

# PPOrphIt: PPO Robot Manipulator Morphology Design Literature Review

Steve Gillet

*University of Colorado Boulder Robotics*

*Email: steve.gillet@colorado.edu*

**Abstract**—This paper aims to address the cost and overactuation of robotics used in manufacturing by leveraging reinforcement learning (RL) to optimize robot morphologies to specific use cases. RL is the right tool for this job because it can optimize over the large search space of possible robot morphologies and effectively explore large varieties of embodiments. The core methodology involves using a PPO model to iterate on the different morphologies and MuJoCo to evaluate their effectiveness in simulation. The evaluation signals will be success, path and energy efficiency, simplicity and manipulability index and these will be used to show the effectiveness of the PPO model at creating morphologies that are faster, more effective, and more efficient for particular tasks.

**Index Terms**—reinforcement learning, robot morphology, manipulator design

## I. INTRODUCTION AND MOTIVATION

High costs, lack of tailorability, and lack of retrofitting are some of the most cited problems for the adoption of the robotics in industrial settings [1]. The most commonly used robotic manipulator in manufacturing is FANUC series of 6-DOF manipulators which cost from \$17,500 to \$400,000 depending on their capacity and customizations [2] [3]. Meanwhile [4] and [5] show that most manipulators are overactuated, less cost-efficient, and less effective for most tasks they are used in. I want to address this problem by creating an automation pipeline where you feed in a use case and you get out a robot specifically optimized for that use case. By matching the robot to the task I can ensure that each robot can perform its task effectively and only the parts needed for that task are used, reducing cost and increasing efficiency. Here I describe the process by which robot manipulator morphologies are optimized to particular tasks using a PPO agent.

I chose reinforcement learning because morphology design is an optimization problem over a complex, high-dimensional search space and RL can explore and learn optimal designs through trial-and-error in simulation [6]. I chose PPO because it is robust and on-policy and excels in continuous action spaces which is ideal for morphology design where there are continuous parameters like link lengths [7]. This is more effective than imitation learning because imitation learning can only use morphologies that already exist severely limiting sample space and innovation [8]. Supervised learning suffers from a very similar issue where labelled data would have to be

provided which would be infeasible and there is no mechanism for exploration [9].

The core method is to use stable\_baselines3 PPO implementation to iterate on the design parameters (number of links, joint types (X, Y, Z hinge, and Z actuation), and link lengths), use MuJoCo to simulate that iteration on a simple task using RRTConnect and damped least squares (DLS) inverse kinematics (IK) for the path planning, feed the results (success, efficiency, and manipulability index) back into the PPO model. I chose RRTConnect because it is simple and has been shown to be faster than other methods [10] which enables faster and more robust evaluation steps. I also chose DLS IK for a similar reason, it has been shown to be robust especially in situations where the target position is out of reach [11] which is inevitable in a situation where an agent is exploring morphologies.

The evaluation signals will be successful path planning, path length, energy cost, and manipulability index. The manipulability index is a measure of a manipulators ability to arbitrarily change the position and orientation of the end effector at a particular point, borrowed from [12]. The hope is that the better morphologies will be able to do the tasks more successfully, efficiently, and quickly while still maintaining some robustness and then I can vary the tasks slightly to find better morphologies for particular tasks. The tasks will be kept simple but varied in ways to test different types of movements like moving near the base, moving from near to far, different elevations, more linear movements. You can imagine that a lot of pick and place tasks are mostly linear and so might largely be done by linear actuators more efficiently.

The research questions I seek to answer: Can PPO be used to generate more efficient morphologies? What are the best evaluation signals to use? How does varying evaluation signals yield different results (does optimizing for path shortness yield more simple manipulators while optimizing for manipulability measure yield more complex manipulators?)?

## II. BACKGROUND AND RELATED WORK

Most of the research in this area codesigns morphology and control of modular, atomic robots like [13] which uses PPO, [14] which uses TD3, [15], and [16] which uses PPO and A3C for configuration and control of a reconfigurable, modular robot. More pertinent [17] uses DDQN to co-optimize

morphology and control for modular robotic manipulators and [18] which uses soft actor critic for morphology and control of quadruped robots to avoid evaluating performance in sim or real to save time. My method will focus on manipulators with simple controllers to allow me to focus on morphology design. The method will use standard 6-DOF FANUC and 7-DOF Franka Emika Panda morphology as a baseline and contrast with PPO generated morphologies. PPO has been used effectively in some of the relevant research and allows for searching over the continuous space of manipulator parameters instead of the discrete space of different module types in the atomic, reconfigurable manipulator optimization. In the future, I would also like to try other RL approaches like Dreamer. Dreamer uses a 'world model' to predict outcomes and plan ahead which improves sample efficiency which might be even more effective by reducing computation needs [19].

## REFERENCES

- [1] McKinsey & Company, "Industrial robotics: Insights into the sector's future growth dynamics," McKinsey & Company, Tech. Rep., 2019. [Online]. Available: <https://www.mckinsey.com/business-functions/operations/our-insights/industrial-robotics-insights-into-the-sectors-future-growth-dynamics>
- [2] Standard Bots, "Fanuc robot prices: Cost, models, and buying insights in 2025," 2025. [Online]. Available: <https://standardbots.com/blog/fanuc-robot-price>
- [3] PatentPC, "Top robotics vendors by market share & installations," September 2025. [Online]. Available: <https://patentpc.com/blog/top-robotics-vendors-by-market-share-installations>
- [4] M. Russo, L. Raimondi, X. Dong, and D. Axinte, "Task-oriented optimal dimensional synthesis of robotic manipulators with limited mobility," *Robotics and Computer-Integrated Manufacturing*, vol. 69, p. 102098, 2021.
- [5] B. He, S. Wang, and Y. Liu, "Underactuated robotics: A review," *International Journal of Advanced Robotic Systems*, vol. 16, no. 4, pp. 1–14, 2019.
- [6] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine Learning*, vol. 3, no. 1, pp. 9–44, 1988.
- [7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [8] J. M. D. Delgado and L. Oyedele, "Robotics in construction: A critical review of the reinforcement learning and imitation learning paradigms," *Advanced Engineering Informatics*, vol. 54, p. 101787, 2022.
- [9] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.
- [10] J. Kuffner and S. LaValle, "Rrt-connect: An efficient approach to single-query path planning," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, vol. 2, 2000, pp. 995–1001 vol.2.
- [11] S. R. Buss, "Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods," *IEEE Journal of Robotics and Automation*, vol. 17, no. 1-19, p. 16, 2004.
- [12] T. Yoshikawa, "Manipulability of robotic mechanisms," *The International Journal of Robotics Research*, vol. 4, no. 2, pp. 3–9, 1985.
- [13] J. Bhatia *et al.*, "Reinforcement learning for freeform robot design," *arXiv preprint arXiv:2310.05670*, 2023.
- [14] B. Tjanaka *et al.*, "Co-optimization of morphology and behavior of modular robots via hierarchical deep reinforcement learning," in *Robotics: Science and Systems (RSS)*, 2023.
- [15] A. Spielberg *et al.*, "Accelerated co-design of robots through morphological pretraining," *arXiv preprint arXiv:2502.10862*, 2025.
- [16] M. Kalimuthu, A. A. Hayat, T. Pathmakumar, M. R. Elara, and K. L. Wood, "A deep reinforcement learning approach to optimal morphologies generation in reconfigurable tiling robots," *Mathematics*, vol. 11, no. 18, p. 3893, 2023.
- [17] Z. Ding, H. Tang, H. Wan, C. Zhang, and R. Sun, "A modular robotic arm configuration design method based on double dqn with prioritized experience replay," *Symmetry*, vol. 16, no. 6, p. 714, 2024. [Online]. Available: <https://www.mdpi.com/2073-8994/16/6/714>
- [18] K. S. Luck, H. B. Amor, R. Calandra, and J. Peters, "Data-efficient co-adaptation of morphology and behaviour with deep reinforcement learning," in *Conference on Robot Learning (CoRL)*, 2019.
- [19] D. Hafner, T. Lillicrap, I. Fischer, R. Villegas, D. Ha, H. Lee, and J. Davidson, "Dream to control: Learning behaviors by latent imagination," *arXiv preprint arXiv:1912.01603*, 2019.