Random Expert Distillation For Imitation Learning

Ruohan Wang, Carlo Ciliberto, Pierluigi Amadori, Yiannis Demiris

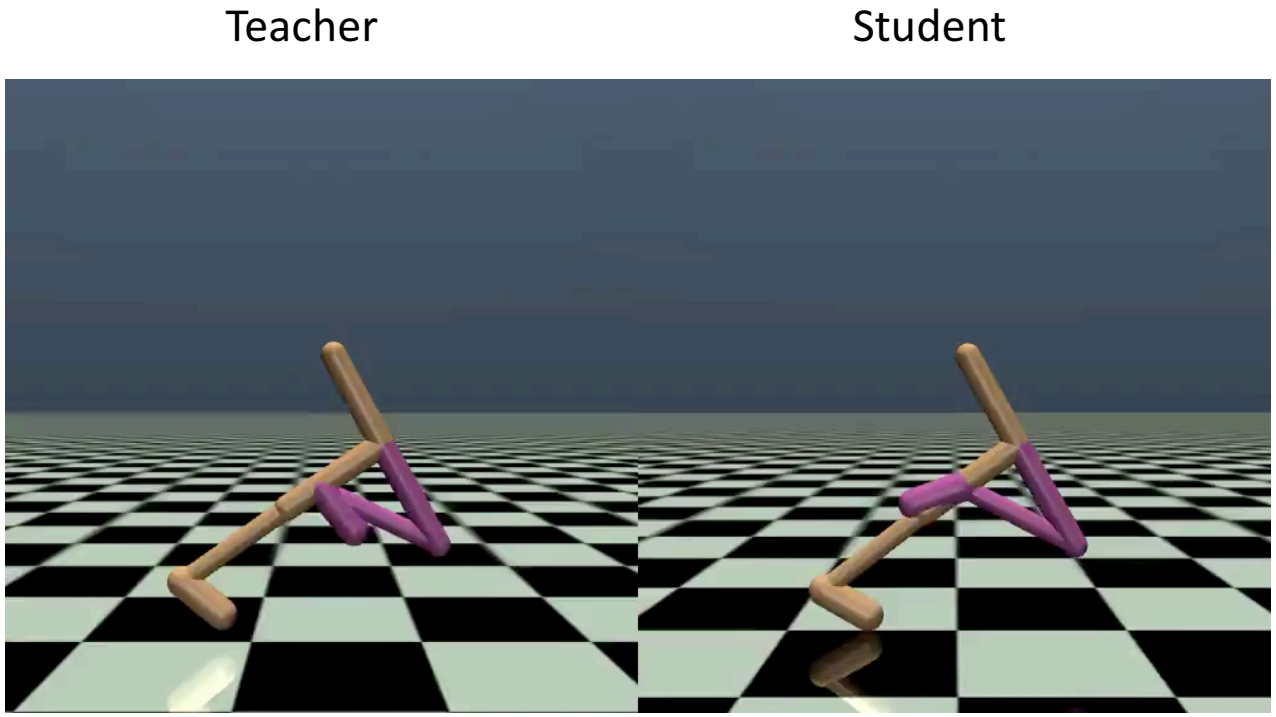
ICML 2019

Imperial College London

PERSONAL ROBOTICS LABORATORY

Imitation Learning

- Policy learning from a limited set of expert demonstrations
- Intuitive & efficient skills transfer
- Captures styles & preferences



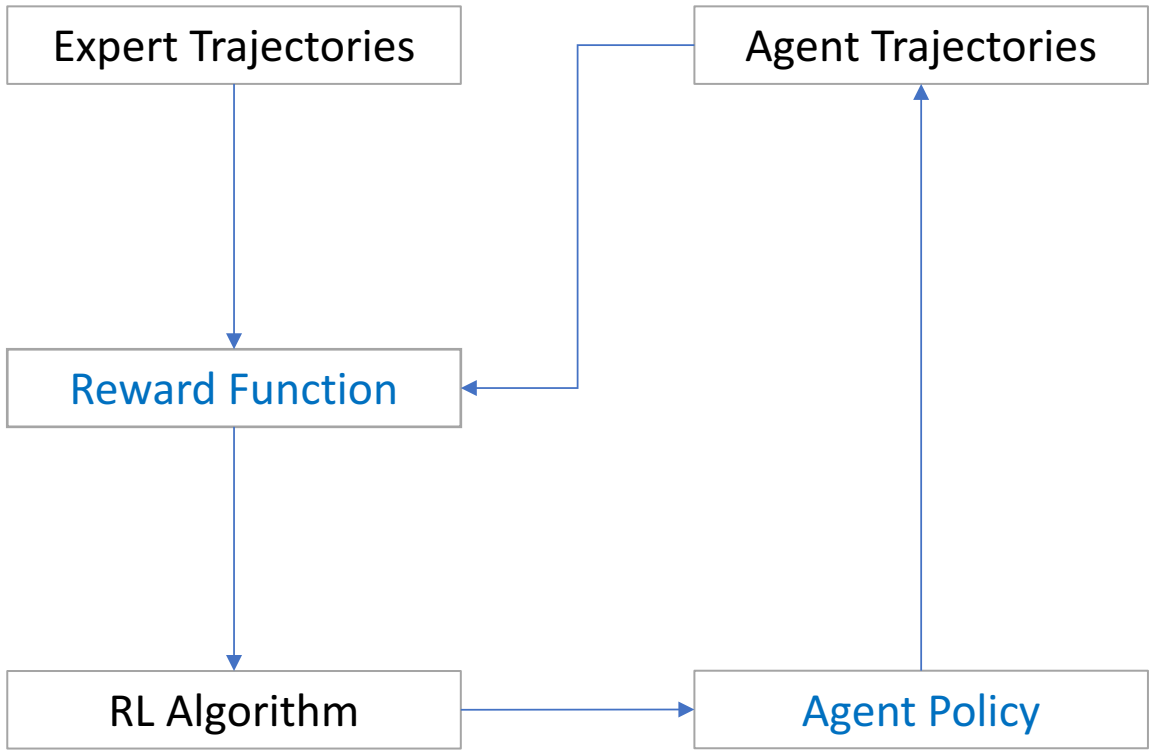
Imperial College London

PERSONAL ROBOTICS LABORATORY

Inverse Reinforcement Learning

- Generative Adversarial Imitation Learning (Ho et al., 2015)

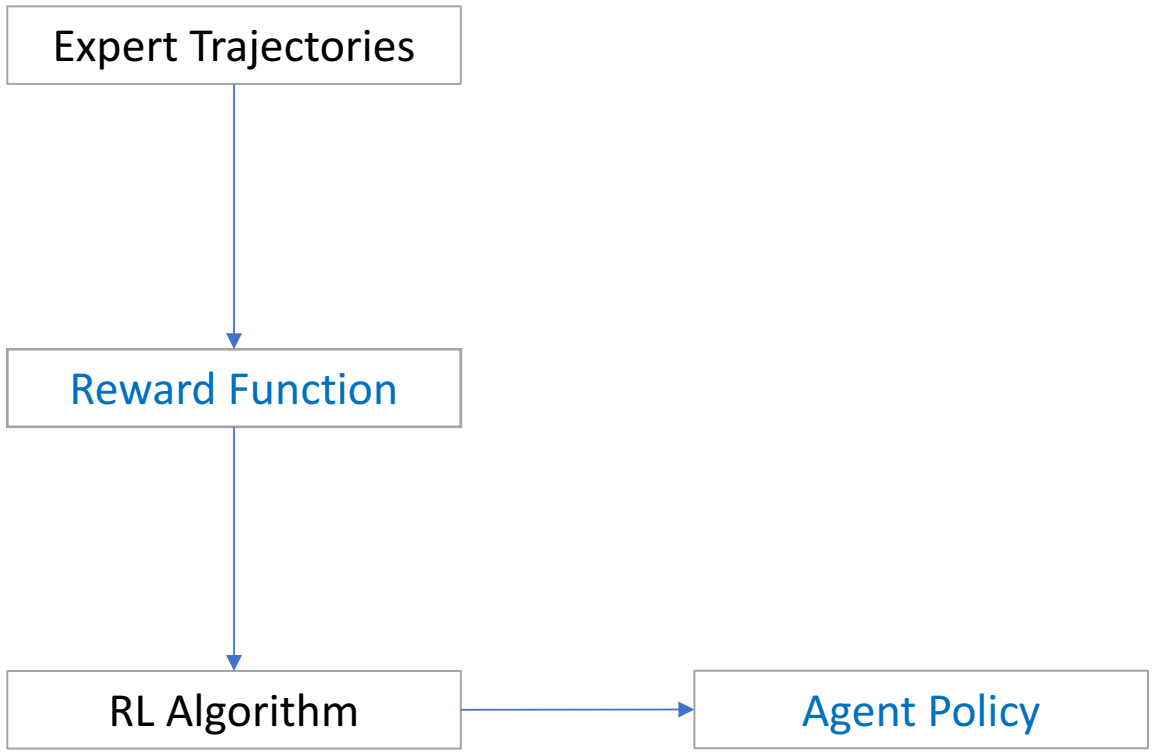
- Optimization challenges
  - Training instability
  - sample inefficiency



Random Expert Distillation (RED)

- Directly learns a reward function with Random Network Distillation (RND) (Burda et al., 2018)

- Considers how "similar" is the agent to the expert, instead of how "different"



Imperial College London

PERSONAL ROBOTICS LABORATORY

Reward Function

Over expert trajectories $D = \{s_i, a_i\}_{i=1}^{N}$ and $f_\theta: \mathbb{R}^m \rightarrow \mathbb{R}^n$

$$\theta^* = \min_\theta ||f_\theta(s,a) - f_{rnd}(s,a)||_2^2 .$$

Define the reward as

$$r(s,a) = \exp(-\sigma||f_{\theta^*}(s,a) - f_{rnd}(s,a)||_2^2)$$

The reward asymptotically estimates the support of the expert policy

Mujoco Experiments

Ant-v2
Make a 3D four-legged robot walk.

HalfCheetah-v2
Make a 2D cheetah robot run.

Hopper-v2
Make a 2D robot hop.

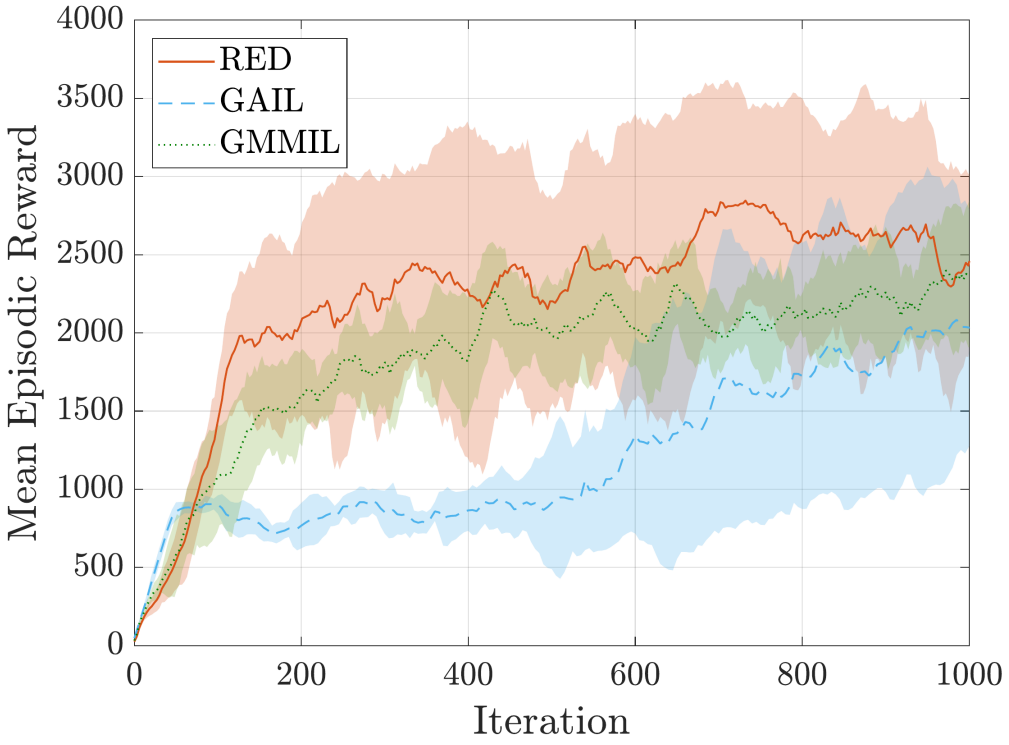
	Hopper	HalfCheetah	Walker2d	Reacher	Ant
GAIL	3614.2 ± 7.2	4515.7 ± 549.5	4878.0 ± 2848.3	-32.4 ± 39.8	3186.8 ± 903.6
GMMIL	3309.3 ± 26.3	3464.2 ± 476.5	2967.1 ± 702.0	-11.89 ± 5.27	991 ± 2.6
RED	3626.0 ± 4.3	3072.0 ± 84.7	4481.4 ± 20.9	-10.43 ± 5.2	3552.8 ± 348.7

Image ref: https://creativestudio2019spring.files.wordpress.com/2019/02/openaigym.png

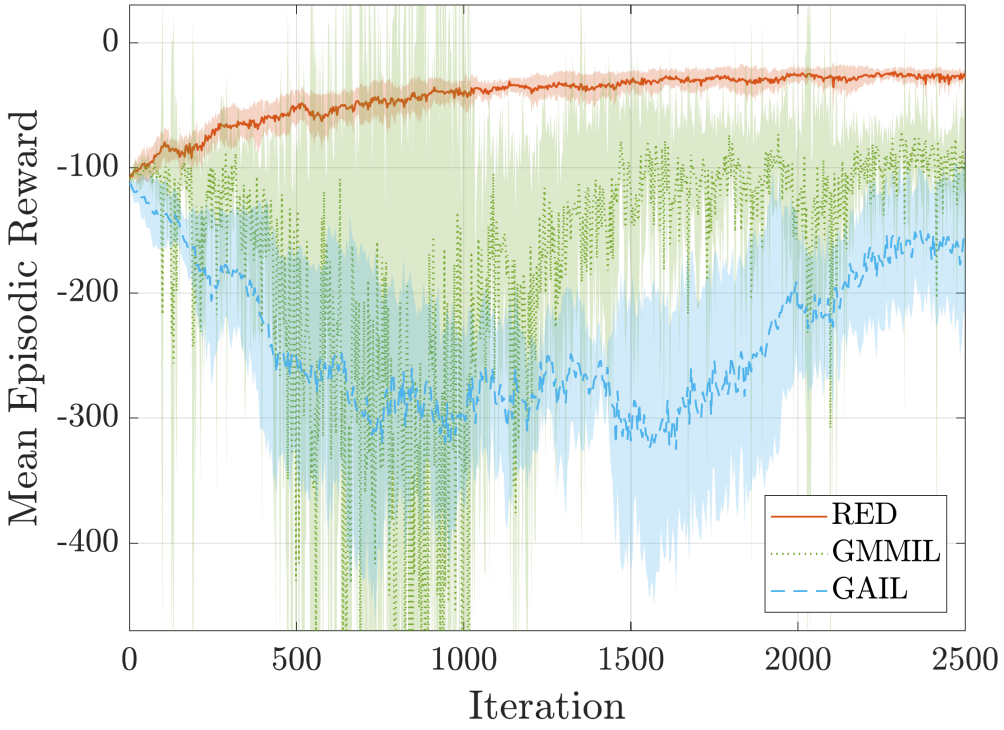
Imperial College London

PERSONAL ROBOTICS LABORATORY

Training Stability & Sample Efficiency



Hopper



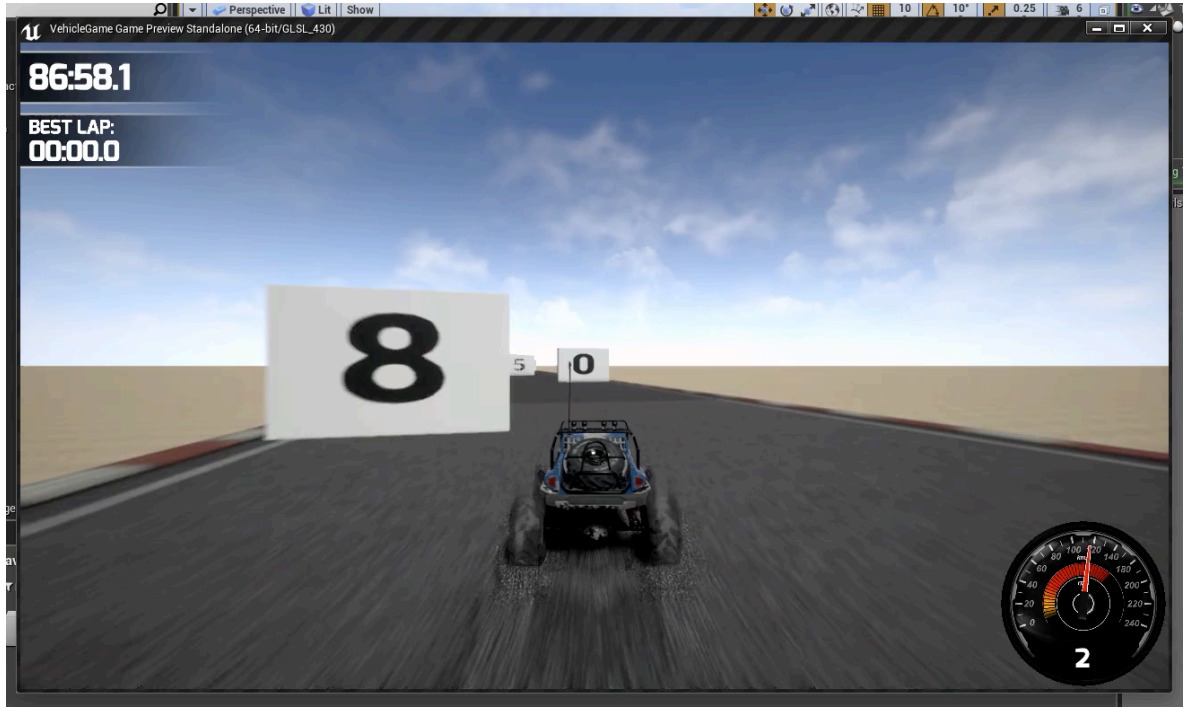
Reacher

Imperial College London

PERSONAL ROBOTICS LABORATORY

Driving Task

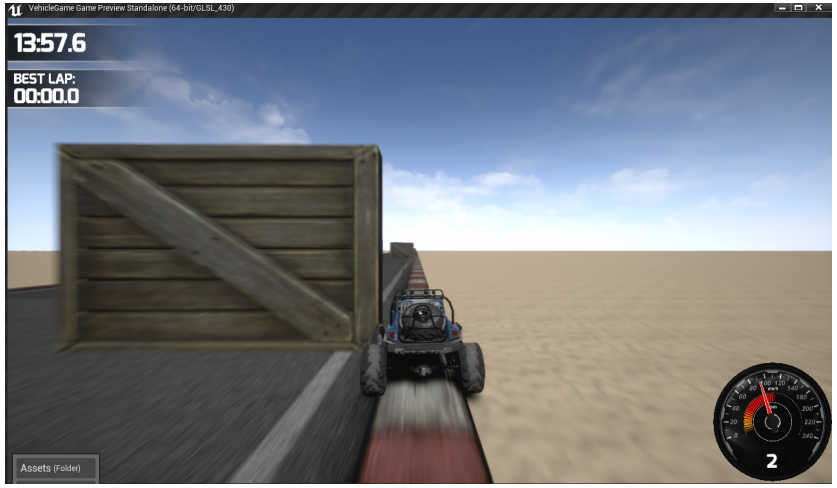
	Average	Best
BC	1033 ± 474	1956
GAIL	795 ± 395	1576
GMMIL	2024 ± 981	3624
RED	4825 ± 1552	7485
Expert	7485 ± 0	7485

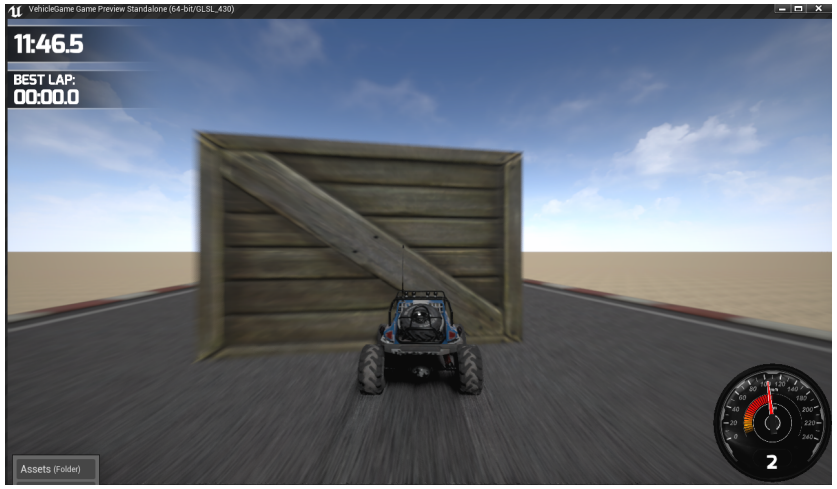


Imperial College London

PERSONAL ROBOTICS LABORATORY

Reward function penalizes dangerous driving





Imperial College London

PERSONAL ROBOTICS LABORATORY

Summary

- Random Expert Distillation is a new framework for imitation learning, using the estimated support of the expert policy as reward.

- Our results suggest that RED is viable, robust and attains good performance.

- Future works: combining different sources of expert information for more robust algorithms.

Imperial College London

PERSONAL ROBOTICS LABORATORY

Thank you

- Code: https://github.com/RuohanW/RED

- Check out our poster:

Pacific Ballroom #39

6:30 to 9:00 pm today

Imperial College London

PERSONAL ROBOTICS LABORATORY