Extended Abstract: An Imitation from Observation Approach to Sim-to-Real Transfer

Siddharth Desai[1§], Ishan Durugkar[1§], Haresh Karnan[1§], Garrett Warnell[2], Josiah Hanna[3], Peter Stone[14]
[1] *University of Texas at Austin,* [2] *Army Research Lab,* [3] *University of Edinburgh,* [4] *Sony AI*
sidrdesai@utexas.edu, ishand@cs.utexas.edu, haresh.miriyala@utexas.edu
warnellg@cs.utexas.edu, josiah.hanna@ed.ac.uk, pstone@cs.utexas.edu

***Abstract*—One approach to sim-to-real transfer is using interactions with the real world to make the simulator more realistic, called grounded sim-to-real transfer. In this extended abstract, we hypothesize that a particular black-box grounded sim-to-real approach, grounded action transformation, is closely related to the problem of imitation from observation (IfO): learning behaviors that mimic the observations of behavior demonstrations. We then consider that recent state-of-the-art approaches from the IfO literature can be effectively repurposed for such grounded sim-to-real transfer. To validate our hypothesis we propose a new sim-to-real transfer algorithm – generative adversarial reinforced action transformation (GARAT) – based on adversarial imitation from observation techniques. We run experiments in several simulation domains with mismatched dynamics, and find that agents trained with GARAT achieve higher returns in the real world compared to existing black-box sim-to-real methods.**

1. Introduction

Sim-to-real approaches seek to leverage inexpensive simulation experience to more efficiently learn control policies that perform well in the real world. Sim-to-real transfer has been effectively used to learn a fast humanoid walk [1], dexterous manipulation [2], and agile locomotion skills [3]. In this work, we focus on the paradigm of simulator grounding [4, 1, 5, 6, 7], which modifies a simulator's dynamics to more closely match the real world dynamics by using some real world data. Specifically, we consider grounded action transformation (GAT) [1], which grounds the simulator to the real world without needing to modify the simulator itself. We seek to improve the grounding process used in this approach for better transfer from sim to real.

We improve the process of learning the action transformation by considering it an *imitation from observation* (IfO) [8] problem, a special case of imitation learning [9]. In IfO, an imitator mimics the expert's behavior without knowing which actions the expert took, only the outcomes of those actions (i.e. state-only demonstrations). While the lack of action information presents an additional challenge, recently-proposed approaches have suggested solutions [10, 11].

§. Equal contribution

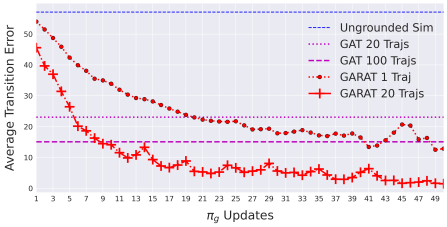

Figure 1: L2 norm of per step transition errors (lower is better) between different simulator environments and the target environment, shown over number of action transformation policy updates for GARAT.

We utilize a distribution-matching objective [12] similar to GAIL [13] and GAIfO [14], and we propose a novel algorithm, generative adversarial reinforced action transformation (GARAT), to ground the simulator by reducing the distribution mismatch between the simulator and the real world. Our experiments confirm our hypothesis by showing that GARAT reduces the difference in the dynamics between two environments more effectively than GAT. Moreover, our experiments show that, in several domains, this improved grounding translates to better transfer of policies from one environment to the other. An extended version of this paper addresses framing the action transformation as an IfO problem as well as the derivation of the GARAT algorithm.

2. GARAT

We propose GARAT to train the action transformation as a policy to minimize the distribution mismatch between the grounded simulator and the real world. Algorithm 1 lays out the steps in the GARAT algorithm.

3. Experiments

In this section, we summarize our experiments showing that GARAT leads to improved sim-to-real transfer compared to previous methods. We evaluate on various MuJoCo [15] and PyBullet [16] environments using the OpenAI Gym interface [17]. We highlight the Minitaur domain as a particularly useful test since there exist two simulators, one of which has been carefully engineered for high fidelity

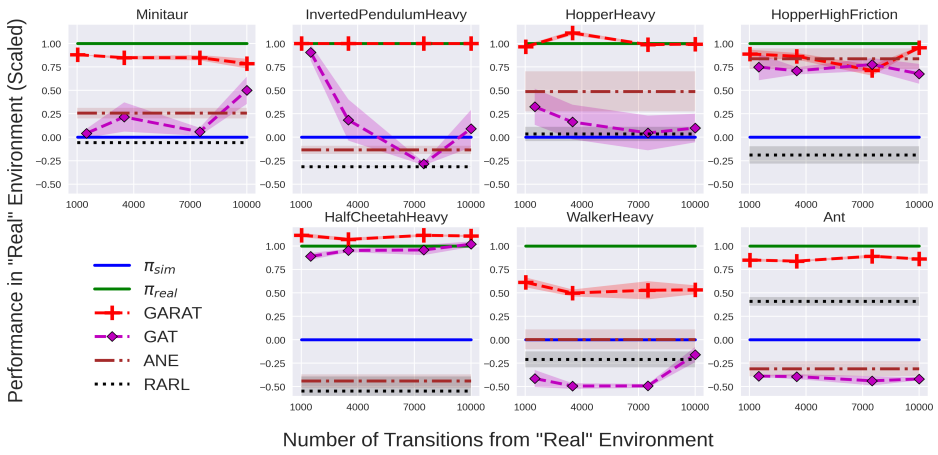

Figure 2: Performance of different techniques evaluated in "real" environment. Environment return on the y-axis is scaled such that π_{real} achieves 1 and π_{sim} achieves 0.

Algorithm 1 GARAT

Input: Agent policy π with parameters η pre-trained in simulator, Initial action transformation policy π_g with parameters θ, Initialize discriminator D_ϕ with parameters ϕ

while *performance of policy π in real world unsatisfactory* **do**

  Rollout policy π in real world to obtain trajectories $\{\tau_{real,1}, \tau_{real,2}, \ldots\}$

  for $i = 0, 1, 2, \ldots N$ **do**

    Rollout Policy π in grounded simulator and obtain trajectories $\{\tau_{gsim,1}, \tau_{gsim,2}, \ldots\}$

    Update parameters ϕ of D_ϕ using gradient descent to minimize $-(\mathbb{E}_{\tau_{gsim}}[\log(D_\phi(s,a,s'))] + \mathbb{E}_{\tau_{real}}[\log(1 - D_\phi(s,a,s'))]$

    Update parameters θ of π_g using policy gradient with reward $-[\log D_\phi(s,a,s')]$

  end

  Optimize parameters η of π in simulator grounded with action transformer π_g

end

to the real robot [18]. For other environments, the "real" environment is the simulator modified in different ways such that a policy trained in the simulator does not transfer well to the "real" environment. Apart from a thorough evaluation across multiple different domains, this sim-to-"real" setup also allows us to compare GARAT and other algorithms against a policy trained directly in the target domain with millions of interactions, which is prohibitively expensive on a real robot. This setup also allows us to thoroughly evaluate sim-to-real algorithms across multiple different domains.

In Figure 1, we evaluate how well GARAT grounds the simulator to the "real" environment both quantitatively and qualitatively. These experiments are in the *InvertedPendulum* domain, where the "real" environment has a heavier pendulum than the simulator. From the figure, it is evident that GARAT leads to a grounded simulator with lower error on average compared to GAT.

Comparison of GARAT on the actual sim-to-"real" transfer task is shown in Figure 2. The agent policy π and action transformation policy π_g are trained with TRPO [19] and PPO [20] respectively. GARAT is compared to GAT [1], RARL [21] adapted for a black-box simulator, and action-noise-envelope (ANE) [22]. π_{real} and π_{sim} denote policies trained in the "real" environment and simulator respectively until convergence. We use the best performing hyperparameters for these methods.

Figure 2 shows that, in most of the domains, GARAT with just a few thousand transitions from the "real" environment facilitates transfer of policies that perform on par with policies trained directly in the "real" environment using 1 million transitions. GARAT also consistently performs better than previous methods on most domains. The shaded envelope denotes the standard error across five experiments with different random seeds for all the methods.

In the Minitaur domain [18] as well, a policy trained only in simulation does not directly transfer well to the "real" environment [23]. In this realistic setting, we see that GARAT learns a policy that obtains more than 80% of the optimal "real" environment performance with just 1000 "real" environment transitions whereas the next best baseline (GAT) obtains at most 50% while also requiring ten times more "real" environment data.

4. Conclusion

We propose a new algorithm (GARAT) for black-box sim-to-real transfer. GARAT leads to simulator transitions that are more similar to the real world's and, in turn, leads to improved transfer of agent policies from simulation to "real" environments in our experiments.

References

[1] J. P. Hanna and P. Stone, "Grounded action transformation for robot learning in simulation," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[2] OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, and L. Zhang, "Solving rubik's cube with a robot hand," 2019.

[3] X. B. Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," *arXiv preprint arXiv:2004.00784*, 2020.

[4] A. Farchy, S. Barrett, P. MacAlpine, and P. Stone, "Humanoid robots learning to walk faster: From the real world to simulation and back," in *Proc. of 12th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS)*, May 2013.

[5] Y. Chebotar, A. Handa, V. Makoviychuk, M. Macklin, J. Issac, N. Ratliff, and D. Fox, "Closing the sim-to-real loop: Adapting simulation randomization with real world experience," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8973–8979.

[6] A. Allevato, E. S. Short, M. Pryor, and A. L. Thomaz, "Tunenet: One-shot residual tuning for system identification and sim-to-real robot task transfer," in *Conference on Robot Learning (CoRL)*, 2019.

[7] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, p. eaau5872, 2019.

[8] Y. Liu, A. Gupta, P. Abbeel, and S. Levine, "Imitation from observation: Learning to imitate behaviors from raw video via context translation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1118–1125.

[9] P. Bakker and Y. Kuniyoshi, "Robot see, robot do: An overview of robot imitation," in *AISB96 Workshop on Learning in Robots and Animals*, 1996, pp. 3–11.

[10] F. Torabi, G. Warnell, and P. Stone, "Behavioral cloning from observation," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 4950–4957.

[11] F. Torabi, G. Warnell, and P. Stone, "Recent advances in imitation learning from observation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, Aug 2019.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[13] J. Ho and S. Ermon, "Generative adversarial imitation learning," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 4565–4573. [Online]. Available: http://papers.nips.cc/paper/6391-generative-adversarial-imitation-learning.pdf

[14] F. Torabi, G. Warnell, and P. Stone, "Generative adversarial imitation from observation," *arXiv preprint arXiv:1807.06158*, 2018.

[15] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5026–5033.

[16] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," *GitHub repository*, 2016.

[17] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[18] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," *CoRR*, vol. abs/1804.10332, 2018. [Online]. Available: http://arxiv.org/abs/1804.10332

[19] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel, "Trust region policy optimization," *CoRR*, vol. abs/1502.05477, 2015. [Online]. Available: http://arxiv.org/abs/1502.05477

[20] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[21] L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, "Robust adversarial reinforcement learning," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 2817–2826.

[22] N. Jakobi, P. Husbands, and I. Harvey, "Noise and the reality gap: The use of simulation in evolutionary robotics," in *Advances in Artificial Life*, F. Morán, A. Moreno, J. J. Merelo, and P. Chacón, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995, pp. 704–720.

[23] W. Yu, C. K. Liu, and G. Turk, "Policy transfer with strategy optimization," *CoRR*, vol. abs/1810.05751, 2018. [Online]. Available: http://arxiv.org/abs/1810.05751