

RSL-RL: A Learning Library for Robotics Research

Clemens Schwarke^{1,2}

CSCHWARKE@ETHZ.CH

Mayank Mittal^{1,2}

MITTALMA@ETHZ.CH

Nikita Rudin^{1,2,3}

NIKITA@FLEXION.AI

David Hoeller^{1,2,3}

DAVID@FLEXION.AI

Marco Hutter¹

MAHUTTER@ETHZ.CH

¹ETH Zürich, ²NVIDIA, ³Flexion Robotics

Abstract

RSL-RL is an open-source Reinforcement Learning library tailored to the specific needs of the robotics community. Unlike broad general-purpose frameworks, its design philosophy prioritizes a compact and easily modifiable codebase, allowing researchers to adapt and extend algorithms with minimal overhead. The library focuses on algorithms most widely adopted in robotics, together with auxiliary techniques that address robotics-specific challenges. Optimized for GPU-only training, RSL-RL achieves high-throughput performance in large-scale simulation environments. Its effectiveness has been validated in both simulation benchmarks and in real-world robotic experiments, demonstrating its utility as a lightweight, extensible, and practical framework to develop learning-based robotic controllers. The library is open-sourced at: https://github.com/leggedrobotics/rsl_rl.

Keywords: Reinforcement Learning, Distillation, Robotics, PyTorch

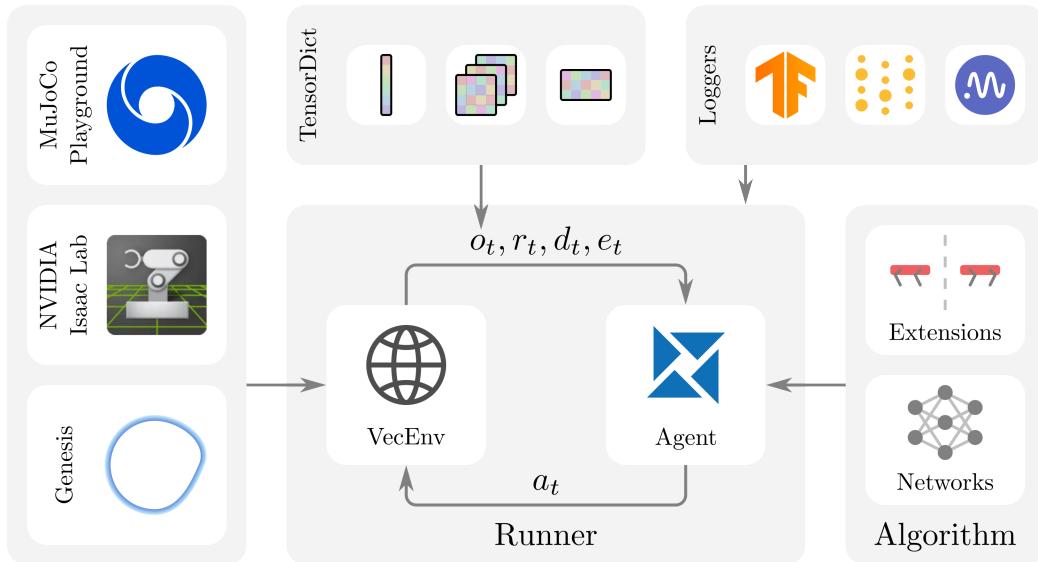


Figure 1: **Overview of the framework.** RSL-RL consists of three main components: **Runners**, **Algorithms**, and **Networks**, which can be easily modified independently. The framework comes with support for common logging choices and useful extensions for robotics.

1 Introduction

Reinforcement Learning (RL) has proven highly effective in tackling complex planning and control problems, often surpassing traditional model-based approaches (Hwangbo et al., 2019; Miki et al., 2022; Handa et al., 2023). These advances have been supported by both the development of better learning algorithms (Schulman et al., 2017) and the emergence of GPU-accelerated physics simulators (Makoviychuk et al., 2021; Mittal et al., 2023; Zakka et al., 2025) that enable large-scale parallelized training on consumer-grade hardware.

Although numerous RL libraries are available (Raffin et al., 2021; Liang et al., 2018; Huang et al., 2022b; Serrano-Munoz et al., 2023; Makoviichuk and Makoviychuk, 2021; Bou et al., 2023), most are developed for the larger machine learning community. Their breadth of algorithms facilitates benchmarking, but the resulting modularity can make the code harder to adapt and extend. Consequently, adoption in robotics has been limited. Although robotics benefits from advances in fundamental algorithmic research, researchers often prioritize a compact and easily modifiable codebase that can efficiently benefit from large-scale simulation. At the same time, robotics applications frequently require specialized features, such as tools for distilling learned policies for real-world deployment.

RSL-RL addresses these limitations by deliberately adopting a minimalist yet powerful design, centered on a curated set of widely adopted state-of-the-art algorithms and features tailored for roboticists. The core characteristics of the framework are as follows:

- **Minimalist design:** A compact, readable codebase with clear extension points (runners, algorithms, networks) for rapid research prototyping.
- **Robotics-first methods:** Proximal Policy Optimization (PPO) and DAgger-style Behavior Cloning (BC), along with auxiliary techniques (symmetry augmentation and curiosity-driven exploration) that are often absent from general RL libraries.
- **High-throughput training:** A GPU-only pipeline for large-scale batched training with native multi-GPU/multi-node support.
- **Proven and packaged:** Used in numerous robotics research publications, and integrated with various GPU-accelerated robotic simulation frameworks such as NVIDIA Isaac Lab, MuJoCo Playgroun, and Genesis for out-of-the-box use.

2 Features

In the following, we introduce the main functionalities of RSL-RL at the time of writing. While more algorithms and features are planned for the future, the library is meant to be adaptable, and our main priority is to keep it simple and easy to use.

2.1 Algorithms

Currently, RSL-RL includes two algorithms: PPO (Schulman et al., 2017) and a BC algorithm similar to DAgger (Ross et al., 2011). PPO is a model-free, on-policy RL method that has become a standard in robot learning for its robustness and simplicity (Shakya et al., 2023). It can be used to learn complex tasks from scratch, without requiring prior

knowledge or demonstrations. In robotics, PPO is most commonly applied to continuous control problems such as locomotion, manipulation, planning, and navigation.

The BC algorithm is a supervised learning method used to distill the behavior of an expert policy into a student policy. It iteratively collects data by rolling out the student policy, relabels the data with expert actions, and trains the student on it. This algorithm is particularly useful after RL training with PPO, if the requirements for training and hardware deployment differ. In such cases, the behavior of an RL agent relying on information only available in simulation can be distilled into a policy that does not rely on it.

2.2 Auxiliary Techniques

Alongside the main algorithms, the library includes techniques to improve performance in robotics applications. Currently, two such extensions are implemented, but this selection will be expanded in the future. The first technique, symmetry augmentation, accelerates sample generation by augmenting the collected data with mirrored states, exploiting the robot’s physical symmetries (Mittal et al., 2024). This leads to more symmetric behaviors, an effect that can be reinforced with an additional symmetry loss.

A second technique improves exploration in sparse reward settings using a curiosity-driven intrinsic reward based on Random Network Distillation (RND) (Burda et al., 2019). However, in contrast to the original RND formulation, the included implementation computes the reward using only a subset of the full state of the system. This modification allows the agent’s curiosity to be focused on specific parts of the state space. Curiosity-driven exploration can substantially reduce the need for hand-engineered dense rewards. For instance, Schwarke et al. (2023) trained a robot to open a door while balancing on two legs with a single binary task reward.

2.3 Utilities

RSL-RL provides several options for logging and evaluating experiments. The most straightforward option is TensorBoard (Abadi et al., 2016), which facilitates local experiment logging. For a more sophisticated evaluation, the library supports Weights & Biases (Biewald, 2020) and Neptune (neptune.ai, 2024). These are excellent choices for training on compute clusters, as they allow experiments to be monitored live online. However, these services require a user account and involve uploading data to their respective cloud platforms. The library also supports distributed training over multiple nodes and multi-GPUs for both of its algorithms, enabling large-scale experimentation.

3 Implementation Details

The framework is organized into three main components, outlined in Fig. 1: (i) the `Runner` that manages environment stepping and agent learning, (ii) the `Algorithm` that defines the learning agent, and (iii) the `Network` architectures used by the algorithm. For most applications, users only need to modify these three files. Implemented in PyTorch (Paszke et al., 2019) and Python, the framework is designed to be intuitive and easily extensible.

The framework defines its own `VecEnv` interface, which requires environments to implement the same-step reset mode (Towers et al., 2024). The `step` method must return

PyTorch tensors, and the observations must be structured as a *TensorDict* (Bou et al., 2023). TensorDicts provide a flexible dictionary-like container for batched tensors. They enable selective routing of observation components to different network modules and naturally support auxiliary techniques such as RND or latent reconstruction for regularization.

The included PPO algorithm incorporates several implementation subtleties highlighted in Huang et al. (2022a), such as the proper handling of episodic timeouts. In large-batch training, many environments tend to terminate simultaneously at the beginning of training, producing correlated rollouts. To mitigate this issue, the framework supports random early termination of episodes during initialization, which improves sample diversity and stabilizes learning. For both RL and distillation settings, recurrent networks are supported through explicit management of hidden states across rollouts and correct handling of Backpropagation Through Time (BPTT).

4 Applications in Research

RSL-RL was originally introduced to demonstrate the benefits of massively parallel simulation and GPU-based training for legged locomotion, achieving walking policies in only a few minutes (Rudin et al., 2022). Since then, it has become a foundation for a wide range of robotics research.

Building on this initial work, researchers have applied the framework for sim-to-real agile locomotion (Cheng et al., 2023; Margolis and Agrawal, 2023) and whole-body control (Fu et al., 2023; He et al., 2024; Arm et al., 2024). Hoeller et al. (2024) extended the framework to use mixed action distributions for training a high-level navigation policy that orchestrates multiple expert skills. Rudin et al. (2025) used teacher-student distillation to train a generalist locomotion policy. More recently, attention-based network architectures have been employed for locomotion on sparse terrain (He et al., 2025) and depth-based navigation (Yang et al., 2025). By incorporating demonstrations, several works have extended the framework for adversarial style (Vollenweider et al., 2022) and DeepMimic style training (Sleiman et al., 2024). Other studies (Lee et al., 2024; Dadiotis et al., 2025) have enhanced the PPO implementation to handle constraints during training through P3O (Zhang et al., 2022) and CaT (Chane-Sane et al., 2024). Finally, the symmetry-based augmentation methods included in the framework have been leveraged for multi-agent coordination (Li et al., 2025) and low-level control (Zhang et al., 2024a,b).

5 When and When Not to Use RSL-RL

RSL-RL is designed for robotics researchers who need a compact, modifiable, and well-validated learning codebase. It supports research aimed at advancing the state-of-the-art in robotics via RL, allowing users to either implement new ideas or apply the library directly to enhance real-robot capabilities. Its features are specifically built to improve performance and facilitate successful sim-to-real transfer, a critical step for validating new methods.

The library is not intended for fundamental or general-purpose machine learning research. By design, it includes a limited set of algorithms, making it unsuitable for straightforward benchmarking of other RL algorithms. Furthermore, RSL-RL does not provide native support for pure imitation learning.

Acknowledgments and Disclosure of Funding

Nikita Rudin and David Hoeller developed the initial version (v1.0.2) of the library, as used in Rudin et al. (2022), during their time at ETH Zürich and NVIDIA. Clemens Schwarke and Mayank Mittal have since maintained and extended the library. The authors thank all open-source contributors. This work is supported by NVIDIA.

References

- Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv e-prints*, pages arXiv–1603, 2016.
- Philip Arm, Mayank Mittal, Hendrik Kolenbach, and Marco Hutter. Pedipulate: Enabling manipulation skills using a quadruped robot’s leg. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 5717–5723. IEEE, 2024.
- Lukas Biewald. Experiment tracking with weights and biases, 2020. URL <https://www.wandb.com/>. Software available from wandb.com.
- Albert Bou, Matteo Bettini, Sebastian Dittert, Vikash Kumar, Shagun Sodhani, Xiaomeng Yang, Gianni De Fabritiis, and Vincent Moens. Torchrl: A data-driven decision-making library for pytorch, 2023.
- Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. Exploration by random network distillation. In *International Conference on Learning Representations (ICLR)*, pages 1–17, 2019.
- Elliot Chane-Sane, Pierre-Alexandre Leziart, Thomas Flayols, Olivier Stasse, Philippe Souères, and Nicolas Mansard. Cat: Constraints as terminations for legged locomotion reinforcement learning. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 13303–13310. IEEE, 2024.
- Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme parkour with legged robots. *Conference on Robot Learning (CoRL)*, 2023.
- Ioannis Dadiotis, Mayank Mittal, Nikos Tsagarakis, and Marco Hutter. Dynamic object goal pushing with mobile manipulators through model-free constrained reinforcement learning. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2025.
- Zipeng Fu, Xuxin Cheng, and Deepak Pathak. Deep whole-body control: learning a unified policy for manipulation and locomotion. In *Conference on Robot Learning (CoRL)*, pages 138–149. PMLR, 2023.
- Ankur Handa, Arthur Allshire, Viktor Makoviychuk, Aleksei Petrenko, Ritvik Singh, Jingzhou Liu, Denys Makoviichuk, Karl Van Wyk, Alexander Zhurkevich, Balakumar

Sundaralingam, et al. Dextreme: Transfer of agile in-hand manipulation from simulation to reality. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 5977–5984. IEEE, 2023.

Junzhe He, Chong Zhang, Fabian Jenelten, Ruben Grandia, Moritz Bächer, and Marco Hutter. Attention-based map encoding for learning generalized legged locomotion. *Science Robotics*, 10(105):eadv3604, 2025.

Tairan He, Wenli Xiao, Toru Lin, Zhengyi Luo, Zhenjia Xu, Zhenyu Jiang, Jan Kautz, Changliu Liu, Guanya Shi, Xiaolong Wang, et al. Hover: Versatile neural whole-body controller for humanoid robots. *arXiv preprint arXiv:2410.21229*, 2024.

David Hoeller, Nikita Rudin, Dhionis Sako, and Marco Hutter. Anymal parkour: Learning agile navigation for quadrupedal robots. *Science Robotics*, 9(88):eadi7566, 2024.

Shengyi Huang, Rousslan Fernand Julien Dossa, Antonin Raffin, Anssi Kanervisto, and Weixun Wang. The 37 implementation details of proximal policy optimization. In *ICLR Blog Track*, 2022a. URL <https://iclr-blog-track.github.io/2022/03/25/ppo-implementation-details/>.

Shengyi Huang, Rousslan Fernand Julien Dossa, Chang Ye, Jeff Braga, Dipam Chakraborty, Kinal Mehta, and JoĂGo GM AraĂšjo. Cleanrl: High-quality single-file implementations of deep reinforcement learning algorithms. *Journal of Machine Learning Research (JMLR)*, 23(274):1–18, 2022b.

Jemin Hwangbo, Joonho Lee, Alexey Dosovitskiy, Dario Bellicoso, Vassilios Tsounis, Vladlen Koltun, and Marco Hutter. Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 4(26):eaau5872, 2019.

Joonho Lee, Marko Bjelonic, Alexander Reske, Lorenz Wellhausen, Takahiro Miki, and Marco Hutter. Learning robust autonomous navigation and locomotion for wheeled-legged robots. *Science Robotics*, 9(89):eadi9641, 2024.

Zichong Li, Filip Bjelonic, Victor Klemm, and Marco Hutter. Marladona-towards cooperative team play using multi-agent reinforcement learning. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

Eric Liang, Richard Liaw, Robert Nishihara, Philipp Moritz, Roy Fox, Ken Goldberg, Joseph Gonzalez, Michael Jordan, and Ion Stoica. Rllib: Abstractions for distributed reinforcement learning. In *International Conference on Machine Learning (ICML)*, pages 3053–3062, 2018.

Denys Makoviichuk and Viktor Makoviychuk. rl-games: A high-performance framework for reinforcement learning. https://github.com/Denys88/rl_games, May 2021.

Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State.

Isaac gym: High performance gpu based physics simulation for robot learning. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

Gabriel B. Margolis and Pulkit Agrawal. Walk these ways: Tuning robot control for generalization with multiplicity of behavior. In *Conference on Robot Learning (CoRL)*, volume 205, pages 22–31. PMLR, 2023.

Takahiro Miki, Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science Robotics*, 7(62), 2022.

Mayank Mittal, Calvin Yu, Qinx Yu, Jingzhou Liu, Nikita Rudin, David Hoeller, Jia Lin Yuan, Ritvik Singh, Yunrong Guo, Hammad Mazhar, et al. Orbit: A unified simulation framework for interactive robot learning environments. *IEEE Robotics and Automation Letters (RA-L)*, 8(6):3740–3747, 2023.

Mayank Mittal, Nikita Rudin, Victor Klemm, Arthur Allshire, and Marco Hutter. Symmetry considerations for learning task symmetric robot policies. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 7433–7439. IEEE, 2024.

neptune.ai. neptune.ai: experiment tracker, 2024. URL <https://neptune.ai>.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research (JMLR)*, 22(268):1–8, 2021.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In *Conference on Robot Learning (CoRL)*, pages 91–100, 2022.

Nikita Rudin, Junzhe He, Joshua Aurand, and Marco Hutter. Parkour in the wild: Learning a general and extensible agile locomotion policy using multi-expert distillation and rl fine-tuning. *International Journal of Robotics Research (IJRR)*, 2025.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv e-prints*, pages arXiv–1707, 2017.

- Clemens Schwarke, Victor Klemm, Matthijs Van der Boon, Marko Bjelonic, and Marco Hutter. Curiosity-driven learning of joint locomotion and manipulation tasks. In *Conference on Robot Learning (CoRL)*, volume 229, pages 2594–2610, 2023.
- Antonio Serrano-Munoz, Dimitrios Chrysostomou, Simon Bøgh, and Nestor Arana-Arexolaleiba. skrl: Modular and flexible library for reinforcement learning. *Journal of Machine Learning Research (JMLR)*, 24(254):1–9, 2023.
- Ashish Kumar Shakya, Gopinatha Pillai, and Sohom Chakrabarty. Reinforcement learning algorithms: A brief survey. *Expert Systems with Applications*, 231:120495, 2023.
- Jean-Pierre Sleiman, Mayank Mittal, and Marco Hutter. Guided reinforcement learning for robust multi-contact loco-manipulation. In *Conference on Robot Learning (CoRL)*, 2024.
- Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, et al. Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*, 2024.
- Eric Vollenweider, Marko Bjelonic, Victor Klemm, Nikita Rudin, Joonho Lee, and Marco Hutter. Advanced skills through multiple adversarial motion priors in reinforcement learning. *arXiv preprint arXiv:2203.14912*, 2022.
- Fan Yang, Per Frivik, David Hoeller, Chen Wang, Cesar Cadena, and Marco Hutter. Improving long-range navigation with spatially-enhanced recurrent memory via end-to-end reinforcement learning. *arXiv preprint arXiv:2506.05997*, 2025.
- Kevin Zakka, Baruch Tabanpour, Qiayuan Liao, Mustafa Haiderbhai, Samuel Holt, Jing Yuan Luo, Arthur Allshire, Erik Frey, Koushil Sreenath, Lueder A Kahrs, et al. Mujoco playground. *Robotics: Science and Systems (RSS)*, 2025.
- Chong Zhang, Nikita Rudin, David Hoeller, and Marco Hutter. Learning agile locomotion on risky terrains. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11864–11871. IEEE, 2024a.
- Chong Zhang, Wenli Xiao, Tairan He, and Guanya Shi. Wococo: Learning whole-body humanoid control with sequential contacts. *arXiv preprint arXiv:2406.06005*, 2024b.
- Linrui Zhang, Li Shen, Long Yang, Shixiang Chen, Bo Yuan, Xueqian Wang, and Dacheng Tao. Penalized proximal policy optimization for safe reinforcement learning. *arXiv preprint arXiv:2205.11814*, 2022.