Object-Aware Regularization for Addressing Causal Confusion in Imitation Learning

Jongjin Park[1*], Younggyo Seo[1*], Chang Liu[2], Li Zhao[2],
Tao Qin[2], Jinwoo Shin[1], Tie-Yan Liu[2].

[*] Equal contribution, in alphabetical order.

[1] Korea Advanced Institute of Science and Technology

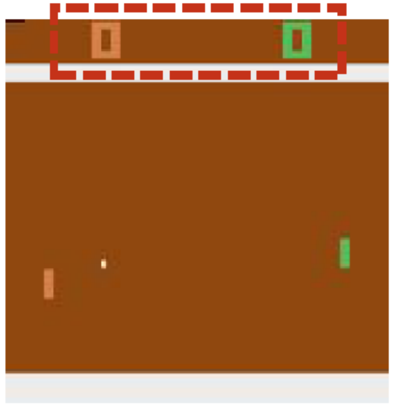
[2] Microsoft Research Asia

Introduction

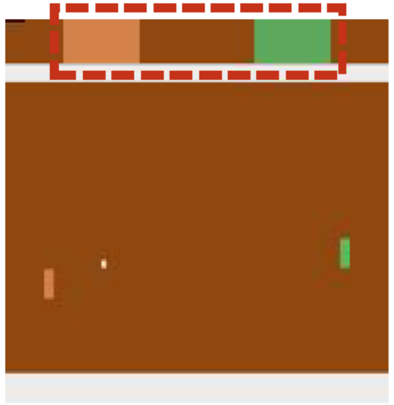
- Behavioral Cloning (BC)
  - Imitation Learning (IL) as a supervised learning problem.
  - (+) Simple. No need of environment interaction.
  - (-) The **causal confusion** problem:

    BC policy may find a “lazy way” to only focus on the noticeable **effect** of the action, but ignore the **cause** when it is subtle or complicated.

    E.g., BC policy fails when the **score** is present in the training states.



(a) Original



(b) Masked

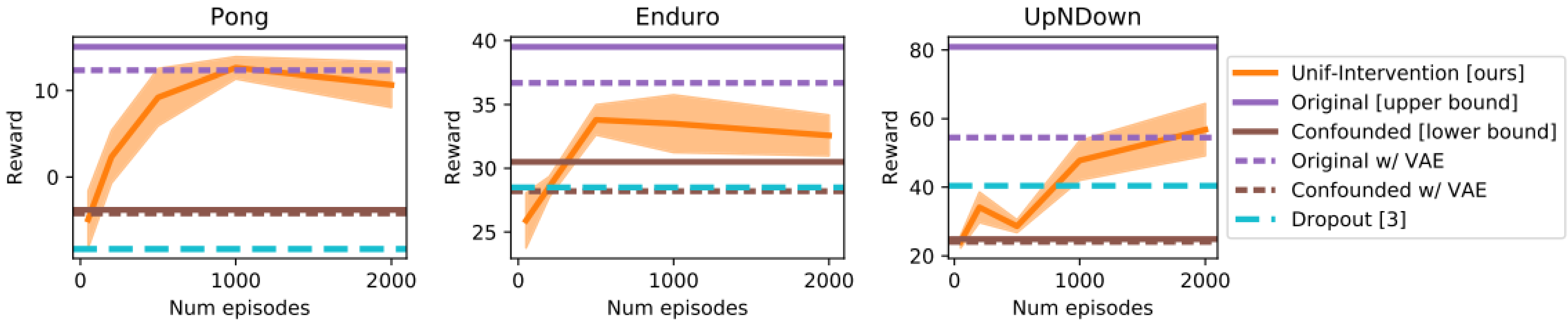
Setup: Train	Setup: Eval	Scores
Original	Original	3.1 $\pm$ 1.4
	Masked	-15.6 $\pm$ 9.2
Masked	Original	**15.9** $\pm$ 0.4
	Masked	**16.6** $\pm$ 0.6

(c) Performance of behavioral cloning

Introduction

- Possible solutions:
  a. Observational causal discovery:
    - Requires tabular/structured data, not suitable for sensory data like images.

  b. Interventional causal discovery [de Haan'19]:
    - Disentangled representation learned using beta-VAE.
    - Requires expert/environment interaction to infer the causal graph.



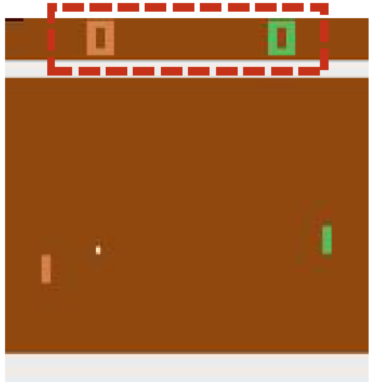
[de Haan'19] de Haan, Pim, Jayaraman, Dinesh, and Levine, Sergey. Causal confusion in imitation learning. In *Advances in Neural Information Processing Systems*, 2019.

Method: Main Idea

- Why BC fails, exactly?

  BC policy focused region often collapses onto a small region, usually the most noticeable effect.

BC policy
focused region

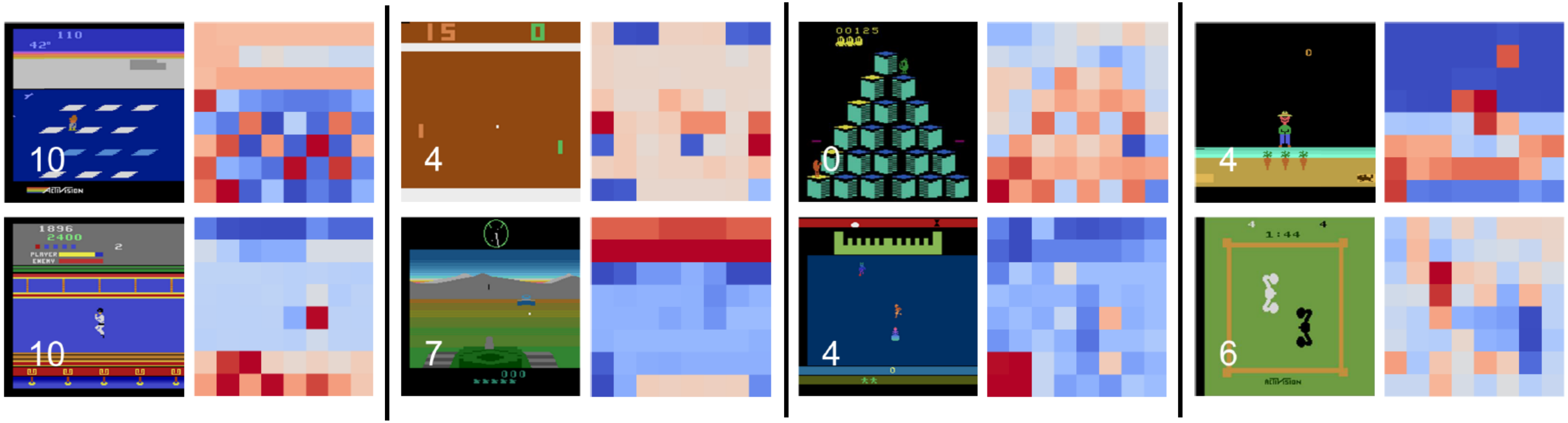


- Main idea:

  Encourage the policy to **uniformly attend to all semantic objects** in the image.

Method: OREO

- **OREO**: **O**bject-aware **RE**gularizati**O**n
  a. Extract semantic objects in an image:
  Leverage the discrete code of **VQ-VAE** [v.d. Oord'17].

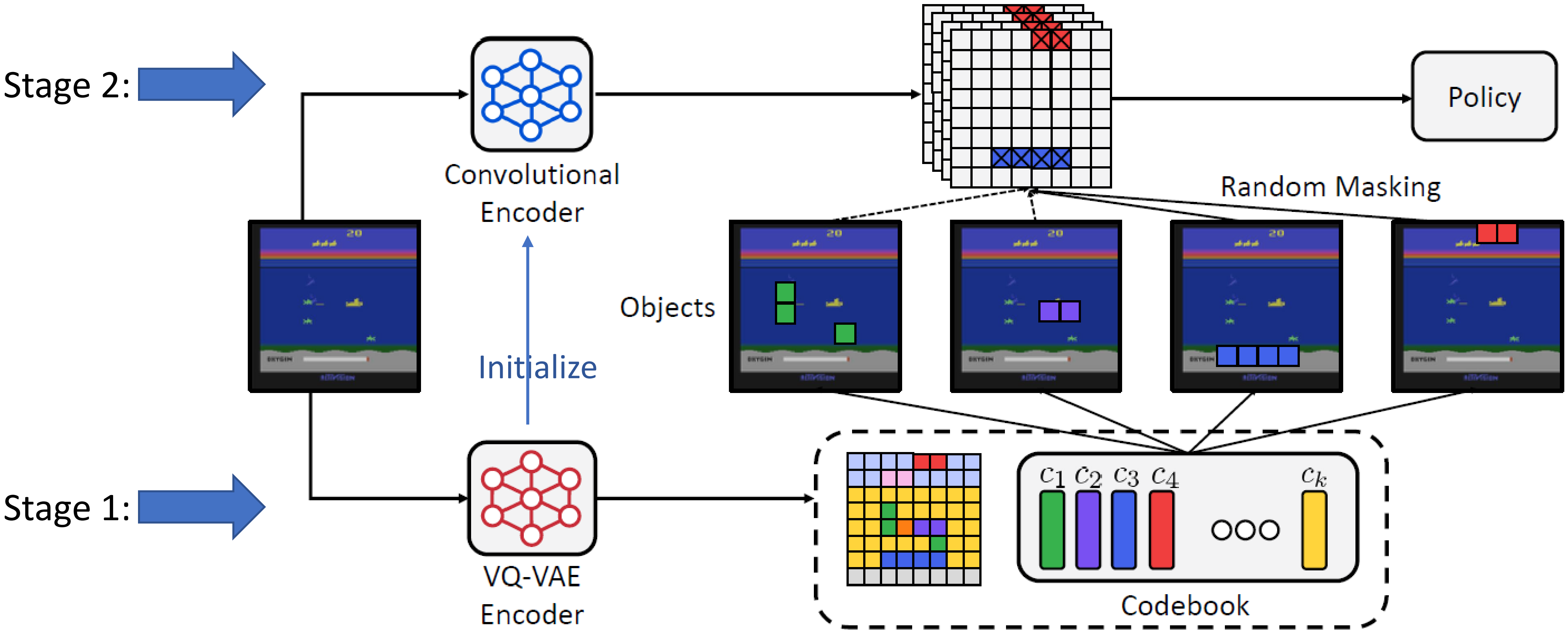


- The latent vector itself still keeps **spacial** information,
- but similar discrete code values mark similar **semantic objects**.

[v.d. Oord'17] van den Oord, A., Vinyals, O., & Kavukcuoglu, K. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, 2017.

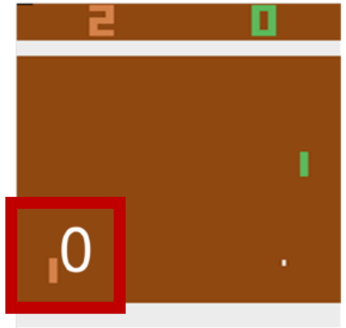
Method: OREO

- **OREO**: **O**bject-aware **RE**gularizati**O**n
  b. Enforcing attendance to all semantic objects:
  Randomly masking out an object, i.e. units that share the same discrete code.



Experiments

- Confounded Atari Environments
  [de Haan'19]



Previous action as a noticeable effect

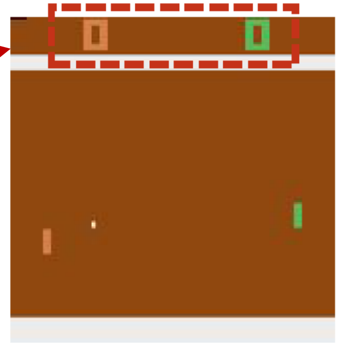
interaction-free version

a direct causal method

Environment	BC	Dropout	DropBlock	Cutout	RandomShift	CCIL†	CRLR	OREO
Alien	954.1	1003.8	926.4	973.3	806.5	820.0	82.5	**1056.2**
Amidar	95.8	89.4	110.1	**118.7**	98.0	74.9	12.0	105.7
Assault	793.8	820.4	815.0	687.6	828.9	683.3	0.0	**840.9**
Asterix	292.2	313.8	345.4	212.4	135.5	643.2	**650.0**	180.8
BankHeist	442.1	485.7	508.4	486.1	367.2	**653.5**	0.0	493.9
BattleZone	11921.2	12457.5	12025.0	11107.5	9180.0	6370.0	1468.8	**12700.0**
Boxing	18.8	20.3	32.2	20.5	**38.3**	34.8	-43.0	36.4
Breakout	**5.7**	5.4	4.8	1.0	2.0	0.5	0.0	4.2
ChopperCommand	874.2	921.4	919.4	1016.1	936.4	760.6	**1077.2**	977.4
CrazyClimber	45372.9	39501.6	38345.6	44523.2	41924.0	22616.8	112.5	**55523.4**
DemonAttack	157.2	180.5	167.8	173.1	**241.8**	171.3	0.0	224.5
Enduro	241.4	250.4	341.8	119.6	316.4	143.1	3.9	**522.8**
Freeway	32.3	32.4	32.7	32.5	33.0	**33.1**	21.4	32.7
Frostbite	116.3	124.5	128.2	**139.4**	121.6	53.3	80.0	129.9
Gopher	1713.9	1819.1	1818.2	1481.0	1995.0	1404.5	0.0	**2515.0**
Hero	11923.1	14109.7	14711.4	14896.6	12816.0	6567.8	346.2	**15219.8**
Jamesbond	419.0	451.0	473.8	381.8	428.4	387.2	0.0	**502.8**
Kangaroo	2781.5	2912.9	3217.1	2824.0	1923.9	1670.5	122.8	**3700.2**
Krull	3634.3	3892.1	3832.1	3656.4	3788.7	3090.8	0.1	**4051.6**
KungFuMaster	15074.8	14452.1	15753.0	11405.6	13389.9	13394.9	0.0	**18065.6**
MsPacman	1432.9	1733.1	1446.4	1711.0	1223.5	1084.2	105.3	**1898.4**
Pong	3.2	10.2	11.5	6.8	-0.1	-2.7	-21.0	**14.2**
PrivateEye	2681.8	2599.1	2720.6	2670.6	**3969.2**	305.3	-1000.0	3124.9
Qbert	5438.4	6469.0	6140.3	5748.6	3921.4	5138.0	125.0	**6966.4**
RoadRunner	18381.5	21470.9	22265.4	12417.1	16210.0	11834.1	1022.9	**24644.2**
Seaquest	454.4	471.3	486.8	330.1	**1016.8**	271.2	172.5	753.1
UpNDown	4221.1	4147.1	**4789.2**	4159.6	3880.2	2631.1	20.0	4577.9
Median HNS	44.1%	47.4%	49.8%	42.0%	47.6%	36.2%	-1.5%	**51.2%**
Mean HNS	73.2%	79.0%	91.7%	69.5%	88.1%	71.7%	-45.9%	**105.6%**

Experiments

- Original Atari Environments



Naturally-existing causal confusion

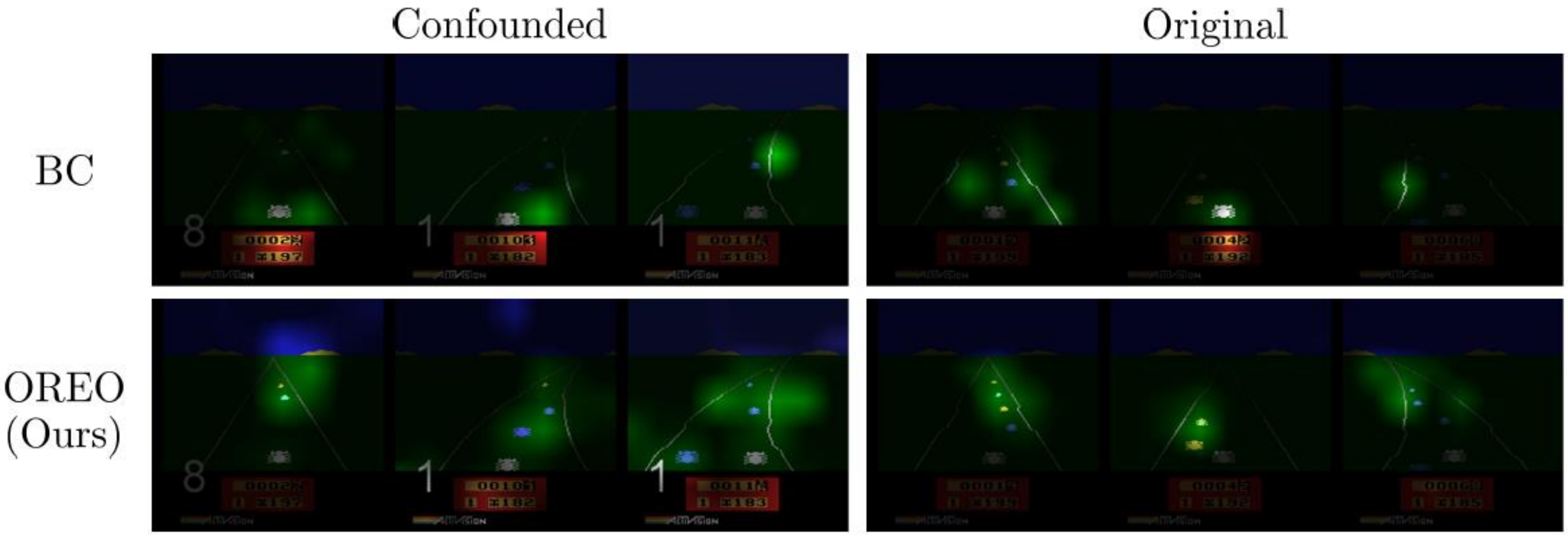
interaction-free version

a direct causal method

Environment	BC	Dropout	DropBlock	Cutout	RandomShift	CCIL†	CRLR	OREO
Alien	986.5	1117.2	1094.8	1104.4	863.5	1050.4	100.0	**1222.2**
Amidar	90.8	81.6	113.5	125.0	78.2	78.6	12.0	**130.5**
Assault	816.8	901.1	829.9	694.1	848.7	755.5	0.0	**905.2**
Asterix	249.0	176.6	252.2	195.0	99.1	314.1	**592.5**	212.5
BankHeist	399.0	476.6	471.2	442.5	354.8	**606.1**	0.0	448.4
BattleZone	10933.8	11621.2	**12067.5**	10641.2	8748.8	11191.2	5615.0	11703.8
Boxing	21.8	25.7	32.1	21.2	35.8	34.2	-43.0	**39.9**
Breakout	**6.4**	2.9	6.0	3.1	4.4	2.1	0.0	5.4
ChopperCommand	1163.0	1162.0	1161.8	1183.9	1026.2	1027.2	1070.2	**1282.9**
CrazyClimber	54142.2	54965.4	55854.0	47456.4	60465.9	39015.2	885.5	**69380.1**
DemonAttack	238.8	**359.3**	225.6	217.8	294.8	194.6	22.7	0.0
Enduro	226.2	304.6	359.1	132.9	282.2	182.8	0.8	**514.4**
Freeway	32.3	32.6	32.6	32.8	33.0	**33.1**	21.4	32.9
Frostbite	153.6	149.2	**165.7**	135.2	133.2	96.7	78.1	152.7
Gopher	1874.4	2220.4	2040.5	1588.2	1456.2	1301.9	0.0	**2903.9**
Hero	15100.4	15994.4	17058.6	15971.8	14867.2	**17487.6**	0.0	16370.3
Jamesbond	447.6	492.3	481.9	418.9	452.1	460.4	0.0	**527.9**
Kangaroo	3162.8	2860.4	**3638.6**	3242.6	2202.1	2938.1	0.0	3602.9
Krull	4447.9	**4764.7**	4526.5	4270.6	4611.6	4247.1	0.0	4633.6
KungFuMaster	12900.6	14994.5	14819.0	9956.9	11698.0	12876.9	0.0	**16955.5**
MsPacman	1921.9	2022.6	2151.7	1949.7	1046.3	1160.6	70.0	**2263.8**
Pong	3.7	10.0	11.6	7.8	0.8	-19.8	-21.0	**12.5**
PrivateEye	3035.4	3396.3	3057.6	3092.2	**3578.9**	1016.4	-1000.0	3162.6
Qbert	5925.4	**6363.1**	5904.3	6174.8	4100.1	5056.3	125.0	5763.4
RoadRunner	18010.1	20137.8	22522.5	12698.9	15615.4	18985.2	1528.6	**27303.9**
Seaquest	527.5	644.4	622.3	376.6	**948.0**	402.4	169.8	921.0
UpNDown	3782.1	3504.3	3886.4	3675.9	3500.4	3062.3	20.0	**4186.8**
Median HNS	46.7%	53.3%	47.7%	42.9%	47.3%	36.8%	-1.5%	**53.6%**
Mean HNS	82.0%	91.5%	99.0%	75.0%	91.7%	85.4%	-45.4%	**114.9%**

Experiments

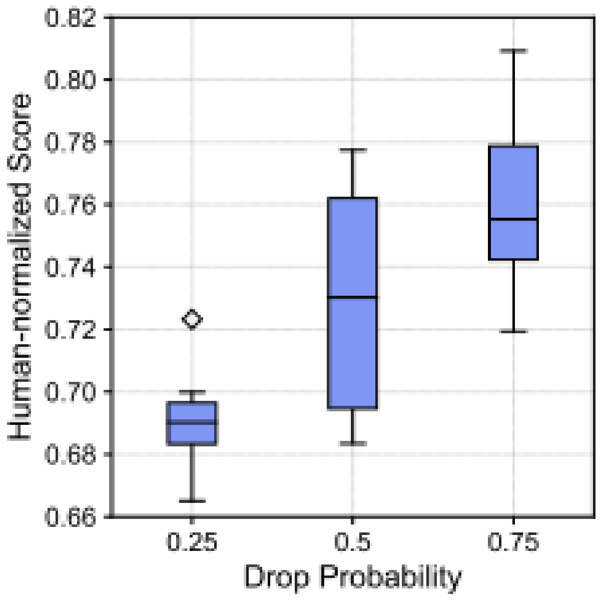
- Visualization
  - OREO attends to more relevant objects, even in the original environment.

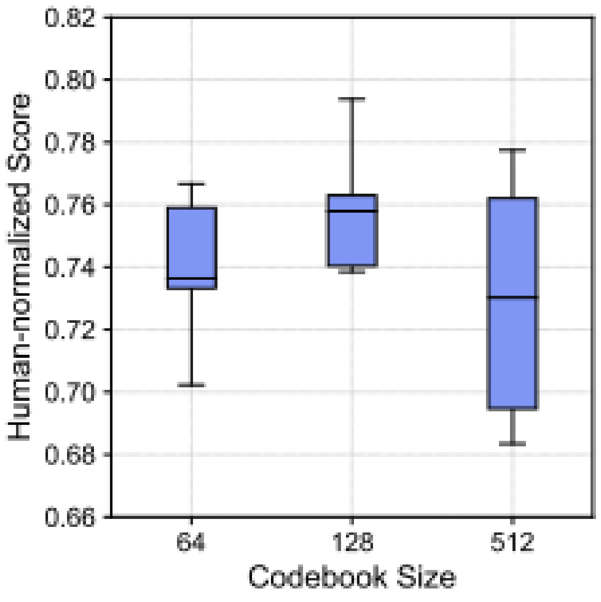


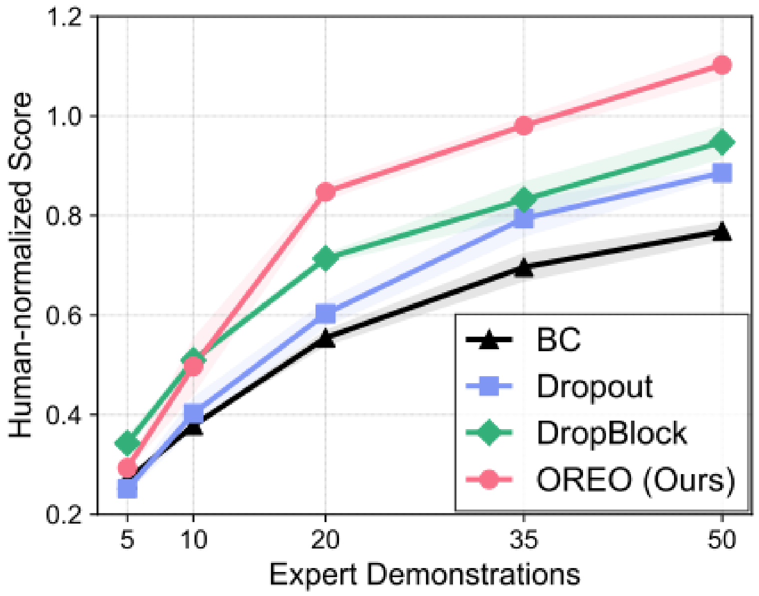
Experiments

- Sensitivity analysis

OREO







Object-aware dropout matters.

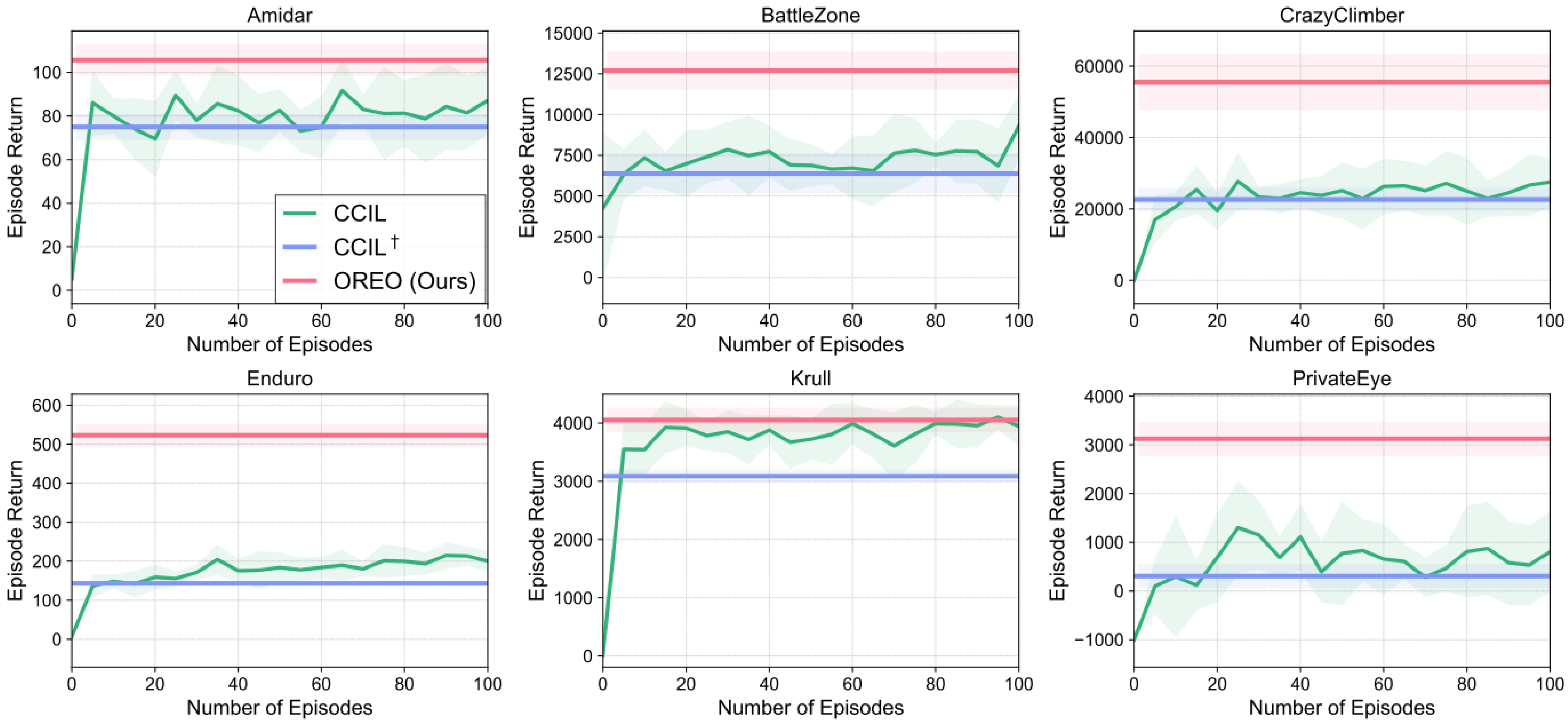
Experiments

- Ablation study
  - The gain comes from the **object-aware dropout design**, but not naively leveraging VQ-VAE.

Environment	BC	VQ-VAE + BC	VQ-VAE + Dropout	VQ-VAE + DropBlock	OREO
BankHeist	442.1± 20.7	358.8± 25.8	491.1± 28.9	488.0± 49.7	**493.9± 17.6**
Enduro	241.4± 28.4	154.6± 10.7	57.1± 12.6	111.2± 16.4	**522.8± 29.1**
KungFuMaster	15074.8± 275.5	11055.1± 867.2	13323.0± 1390.0	14861.1± 1561.5	**18065.6± 1411.5**
Pong	3.2± 0.7	3.6± 1.8	10.4± 0.8	13.6± 0.3	**14.2± 0.4**
PrivateEye	2681.8± 270.2	2255.8± 569.5	390.2± 300.9	746.8± 527.8	**3124.9± 349.6**
RoadRunner	18381.5± 1519.9	5783.2± 403.6	6633.8± 716.8	7771.1± 843.6	**24644.2± 2235.1**
Seaquest	454.4± 53.5	344.9± 35.2	325.6± 28.2	396.6± 36.8	**753.1± 63.6**
UpNDown	4221.1± 214.5	2676.9± 268.9	3310.8± 536.2	4073.9± 760.9	**4577.9± 307.6**
Median HNS	62.7%	47.9%	45.3%	53.2%	**72.9%**
Mean HNS	70.8%	41.3%	45.7%	53.0%	**100.1%**

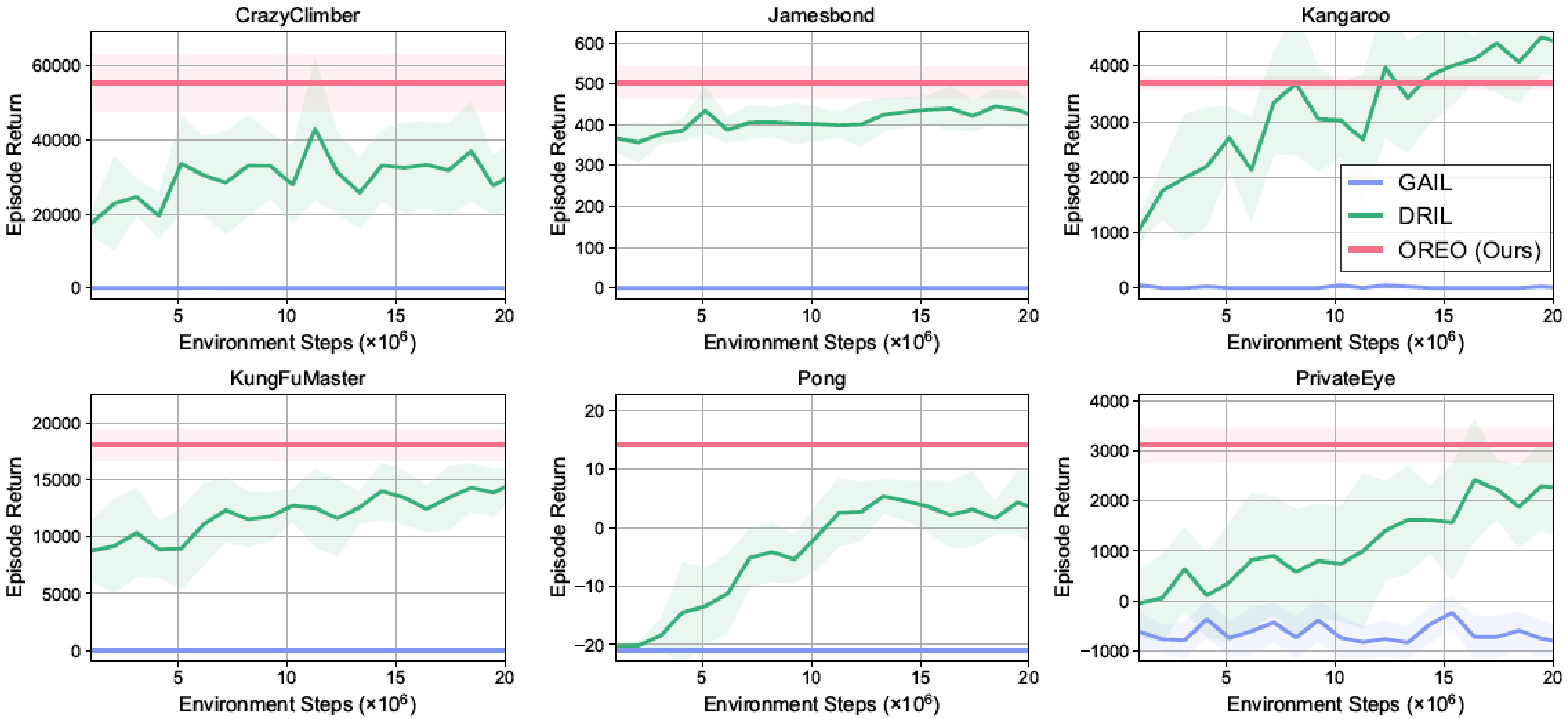
Experiments

- Comparison with CCIL with env. interaction



Experiments

- Comparison with Inverse Reinforcement Learning (with env. interaction)



Experiments

- Real-world application: the CARLA self-driving environment.

Table 3: Performance of policies trained on 150 expert demonstrations from the CARLA driving dataset, under a weather condition of daytime. The results for each environment report the mean and standard deviation of success rates over four runs. OREO achieves the best success rate on all tasks.

Task	BC	Dropout	DropBlock	OREO
Straight	75.0± 1.7	82.0± 8.3	74.0± 3.5	**87.0**± **4.4**
One turn	43.0± 9.1	59.0± 3.3	53.0± 5.2	**70.0**± **7.2**
Navigation	16.9± 7.6	30.4± 10.7	21.7± 9.2	**35.7**± **10.2**
Navigation w/ dynamic obstacles	18.0± 4.5	26.0± 6.0	19.0± 5.2	**30.0**± **4.5**

Thanks!

https://arxiv.org/abs/2110.14118