universität freiburg

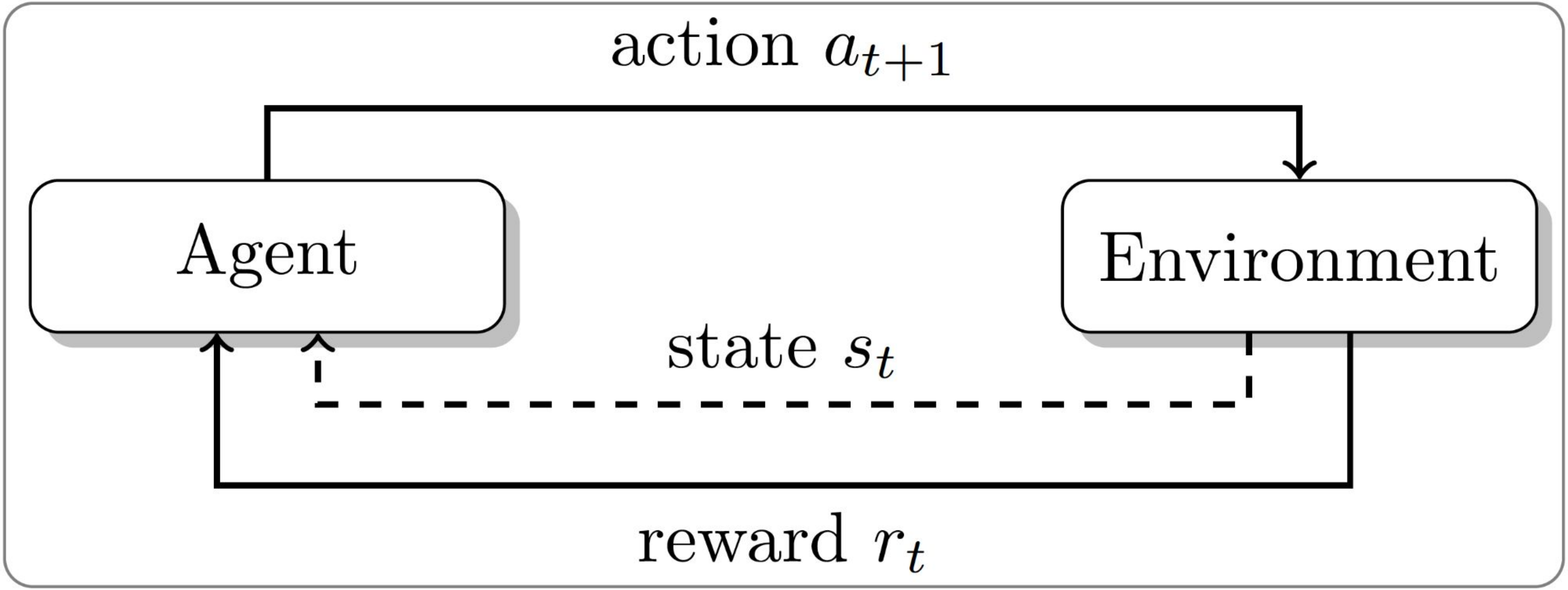
HPO-RL-Bench

A Zero-Cost Benchmark for HPO in Reinforcement Learning

Gresa Shala, Sebastian Pineda Arango,
André Biedenkapp, Frank Hutter and Josif Grabocka

Reminder about Reinforcement Learning

Agents learn by ***interacting*** with their world



RL is a Simple Yet Extremely Powerful Paradigm

Many, well publicised
success stories

RL is Extremely Sensitive to Hyperparameters

Various studies have highlighted this fact and lack of reproducibility in RL:

- Islam et al., RML@ICML’17
- Henderson et al., AAAI’18
- Engstrom et al., ICLR’20
- Andrychowicz et al., ICLR’21
- Agarwal et al., NeurIPS’21
- Eimer et al., ICML’23

AutoML to the Rescue!

If Only It Were That Easy

Why can we not simply apply AutoML tools?

RL training can be prohibitively expensive!
Take the common Atari training protocol:

- 50x10^6 training steps
- per game

If Only It Were That Easy

Direct Quote from Mnih et al., Nature 2015:

“The values of all the hyperparameters and optimization parameters were selected by performing an informal search on the games Pong, Breakout, Seaquest, Space Invaders and Beam Rider. **We did not perform a systematic grid search owing to the high computational cost**. These parameters were then held fixed across all other games.”

Have We Not Made Any Progress?

- We did!
- But benchmarking of AutoML/AutoRL solutions still remains an open challenge
- In particular, there was no comparison of different solution approaches until now

Journal of Artificial Intelligence Research 74 (2022) 517-568 Submitted 01/2022; published 06/2022

Automated Reinforcement Learning (AutoRL): A Survey and Open Problems

Jack Parker-Holder JACKPH@ROBOTS.OX.AC.UK
University of Oxford

Raghu Rajan RAJANR@CS.UNI-FREIBURG.DE
University of Freiburg

Xingyou Song XINGYOUSONG@GOOGLE.COM
Google Research, Brain Team

André Biedenkapp BIEDENKA@CS.UNI-FREIBURG.DE
University of Freiburg

Yingjie Miao YINGJIEMIAO@GOOGLE.COM
Google Research, Brain Team

Theresa Eimer EIMER@TNT.UNI-HANNOVER.DE
Leibniz University Hannover

Baohe Zhang ZHANGB@CS.UNI-FREIBURG.DE
University of Freiburg

Vu Nguyen VUTNGN@AMAZON.COM
Amazon Australia

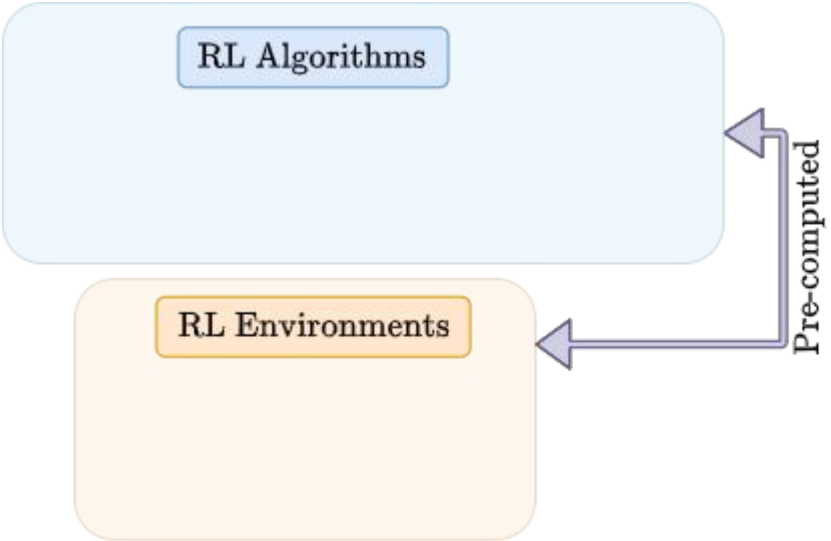
Roberto Calandra RCALANDRA@FB.COM
Meta AI

Aleksandra Faust SANDRAFAUST@GOOGLE.COM
Google Research, Brain Team

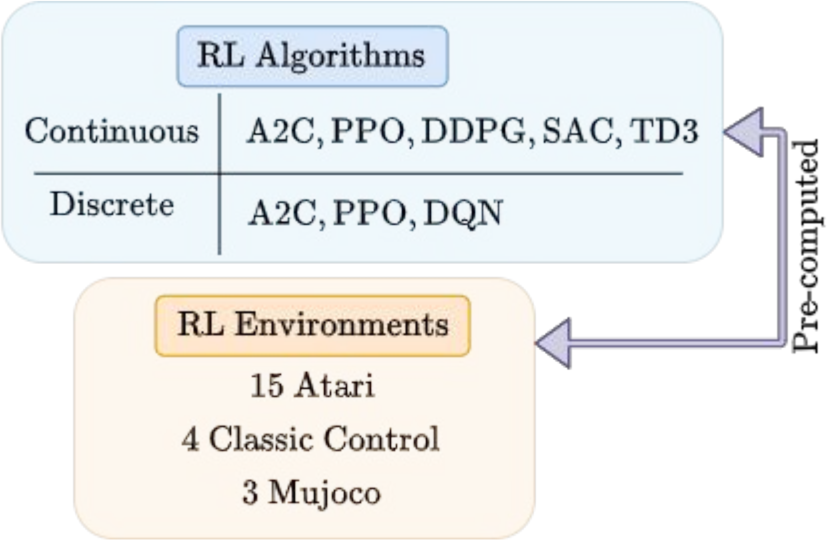
Frank Hutter FH@CS.UNI-FREIBURG.DE
University of Freiburg & Bosch Center for Artificial Intelligence

Marius Lindauer LINDAUER@TNT.UNI-HANNOVER.DE
Leibniz University Hannover

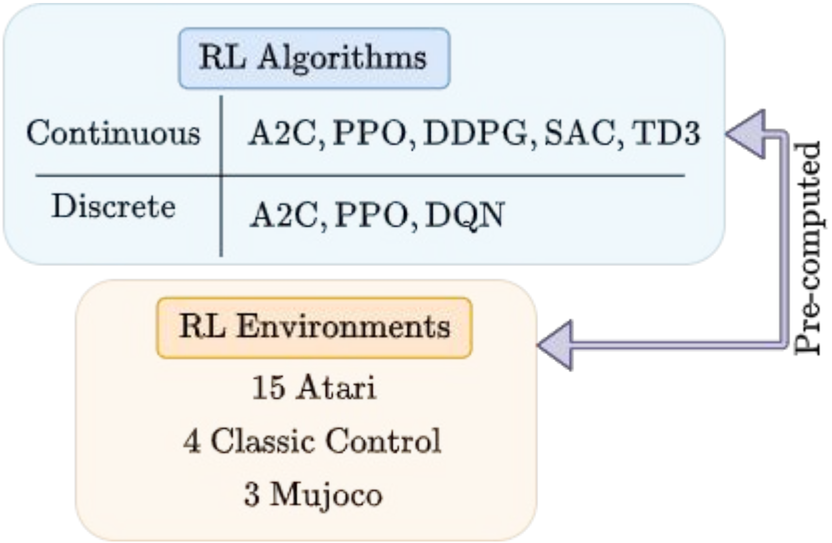
HPO-RL-Bench In a Nutshell



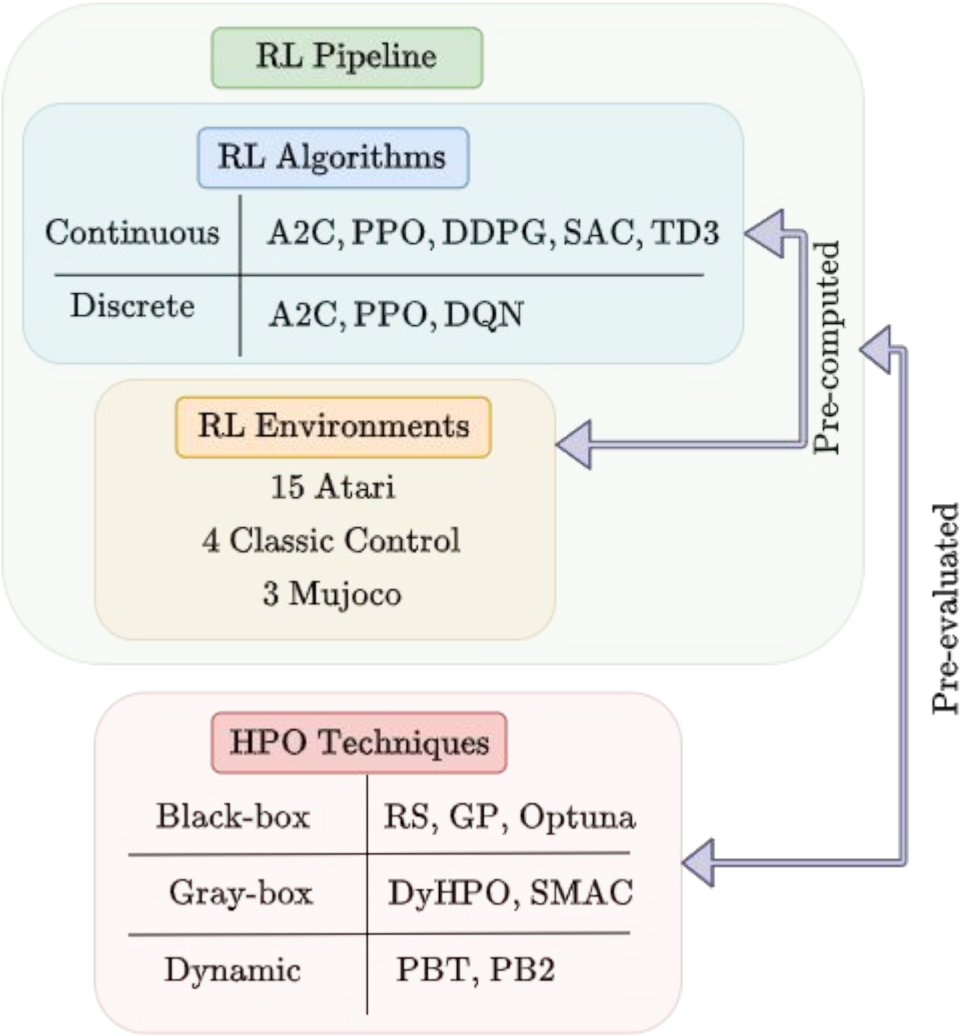
HPO-RL-Bench In a Nutshell



HPO-RL-Bench In a Nutshell



HPO-RL-Bench In a Nutshell



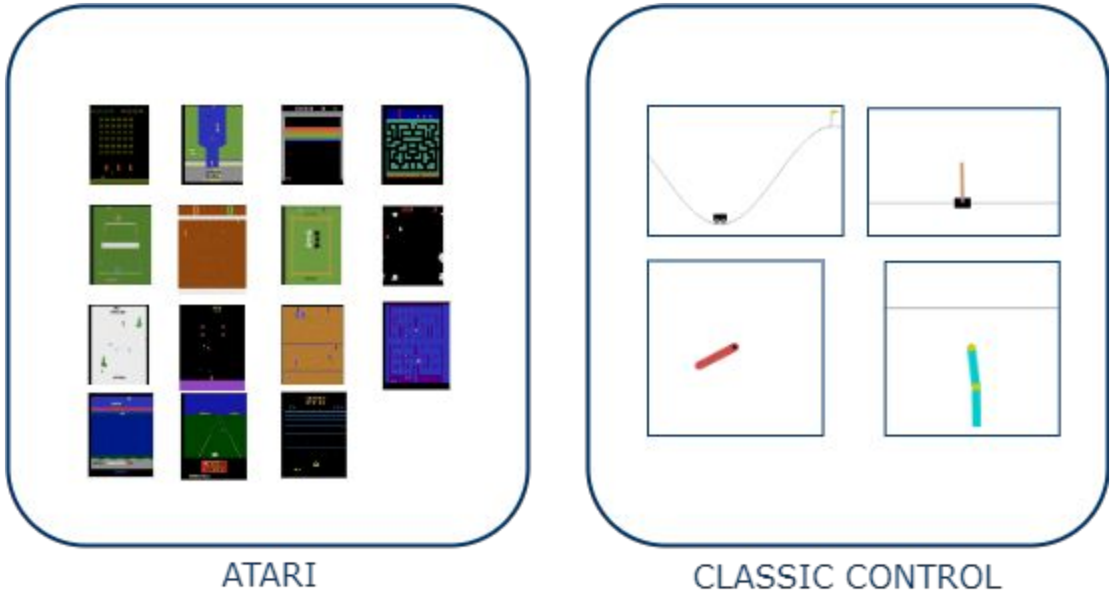
HPO-RL-Bench

Environments

ATARI

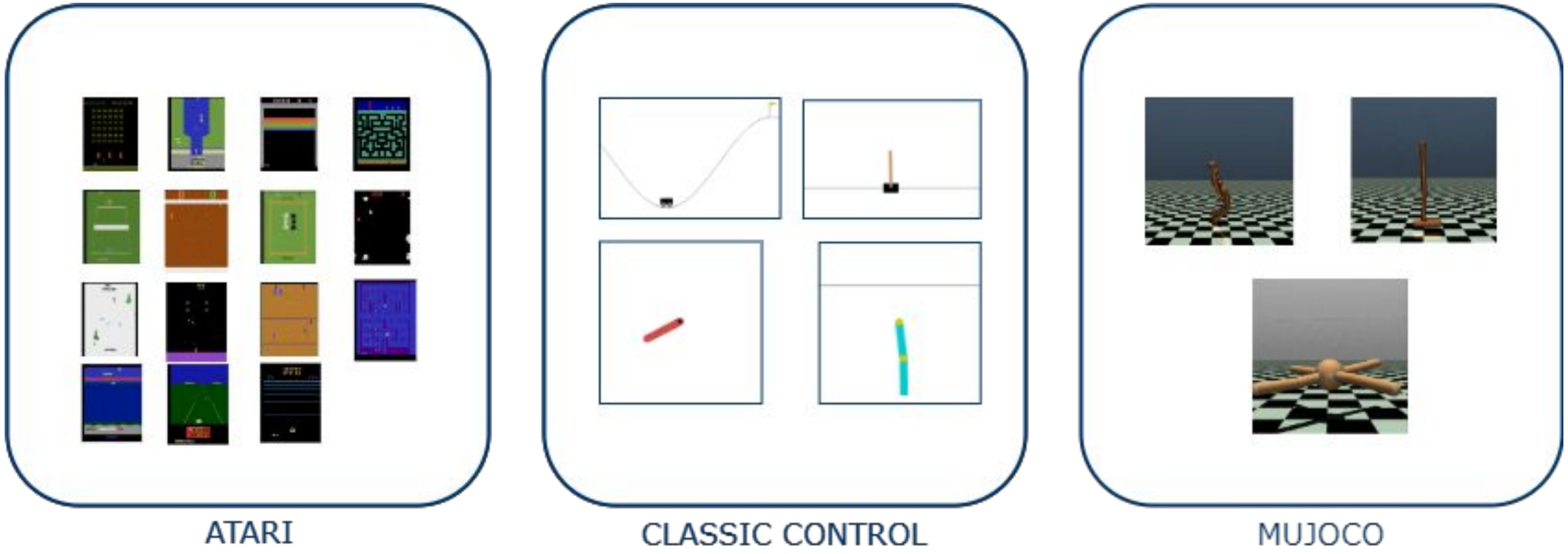
HPO-RL-Bench

Environments



HPO-RL-Bench

Environments



Static Benchmark Search Spaces

Algorithm	Hyperparameter name	Hyperparameter values
	clip range	0.2, 0.3, 0.4
	lr (log_{10})	-1, -2, -3, -4, -5, -6
TD3, SAC, DDPG; DQN; PPO; A2C	*gamma*	0.8, 0.9, 0.95, 0.98, 0.99, 1.0
	n_layers	1, 2, 3
	n_units	32, 64, 128
	epsilon	0.2, 0.3, 0.4
	tau	0.0001, 0.001, 0.005

Static Benchmark
Validating Usefulness

- RL Zoo-3 [Raffin, 2020] uses Optuna to provide tuned hyperparameters for RL algorithms included in stable-baselines3 [Raffin, 2021].

Static Benchmark

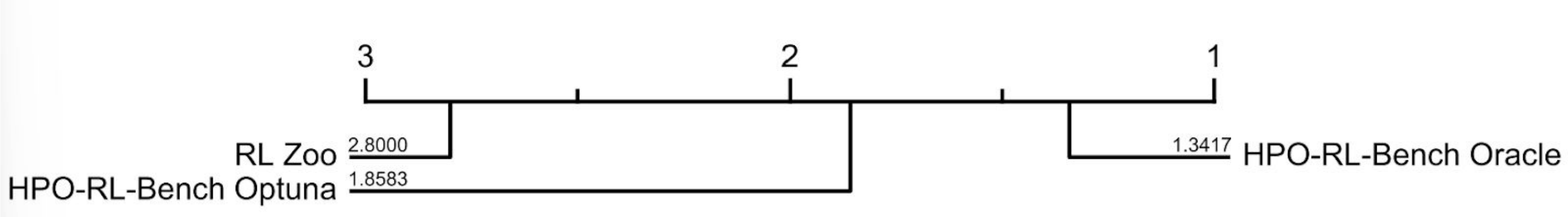
Validating Usefulness

- RL Zoo-3 [Raffin, 2020] uses Optuna to provide tuned hyperparameters for RL algorithms included in stable-baselines3 [Raffin, 2021].
- RL Zoo-3 search spaces contain 9-13 hyperparameters.

Static Benchmark

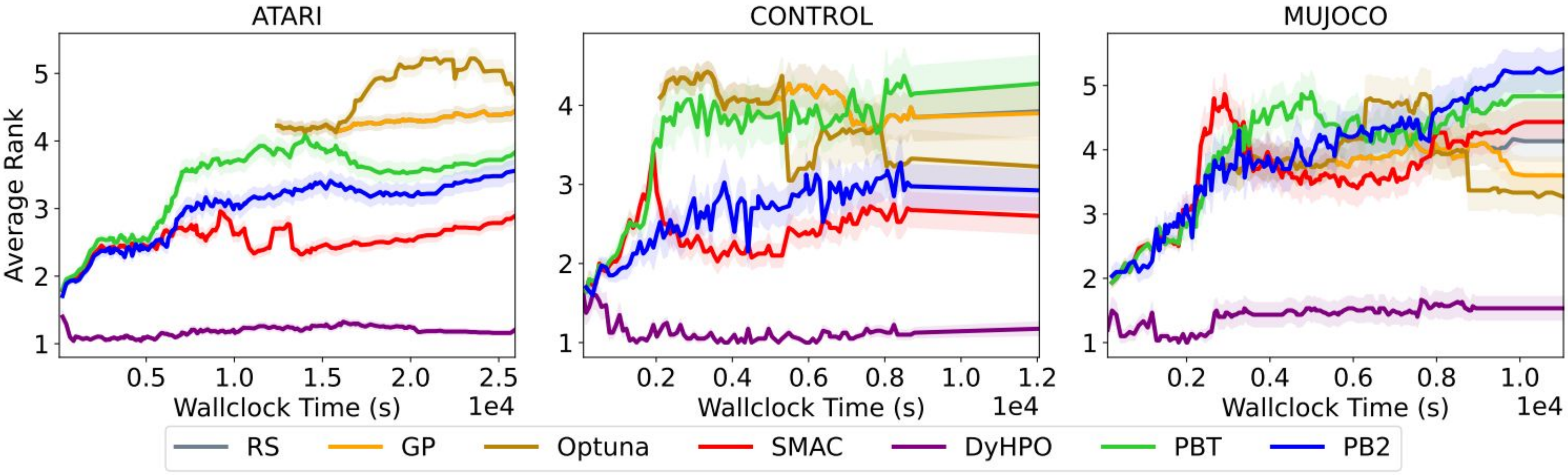
Validating Usefulness

- RL Zoo-3 [Raffin, 2020] uses Optuna to provide tuned hyperparameters for RL algorithms included in stable-baselines3 [Raffin, 2021].
- RL Zoo-3 search spaces contain 9-13 hyperparameters.



Static Benchmark

Results



Dynamic Benchmark

Search Spaces

Algorithm	Hyperparameter name	Hyperparameter values
PPO, TD3, SAC	*lr (log_{10})*	-3, -4, -5
	gamma	0.95, 0.98, 0.99

- HPO-RL-Bench includes evaluations and learning curves of performance of hyperparameter schedules for 5 environments and 3 algorithms.

Dynamic Benchmark

Search Spaces

Algorithm	Hyperparameter name	Hyperparameter values
PPO, TD3, SAC	*lr (log_{10})*	-3, -4, -5
	gamma	0.95, 0.98, 0.99

- HPO-RL-Bench includes evaluations and learning curves of performance of hyperparameter schedules for 5 environments and 3 algorithms.
- Evaluating hyperparameter schedules with 2 switching points already amounts to (3^2)^3=729 different configurations.

Dynamic Benchmark

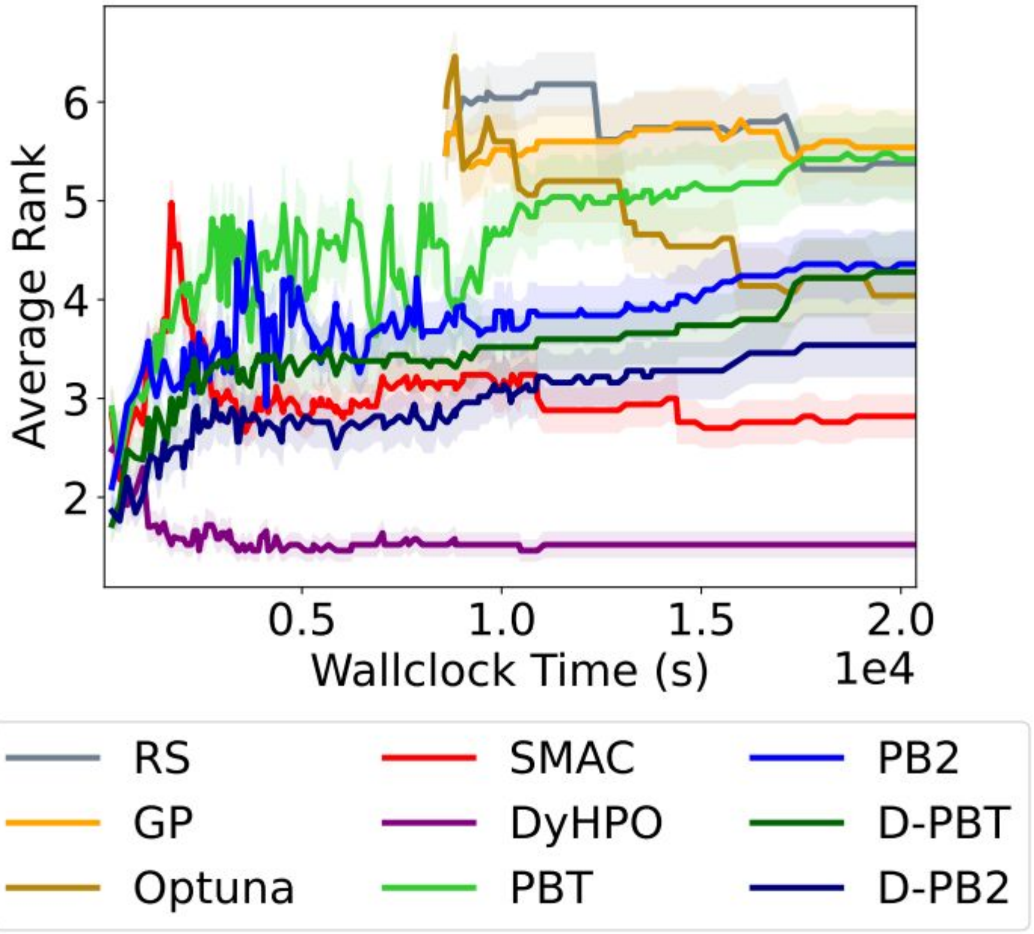
Search Spaces

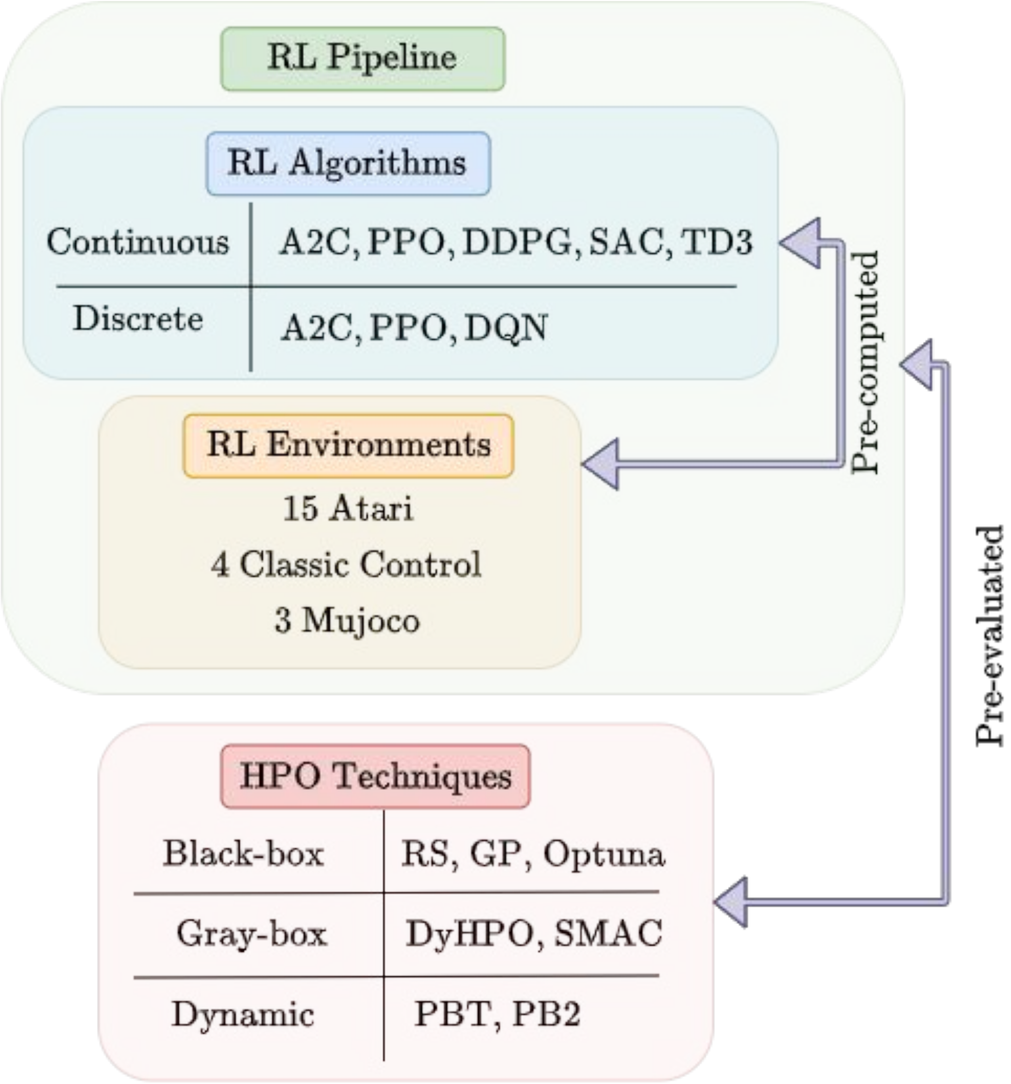
Algorithm	**Hyperparameter name**	**Hyperparameter values**
PPO, TD3, SAC	*lr (log_{10})*	-3, -4, -5
	gamma	0.95, 0.98, 0.99

- HPO-RL-Bench includes evaluations and learning curves of performance of hyperparameter schedules for 5 environments and 3 algorithms.
- Evaluating hyperparameter schedules with 2 switching points already amounts to (3^2)^3=729 different configurations.
- Using the original spaces for PPO, TD3, and SAC would have resulted in (6·6·3·3·3)^3>9·10^8 different configurations per algorithm and environment.

Dynamic Benchmark

Results





HPO-RL-Bench

- HPO RL-Bench drastically **reduces computational requirements** for evaluating HPO methods for RL.
- It includes evaluations across **22 environments** and **6 RL algorithms**, with episodic reward curve information.
- In addition to **static hyperparameter configurations**, it includes performance evaluations of **hyperparameter schedules** with distinct switching points.

Thank You!

Come meet us in poster session 2!

Come to the AutoRL Tutorial this afternoon!