

REFERENCES

- [1] Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. 2013. The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research* 47 (2013), 253–279.
- [2] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. *arXiv:1606.01540* (2016).
- [3] Thomas G Dietterich. 2000. Hierarchical Reinforcement Learning with the MAXQ Value Function Decomposition. *Journal of Artificial Intelligence Research* 13 (2000), 227–303.
- [4] William Fedus, Prajit Ramachandran, Rishabh Agarwal, Yoshua Bengio, Hugo Larochelle, Mark Rowland, and Will Dabney. 2020. Revisiting Fundamentals of Experience Replay. *arXiv:2007.06700* (2020).
- [5] Scott Fujimoto, Herke Van Hoof, and David Meger. 2018. Addressing Function Approximation Error in Actor-Critic Methods. *arXiv:1802.09477* (2018).
- [6] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. *arXiv:1801.01290* (2018).
- [7] Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980* (2014).
- [8] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous Control with Deep Reinforcement Learning. *arXiv:1509.02971* (2015).
- [9] Long-Ji Lin. 1992. Self-Improving Reactive Agents Based on Reinforcement Learning, Planning and Teaching. *Machine Learning* 8, 3-4 (1992), 293–321.
- [10] Ruishan Liu and James Zou. 2018. The Effects of Memory Replay in Reinforcement Learning. In *Allerton Conference on Communication, Control, and Computing*. IEEE, 478–485.
- [11] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-Level Control through Deep Reinforcement Learning. *Nature* 518, 7540 (2015), 529–533.
- [12] Georg Ostrovski, Marc G Bellemare, Aaron Oord, and Rémi Munos. 2017. Count-Based Exploration with Neural Density Models. In *International Conference on Machine Learning*. PMLR, 2721–2730.
- [13] Anil Sharma, Mayank K Pal, Saket Anand, and Sanjit K Kaul. 2020. Stratified Sampling Based Experience Replay for Efficient Camera Selection Decisions. In *IEEE International Conference on Multimedia Big Data*. IEEE, 144–151.
- [14] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement Learning: An Introduction*. MIT Press.
- [15] Ziyu Wang, Victor Bapst, Nicolas Heess, Volodymyr Mnih, Remi Munos, Koray Kavukcuoglu, and Nando de Freitas. 2016. Sample Efficient Actor-Critic with Experience Replay. *arXiv:1611.01224* (2016).
- [16] Christopher John Cornish Hellaby Watkins. 1989. *Learning from Delayed Rewards*. Ph.D. Dissertation. King’s College.
- [17] Shangdong Zhang and Richard S Sutton. 2017. A Deeper Look at Experience Replay. *arXiv:1712.01275* (2017).