

REFERENCES

- [1] Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. 2012. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games* 4, 1 (2012), 1–43.
- [2] S. Calinon, F. D'halluin, E. L. Sauser, D. G. Caldwell, and A. G. Billard. 2010. Learning and Reproduction of Gestures by Imitation. *IEEE Robotics Automation Magazine* 17, 2 (June 2010), 44–54. <https://doi.org/10.1109/MRA.2010.936947>
- [3] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980 <http://arxiv.org/abs/1412.6980>
- [4] Levente Kocsis and Csaba Szepesvári. 2006. Bandit based monte-carlo planning. *Machine learning: ECML 2006* (2006), 282–293.
- [5] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. 2016. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research* 17, 39 (2016), 1–40.
- [6] Sergey Levine, Peter Pastor, Alex Krizhevsky, Julian Ibarz, and Deirdre Quillen. 2016. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research* (2016), 0278364917710318.
- [7] 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 preprint arXiv:1509.02971* (2015).
- [8] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning*. 1928–1937.
- [9] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fiedjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (26 02 2015), 529–533. <http://dx.doi.org/10.1038/nature14236>
- [10] F. Riccio, R. Capobianco, and D. Nardi. 2018. DOP: Deep Optimistic Planning with Approximate Value Function Evaluation. *ArXiv e-prints* (mar 2018). arXiv:cs.RO/1803.08501
- [11] Francesco Riccio, Roberto Capobianco, and Daniele Nardi. 2018. Q-CP: Learning Action Values for Cooperative Planning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (2018-01-30). –.
- [12] Stephane Ross and J Andrew Bagnell. 2014. Reinforcement and imitation learning via interactive no-regret learning. *arXiv preprint arXiv:1406.5979* (2014).
- [13] Stéphane Ross, Geoffrey J Gordon, and Drew Bagnell. 2011. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. In *International Conference on Artificial Intelligence and Statistics*. 627–635.
- [14] Andrei A Rusu, Matej Vecerik, Thomas Rothörl, Nicolas Heess, Razvan Pascanu, and Raia Hadsell. 2016. Sim-to-real robot learning from pixels with progressive nets. *arXiv preprint arXiv:1610.04286* (2016).
- [15] David Silver, Richard S Sutton, and Martin Müller. 2012. Temporal-difference search in computer Go. *Machine learning* 87, 2 (2012), 183–219.
- [16] Matej Večerik, Todd Hester, Jonathan Scholz, Fumin Wang, Olivier Pietquin, Bilal Piot, Nicolas Heess, Thomas Rothörl, Thomas Lampe, and Martin Riedmiller. 2017. Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards. *arXiv preprint arXiv:1707.08817* (2017).
- [17] Mark H. Winands, Yngvi Björnsson, and Jahn-Takeshi Saito. 2008. Monte-Carlo Tree Search Solver. In *Proceedings of the 6th International Conference on Computers and Games (CG '08)*. Springer-Verlag, Berlin, Heidelberg, 25–36. https://doi.org/10.1007/978-3-540-87608-3_3