Efficient Vision-Based Reinforcement Learning for Physical Robots

Minghui Ye[1]

1 Outline of Proposed Research

Reinforcement Learning (RL) agents have enjoyed significant success—learning to solve challenging continuous control tasks [1, 2] and achieving superhuman performance on Atari [3] and Go [4, 5]. These progresses [6] have enabled robots to learn task-specific policies to perform complex control tasks, instead of designing operation rules for single ones. However, the current state-of-the-art methods primarily rely on state-based features, such as the relative position and velocity of the object to be manipulated, showing limited capability in addressing vision-based continuous control tasks [7, 8, 9, 1, 10]. While state-based learning algorithms have achieved excellent results in physical robot settings, they require precise state parameters in both learning and deployment stages, which is not always available in real-world scenarios. Furthermore, the tedious work of accurate measurements of state parameters does not guarantee enhanced performance [11]. Therefore, **the primary objective of my research project is to enable robots to learn vision-based control efficiently in real-world environments using RL techniques.** To achieve this goal, three lines of research will be explored: (1) The first research direction is to leverage offline datasets (i.e. historical suboptimal learning trajectories and expert demonstration) collected from real robots to facilitate online robot learning. (2) The second approach is to utilize imperfect simulators for policy learning and develop Sim2Real techniques that enable the successful deployment of learned policies on physical robots. (3) The last approach is to incorporate imperfect simulators and offline datasets for physical robot learning. Given the exceptional data efficiency demonstrated by Model-based RL in both online [12] and offline [13] settings, particularly in the context of physical robots, my research project will employ Model-based RL to harness imperfect simulators and offline datasets for vision-based control on physical robots. To explore this, my project draws upon ideas from representation learning [14, 15, 16], offline RL [17, 18], transfer learning [19, 20], Sim2Real [21, 22] and RL [23]. By exploring these three research directions, my project seeks to advance the field of physical robot learning in vision-based control tasks, ultimately enabling robots to acquire skills more effectively and perform tasks in real-world environments. Below I outline related works and subsequently detail the specific aspects I intend to pursue in my research.

2 Background & Related Work

Reinforcement learning has been introduced to tackle vision-based control challenges [24, 25, 26, 27, 28, 29]. However, many of these approaches face limitations in terms of testing solely within simulated environments. Moreover, due to the computational affordability of simulators, a majority of algorithms are trained from scratch for each instance, disregarding the potential benefits of leveraging prior interactions, which may not be conducive to physical robot platforms. Notably, training from scratch on real robots has been shown to be inefficient [24]. To mitigate the reliance on extensive online interactions, a range of valuable resources, including simulators, trajectories from similar tasks, trajectories from similar environments, and expert demonstrations, were leveraged to facilitate the reinforcement learning process.

Offline Reinforcement Learning A highly efficient and effective learning approach for real robots is imitation learning [31, 32], wherein policies are directly learned from expert demonstrations. Remarkable results have been achieved by combining pretrained visual encoders and imitation learning [33]. However, a limitation of this approach is its poor generalization capability, as expert data typically covers only a subset of the environment state. Moreover, expert demonstrations are also prohibitively expensive, while historical trajectories from similar tasks offer a viable alternative [34]. Offline RL [35, 36, 37] provides methods to leverage previous environment interactions, significantly reducing the number of interactions required for learning new tasks. However, current Offline RL algorithms are almost entirely testing in simulation [13, 38], leaving a notable research gap when it comes to their practical application on physical robots. Few works were done on real robot platforms, particularly

[1] Email: `yeminghui1@gmail.com`

in the context of manipulation [39, 40, 36]. Furthermore, it is beneficial to combine different techniques(such as Offline RL and meta RL) for efficient robot learning in the real world [41, 35].

Sim2Real Apart from historical trajectories collected from real robots, simulators can also serve as valuable resources for physical robot learning [42]. However, the inherent inaccuracies of simulators give rise to the reality-gap problem. To address this, various sim2real methods, such as domain[43] and dynamic [9, 8] randomization, domain adaption [30, 44, 45], and hybrid models [21, 22], have been proposed to assist policies trained in the simulator in adapting to the real world. The methods that will be utilized on this project is the hybrid model approach and domain adaption. The hybrid model approach is to use augmented simulation which can model the error between the physics engine and the real system and thus alleviate the problem of reality gap. However, most of these works were done on state-based RL [22]. Moreover, the prevailing method that transfers the vision-based policy learned in simulations to the real world [45] is domain adaption, which is to learn a mapping from the source domain(i.e. simulator) to the target domain(i.e. real-world) in pixel-level or feature-level. Pixel-level domain adaptation [46] involves learning transformation in the pixel space from one domain to the other, while feature-level adaptation [47] is to transform the encoded representation to another domain.

Representation Learning for RL One approach to address high-dimensional visual spaces is to pre-train representation models using in-domain data [48, 49], such as trajectories from the same environment, and out-of-domain data [33, 50], such as the ImageNet dataset. In this approach, the trained visual encoder is frozen and subsequently connected to a policy network which can be trained through imitation learning or reinforcement learning to execute specific tasks. Notably, despite being task-agnostic, the pre-trained visual encoder has been proven to be powerful across various robotic tasks [33], performing no worse than end-to-end trained encoders. Alternatively, another solution involves training the visual encoder from scratch by integrating representation learning [29] or world modeling [51] techniques to facilitate the learning of the visual encoder. Representation learning techniques, including autoencoders [52], contrastive learning [53], and data augmentation [29], were successfully integrated into reinforcement learning algorithms to enable vision-based control. Moreover, world modeling techniques exhibited remarkable data efficiency [13] and modeling capabilities, enabling model-based RL to achieve superhuman performance in visual tasks such as Atari [54, 55].

3 Hypotheses, Research Objectives, and Methodology

Boosting Robot Learning with Offline Datasets Unlike prevailing reinforcement learning algorithms that typically train the policy from scratch, physical robot learning prioritizes the utilization of prior experience to reduce the data budget required for training [56]. Offline RL was proposed to deal with this problem. However, existing offline RL research primarily focuses on state-based settings. Only recently has the RL community put forward benchmarks for the vision-based offline RL problem [13] and the physical robot setting [36]. The significant attributes of physical robot learning in the context of offline learning are: (1) offline data is also expensive, and (2) the generalization capability of the learned policy is highly valued. To address these considerations, **I would like to extend the Model-based offline RL algorithms [57] to vision-based tasks and physical robots by incorporating representation learning into current algorithms.** To be more specific, since offline data is still expensive, I will introduce visual encoders pre-trained on the ImageNet dataset to the world model and finetune the world model with offline data to approximate the dynamics of the physical world. In this way, the world model skips the representation learning process, thus enabling it to possess strong modeling ability with only a limited amount of real-world trajectories. As for policy learning, I will put more emphasis on developing algorithms that can generalize vision-based policy to new tasks based on previous experience in similar tasks. By delving into these research problems, this study will provide valuable insights into the development of offline RL algorithms for physical robots. The goal of this research is to enable robots to adapt to novel tasks more swiftly and intelligently by reusing previous experience.

Leveraging Imperfect Simulators for Physical Robot Learning Model-based reinforcement learning has demonstrated remarkable data efficiency, successfully enabling physical robots to learn vision-based control tasks from scratch [12]. To push the boundaries of this approach, an intriguing question arises: **can we enable**

robots to learn more complex policies with model-based RL by leveraging imperfect simulators? I hypothesized that learning a world model from imperfect simulators and subsequently finetune it with real-world trajectories would be more efficient than directly train a world model with interaction data in the real-world environments. The idea is that a simulator is better than nothing, even if it is not perfect. I intend to follow the Sim2Real pipeline which is to train a policy in the simulator and subsequently transfer the learned policy to the physical environment. However, unlike previous methods that commonly use model-free RL in Sim2Real which learns a black-box policy network that can only output actions, I choose to explore the model-based RL technique which builds a world model to approximate how the world evolves in response to each action. In this way, the dynamics error of the simulator can be modeled explicitly in the learning process, increasing the explainability of the algorithm and paving the way for finetuning the learned world model with online exploration in physical robots. Within this context, two difficulties remain to be solved: (1) how to deal with the discrepancy between the scene rendering and the real observation, and (2) how to finetune the inaccurate pretrained world model to approximate the real-world in the dynamic aspect. I will address the first problem by domain adaptation (e.g. translating images via CycleGAN). For the second question, I will employ suitable explore strategies to collect valuable trajectories for world model correction and develop methods for finetuning the pre-trained world model [58, 59, 60].

Robot Learning with Vision-based Hybrid Model In this setting, imperfect simulators and offline datasets collected from similar tasks are assumed to be available. Given that differences between the simulator and the real world exist not only in scene rendering but also in dynamic properties, I argue that the incorporation of simulators and offline datasets can significantly enhance the efficiency of physical robot learning. This can be achieved by learning both visual adaptation and dynamic adaptation at the same time. Hybrid model methods can be improved to deal with this problem setting. While previous Hybrid model methods are mostly state-based and learn the dynamics discrepancy between the simulation and the real-world online, I would like to extend these methods to vision-based robotic tasks in offline learning settings. Therefore, **I will investigate vision-based hybrid model methods, which is to build a world model to simultaneously learn the dynamics residual as well as the visual representation residual between the simulator and the real world with offline data.** Different from the setting in the last paragraph where no offline data was provided and the dynamics of the world model has to be finetuned online, this setting is to learn a policy totally offline or with minimal online interaction. Overall, this research is to develop advanced world modeling techniques that can leverage simulation and offline data, thereby enhancing the efficiency and effectiveness of online robot learning.

In summary, My research focuses on addressing the limitations of current RL methods in vision-based control tasks, specifically in the context of physical robot. By leveraging offline datasets, imperfect simulators, and model-based RL techniques, the project aims to contribute to the development of more efficient and effective learning methods for physical robots, thus paving the way for the widespread deployment of robots in the real world.

References

[1] Timothy P Lillicrap et al. "Continuous control with deep reinforcement learning". In: *arXiv preprint arXiv:1509.02971* (2015).

[2] Matthias Plappert et al. "Multi-goal reinforcement learning: Challenging robotics environments and request for research". In: *arXiv preprint arXiv:1802.09464* (2018).

[3] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *nature* 518.7540 (2015), pp. 529–533.

[4] David Silver et al. "Mastering the game of Go with deep neural networks and tree search". In: *nature* 529.7587 (2016), pp. 484–489.

[5] David Silver et al. "Mastering the game of go without human knowledge". In: *nature* 550.7676 (2017), pp. 354–359.

[6] Richard S Sutton and Andrew G Barto. “Reinforcement learning: an introduction MIT Press”. In: *Cambridge, MA* 22447 (1998).

[7] Robert Kirk et al. “A survey of generalisation in deep reinforcement learning”. In: *arXiv preprint arXiv:2111.09794* (2021).

[8] Jonah Siekmann et al. “Blind bipedal stair traversal via sim-to-real reinforcement learning”. In: *arXiv preprint arXiv:2105.08328* (2021).

[9] Yandong Ji et al. “Hierarchical reinforcement learning for precise soccer shooting skills using a quadrupedal robot”. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2022, pp. 1479–1486.

[10] John Schulman et al. “Proximal policy optimization algorithms”. In: *arXiv preprint arXiv:1707.06347* (2017).

[11] Kyle Hsu et al. “Vision-based manipulators need to also see from their hands”. In: *arXiv preprint arXiv:2203.12677* (2022).

[12] Philipp Wu et al. “Daydreamer: World models for physical robot learning”. In: *Conference on Robot Learning*. PMLR. 2023, pp. 2226–2240.

[13] Cong Lu et al. “Challenges and opportunities in offline reinforcement learning from visual observations”. In: *arXiv preprint arXiv:2206.04779* (2022).

[14] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. “Representation learning with contrastive predictive coding”. In: *arXiv preprint arXiv:1807.03748* (2018).

[15] Kaiming He et al. “Momentum contrast for unsupervised visual representation learning”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 9729–9738.

[16] Ting Chen et al. “A simple framework for contrastive learning of visual representations”. In: *International conference on machine learning*. PMLR. 2020, pp. 1597–1607.

[17] Aviral Kumar et al. “Conservative q-learning for offline reinforcement learning”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1179–1191.

[18] Sergey Levine et al. “Offline reinforcement learning: Tutorial, review, and perspectives on open problems”. In: *arXiv preprint arXiv:2005.01643* (2020).

[19] Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov. “Actor-mimic: Deep multitask and transfer reinforcement learning”. In: *arXiv preprint arXiv:1511.06342* (2015).

[20] Fuzhen Zhuang et al. “A comprehensive survey on transfer learning”. In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.

[21] Anurag Ajay et al. “Combining physical simulators and object-based networks for control”. In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 2019, pp. 3217–3223.

[22] Kei Ota et al. “Data-efficient learning for complex and real-time physical problem solving using augmented simulation”. In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 4241–4248.

[23] Yutaka Matsuo et al. “Deep learning, reinforcement learning, and world models”. In: *Neural Networks* (2022).

[24] Dmitry Kalashnikov et al. “Scalable deep reinforcement learning for vision-based robotic manipulation”. In: *Conference on Robot Learning*. PMLR. 2018, pp. 651–673.

[25] Danijar Hafner et al. “Learning latent dynamics for planning from pixels”. In: *International conference on machine learning*. PMLR. 2019, pp. 2555–2565.

[26] Michael Laskin, Aravind Srinivas, and Pieter Abbeel. “Curl: Contrastive unsupervised representations for reinforcement learning”. In: *International Conference on Machine Learning*. PMLR. 2020, pp. 5639–5650.

[27] Amy Zhang et al. “Learning invariant representations for reinforcement learning without reconstruction”. In: *arXiv preprint arXiv:2006.10742* (2020).

[28] Ilya Kostrikov, Denis Yarats, and Rob Fergus. “Image augmentation is all you need: Regularizing deep reinforcement learning from pixels”. In: *arXiv preprint arXiv:2004.13649* (2020).

[29] Denis Yarats et al. “Mastering visual continuous control: Improved data-augmented reinforcement learning”. In: *arXiv preprint arXiv:2107.09645* (2021).

[30] Kanishka Rao et al. “Rl-cyclegan: Reinforcement learning aware simulation-to-real”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 11157–11166.

[31] Eric Jang et al. “Bc-z: Zero-shot task generalization with robotic imitation learning”. In: *Conference on Robot Learning*. PMLR. 2022, pp. 991–1002.

[32] Yuke Zhu et al. “Reinforcement and imitation learning for diverse visuomotor skills”. In: *arXiv preprint arXiv:1802.09564* (2018).

[33] Ilija Radosavovic et al. “Real-world robot learning with masked visual pre-training”. In: *Conference on Robot Learning*. PMLR. 2023, pp. 416–426.

[34] Ilya Kostrikov et al. “Offline reinforcement learning with fisher divergence critic regularization”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 5774–5783.

[35] Alex X Lee et al. “How to spend your robot time: Bridging kickstarting and offline reinforcement learning for vision-based robotic manipulation”. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2022, pp. 2468–2475.

[36] Nico Gürtler et al. “Benchmarking offline reinforcement learning on real-robot hardware”. In: *The Eleventh International Conference on Learning Representations*. 2023.

[37] Jinxin Liu, Hongyin Zhang, and Donglin Wang. “Dara: Dynamics-aware reward augmentation in offline reinforcement learning”. In: *arXiv preprint arXiv:2203.06662* (2022).

[38] Justin Fu et al. “D4rl: Datasets for deep data-driven reinforcement learning”. In: *arXiv preprint arXiv:2004.07219* (2020).

[39] Yevgen Chebotar et al. “Actionable models: Unsupervised offline reinforcement learning of robotic skills”. In: *arXiv preprint arXiv:2104.07749* (2021).

[40] Gaoyue Zhou et al. “Real World Offline Reinforcement Learning with Realistic Data Source”. In: *arXiv preprint arXiv:2210.06479* (2022).

[41] Vitchyr H Pong et al. “Offline meta-reinforcement learning with online self-supervision”. In: *International Conference on Machine Learning*. PMLR. 2022, pp. 17811–17829.

[42] Konstantinos Dimitropoulos, Ioannis Hatzilygeroudis, and Konstantinos Chatzilygeroudis. “A brief survey of Sim2Real methods for robot learning”. In: *Advances in Service and Industrial Robotics: RAAD 2022* (2022), pp. 133–140.

[43] Joanne Truong, Sonia Chernova, and Dhruv Batra. “Bi-directional domain adaptation for sim2real transfer of embodied navigation agents”. In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 2634–2641.

[44] Manish Sahu et al. “Endo-Sim2Real: Consistency learning-based domain adaptation for instrument segmentation”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23*. Springer. 2020, pp. 784–794.

[45] Fangyi Zhang et al. “Adversarial discriminative sim-to-real transfer of visuo-motor policies”. In: *The International Journal of Robotics Research* 38.10-11 (2019), pp. 1229–1245.

[46] Konstantinos Bousmalis et al. “Unsupervised pixel-level domain adaptation with generative adversarial networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3722–3731.

[47] Ajay Tanwani. “DIRL: Domain-invariant representation learning for sim-to-real transfer”. In: *Conference on Robot Learning*. PMLR. 2021, pp. 1558–1571.

[48] Max Schwarzer et al. “Pretraining representations for data-efficient reinforcement learning”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 12686–12699.

[49] Younggyo Seo et al. “Reinforcement learning with action-free pre-training from videos”. In: *International Conference on Machine Learning*. PMLR. 2022, pp. 19561–19579.

[50] Simone Parisi et al. “The unsurprising effectiveness of pre-trained vision models for control”. In: *International Conference on Machine Learning*. PMLR. 2022, pp. 17359–17371.

[51] Danijar Hafner et al. “Dream to control: Learning behaviors by latent imagination”. In: *arXiv preprint arXiv:1912.01603* (2019).

[52] Herke Van Hoof et al. “Stable reinforcement learning with autoencoders for tactile and visual data”. In: *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE. 2016, pp. 3928–3934.

[53] Masashi Okada and Tadahiro Taniguchi. “Dreaming: Model-based reinforcement learning by latent imagination without reconstruction”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 4209–4215.

[54] Danijar Hafner et al. “Mastering atari with discrete world models”. In: *arXiv preprint arXiv:2010.02193* (2020).

[55] Danijar Hafner et al. “Mastering Diverse Domains through World Models”. In: *arXiv preprint arXiv:2301.04104* (2023).

[56] Julian Ibarz et al. “How to train your robot with deep reinforcement learning: lessons we have learned”. In: *The International Journal of Robotics Research* 40.4-5 (2021), pp. 698–721.

[57] Rahul Kidambi et al. “Morel: Model-based offline reinforcement learning”. In: *Advances in neural information processing systems* 33 (2020), pp. 21810–21823.

[58] Xingyou Song et al. “Rapidly adaptable legged robots via evolutionary meta-learning”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 3769–3776.

[59] Andrei A Rusu et al. “Sim-to-real robot learning from pixels with progressive nets”. In: *Conference on robot learning*. PMLR. 2017, pp. 262–270.

[60] Laura Smith et al. “Legged robots that keep on learning: Fine-tuning locomotion policies in the real world”. In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE. 2022, pp. 1593–1599.