test-time and training-time

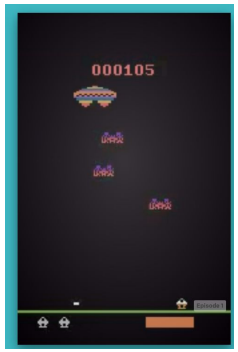
Contributions

We experimentally verify that:

- Parameter-space noise reduces the transferability of adversarial examples in policy manipulation attacks
- Also, enhances the resilience and robustness of DQNs
  - to both whitebox and blackbox attacks
  - to both test-time and training-time attacks

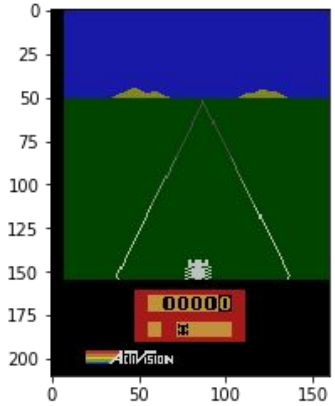
Experiment Setup

- DQNs - Same parameters and architecture as Mnih et al. (2015)
- 3 Atari Games: Assault, Breakout, *Enduro*



Assault

Breakout

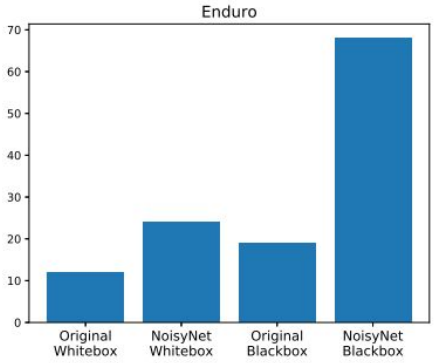


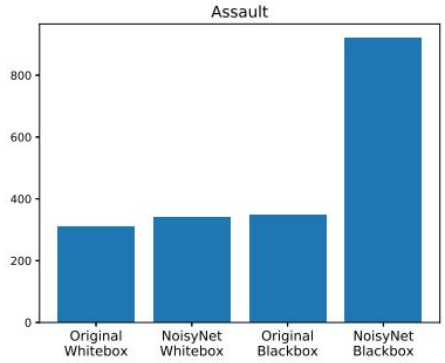
Enduro

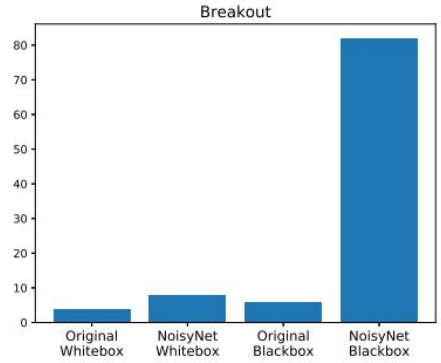
- Attack model similar to that of Behzadan & Munir (2017)

Results
Test-time Attack

- Epsilon-greedy exploration vs. NoisyNet
- Whitebox vs. Blackbox

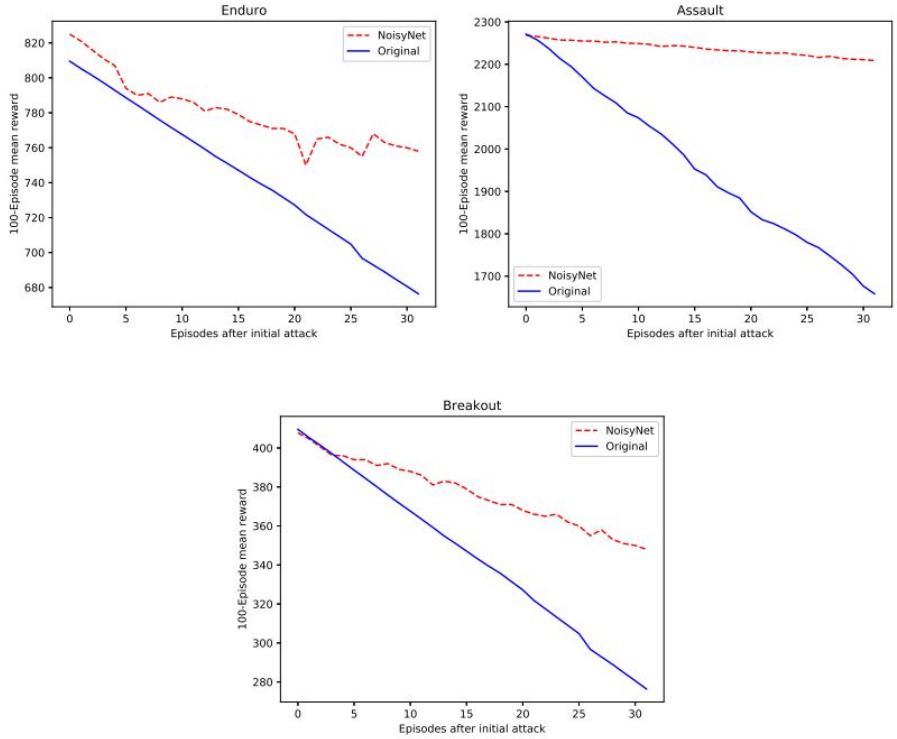






Results
Training-time Attack

- Epsilon-greedy exploration vs. NoisyNet
- Whitebox vs. Blackbox



Conclusions (1)

- RL and deep RL are shown to be vulnerable to adversarial perturbations at both test-time and training-time
- Current defensive techniques tend to fail vs. RL attacks
- NoisyNet exploration greatly enhances the resilience of DQN to whitebox and blackbox attacks at test-time
- NoisyNet significantly enhances the resilience of DQN to training-time attacks

Conclusions (2)

- NoisyNet provides better generalization, thus alleviates susceptibility of model to adversarial examples
- NoisyNet introduces adaptive randomness, thereby reduces susceptibility of model to transferability of adversarial examples
- Urgent need for AI security research

Questions?