1 APPENDIX

1.1 Novel environment definition

Algorithms 1, 2, and 3 display pseudocodes for generating ARTS, ARNO, and ARNS environments, respectively.

Algorithm 1 Pseudocode for ARTS environment

Initialization at $t = 0$:
generate AR_NOISE of size $(num\_dimensions = 1, max\_steps)$
Agent step
sample action $a_t \sim \pi_A(a_t|s_t)$
Environment step
function STEP(a_t)
    $o_{t+1} \leftarrow$ AR_NOISE(t)
    return $o_{t+1}, r_t, terminated$
end function

Algorithm 2 Pseudocode for ARNO environment

Initialization at $t = 0$:
generate AR_NOISE of size $(num\_dimensions, max\_steps)$
Agent step
sample action $a_t \sim \pi_A(a_t|s_t)$
Environment step
function STEP(a_t)
    $s_{t+1}, r_t, terminated \leftarrow$ ENVIRONMENT.STEP(a_t, s_t)
    $o_{t+1} \leftarrow s_{t+1}$ + AR_NOISE(t)
    return $o_{t+1}, r_t, terminated$
end function

Algorithm 3 Pseudocode for ARNS environment

Initialization at $t = 0$:
generate AR_NOISE of size $(num\_dimensions, max\_steps)$
Agent step
sample action $a_t \sim \pi_A(a_t|s_t)$
Environment step
function STEP(a_t)
    $s'_t \leftarrow s_t$ + AR_NOISE(t)
    $s_{t+1}, r_t, terminated \leftarrow$ ENVIRONMENT.STEP(a_t, s'_t)
    return $s_{t+1}, r_t, terminated$
end function

1.2 Out-of-Distribution Detection Terminology

Müller et al. [7] refer to *anomaly detection* as the general task of identifying datapoints that substantially differ from normality in RL environments. The authors do not use the term out-of-distribution (OOD) detection at all. Danesh and Fern [4] coin the term *out-of-distribution dynamics (OODD) detection* to define the general task of "*detecting when the dynamics of a temporal process change compared to the training-distribution dynamics*". The authors differentiate between *sensor-injected* anomalies, which "corrupt the environment observations received by the agent", and *dynamics-injected* anomalies, which *"directly change the dynamics of the environment by modifying key physical parameters of the environment simulator"*.

Haider et al. [5] provide a more comprehensive review of terminology issues in the field of OOD detection in reinforcement learning. To our knowledge, they also propose the most recent re-definition of out-of-distribution in reinforcement learning as *"severe perturbations of the Markov Decision Process, which effectively change the semantics of the system"*, instead of simply changing the observations that the agent receives. The authors draw on the aforementioned review by Yang et al. [9] to argue that a semantic anomaly should effectively shift the transition function of an MDP, i.e. introduce new semantic concepts or changing the environment dynamics. In effect, this corresponds to renaming *dynamics-injected* anomalies from Danesh and Fern [4] to *semantic anomalies*.

While we try to follow the existing terminologies, we believe that the most recent definitions of OOD detection offered by Danesh and Fern [4] and Haider et al. [5] could be improved for several reasons. First, Haider et al. [5] define OOD detection as the identification of semantic anomalies, but do not provide a meaningful term for detecting anomalies that target the observations that the agent receives, which has been a focus of many previous works in the literature on OOD detection in RL [8] [6] [4].

Second, in many real-world environments, reinforcement learning agents could face both sensory and semantic anomalies simultaneously. We can take the example of a self-driving car that is trained to drive in normal conditions, and suddenly exposed to a heavy hailstorm. On the one hand, the ice will change the observations that the car receives from its cameras, introducing observational noise. On the other, the ice will also make the road more slippery, effectively changing the environment dynamics. However, the terminology offered by [5] does not allow to effectively distinguish between these anomalies. We think it is reasonable to suggest there should be a common terminology to differentiate between these two types of anomalies in the environment, and that ideally, a well-performing detector should be able to detect both of them.

Lastly, while the terminology of Danesh and Fern [4] offers a distinction between environmental anomalies, the suggested terms do not correspond to the terminology used by OOD detection outside of reinforcement learning.

As we discuss in Section 4 of the paper, it is relatively simple to align the labels for anomalies in reinforcement learning with the standardized terminology from Yang et al. [9].

For this reason, Section 4 proposes a clarification of terminology for OOD detection in reinforcement learning, which aligns it with literature from other machine learning domains.

1.3 Implementation details

1.3.1 ARTS, ARNO, ARNS scenarios. To generate noise with varied orders of correlation, we use the implementation of Autoregressive Process from `statsmodels`[1] library in Python.

1.3.2 ARNS Acrobot environment. We implement ARNS scenarios with Light, Medium, and Strong levels of noise on Cartpole

[1]More information about the library can be found here: https://www.statsmodels.org/stable/index.html

and Reacher, as the implementation on Acrobot leads to inconsistent agent policies. In the implementation of Acrobot environment used in this project [1], adding additional noise to the underlying states tends to increase the agent reward, which is inconsistent with our definition of Light, Medium, and Strong noise in terms of the reduction in the average cumulative average reward over an episode.

1.3.3 Changepoint detectors. To implement the changepoint detectors from Chen et al. [3] and Chan [2], we use the ocd package[2] for R, published by Chen et al. [3]. The package contains the implementation for both detectors. Both algorithms rely on the patience parameter, which is defined as the average run length under the null hypothesis [3]. Following the definition, we set patience to the average length of the episode under uncorrelated noise in a given setting. For detection threshold, we use the option of automatically calculating the threshold using Monte Carlo simulations, implemented in the ocd package.

Since this project is built on Python codebase, we also had to adapt these detectors from R to Python. Therefore, a small contribution that we provide is, to our knowledge, the first implementation of the algorithms of Chen et al. [3] and Chan [2] in Python.

1.4 AUROC results

When reporting the performance of detectors using the Area under the Receiver Operator Characteristic (AUROC), we consider a two-sided test. In other words, the victim agent tests whether the predicted anomaly scores are either higher or lower than the anomaly scores predicted for the unperturbed observations. Therefore, we report the AUC score as $max(AUROC, 1 - AUROC)$.

REFERENCES

[1] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. arXiv:arXiv:1606.01540

[2] Hock Peng Chan. 2017. Optimal sequential detection in multi-stream data. (2017).

[3] Yudong Chen, Tengyao Wang, and Richard J Samworth. 2022. High-dimensional, multiscale online changepoint detection. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84, 1 (2022), 234–266.

[4] Mohamad H Danesh and Alan Fern. 2021. Out-of-Distribution Dynamics Detection: RL-Relevant Benchmarks and Results. *arXiv preprint arXiv:2107.04982* (2021).

[5] Tom Haider, Karsten Roscher, Felippe Schmoeller da Roza, and Stephan Günnemann. 2023. Out-of-Distribution Detection for Reinforcement Learning Agents with Probabilistic Dynamics Models. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 851–859.

[6] Aaqib Parvez Mohammed and Matias Valdenegro-Toro. 2021. Benchmark for out-of-distribution detection in deep reinforcement learning. *arXiv preprint arXiv:2112.02694* (2021).

[7] Robert Müller, Steffen Illium, Thomy Phan, Tom Haider, and Claudia Linnhoff-Popien. 2022. Towards Anomaly Detection in Reinforcement Learning. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 1799–1803.

[8] Andreas Sedlmeier, Thomas Gabor, Thomy Phan, Lenz Belzner, and Claudia Linnhoff-Popien. 2019. Uncertainty-based out-of-distribution classification in deep reinforcement learning. *arXiv preprint arXiv:2001.00496* (2019).

[9] Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. 2021. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334* (2021).

[2] More information about the ocd package can be found at: https://cran.r-project.org/web/packages/ocd/index.html