Regularized Behavior Cloning for Blocking the Leakage of Past Action Information

NeurIPS 2023 | Spotlight

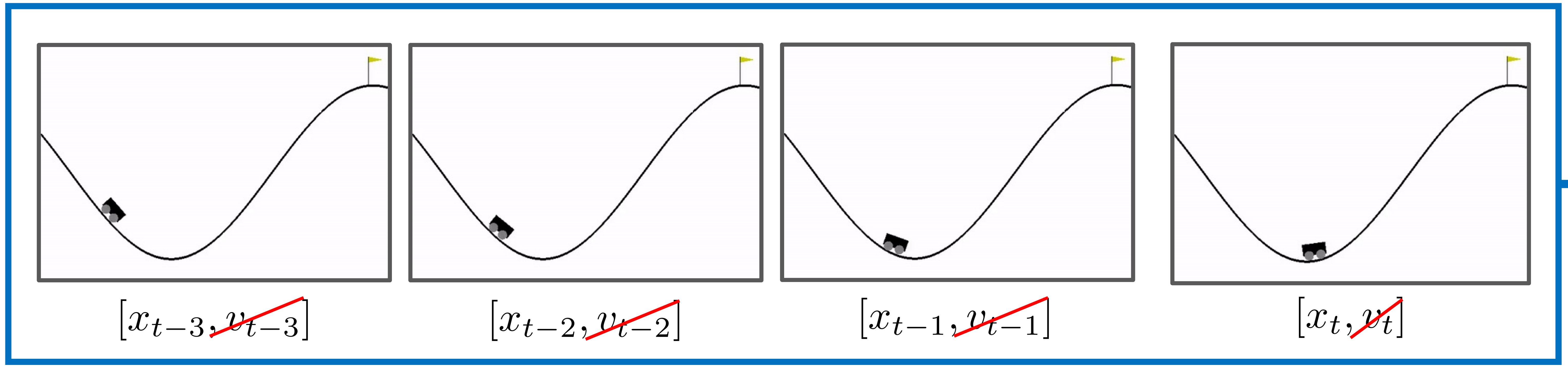
Seokin Seo, HyeongJoo Hwang,
Hongseok Yang, and Kee-Eung Kim

KAIST

Problem

Imitation Learning with Observation Histories (ILOH)

…

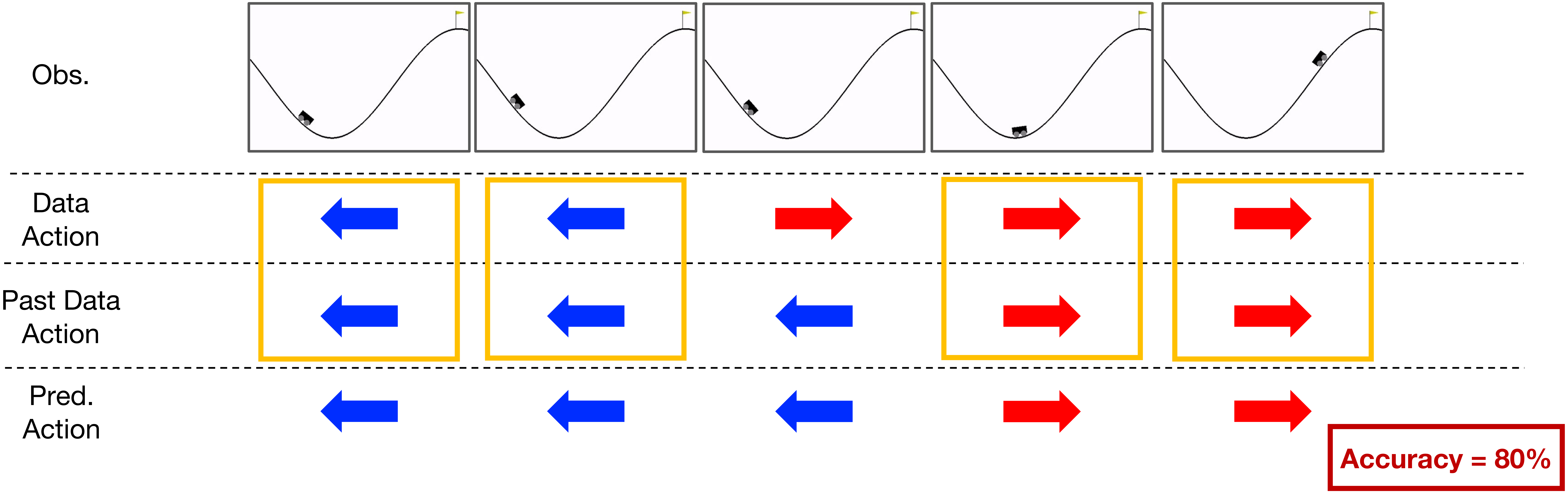


Use the observation history
as a policy input.

- Individual observation does not contain sufficient control-relevant information (e.g. velocity, …).
- ILOH use observation histories as policy inputs to imitate the expert actions.

Problem

Past Action Information in ILOH



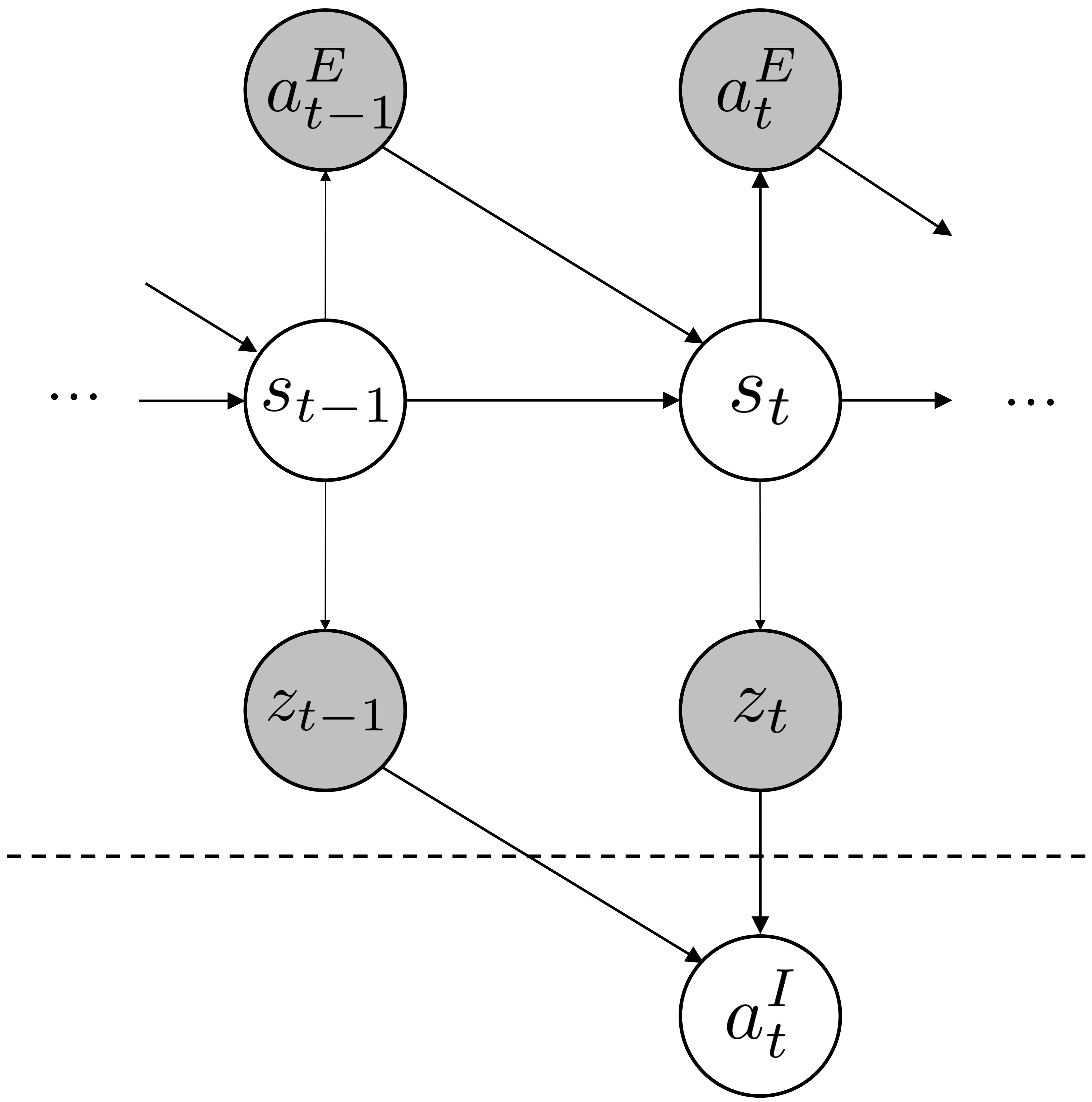
- The unnecessary **past action information** in ILOH is harmful!

  May repeat only past actions: known as *copycat problem* [1], *inertia problem* [2], *latching effect* [3], …

[1] Wen et al., “Fighting Copycat Agents from Observation Histories”, NeurIPS 2020.
[2] Codevilla et al., “Exploring the Limitations of Behavior Cloning for Autonomous Driving”, ICCV 2019.
[3] Swamy et al., “Sequence Model Imitation Learning with Unobserved Contexts”, NeurIPS 2022.

NEURAL INFORMATION PROCESSING SYSTEMS

Past Action Information Leakage



Q. What is the leaked past action information?

Information leaked from observation history that hinders to accomplish:

$$a^I_t \perp\!\!\!\perp a^E_{t-1} | a^E_t$$

Hence, amount of leaked past action information can be measured by a conditional dependence metric.

We use the kernel-based metric HSCIC(Hilbert-Schmidt Conditional Independence Criterion)[4].

$$\mathrm{HSCIC}(X, Y|z) := \mathrm{MMD}^2(P_{XY|z}, P_{X|z}P_{Y|z}) \\ = \|\mu_{P_{XY|z}} - \mu_{P_{X|z}} \otimes \mu_{P_{Y|z}}\|^2_{\mathcal{H}_X \otimes \mathcal{H}_Y}$$

Can we apply the metric to training?

KAIST

[4] Park et al., "A Measure-theoretic Approach to Kernel Conditional Mean Embeddings", NeurIPS 2020.

Solution

Regularized BC Framework

- We regularize φ_t , the representation of the observation history:

$$\mathcal{L}(\pi) := \mathcal{L}_{\text{bc}}(\pi; a_t^E) + \alpha \cdot \mathcal{L}_{\text{reg}}(\varphi_t; a_{t-1}^E, a_t^E)$$

standard BC loss

regularize to satisfy $\varphi_t \perp\!\!\!\perp a_{t-1}^E | a_t^E$

Q. How to regularize? $\text{HSCIC}(\varphi_t; a_{t-1}^E | a_t^E)$

Advantages

HSCIC can be estimated in a *closed-form* solution. → (1) **No** nested optimization; (2) **No** additional neural network

HSCIC is based on *non-parametric* statistics. → (3) **No** assumption on distribution

Experiment

Performance Comparison on D4RL Dataset

- We use expert demonstrations provided by D4RL benchmark [5] for all experiments.
- 4 continuous control task (MuJoCo) + 1 pixel-based autonomous driving task (CARLA)

Task	w	BC	KF	PrimeNet	RAP	FCA	MINE	PALR (Ours)
hopper	2	32.5 ± 2.9	32.0 ± 1.9	30.0 ± 1.6	20.2 ± 1.4	31.9 ± 2.5	25.0 ± 1.9	$\mathbf{42.0 \pm 2.4}$
	4	47.7 ± 3.4	45.7 ± 1.0	45.3 ± 2.8	32.6 ± 2.6	36.9 ± 2.4	37.6 ± 3.1	$\mathbf{58.4 \pm 2.8}$
walker2d	2	53.0 ± 2.7	50.0 ± 2.3	48.5 ± 3.3	15.8 ± 2.0	63.1 ± 2.7	58.6 ± 5.5	$\mathbf{79.8 \pm 2.3}$
	4	63.2 ± 6.3	77.4 ± 2.0	79.2 ± 3.3	25.4 ± 2.1	$\mathbf{81.9 \pm 3.3}$	68.7 ± 6.7	$\mathbf{83.4 \pm 5.4}$
halfcheetah	2	74.1 ± 2.3	64.3 ± 1.4	61.5 ± 1.9	63.9 ± 2.1	78.2 ± 2.8	76.3 ± 1.9	$\mathbf{86.4 \pm 1.1}$
	4	68.4 ± 2.6	55.7 ± 4.1	45.5 ± 1.7	59.0 ± 2.7	69.9 ± 2.6	73.4 ± 2.4	$\mathbf{79.1 \pm 4.3}$
ant	2	56.3 ± 3.5	54.9 ± 1.7	51.7 ± 2.4	44.1 ± 1.2	51.1 ± 2.2	53.9 ± 1.9	$\mathbf{59.6 \pm 3.0}$
	4	$\mathbf{64.4 \pm 1.8}$	48.6 ± 3.8	58.2 ± 1.9	48.6 ± 2.6	57.7 ± 1.3	56.6 ± 1.8	$\mathbf{64.6 \pm 2.5}$
carla-lane	3	52.5 ± 6.2	66.6 ± 2.1	58.2 ± 2.2	25.3 ± 5.4	57.1 ± 3.1	60.1 ± 4.1	$\mathbf{72.9 \pm 2.6}$

* Video sources : https://gymnasium.farama.org/environments/mujoco, https://www.youtube.com/watch?v=om8klsBj4rc

[5] Fu et al., “D4RL: Datasets for deep data-driven reinforcement learning.” In: arXiv preprint arXiv:2004.07219 (2020).

Summary

1. Past Action Information Leakage : $a_t^I \not\perp\!\!\!\perp a_{t-1}^E | a_t^E$
2. Past Action Leakage Regularization (PALR) :

   a simple HSCIC-regularized BC can effectively prevent the leakage and can improve imitation learning performance.

Poster Great Hall & Hall B1+B2 #1419

Wed 13 Dec 5 PM— 7 PM

Code & Paper