Model-Based Reinforcement Learning Robots for Transparent Vessel Identification and Manipulation in Self-Driving Labs

Colin Bellinger, Rupam Mahmood, Gautham Vasan, Fahim Shahriar

November 14, 2023

1 Background & Motivation

The design and validation of new materials and drugs is an iterative process involving hypothesis generation and confirmation in a physical lab. Traditional hypothesis generation is limited by the research teams's ability to analyze high-dimensional search spaces and the ever-growing volume of available data. Moreover, progress is limited by the need to efficiently carryout high-precision, repeatable experimental validations. As a result, there is a growing interest in so-called self-driving labs [8]. Self-driving labs combine recent developments in Artificial Intelligence (AI) and Machine Learning (ML) for data analysis and hypothesis generation with robotic systems to carry out the physical experiments and provide feedback [31, 29, 12]. Whilst the vast majority of the current research into digital chemistry has focused on the first component of the self-driving lab, automating the physical experiments is critical step in the AI4D process. It enables the design and discovery process to quickly and efficiently accept hypotheses, or reject them and move on to the next experiment. In the cases of rejection, automated labs create an opportunity for a real-time feedback loop in hypothesis generation, which is not viable at the rate of a human experimenter.

In recent years, progress in lab automation has been predominantly directed toward large-scale, high-throughput laboratories that are expensive to build and access. Such large-scale laboratories are out of reach of many university research groups and small and medium-sized companies. With the steady decrease in the cost of robotic arms, automation of human-centric labs is starting to emerge as the next step in scientific automation [27, 24, 1]. This has the potential to accelerate research in smaller and lower budget labs, and thus, help to democratize the discovery process.

Recently, authors in [6, 37] demonstrated the great potential for the integration of data-driven experimental design with a mobile robotic arm in small-scale and human-centric chemistry labs. In [6], for example, a KUKA Mobile Robot was developed to receive instructions and execute tasks, such as measurements to make and vessels to transport. The results are impressive; the robot operated continuously and autonomously over eight days, enabling the self-driving lab to identify photocatalyst mixtures that were six times more active than the initial formulations. Nonetheless, the integration of the robot into the lab required over a year of development involving highly technical tasks, including the installation of sensors, the meticulous specification of all relevant experimental objects in the lab, and robot programming. In particular, the robot must be pre-programmed with the coordinates of the necessary vessels and equipment required to carry out each experiment. Moreover, the entire system must be manually updated if new vessels, instruments, or experiments are required. Likewise, the system must be reprogrammed if it is transferred to a new lab with a different setup or orientation. We postulate that to achieve widespread use, robot chemists must be easily and efficiently transferred between settings and adaptable to new sample types, operations, instruments, and measurements. To this end, our work will research and develop general-purpose low-cost lab robots with these qualities through the use of Reinforcement Learning (RL).

Recent successes in the incorporation of ML into the robotic system suggest a path to better automation. Indeed, ML has fuelled advances in robots that can perform specialized tasks such as vacuum cleaning, assembly, welding, and pick-and-place items in warehouses. However, most of these robots cannot accomplish much beyond what they were explicitly designed to do, often operating within a very narrow set of conditions [25]. RL is a form of machine learning for sequential decision-making that learns general policies through experience by observing the outcomes of its chosen actions [30]. Thus, it is a natural fit for robots learning to navigate and operate in the real-world. Moreover, RL offers advantages over traditional robot programming, such as the ability to generalize, learn new skills, and refine existing skills. By learning a behaviour policy rather than requiring direct programming and reprogramming, RL has the potential to reduce costs and simplify the incorporation of robotic arms into chemistry labs [21].

In the last five years, RL for robotics and object manipulation as achieved significant advances [17, 28]. Nonetheless, many general questions remain open in the context of real-world robotics. To date, none of these have been explored in the context of lab robots. This work will fill part of the void by developing the research on low-cost RL robots for the identification and manipulation of transparent lab objects. Our experimental design will focus on sample efficiency, transferability, adaptability, and high-frequency control in physical robots, as these are essential components to achieve widespread adoption. From a theoretical perspective, we will explore the hypothesis that model-based methods are superior to model-free methods with respect to transferability and adaptability in complex real-world settings, such as automated chemistry labs.

2 Research Objective

Recently, RL has produced stunning performance in complex robotics tasks such as humanoid locomotion [13, 10] and solving a Rubik's Cube with a robot hand [3]. However, these results are produced in simulated environments and the algorithms have limited applicability to real-time learning setting with physical robots due to the large requirements for training data and computational resources. To be useful in real-world chemistry labs, and in a wide range of other applications, RL robots must have minimal real-world data requirements. This limits cost, wear-and-tear on the robot, and lab downtime due to robot training.

Learning from demonstrations [9, 33], sim-to-real [22, 5], and offline RL [16] have been widely studied to improve sample efficiency and reduce the costs of learning in the real-world. Pretrained agents, however, often encounter uncertainties, sensory noise, and unforeseen challenges that were not present in the simulation or offline dataset. Model-based RL presents itself as an alternative to model-free methods for sample efficiency [23]. By exploiting a learned (or given) dynamics and reward model for policy learning, model-based RL can reduce training time and the number of interactions with the physical robot. As with sim-to-real transfer, however, small errors in the model can cause errors in the learned policy, leading to reduced performance on the physical robot [22].

The model-based RL method denoted *DreamerV2* learns to encode observations and actions into discrete embedded states from experience. It was demonstrated to learn from small amounts of interaction in simulated game environments by planning within a learned world model [11]. Dreamer learns to predict actions and subsequent states by envisioning potential trajectories through its model of the world in order to optimize forthcoming imagined rewards. Recently, Dreamer was validated for the control of online physical robot tasks, such as grasping opaque, rigid objects. It was found to produce near human-level teleoperation performance from camera images and sparse rewards after 8 hours of training [36]. Alternatively, state-of-the-art model-free RL algorithms fail to learn to complete the task in this time.

Despite the recent success of DreamerV2, critical research questions related to the use of model-based RL robots in real-world labs (and in general) remain open. This work will focus on following questions:

- **Generality:** Automated labs are expected to adopt a wide variety of robotic arms. However, the generality of RL algorithms and architectures to diverse robotics platforms is not well studied. Therefore, our research will develop RL methods that generalize across robotics platforms.

- **Transferability:** Whilst each lab has subtle differences, the tasks that are expected to be completed, such as identifying and transferring vessels and measuring, will be common to a wide variety of lab settings. Therefore, robust transferable RL agents are essential to minimizing the cost and complexity of lab robots. The transferability of RL robots in physical environments, however, is poorly studied in general, and particularly in the context of model-based RL. This work will develop transferable model-based RL agents, and assess their efficacy in new settings in terms of safety, robustness and fine-tuning requirements.

- **Adaptability:** As lab research progresses, new vessels, instruments and skills are likely to be required. Similarly to transferability, adaptability is essential to minimizing the cost and complexity of laboratory robots. Therefore, our research will include a focus on designing, developing and assessing the adaptability of model-based RL. This research will include addressing the problems of catastrophic forgetting and the loss of plasticity [7] in continual learning RL in non-stationarities physical environments.

- **Transparent objects:** vessels, such as beakers and test-tubes, in chemistry labs are typically transparent or translucent depending on their contents. This pose a unique challenge for systems with visual perceptions and is known to lead to failures in grasp prediction [26, 20]. Existing work utilizes colour and depth sensing and relies on transfer learning from established depth-based grasping models built on large labelled datasets [35, 14]. Our work will research the manipulation of transparent, specular, or low-reflectance objects with RL, which does not require labelled training data.

- **High-frequency control:** rapid reaction times in RL robots are essential for ensuring agility, responsiveness, and adaptability. Existing work, however, typically operates a lower frequency control cycle due to the inherent complexities and computational demands of the RL algorithms [2, 36]. This work will focus on high-frequency control based on the real-time learning system called the Remote-Local Distributed (ReLoD) [34]. In particular, we will extend ReLoD to model-based RL.

3 Experimental Approach & Methods

Our experimental setup will utilize the Franka Emika Panda Arm, UR5 and UR10e industrial robot arms as physical test-beds. This will enable us to assess the developed algorithms on a diverse set of platforms and builds on our extensive experience prototyping applications and publishing research papers with them [18, 34, 4]. Since experiments are costly to perform on real robots, we will also use simulated experiments on Gazebo [15], MuJoCo [32] and Isaac Gym [19]. The simulated experiments will serve as litmus tests to rule out ineffective approaches.

The developed model-based RL algorithm will be evaluated on chemistry vessel identification and manipulation tasks that are inspired by the KUKA mobile robot's tasks in [6] and the Universal Robot's tasks in [37]. As discussed above, the developed algorithm and agents will be assessed in terms of generality, transferability, adaptability, accuracy with transparent and translucent objects, and high-frequency control. See Section 4 Statement of Work in the main document for more details on the individual activities included in our proposed experimental method.

References

[1] Abolhasani, M., Kumacheva, E.: The rise of self-driving labs in chemical and materials sciences. Nature Synthesis pp. 1–10 (2023)

[2] Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C., Gopalakrishnan, K., Hausman, K., et al.: Do as i can, not as i say: Grounding language in robotic affordances. arXiv preprint arXiv:2204.01691 (2022)

[3] Andrychowicz, O.M., Baker, B., Chociej, M., Jozefowicz, R., McGrew, B., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., et al.: Learning dexterous in-hand manipulation. The International Journal of Robotics Research **39**(1), 3–20 (2020)

[4] Bellinger, C., Lamarche-Cliche, L.: Learning visual tracking and reaching with deep reinforcement learning on a ur10e robotic arm (2023)

[5] Bousmalis, K., Irpan, A., Wohlhart, P., Bai, Y., Kelcey, M., Kalakrishnan, M., Downs, L., Ibarz, J., Pastor, P., Konolige, K., et al.: Using simulation and domain adaptation to improve efficiency of deep robotic grasping. In: 2018 IEEE international conference on robotics and automation (ICRA). pp. 4243–4250. IEEE (2018)

[6] Burger, B., Maffettone, P.M., Gusev, V.V., Aitchison, C.M., Bai, Y., Wang, X., Li, X., Alston, B.M., Li, B., Clowes, R., et al.: A mobile robotic chemist. Nature **583**(7815), 237–241 (2020)

[7] Dohare, S., Hernandez-Garcia, J., Rahman, P., Sutton, R., Mahmood, A.R.: Loss of plasticity in deep continual learning (2023)

[8] Editorial: For chemists, the ai revolution has yet to happen. Nature **617**(438) (2023)

[9] Gupta, A., Eppner, C., Levine, S., Abbeel, P.: Learning dexterous manipulation for a soft robotic hand from human demonstrations. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3786–3793. IEEE (2016)

[10] Haarnoja, T., Zhou, A., Abbeel, P., Levine, S.: Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: International conference on machine learning. pp. 1861–1870. PMLR (2018)

[11] Hafner, D., Lillicrap, T., Norouzi, M., Ba, J.: Mastering atari with discrete world models. arXiv preprint arXiv:2010.02193 (2020)

[12] Häse, F., Roch, L.M., Aspuru-Guzik, A.: Next-generation experimentation with self-driving laboratories. Trends in Chemistry **1**(3), 282–291 (2019)

[13] Heess, N., Tb, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S., et al.: Emergence of locomotion behaviours in rich environments. arXiv preprint arXiv:1707.02286 (2017)

[14] Hoffman, J., Gupta, S., Leong, J., Guadarrama, S., Darrell, T.: Cross-modal adaptation for rgb-d detection. In: 2016 IEEE international conference on robotics and automation (ICRA). pp. 5032–5039. IEEE (2016)

[15] Koenig, N., Howard, A.: Design and use paradigms for gazebo, an open-source multi-robot simulator. In: 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566). vol. 3, pp. 2149–2154. IEEE (2004)

[16] Levine, S., Kumar, A., Tucker, G., Fu, J.: Offline reinforcement learning: Tutorial, review, and perspectives on open problems. arXiv preprint arXiv:2005.01643 (2020)

[17] Lobbezoo, A., Qian, Y., Kwon, H.J.: Reinforcement learning for pick and place operations in robotics: A survey. Robotics **10**(3), 105 (2021)

[18] Mahmood, A.R., Korenkevych, D., Vasan, G., Ma, W., Bergstra, J.: Benchmarking reinforcement learning algorithms on real-world robots. In: Conference on robot learning. pp. 561–591. PMLR (2018)

[19] Makoviychuk, V., Wawrzyniak, L., Guo, Y., Lu, M., Storey, K., Macklin, M., Hoeller, D., Rudin, N., Allshire, A., Handa, A., et al.: Isaac gym: High performance gpu-based physics simulation for robot learning. arXiv preprint arXiv:2108.10470 (2021)

[20] Morrison, D., Corke, P., Leitner, J.: Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach. arXiv preprint arXiv:1804.05172 (2018)

[21] Pedersen, M.R., Nalpantidis, L., Andersen, R.S., Schou, C., Bøgh, S., Krüger, V., Madsen, O.: Robot skills for manufacturing: From concept to industrial deployment. Robotics and Computer-Integrated Manufacturing **37**, 282–291 (2016)

[22] Peng, X.B., Andrychowicz, M., Zaremba, W., Abbeel, P.: Sim-to-real transfer of robotic control with dynamics randomization. In: 2018 IEEE international conference on robotics and automation (ICRA). pp. 3803–3810. IEEE (2018)

[23] Polydoros, A.S., Nalpantidis, L.: Survey of model-based reinforcement learning: Applications on robotics. Journal of Intelligent & Robotic Systems **86**(2), 153–173 (2017)

[24] Pyzer-Knapp, E.O., Pitera, J.W., Staar, P.W., Takeda, S., Laino, T., Sanders, D.P., Sexton, J., Smith, J.R., Curioni, A.: Accelerating materials discovery using artificial intelligence, high performance computing and robotics. npj Computational Materials **8**(1), 84 (2022)

[25] Roy, N., Posner, I., Barfoot, T., Beaudoin, P., Bengio, Y., Bohg, J., Brock, O., Depatie, I., Fox, D., Koditschek, D., et al.: From machine learning to robotics: challenges and opportunities for embodied intelligence. arXiv preprint arXiv:2110.15245 (2021)

[26] Satish, V., Mahler, J., Goldberg, K.: On-policy dataset synthesis for learning robot grasping policies using fully convolutional deep networks. IEEE Robotics and Automation Letters **4**(2), 1357–1364 (2019)

[27] Seifrid, M., Pollice, R., Aguilar-Granda, A., Morgan Chan, Z., Hotta, K., Ser, C.T., Vestfrid, J., Wu, T.C., Aspuru-Guzik, A.: Autonomous chemical experiments: Challenges and perspectives on establishing a self-driving lab. Accounts of Chemical Research **55**(17), 2454–2466 (2022)

[28] Singh, B., Kumar, R., Singh, V.P.: Reinforcement learning in robotic applications: a comprehensive survey. Artificial Intelligence Review pp. 1–46 (2022)

[29] Stein, H.S., Gregoire, J.M.: Progress and prospects for accelerating materials science with automated and autonomous workflows. Chemical science **10**(42), 9640–9649 (2019)

[30] Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018)

[31] Tabor, D.P., Roch, L.M., Saikin, S.K., Kreisbeck, C., Sheberla, D., Montoya, J.H., Dwaraknath, S., Aykol, M., Ortiz, C., Tribukait, H., et al.: Accelerating the discovery of materials for clean energy in the era of smart automation. Nature reviews materials **3**(5), 5–20 (2018)

[32] Todorov, E., Erez, T., Tassa, Y.: Mujoco: A physics engine for model-based control. In: 2012 IEEE/RSJ international conference on intelligent robots and systems. pp. 5026–5033. IEEE (2012)

[33] Vasan, G., Pilarski, P.M.: Learning from demonstration: Teaching a myoelectric prosthesis with an intact limb via reinforcement learning. In: 2017 International Conference on Rehabilitation Robotics (ICORR). pp. 1457–1464. IEEE (2017)

[34] Wang, Y., Vasan, G., Mahmood, A.R.: Real-time reinforcement learning for vision-based robotics utilizing local and remote computers. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 9435–9441. IEEE (2023)

[35] Weng, T., Pallankize, A., Tang, Y., Kroemer, O., Held, D.: Multi-modal transfer learning for grasping transparent and specular objects. IEEE Robotics and Automation Letters **5**(3), 3791–3798 (2020)

[36] Wu, P., Escontrela, A., Hafner, D., Abbeel, P., Goldberg, K.: Daydreamer: World models for physical robot learning. In: Conference on Robot Learning. pp. 2226–2240. PMLR (2023)

[37] Zhu, Q., Huang, Y., Zhou, D., Zhao, L., Guo, L., Yang, R., Sun, Z., Luo, M., Zhang, F., Xiao, H., Tang, X., Zhang, X., Song, T., Li, X., Chong, B., Zhou, J., Zhang, Y., Zhang, B., Cao, J., Zhang, G., Wang, S., Ye, G., Zhang, W., Zhao, H., Cong, S., Li, H., Ling, L.L., Zhang, Z., Shang, W., Jiang, J., Luo, Y.: Automated synthesis of oxygen-producing catalysts from martian meteorites by a robotic ai chemist. Nature Synthesis (2023). https://doi.org/10.1038/s44160-023-00424-1, `https://doi.org/10.1038/s44160-023-00424-1`