State Entropy Maximization with Random Encoders for Efficient Exploration

Younggyo Seo*, **Lili Chen***, Jinwoo Shin, Honglak Lee, Pieter Abbeel, Kimin Lee

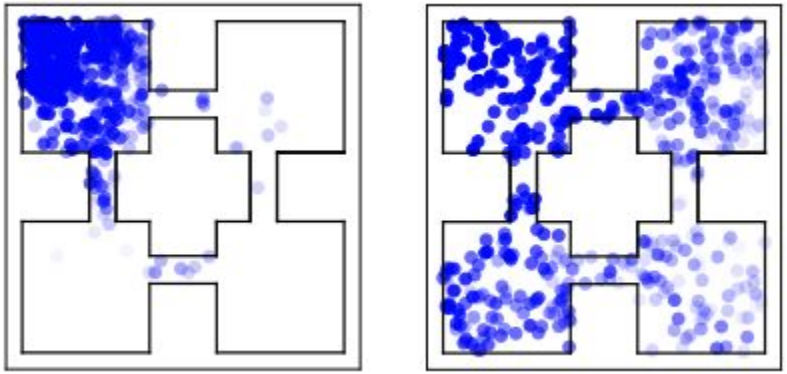
*Equal Contribution

Exploration remains a challenge for deep RL

[Lee'19] Lee, Lisa, Benjamin Eysenbach, Emilio Parisotto, Eric Xing, Sergey Levine, and Ruslan Salakhutdinov. "Efficient exploration via state marginal matching." arXiv preprint, 2019.
[Hazan'19] Hazan, Elad, Sham Kakade, Karan Singh, and Abby Van Soest. "Provably efficient maximum entropy exploration." In ICML, 2019.
[Mutti'21] Mutti, Mirco, Lorenzo Pratissoli, and Marcello Restelli. "Task-Agnostic Exploration via Policy Gradient of a Non-Parametric State Entropy Estimate." In AAAI, 2021.

Exploration remains a challenge for deep RL

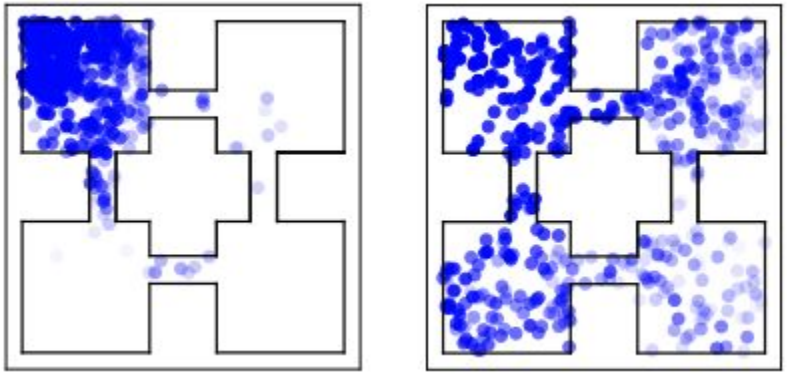
- A promising, principled approach: encourage uniform (i.e., maximum entropy) state space coverage [Lee'19; Hazan'19]



[Lee'19] Lee, Lisa, Benjamin Eysenbach, Emilio Parisotto, Eric Xing, Sergey Levine, and Ruslan Salakhutdinov. "Efficient exploration via state marginal matching." arXiv preprint, 2019.
[Hazan'19] Hazan, Elad, Sham Kakade, Karan Singh, and Abby Van Soest. "Provably efficient maximum entropy exploration." In ICML, 2019.
[Mutti'21] Mutti, Mirco, Lorenzo Pratissoli, and Marcello Restelli. "Task-Agnostic Exploration via Policy Gradient of a Non-Parametric State Entropy Estimate." In AAAI, 2021.

Exploration remains a challenge for deep RL

- A promising, principled approach: encourage uniform (i.e., maximum entropy) state space coverage [Lee'19; Hazan'19]
- In practice: estimate state entropy by measuring distances between states and their *k*-nearest neighbors [Mutti'21]



[Lee'19] Lee, Lisa, Benjamin Eysenbach, Emilio Parisotto, Eric Xing, Sergey Levine, and Ruslan Salakhutdinov. "Efficient exploration via state marginal matching." arXiv preprint, 2019.
[Hazan'19] Hazan, Elad, Sham Kakade, Karan Singh, and Abby Van Soest. "Provably efficient maximum entropy exploration." In ICML, 2019.
[Mutti'21] Mutti, Mirco, Lorenzo Pratissoli, and Marcello Restelli. "Task-Agnostic Exploration via Policy Gradient of a Non-Parametric State Entropy Estimate." In AAAI, 2021.

How to extend this to high-dimensional observations?

Measuring distance between images is non-trivial (cannot directly use pixel space)

[Badia'20] Badia, Adrià Puigdomènech, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Tieleman et al. "Never Give Up: Learning Directed Exploration Strategies." In ICLR, 2020.
[Tao'20] Tao, Ruo Yu, Vincent François-Lavet, and Joelle Pineau. "Novelty Search in representational space for sample efficient exploration." In NeurIPS, 2020.
[Liu'21] Liu, Hao, and Pieter Abbeel. "Behavior from the void: Unsupervised active pre-training." In ICML, 2021.

How to extend this to high-dimensional observations?

Measuring distance between images is non-trivial (cannot directly use pixel space)

- Prior approaches: *k*-NN measured in low-dimensional learned latent space
  - Dynamics learning [Tao'20]
  - Inverse dynamics prediction [Badia'20]
  - Contrastive learning [Liu'21]

[Badia'20] Badia, Adrià Puigdomènech, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Tieleman et al. "Never Give Up: Learning Directed Exploration Strategies." In ICLR, 2020.
[Tao'20] Tao, Ruo Yu, Vincent François-Lavet, and Joelle Pineau. "Novelty Search in representational space for sample efficient exploration." In NeurIPS, 2020.
[Liu'21] Liu, Hao, and Pieter Abbeel. "Behavior from the void: Unsupervised active pre-training." In ICML, 2021.

How to extend this to high-dimensional observations?

Measuring distance between images is non-trivial (cannot directly use pixel space)

- Prior approaches: *k*-NN measured in low-dimensional learned latent space
  - Dynamics learning [Tao'20]
  - Inverse dynamics prediction [Badia'20]
  - Contrastive learning [Liu'21]

- But optimizing these auxiliary losses adds complexity (e.g., hyperparameter tuning), instability, and computational overhead

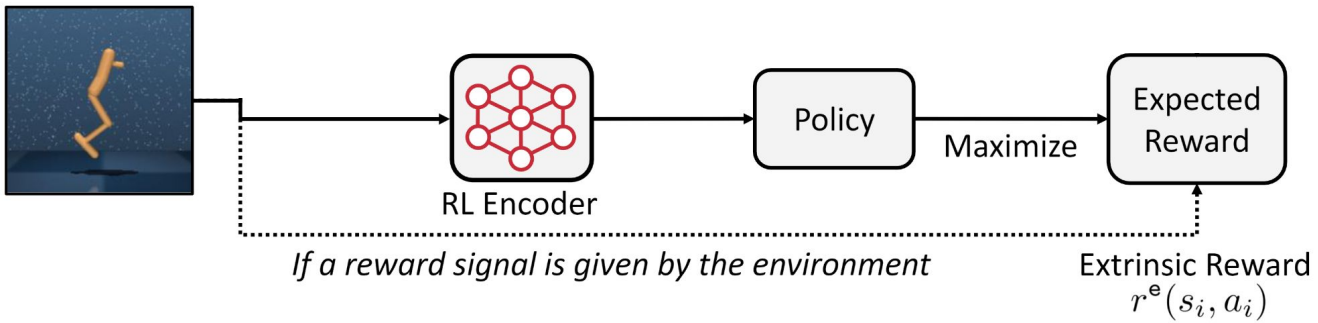
[Badia'20] Badia, Adrià Puigdomènech, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Tieleman et al. "Never Give Up: Learning Directed Exploration Strategies." In ICLR, 2020.
[Tao'20] Tao, Ruo Yu, Vincent François-Lavet, and Joelle Pineau. "Novelty Search in representational space for sample efficient exploration." In NeurIPS, 2020.
[Liu'21] Liu, Hao, and Pieter Abbeel. "Behavior from the void: Unsupervised active pre-training." In ICML, 2021.

Random Encoders for Efficient Exploration (RE3)

Core idea: intrinsic reward via k-NN state entropy estimator in the representation space of a **randomly initialized** encoder, **fixed throughout training**

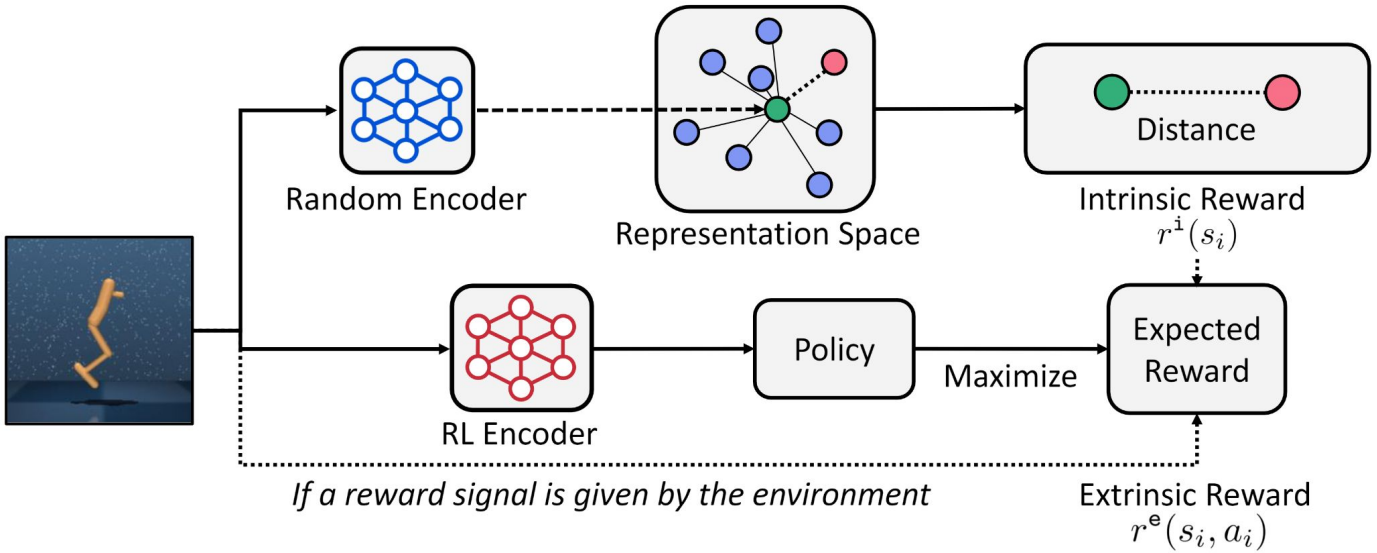
Random Encoders for Efficient Exploration (RE3)

Core idea: intrinsic reward via k-NN state entropy estimator in the representation space of a **randomly initialized** encoder, **fixed throughout training**



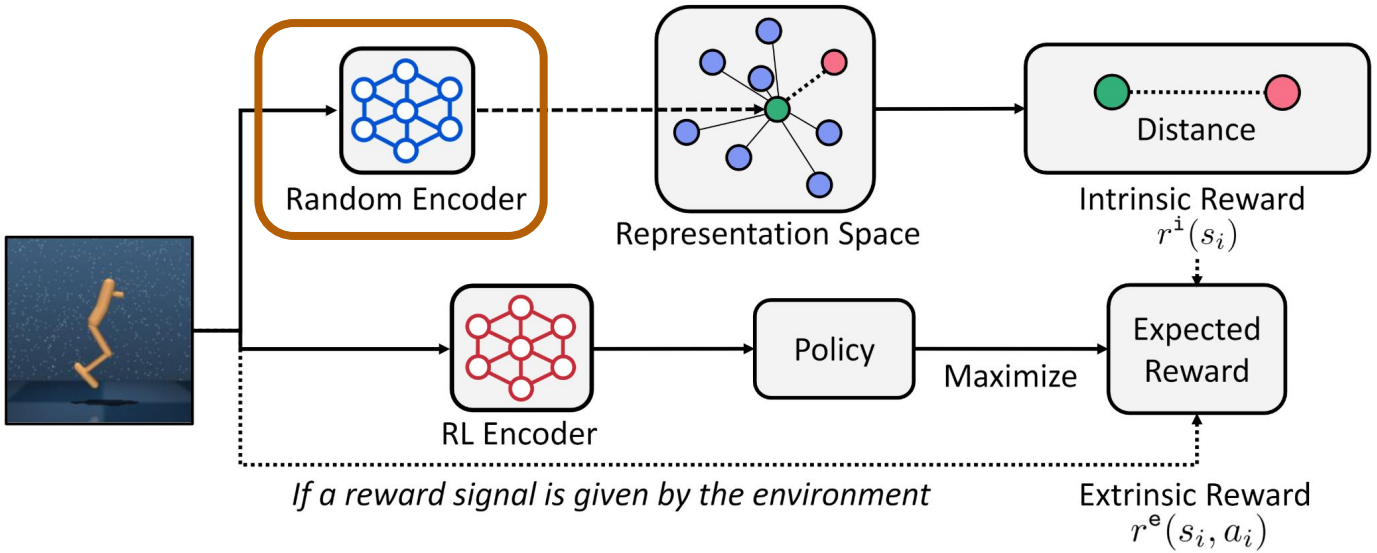
Random Encoders for Efficient Exploration (RE3)

Core idea: intrinsic reward via k-NN state entropy estimator in the representation space of a **randomly initialized** encoder, **fixed throughout training**



Random Encoders for Efficient Exploration (RE3)

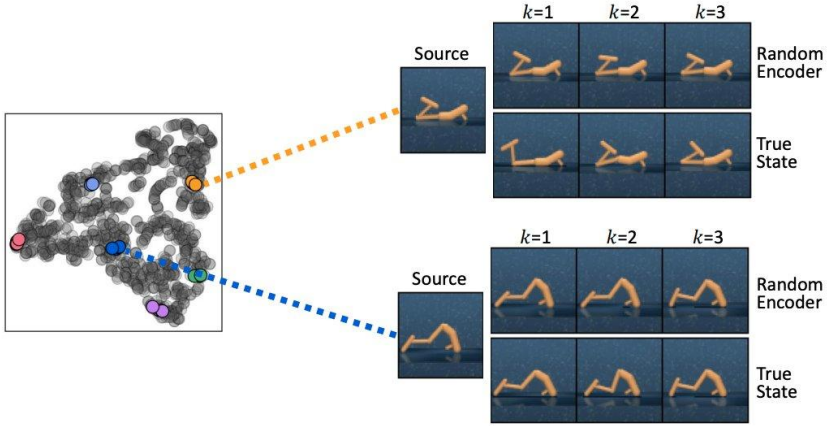
Core idea: intrinsic reward via k-NN state entropy estimator in the representation space of a **randomly initialized** encoder, **fixed throughout training**



Random Encoders for Efficient Exploration (RE3)

Core idea: intrinsic reward via k-NN state entropy estimator in the representation space of a **randomly initialized** encoder, **fixed throughout training**

Hypothesis: the representation space of a random encoder effectively captures information about similarity between states



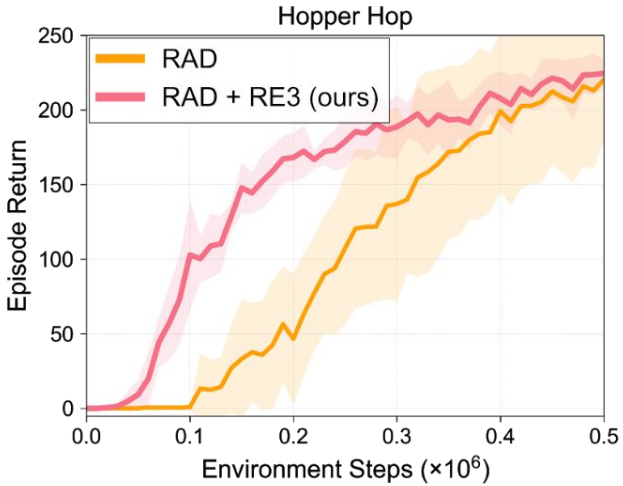
RE3 can be combined with a variety of RL algorithms

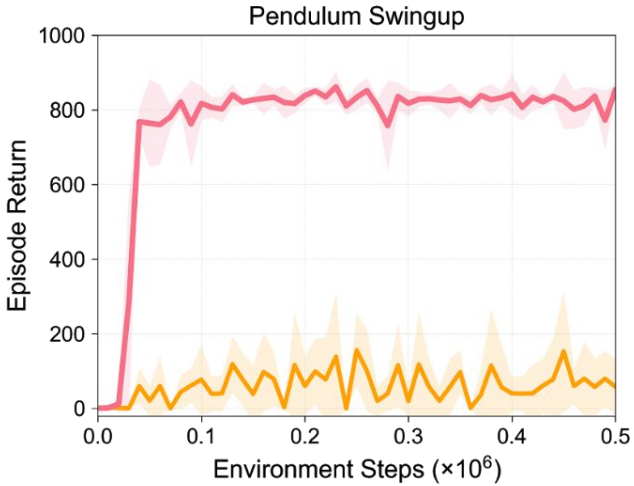
We improve the sample-efficiency of both model-free (RAD [Laskin'20]) and model-based (Dreamer [Hafner'20]) algorithms on DeepMind Control Suite [Tassa'18]

[Laskin'20] Laskin, Michael, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, Aravind Srinivas. "Reinforcement Learning with Augmented Data." in NeurIPS, 2020.
[Hafner'20] Hafner, Danijar, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. "Dream to control: Learning behaviors by latent imagination." in ICLR. 2020.
[Tassa'18] Tassa, Yuval, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden et al. "Deepmind control suite." arXiv preprint. 2018.

RE3 can be combined with a variety of RL algorithms

We improve the sample-efficiency of both model-free (RAD [Laskin'20]) and model-based (Dreamer [Hafner'20]) algorithms on DeepMind Control Suite [Tassa'18]

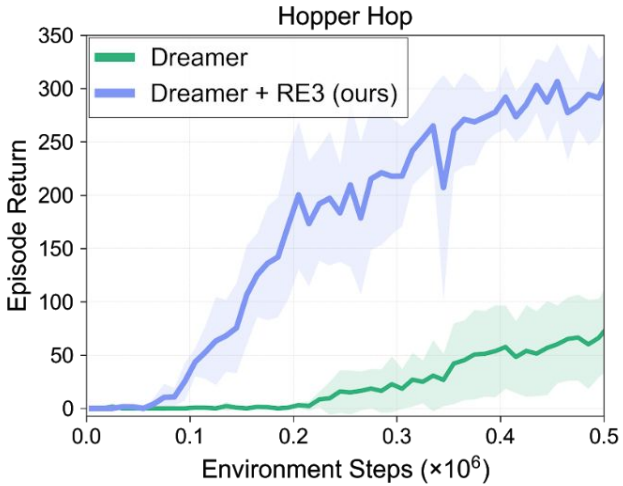


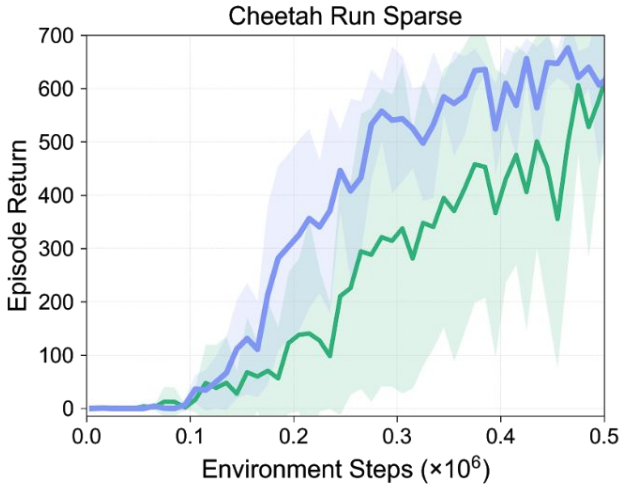


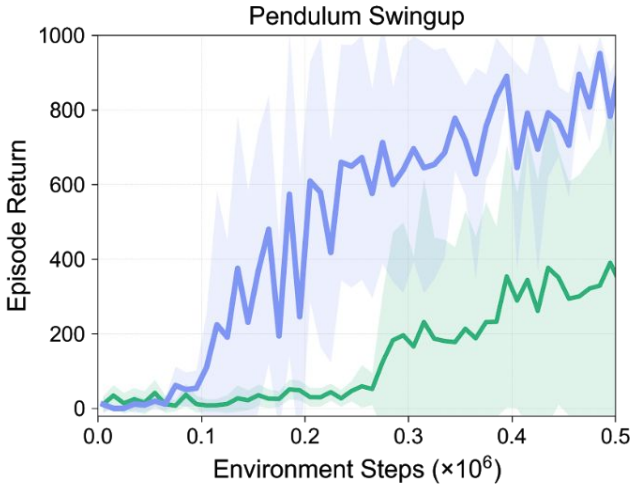
[Laskin'20] Laskin, Michael, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, Aravind Srinivas. "Reinforcement Learning with Augmented Data." in NeurIPS, 2020.
[Hafner'20] Hafner, Danijar, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. "Dream to control: Learning behaviors by latent imagination." in ICLR. 2020.
[Tassa'18] Tassa, Yuval, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden et al. "Deepmind control suite." arXiv preprint. 2018.

RE3 can be combined with a variety of RL algorithms

We improve the sample-efficiency of both model-free (RAD [Laskin'20]) and model-based (Dreamer [Hafner'20]) algorithms on DeepMind Control Suite [Tassa'18]

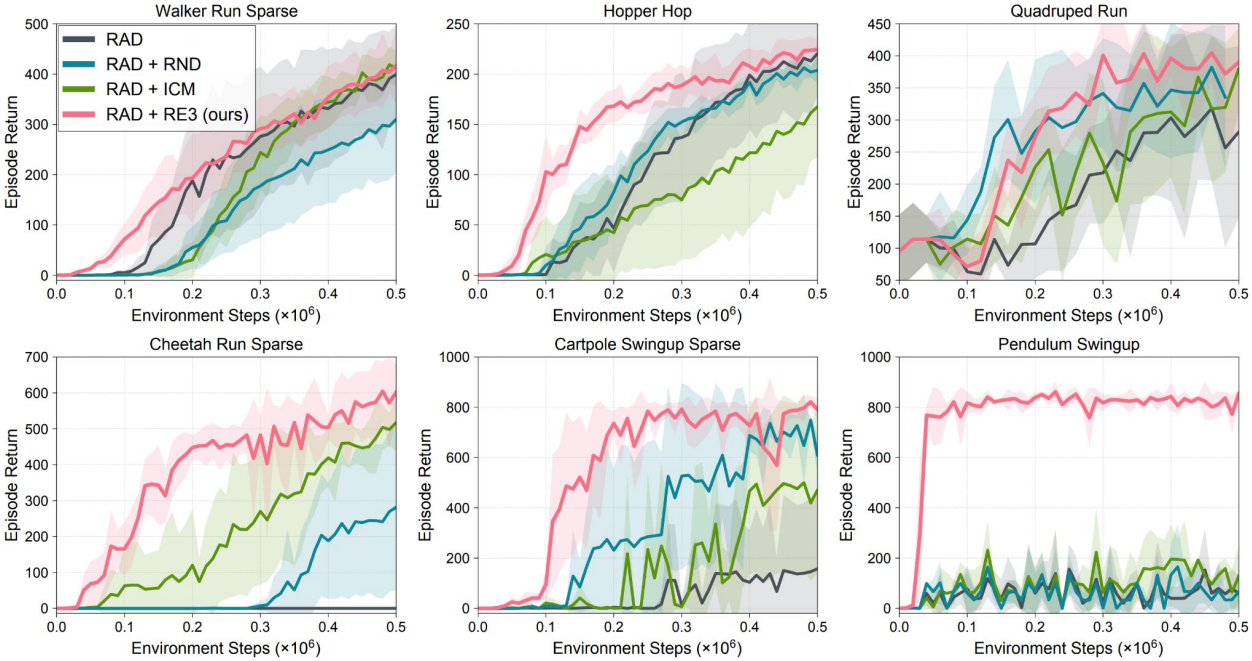






[Laskin'20] Laskin, Michael, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, Aravind Srinivas. "Reinforcement Learning with Augmented Data." in NeurIPS, 2020.
[Hafner'20] Hafner, Danijar, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. "Dream to control: Learning behaviors by latent imagination." in ICLR. 2020.
[Tassa'18] Tassa, Yuval, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden et al. "Deepmind control suite." arXiv preprint. 2018.

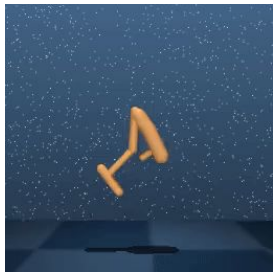
RE3 outperforms other exploration methods



[Pathak'17] Pathak, Deepak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. "Curiosity-driven Exploration by Self-supervised Prediction". In ICML, 2017.
[Burda'19] Burda, Yuri, Harrison Edwards, Amos Storkey, and Oleg Klimov. "Exploration by random network distillation." In ICLR. 2019.

RE3 is also effective for reward-free pre-training

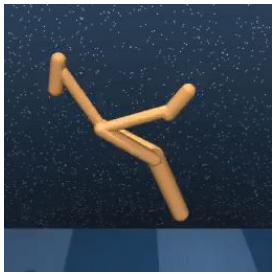
Pre-training with the RE3 objective encourages diverse behaviors



Random exploration

RE3

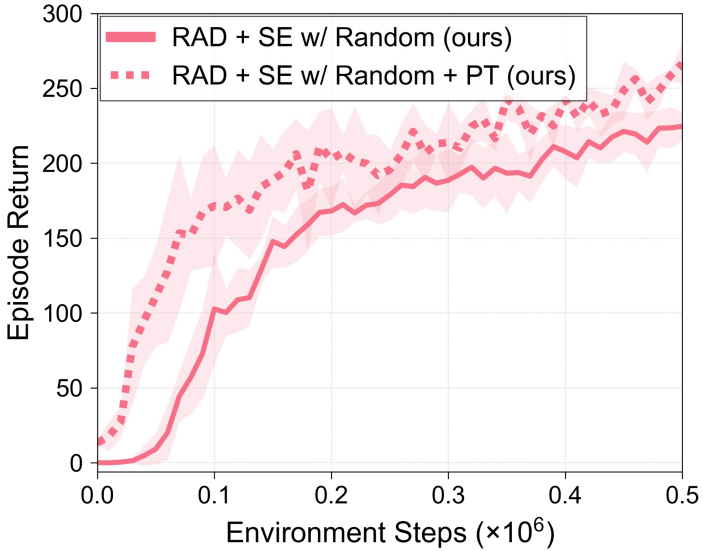
Random exploration



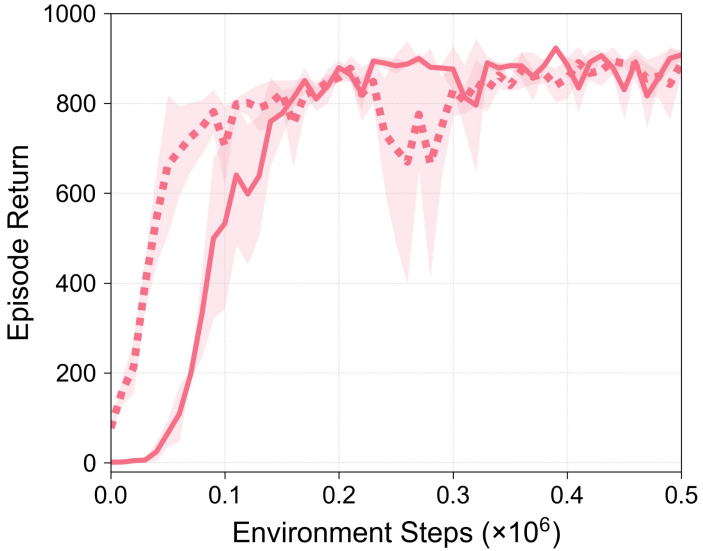
RE3

RE3 is also effective for reward-free pre-training

Pre-training with the RE3 objective encourages diverse behaviors which are useful for downstream tasks



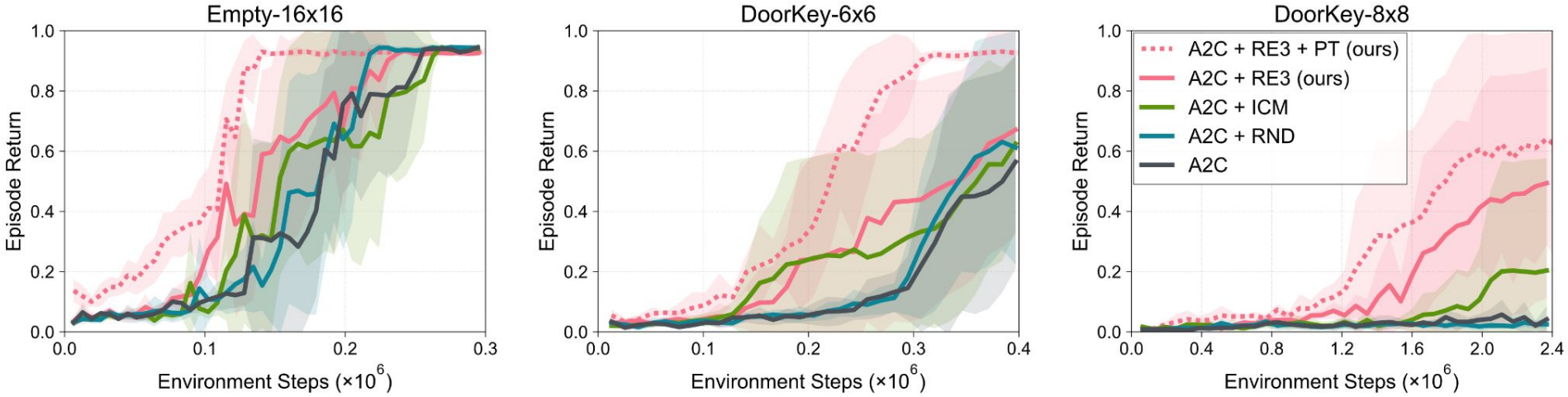
Hopper Hop



Hopper Stand

RE3 can also be combined with on-policy RL algorithms

RE3 improves sample-efficiency of A2C [Mnih'16] and outperforms other exploration methods in Minigrid environments [Chevalier-Boisvert'18]

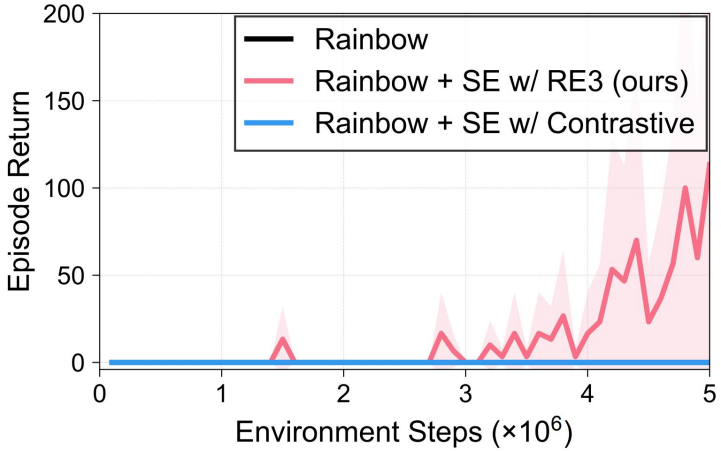


[Mnih'16] Mnih, Volodymyr, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. "Asynchronous methods for deep reinforcement learning." In ICML, 2016.

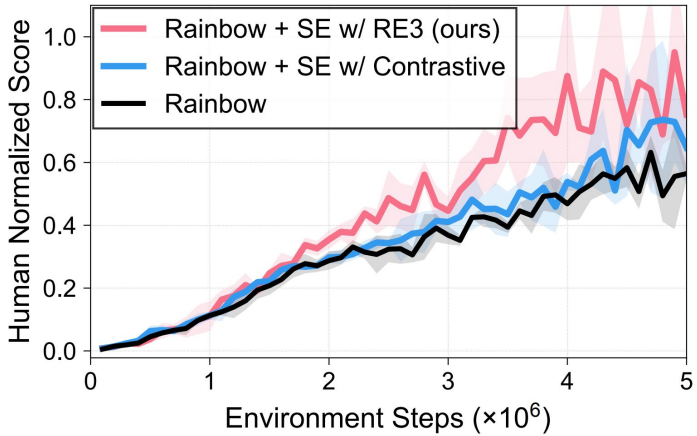
[Chevalier-Boisvert'18] Chevalier-Boisvert, M., Willems, L., and Pal, S. Minimalistic gridworld environment for openai gym. https://github.com/maximecb/gym-minigrid, 2018

RE3 improves learning in hard exploration Atari games

Montezuma's Revenge



Normalized score over 6 games



[Hessel'18] Hessel, Matteo, Joseph Modayil, Hado Van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. "Rainbow: Combining improvements in deep reinforcement learning." In AAAI, 2018.

Thank you!