

ASSESSING REINFORCEMENT LEARNING POLICIES VIA NATURAL CORRUPTIONS AT THE EDGE OF IMPERCEPTIBILITY

Anonymous authors

Paper under double-blind review

1 PERSPECTIVE TRANSFORM

In Section 3 of the main paper we propose a natural perturbation framework and give detailed descriptions of the components. Perspective transform is one of the successful perturbation types in decreasing the performance of the deep neural policy while having the lowest perceptual similarity distance to the unperturbed states (i.e. perceptually more similar to the unperturbed state). In this section we provide more detail on this geometric transformation and how it should be utilized.

1.1 FORMULAS FOR PERSPECTIVE TRANSFORMATION

The perspective transform is a geometric transformation of an image, uniquely determined by the coordinates of four source and four destination pixels. Let $s_i^{\text{dst}_k}$ and $s_j^{\text{dst}_k}$ represent the coordinates of the k -th destination pixel, and let $s_i^{\text{src}_k}$ and $s_j^{\text{src}_k}$ represent the coordinates of the k -th source pixel. The perspective transform maps the given source pixel values to the destination pixel values as follows. First, solve for a matrix M and real numbers t_k such that for all k :

$$t_k \begin{bmatrix} s_i^{\text{dst}_k} \\ s_j^{\text{dst}_k} \\ 1 \end{bmatrix} = M \cdot \begin{bmatrix} s_i^{\text{src}_k} \\ s_j^{\text{src}_k} \\ 1 \end{bmatrix}. \quad (1)$$

Next assign each pixel value of the transformed image $s_{\text{adv}}(i, j)$ using M as follows:

$$s_{\text{adv}}(i, j) = s \left(\frac{M_{11}s_i + M_{12}s_j + M_{13}}{M_{31}s_i + M_{32}s_j + M_{33}}, \frac{M_{21}s_i + M_{22}s_j + M_{23}}{M_{31}s_i + M_{32}s_j + M_{33}} \right). \quad (2)$$

2 FOURIER DOMAIN COMPLEMENTARY RESULTS

In this section we provide complementary results to Section 5 of the main body of the paper. Figure 1 demonstrates the total energy spectrum $\mathcal{E}(f)$ of rotation modification for TimePilot, BankHeist and JamesBond. Note that the rotation modification causes decrease in the high frequencies. In Section 4 we provide an analysis on the resilience of the state-of-the-art adversarial training deep neural policies and vanilla trained deep neural policies. Figure 3 from the main body of the paper demonstrates the parameter analysis of the perturbations for adversarially trained deep neural policies and vanilla trained deep neural policies. In particular, in Figure 3 of the main body of the paper we show that the state-of-the-art adversarially trained deep neural policies are more vulnerable to rotation modification.

In Figure 2 we show the Fourier spectrum of compression artifacts, shifting, brightness and contrast from our proposed natural perturbation framework in TimePilot. In Section 3 of the main body of the paper in Table 1 we show the perceptual similarity distances for the compression artifacts, shifting, brightness and contrast. In particular, Table 2 demonstrates that the shifting, brightness and contrast modifications are perceptually more similar to the unperturbed states compared to adversarially perturbed states.

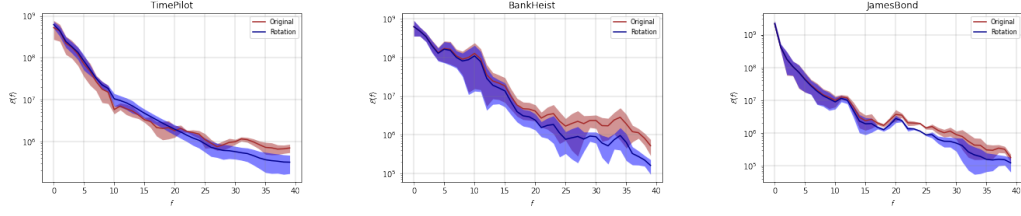


Figure 1: Total energy spectrum $\mathcal{E}(f)$ of rotation modification from natural perturbation framework for TimePilot, BankHeist and JamesBond.

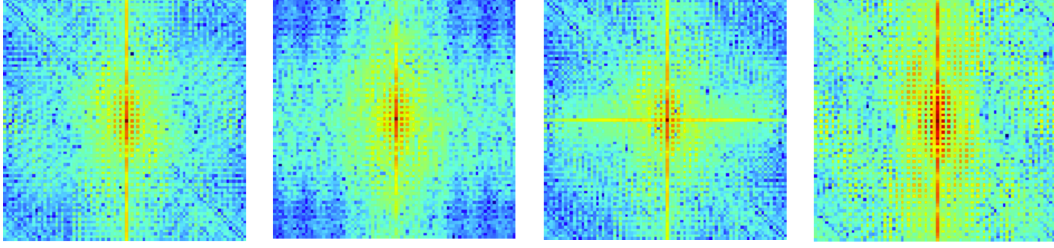


Figure 2: Fourier spectrum $\mathcal{F}(s)$ of the natural perturbation modifications for TimePilot. Left: Unperturbed. Mid-left: Compression artifacts. Mid-right: Shifting. Right: Brightness and contrast.

Figure 3 and Figure 4 show the total energy spectrum of compression artifacts, shifting, brightness and contrast from our proposed natural perturbation framework in TimePilot and JamesBond respectively. Figure 3 demonstrates that the brightness and contrast perturbations cause a tight shift in mid and high frequencies, while compression artifacts cause dramatic decrease in high frequencies.

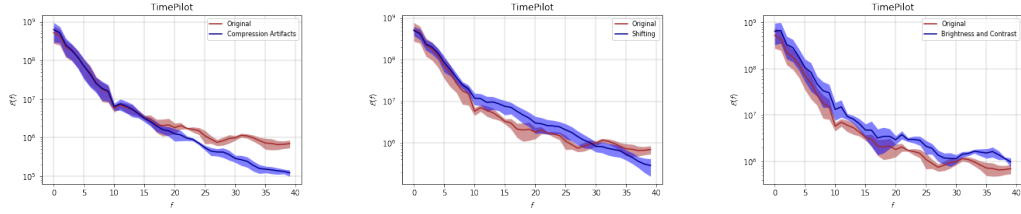


Figure 3: Total energy spectrum $\mathcal{E}(f)$ of natural perturbation modifications compression artifacts, shifting, brightness and contrast for TimePilot.

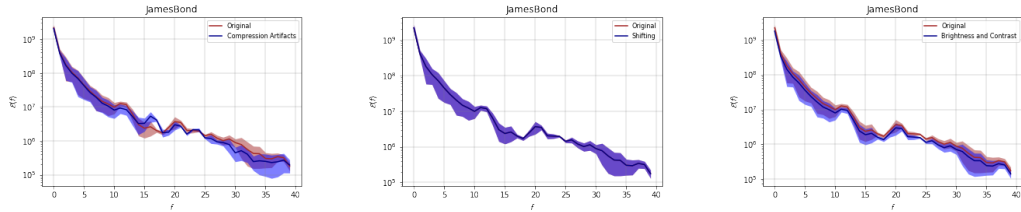


Figure 4: Total energy spectrum $\mathcal{E}(f)$ of natural perturbation modifications compression artifacts, shifting, brightness and contrast for JamesBond.

In Figure 4 the difference between the unperturbed states and the naturally modified states is quite small for JamesBond compared to TimePilot. This can also be seen in Table 1 from the perceptual similarity distance results. In particular, Table 1 from the main body of the paper demonstrates that the naturally modified states are perceptually more similar to the unperturbed states compared to

adversarially perturbed states. In other words, we show in the main body of the paper that the perceptual similarity distance between naturally perturbed states and the unperturbed states is smaller than the perceptual similarity distance between adversarially perturbed states and the unperturbed states.

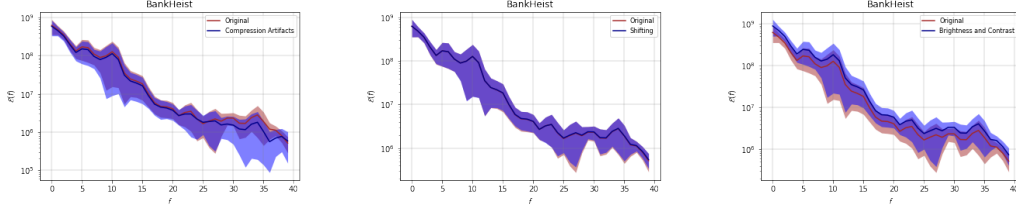


Figure 5: Total energy spectrum $\mathcal{E}(f)$ of natural perturbation modifications compression artifacts, shifting, brightness and contrast for BankHeist.

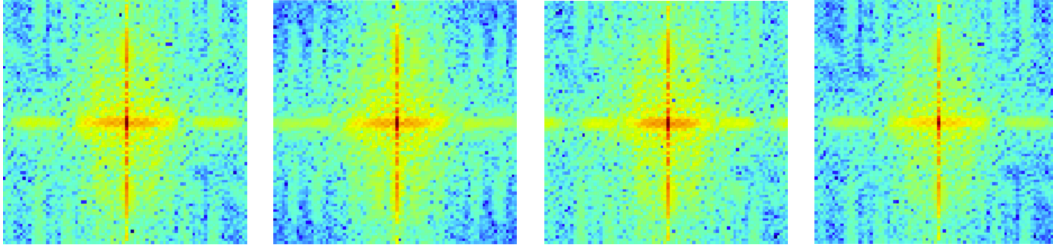


Figure 6: Fourier spectrum $\mathcal{F}(s)$ of the natural perturbation modifications for JamesBond. Left: Unperturbed. Mid-left: Compression artifacts. Mid-right: Shifting. Right: Brightness and contrast.

3 POLICY GRADIENT METHODS UNDER NATURAL PERTURBATION FRAMEWORK

In this section we investigate policy gradient methods under semantically meaningful minimal natural perturbations. In particular, Table 1 shows the perceptual similarities $\mathcal{P}_{\text{similarity}}$, the raw scores and the impact values \mathcal{I} of the agent trained with Asynchronous Advantage Actor-Critic (A3C) Mnih et al. (2016) under our proposed natural perturbation framework with the following observation modifications: brightness & contrast, blurring, rotation, shifting, compression artifacts and perspective transform.

Table 1: Perceptual similarities, raw scores and impacts of the deep neural policy trained with A3C Mnih et al. (2016) algorithm in Pong environment and evaluated with our proposed natural perturbation framework: brightness & contrast, blurring, rotation, shifting, compression artifacts (CA) and perspective transform (PT).

Pong	Bright&Contrast	Blurring	Rotation	Shifting	CA	PT
Raw Scores	-17	-20.35	-19.96	-20.71	-20.89	-19.11
Impacts \mathcal{I}	0.904	0.984	0.974	0.993	0.997	0.954
Perceptual Similarities $\mathcal{P}_{\text{similarity}}$	0.2190	0.0351	0.1020	0.2455	0.2506	0.0140
Natural perturbation hyperparameters	[1.7,40]	3	3	[2,1]	-	3

In Table 1 the exact same hyperparameters have been used as stated in Table 1 of the main the paper for the natural perturbation framework. Note that “natural perturbation hyperparameters” refers for brightness and contrast to $[\alpha, \beta]$, for blurring to the kernel size, for rotation to rotation degree, for shifting to $[t_i, t_j]$, and for perspective transformation to perspective norm. Shifting and

compression artifacts have nearly maximal impact on the performance of the agent trained with A3C, while the other perturbations all have impact at least 0.9. Note that for a direct comparison between A3C deep neural policy and Double Deep Q-Network (DDQN) deep neural policy the hyperparameters for the natural perturbation framework are identical to Table 1 of the main paper. Therefore, although impact is slightly lower for brightness & contrast for A3C than for DDQN, it is possible that choosing different values of α and β while minimizing the perceptual similarity $\mathcal{P}_{\text{similarity}}$ can still result in a higher impact for an agent trained with A3C.

4 PERCEPTUAL SIMILARITIES

In this section we share the perceptual similarities in full detail. In particular, Figure 7 shows the original observation without any modification and with natural perturbation framework observation manipulations.

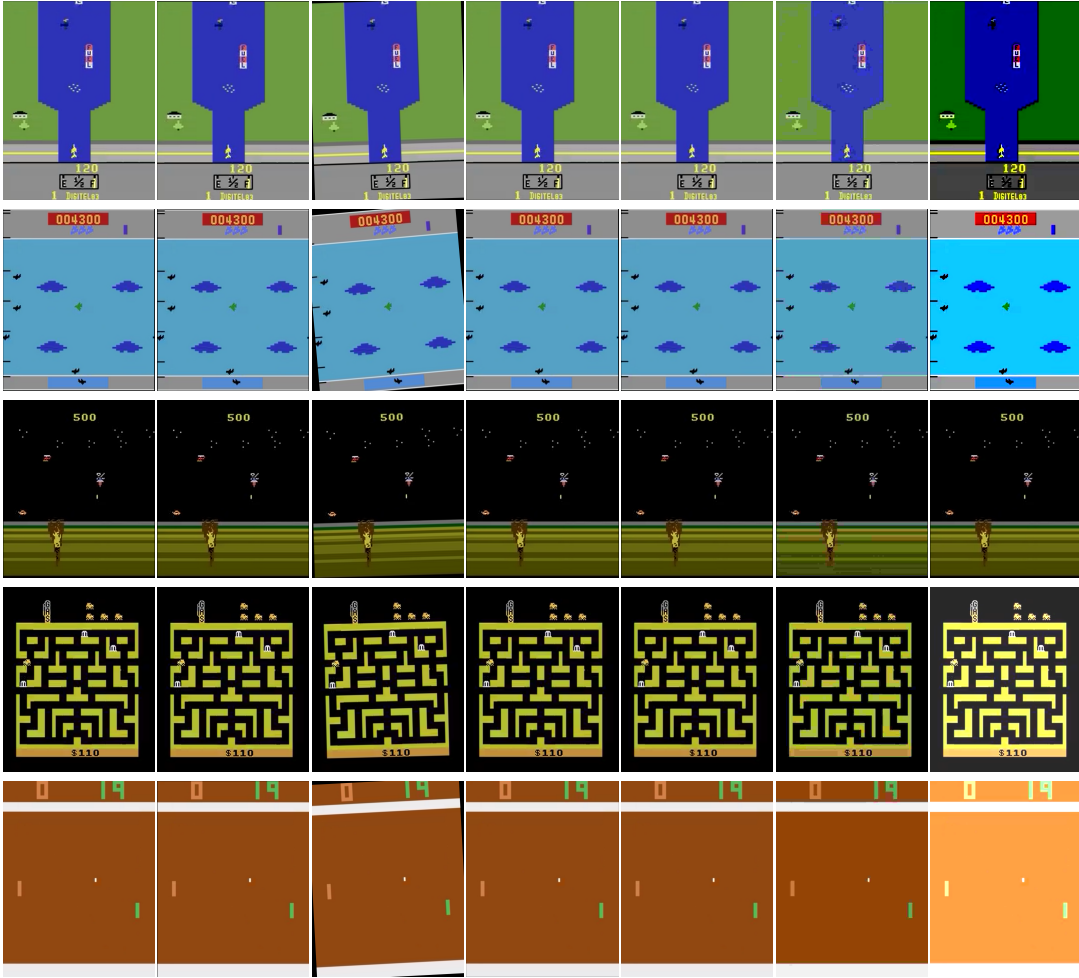


Figure 7: Original frame and environmental modifications. Columns: original frame, shifting, rotation, perspective transformation, blurring, compression artifacts. brightness and contrast. Rows: Riverraid, Timepilot, JamesBond, BankHeist and Pong.

Interestingly, we found that a majority of the games have high robustness against rotation. On the other hand, shifting and perspective transformation can reach a higher impact level than the state-of-the-art targeted attack while not being recognizable by human perception. We observed that in some games, such as Pong and Riverraid, brightness and contrast requires radical changes to cause the agent to fail, while for others the change required is imperceptible. Another thing we observed is

that for games like Pong, which is relatively trivial compared to other games in the Arcade Learning Environment, the threshold values for the environmental modification are higher. When the complexity in the game increases the environment modification thresholds decrease drastically. We think that this issue could become more important as deep reinforcement learning agents are deployed in more complex and realistic scenarios. The rows from Figure 7 are allocated to several games from the Atari environment and columns represent the modification type. Specifically, column 1 represents a state without a modification, column 2 represents shifting, column 3 represents rotation, column 4 represents perspective transformation, column 5 represents blurring, column 6 represents compression artifacts, column 7 represents brightness and contrast. Amongst these observations the most perceptually dissimilar are the TimePilot rotation and Pong rotation modifications. In one way the DDQN neural policy is quite robust to rotation modification in these environments. On the other hand, state-of-the-art adversarially trained neural policies (Huan et al. (2020)) are more sensitive to the rotation modification as we also pointed out in Section 5. Please see more details on perceptual similarities and neural policy performance in our website¹.

5 NATURAL PERTURBATIONS IN TIME DOMAIN

In this section we provide an analysis in the time domain to investigate if there are any additive effects that might effect the performance degradation. In the previous sections the environment modifications were applied to every state that the agent visited for both Carlini & Wagner (2017) and our proposed framework. In this section we will examine the effects of both of the adversarial perturbations and the natural perturbation framework when the perturbations are applied to only a small fraction of states.

Table 2: Impact comparison with the fraction of adversarial observations per episode e_{adv} .

	RiverRaid	TimePilot	BankHeist	Pong	JamesBond
Carlini&Wagner Impact	0.359	0.148	0.249	0.077	0.021
Shifting Impact	0.513	0.374	0.326	0.114	0.165
Perspective Transformation Impact	0.391	0.315	0.338	0.108	0.121
Blurring Impact	0.501	0.155	0.304	0.12	0.319
Brightness & Contrast Impact	0.517	0.188	0.313	0.098	0.154
Rotation Impact	0.417	0.192	0.260	0.079	0.044
Compression Artifacts Impact	0.184	0.262	0.267	0.017	0.198
Carlini&Wagner e_{adv}	0.096	0.020	0.021	0.062	0.081
Shifting e_{adv}	0.100	0.019	0.020	0.062	0.084
Perspective Transform e_{adv}	0.098	0.020	0.020	0.060	0.082
Blurring e_{adv}	0.099	0.019	0.020	0.061	0.083
Brightness e_{adv}	0.101	0.020	0.018	0.062	0.082
Rotation e_{adv}	0.099	0.021	0.020	0.061	0.083
Compression Artifacts e_{adv}	0.097	0.020	0.020	0.056	0.080
s_{adv} observation probability p	0.1	0.02	0.02	0.06	0.08

For this purpose we introduce the adversarial states s_{adv} in randomly sampled states where the observation s_{adv} is observed by the agent with probability p , and the original states s is observed by the agent with probability $1 - p$. We use $n_{s_{adv}}$ to denote the number of states where the agent observed s_{adv} instead of the original state s , and we use n_s to denote the total number of states visited by the agent in the given episode. We use e_{adv} to denote the fraction $n_{s_{adv}}/n_s$ of adversarial perturbations per episode. In Table 2 we show the attack impacts of Carlini & Wagner (2017) and our proposed framework with corresponding adversarial observation probability p averaged over 10 random episodes. Even for low p values our proposed framework obtains higher impact. Thus, to capture a broader view on the robustness of the deep reinforcement learning policies, the prior work on the timing perspective by Sun et al. (2020); Lin et al. (2017) based on worst-case distributional shift, can also be revisited with the natural perturbation framework. Observe that the fraction e_{adv} can differ slightly from p due to random fluctuations, therefore we also report these values in Table 2. Note that e_{adv} varies between games. This was done because each game has a different minimum threshold for e_{adv} to achieve stable impact across episodes.

¹<https://naturalperturbationframework.github.io/>

REFERENCES

- Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. *In 2017 IEEE Symposium on Security and Privacy (SP)*, pp. 39–57, 2017.
- Zhang Huan, Hongge Chen, Chaowe Xiao, Bo Li, Mingyan Liu, Duane S. Boning, and ChoJui Hseh. Robust deep reinforcement learning against adversarial perturbations on state observations. In Hugo Larochelle, Marc’ Aurelo Ranzato, Raia Hadsell, Maria-Florna Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- Yen-Chen Lin, Hong Zhang-Wei, Yuan-Hong Liao, Meng-Li Shih, ing-Yu Liu, and Min Sun. Tactics of adversarial attack on deep reinforcement learning agents. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 3756–3762, 2017.
- Volodymyr Mnih, Adria Badia Puigdomenech, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. *In International Conference on Machine Learning*, pp. 1928–1937, 2016.
- Jianwen Sun, Tianwei Zhang, Lei Xiaofei, Xie Ma, Yan Zheng, Kangjie Chen, and Yang. Liu. Stealthy and efficient adversarial attacks against deep reinforcement learning. *Association for the Advancement of Artificial Intelligence (AAAI)*, 2020.