Do World Models Already Have Built-In Anomaly Detection?

Novelty Detection in Reinforcement Learning with World Models

Georgia Institute of Technology

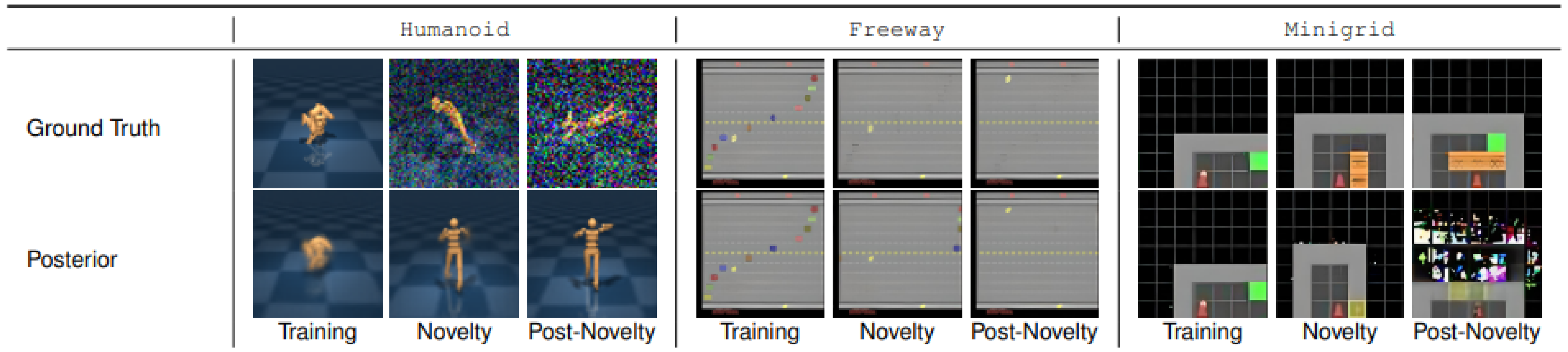
Geigh Zollicoffer, Kenneth Eaton*, Jonathan C Balloch, Julia Kim, Wei Zhou, Robert Wright*, Mark Riedl

Georgia Institute of Technology, Georgia Tech Research Institute*

ICML International Conference On Machine Learning

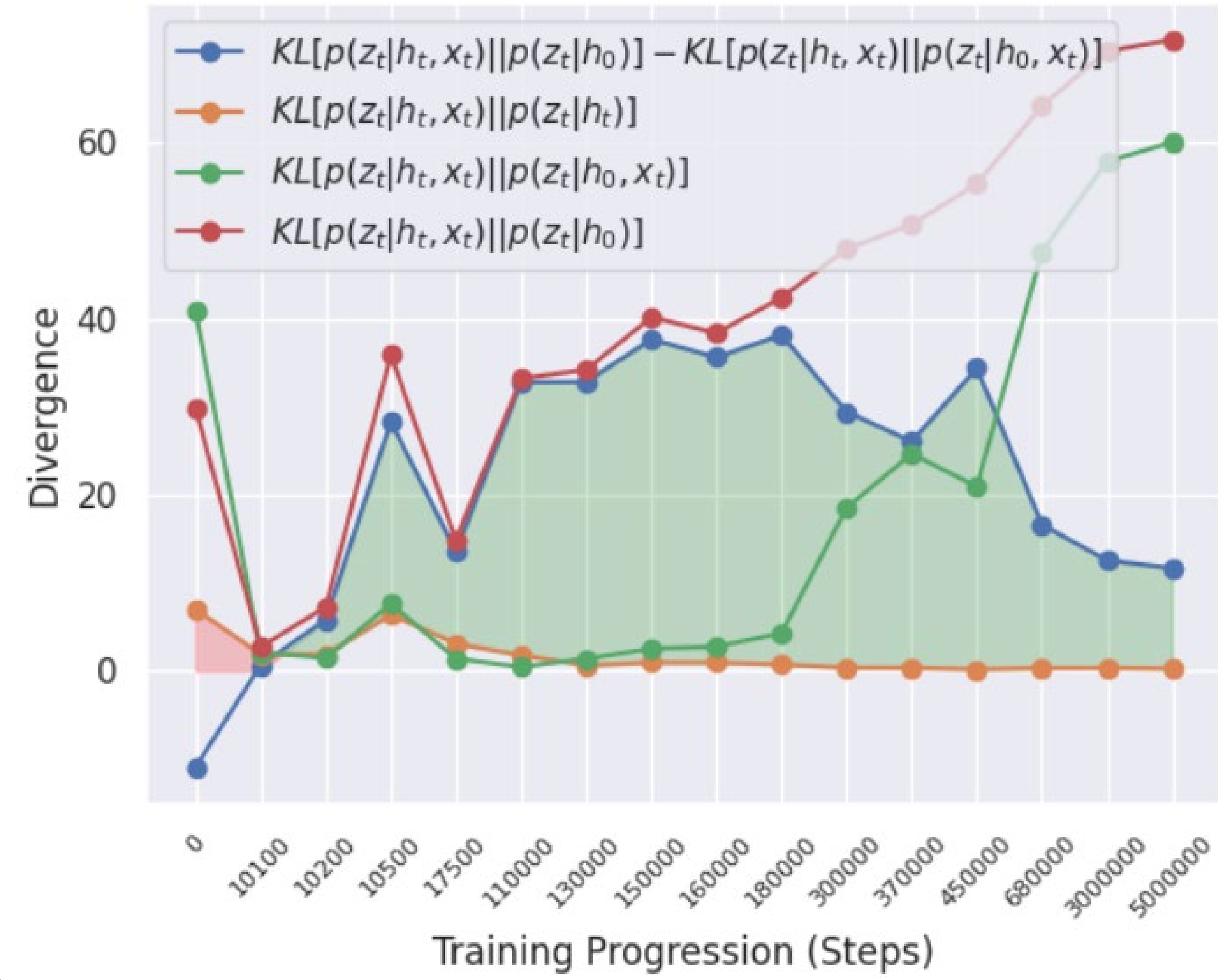
1. Problem Statement

Reinforcement learning (RL) using world models has found significant recent successes. However, when a sudden change to world mechanics or properties occurs then agent performance and reliability can dramatically decline.



Our technique calculates a novelty threshold bound without additional hyperparameters by considering how much the actual world observation deviates from the distribution of world observations that the agent predicts it will encounter.

2. Latent Based Detection



World Model Architecture:

$$\begin{cases} \text{Recurrent model:} h_t = f_\phi(h_{t-1}, z_{t-1}, a_{t-1}) \\ \text{Representation model: } p_\phi(z_t|h_t, x_t) \\ \text{Transition prediction model: } p_\phi(\hat{z}_t|h_t) \\ \text{Image prediction model: } p_\phi(\hat{x}_t|h_t, z_t) \\ \text{Reward prediction model: } p_\phi(\hat{r}_t|h_t, z_t) \\ \text{Discount prediction model: } p_\phi(\hat{\gamma}_t|h_t, z_t) \end{cases}$$

$$\mathcal{L}(\phi) = \mathop{\mathbb{E}}_{p_\phi(z|a,x)}\Big[\sum_t^T -\ln p_\phi(x_t|h_t,z_t) - \ln p_\phi(r_t|h_t,z_t) - \ln p_\phi(\gamma_t|h_t,z_t) + \beta KL\left[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_t)\right]\Big]$$

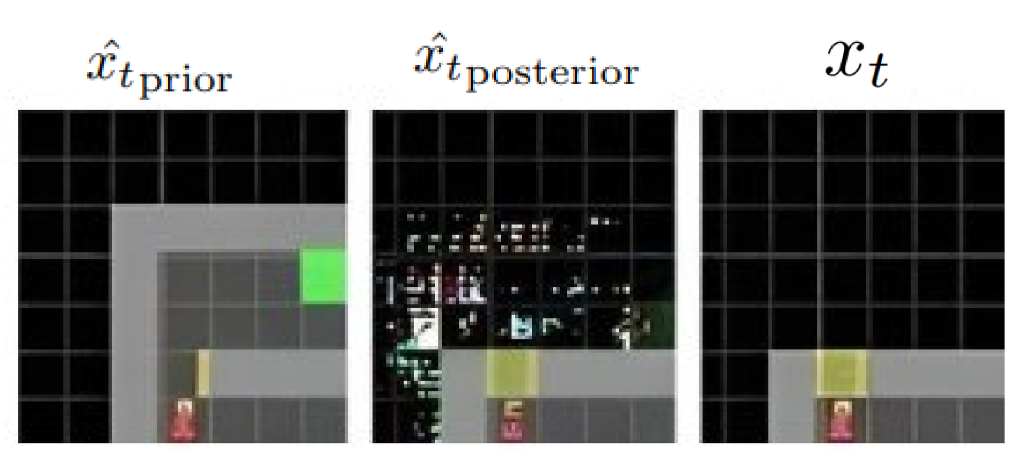
1. Measure the cross-entropy score comparison and derive in terms of KL to use world model components:

$$\begin{aligned} H(p_\phi(z_t|h_t,x_t), p_\phi(z_t|h_0)) - H(p_\phi(z_t|h_t,x_t), p_\phi(z_t|h_0,x_t)) &= \\ (KL[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_0)] + H(p_\phi(z_t|h_t,x_t))) \\ - KL[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_0,x_t)] - H(p_\phi(z_t|h_t,x_t)) &= \\ KL[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_0)] \\ - KL[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_0,x_t)] \end{aligned}$$

2. Derive a bound using the minimization objective of the world model as an anomaly score:

$$\begin{aligned} KL[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_t)] \leq \\ KL[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_0)] \\ - KL[p_\phi(z_t|h_t,x_t)||p_\phi(z_t|h_0,x_t)] \end{aligned}$$

3. Observation based Detection



Derive a bound from difference in pixel values compared to tuned hyperparameter λ:

$$\frac{\sum_i^N |\hat{x}_{t_{\text{prior}}}^{\,i} - \hat{x}_{t_{\text{posterior}}}^{\,i}|}{N} \leq \lambda$$

Note, the prior and posterior images are generated without providing the ground truth, x_t, as input

4. Experimental Evaluation

MuJoCo · Minigrid · The Arcade Learning Environment

Task: Detect a novel transition as soon as the agent experiences it, while minimizing false positives.

Techniques: KL Bound (Latent Based), RIQN (Ensemble Based), PP-MARE (Observation Based).

Results: Agents can utilize world models to detect novel transitions at a more effective rate than previous methods.

Method	False Positive Rate			
Atari	**Boxing**	**Kangaroo**	**Freeway**	**SeaQuest**
KL Bound	$\leq 10^{-2}$	$\leq 10^{-2}$	$\leq 10^{-2}$	$\leq 10^{-2}$
PP-Mare	.04	$\leq 10^{-2}$	$\leq 10^{-2}$	$\leq 10^{-2}$
RIQN	.17	.39	.24	.43
DMC	**Cartpole-B.**	**Quadruped-R.**	**Humanoid-S.**	**Walker-W.**
KL Bound	$\leq 10^{-2}$	$\leq 10^{-2}$	$\leq 10^{-2}$	$\leq 10^{-2}$
PP-Mare	.01	.01	$\leq 10^{-2}$	$\leq 10^{-2}$
RIQN	.68	.19	.05	.18

Method	Average False Positive Rate		Inference Run-time Speedup	
	DMC	Atari	DMC	Atari
RIQN	.275	.308	$\times 1$	$\times 1$
PP-Mare	$\leq 10^{-2}$	$\leq 10^{-2}$	$\times 1.16 \cdot 10^2$	$\times 4.45 \cdot 10^1$
KL bound	$\leq 10^{-2}$	$\leq 10^{-2}$	$\times 1.34 \cdot 10^3$	$\times 5.12 \cdot 10^2$

World Model based detection improves the speed and minimizes the average false positive rate.

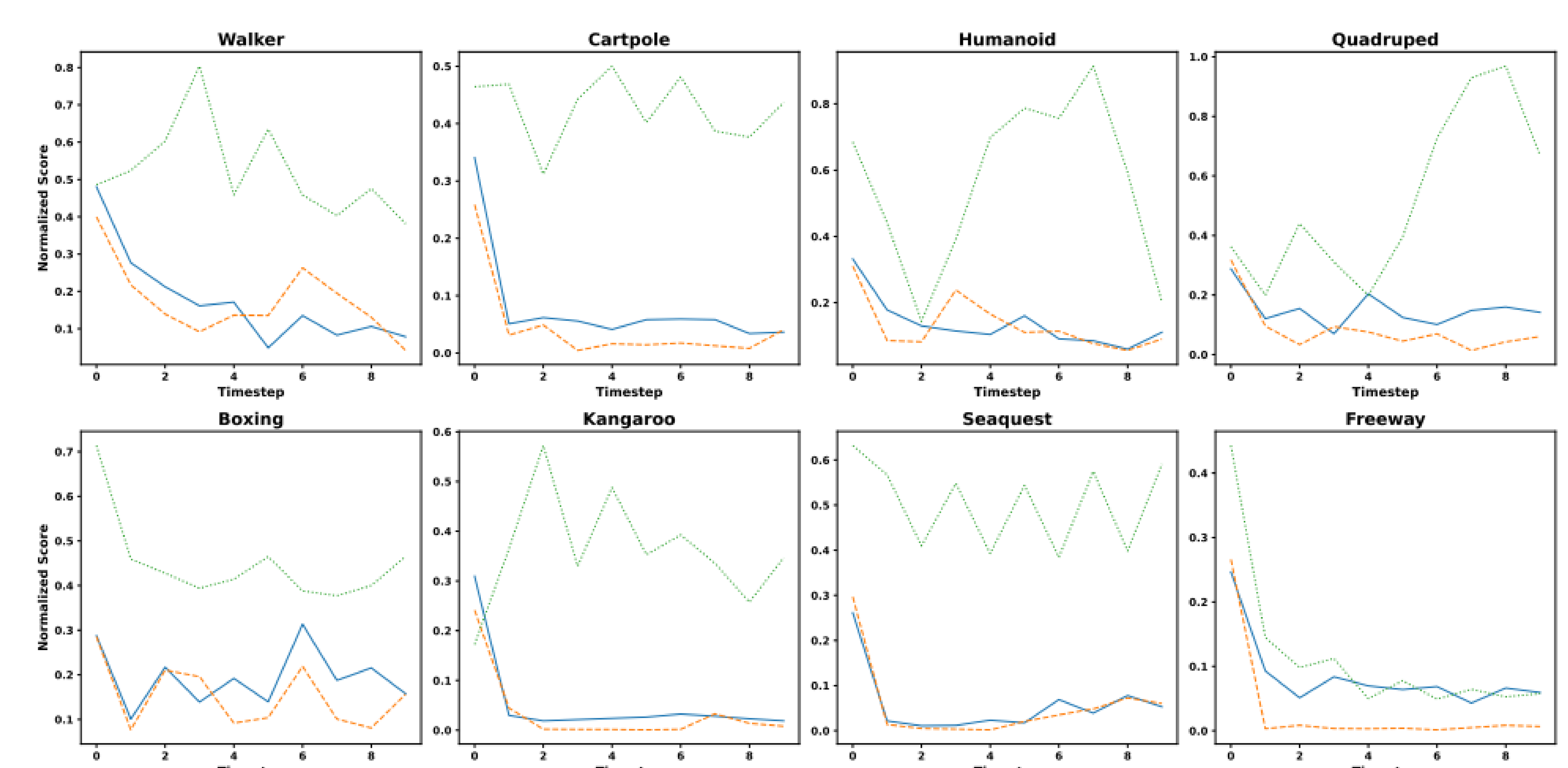
	ADE↓	AUC↑	ADE↓	AUC↑	ADE↓	AUC↑
Boxing	OneArm		BodySwitch		Doppleganger	
KL Bound	52.6	.708	$\leq 10^{-2}$	$\geq .99$	$\leq 10^{-2}$	$\geq .99$
PP-Mare (2)	22.5	.605	6.30	.915	5.4	.862
RIQN ($10^{-5}, 10^{-7}$)	347.5	.505	509.3	.380	103.1	.401
Kangaroo	Floorswap		Difficulty+		DisableMonkey	
KL Bound	9.9	.787	$\leq 10^{-2}$	$\geq .99$	.960	$\geq .99$
PP-Mare (.5)	85.2	.281	42.5	$\geq .99$	41.3	.937
RIQN (10^{-2},10^{-2})	166.3	.541	94.2	$\geq .99$	93.1	.710
Freeway	InvisibleCars		ColorCars		FrozenCars	
KL Bound	$\leq 10^{-2}$	$\geq .99$	$\leq 10^{-2}$	$\geq .99$	$\leq 10^{-2}$	.985
PP-Mare (.5)	2.33	$\geq .99$	1.84	$\geq .99$	2.60	.931
RIQN ($10^{-7}, 10^{-9}$)	2.87	.938	2.62	$\geq .99$	.502	.980
SeaQuest	DisableEnemy		Gravity		UnlimitedOxygen	
KL Bound	.202	.962	45.8	.949	$\leq 10^{-2}$	$\geq .99$
PP-Mare (.7)	111.6	.678	24.1	.882	3.73	.938
RIQN ($10^{-2}, 10^{-3}$)	45.3	.272	157.3	.390	16.0	.701

	ADE↓	AUC↑	ADE↓	AUC↑	ADE↓	AUC↑
DoorKey-6x6	LavaGap		BrokenDoor		DoorGone	
KL Bound	.110	.732	$\leq 10^{-2}$	.939	.120	.940
PP-Mare (1)	.170	.765	.017	.784	.080	.685
RIQN ($10^{-2}, 10^{-2}$)	$\leq 10^{-2}$	.760	2.56	.920	.066	.600
	Teleport		ActionFlip			
KL Bound	$\leq 10^{-2}$	.992	$\leq 10^{-2}$	.991		
PP-Mare (1)	.080	.959	.105	.962		
RIQN ($10^{-2}, 10^{-2}$)	6.29	.950	2.34	.890		

	ADE↓	AUC↑	ADE↓	AUC↑	ADE↓	AUC↑
Cartpole-3D-Balance	LowPerterb		HighPerterb		LowNoise	
KL Bound	51.9	.799	49.2	.774	24.1	.853
PP-Mare (.7)	45.4	.529	43.7	.522	61.3	.600
RIQN (10^{-1},10^{-1})	6.20	.915	5.75	.890	3.67	$\geq .99$
	HighNoise		LowDamp		HighDamp	
KL Bound	2.86	$\geq .99$	20.2	.784	25.0	.770
PP-Mare (0.7)	22.7	.840	110.8	.395	94.5	.448
RIQN (10^{-1},10^{-1})	3.33	.923	5.33	.543	1.66	.520
Quadruped-3D-Run	LowPerterb		HighPerterb		LowNoise	
KL Bound	1.00	.908	1.30	.901	24.8	.839
PP-Mare (1.5)	1.00	.793	1.26	.830	17.3	.645
RIQN (10^{-1},10^{-1})	13.3	.806	12.7	.883	4.70	.850
	HighNoise		LowDamp		HighDamp	
KL Bound	3.82	$\geq .99$	89.8	.512	112.9	.507
PP-Mare (1.5)	3.94	.821	49.1	.448	52.86	.415
RIQN (10^{-1},10^{-1})	4.02	.980	94.3	.426	103.0	.489
Humanoid-3D-Stand	LowPerterb		HighPerterb		LowNoise	
KL Bound	44.4	.925	$\leq 10^{-2}$	$\geq .99$	9.05	.755
PP-Mare (4)	17.9	.833	15.2	.849	157.0	.363
RIQN (10^{-1},10^{-1})	64.8	.854	32.3	.983	2.16	.919
	HighNoise		LowFriction		HighFriction	
KL Bound	$\leq 10^{-2}$	$\geq .99$	59.8	.748	18.75	.845
PP-Mare (4)	258.5	.372	17.3	.853	8.10	.822
RIQN (10^{-1},10^{-1})	1.50	.982	53.1	.549	12.7	.480
Walker-3D-Walk	LowPerterb		HighPerterb		LowNoise	
KL Bound	$\leq 10^{-2}$	$\geq .99$	$\leq 10^{-2}$	$\geq .99$	17.6	.573
PP-Mare (7)	18.7	.725	7.16	.798	50.9	.407
RIQN (10^{-1},10^{-1})	8.11	.743	7.80	.715	2.73	.883
	HighNoise		LowFriction		HighFriction	
KL Bound	1.48	.911	17.25	.961	5.93	$\geq .99$
PP-Mare (7)	21.1	.546	2.37	.943	1.98	.962
RIQN (10^{-1},10^{-1})	2.10	.986	10.55	.700	13.6	.722

Left: Evaluation scores of AUC and Average Delay Error (ADE) on Minigrid and Atari environments

Right: Evaluation scores of AUC and Average Delay Error (ADE) on DeepMind Control environments.



The normalized anomaly score trend of each detection method as the episode progresses in the nominal environment.