

References

- [1] S. Stahlke, A. Nova, and P. Mirza-Babaei, *Artificial Players in the Design Process: Developing an Automated Testing Tool for Game Level and World Design*. New York, NY, USA: Association for Computing Machinery, 2020, p. 267–280. ISBN 9781450380744. [Online]. Available: <https://doi.org/10.1145/3410404.3414249>
- [2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018. ISBN 0262039249
- [3] O. Canal, “Transfer and multi-task reinforcement learning,” <https://campusai.github.io/>, 2020. [Online]. Available: https://campusai.github.io/rl/transfer_and_multitask_rl
- [4] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of Go with deep neural networks and tree search,” *Nature*, vol. 529, no. 7587, pp. 484–489, January 2016. doi: 10.1038/nature16961
- [5] A. P. Badia, B. Piot, S. Kapturowski, P. Sprechmann, A. Vitvitskyi, D. Guo, and C. Blundell, “Agent57: Outperforming the atari human benchmark,” 2020.
- [6] S. Levine, “Deep reinforcement learning course,” <http://rail.eecs.berkeley.edu/deeprlcourse/>, accessed: 2021-04-30.
- [7] M. Botvinick, S. Ritter, J. X. Wang, Z. Kurth-Nelson, C. Blundell, and D. Hassabis, “Reinforcement Learning, Fast and Slow,” pp. 408–422, 2019. [Online]. Available: <http://creativecommons.org/licenses/by/4.0/>

- [8] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” 2018.
- [9] R. Hadsell, D. Rao, A. A. Rusu, and R. Pascanu, “Embracing Change: Continual Learning in Deep Neural Networks,” pp. 1028–1040, dec 2020. [Online]. Available: <https://doi.org/10.1016/j.tics.2020.09.004>
- [10] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, “Improved protein structure prediction using potentials from deep learning,” *Nature*, vol. 577, no. 7792, pp. 706–710, jan 2020. doi: 10.1038/s41586-019-1923-7. [Online]. Available: <https://www.nature.com/articles/s41586-019-1923-7>
- [11] OpenAI, “Openai five,” <https://blog.openai.com/openai-five/>, 2018.
- [12] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, A. S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, J. Quan, S. Gaffney, S. Petersen, K. Simonyan, T. Schaul, H. van Hasselt, D. Silver, T. Lillicrap, K. Calderone, P. Keet, A. Brunasso, D. Lawrence, A. Ekermo, J. Repp, and R. Tsing, “Starcraft ii: A new challenge for reinforcement learning,” 2017.
- [13] A. P. Badia, B. Piot, S. Kapturowski, P. Sprechmann, A. Vitvitskyi, D. Guo, and C. Blundell, “Agent57: Outperforming the atari human benchmark,” 2020.
- [14] Y. Zhao, I. Borovikov, F. de Mesentier Silva, A. Beirami, J. Rupert, C. Somers, J. Harder, J. Kolen, J. Pinto, R. Pourabolghasem, J. Pestrak, H. Chaput, M. Sardari, L. Lin, S. Narravula, N. Aghdaie, and K. Zaman, “Winning isn’t everything: Enhancing game development with intelligent agents,” 2020.
- [15] “Unity,” <https://unity.com/>, accessed: 2021.
- [16] C. Holmgård, M. C. Green, A. Liapis, and J. Togelius, “Automated playtesting with procedural personas through mcts with evolved heuristics,” 2018.

- [17] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [18] B. Baker, I. Kanitscheider, T. Markov, Y. Wu, G. Powell, B. McGrew, and I. Mordatch, “Emergent tool use from multi-agent autocurricula,” 2020.
- [19] M. Jaderberg, W. M. Czarnecki, I. Dunning, L. Marris, G. Lever, A. G. Castañeda, C. Beattie, N. C. Rabinowitz, A. S. Morcos, A. Ruderman, and et al., “Human-level performance in 3d multiplayer games with population-based reinforcement learning,” *Science*, vol. 364, no. 6443, p. 859–865, May 2019. doi: 10.1126/science.aau6249. [Online]. Available: <http://dx.doi.org/10.1126/science.aau6249>
- [20] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, “Mastering chess and shogi by self-play with a general reinforcement learning algorithm,” 2017.
- [21] T. M. Moerland, J. Broekens, and C. M. Jonker, “Model-based reinforcement learning: A survey,” 2021.
- [22] M. Janner, J. Fu, M. Zhang, and S. Levine, “When to trust your model: Model-based policy optimization,” 2019.
- [23] M. Janner, I. Mordatch, and S. Levine, “ γ -models: Generative temporal difference learning for infinite-horizon prediction,” *CoRR*, vol. abs/2010.14496, 2020. [Online]. Available: <https://arxiv.org/abs/2010.14496>
- [24] P. Dayan, “Improving Generalization for Temporal Difference Learning: The Successor Representation,” Tech. Rep. 4, 1993.
- [25] A. Barreto, W. Dabney, R. Munos, J. J. Hunt, T. Schaul, H. van Hasselt, and D. Silver, “Successor features for transfer in reinforcement learning,” 2018.
- [26] A. Barreto, D. Borsa, J. Quan, T. Schaul, D. Silver, M. Hessel, D. Mankowitz, A. Židek, and R. Munos, “Transfer in deep reinforcement learning using successor features and generalised policy improvement,” 2019.
- [27] A. Barreto, S. Hou, D. Borsa, D. Silver, and D. Precup, “Fast reinforcement learning with generalized policy updates,” *Proceedings*

- of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 079–30 087, 2020. doi: 10.1073/pnas.1907370117. [Online]. Available: <https://www.pnas.org/content/117/48/30079>
- [28] Z. Zhu, K. Lin, and J. Zhou, “Transfer learning in deep reinforcement learning: A survey,” 2021.
 - [29] S. Levine, C. Finn, T. Darrell, and P. Abbeel, “End-to-end training of deep visuomotor policies,” 2016.
 - [30] E. Shelhamer, P. Mahmoudieh, M. Argus, and T. Darrell, “Loss is its own reward: Self-supervision for reinforcement learning,” 2017.
 - [31] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, “Sim-to-real transfer of robotic control with dynamics randomization,” *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018. doi: 10.1109/icra.2018.8460528. [Online]. Available: <http://dx.doi.org/10.1109/ICRA.2018.8460528>
 - [32] T. Haarnoja, H. Tang, P. Abbeel, and S. Levine, “Reinforcement learning with deep energy-based policies,” 2017.
 - [33] I. Higgins, A. Pal, A. A. Rusu, L. Matthey, C. P. Burgess, A. Pritzel, M. Botvinick, C. Blundell, and A. Lerchner, “Darla: Improving zero-shot transfer in reinforcement learning,” 2018.
 - [34] A. Rajeswaran, S. Ghotra, B. Ravindran, and S. Levine, “Epopt: Learning robust neural network policies using model ensembles,” 2017.
 - [35] W. Yu, J. Tan, C. K. Liu, and G. Turk, “Preparing for the unknown: Learning a universal policy with online system identification,” 2017.
 - [36] F. Sadeghi and S. Levine, “Cad2rl: Real single-image flight without a single real image,” 2017.
 - [37] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” vol. 60, 01 2009. doi: 10.1145/1553374.1553380 p. 6.
 - [38] E. Parisotto, J. L. Ba, and R. Salakhutdinov, “Actor-mimic: Deep multitask and transfer reinforcement learning,” 2016.
 - [39] T. Schaul, D. Horgan, K. Gregor, and D. Silver, “Universal value function approximators,” in *Proceedings of the 32nd International*

- Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 1312–1320. [Online]. Available: <http://proceedings.mlr.press/v37/schaul15.html>
- [40] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
 - [41] C. Ma, D. R. Ashley, J. Wen, and Y. Bengio, “Universal successor features for transfer reinforcement learning,” 2020.
 - [42] D. Borsa, A. Barreto, J. Quan, D. Mankowitz, R. Munos, H. van Hasselt, D. Silver, and T. Schaul, “Universal successor features approximators,” 2018.
 - [43] J. Andreas, M. Rohrbach, T. Darrell, and D. Klein, “Neural module networks,” 2017.
 - [44] C. Devin, A. Gupta, T. Darrell, P. Abbeel, and S. Levine, “Learning modular neural network policies for multi-task and multi-robot transfer,” 2016.
 - [45] O. Canal, “Meta-reinforcement learning basics,” <https://campusai.github.io/>, 2021. [Online]. Available: <https://campusai.github.io/rl/meta-rl>
 - [46] J. X. Wang, Z. Kurth-Nelson, D. Tirumala, H. Soyer, J. Z. Leibo, R. Munos, C. Blundell, D. Kumaran, and M. Botvinick, “Learning to reinforcement learn,” 2017.
 - [47] N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel, “A simple neural attentive meta-learner,” 2018.
 - [48] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, “Benchmarking deep reinforcement learning for continuous control,” 2016.
 - [49] K. Rakelly, A. Zhou, D. Quillen, C. Finn, and S. Levine, “Efficient off-policy meta-reinforcement learning via probabilistic context variables,” 2019.
 - [50] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” 2017.

- [51] J. Foerster, G. Farquhar, M. Al-Shedivat, T. Rocktäschel, E. P. Xing, and S. Whiteson, “Dice: The infinitely differentiable monte-carlo estimator,” 2018.
- [52] J. Rothfuss, D. Lee, I. Clavera, T. Asfour, and P. Abbeel, “Promp: Proximal meta-policy search,” 2018.
- [53] A. Ghadirzadeh, X. Chen, P. Poklukar, C. Finn, M. Björkman, and D. Kragic, “Bayesian meta-learning for few-shot policy adaptation across robotic platforms.” [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-285972>
- [54] O. Canal and F. Taschin, “From expectation maximization to variational inference,” <https://campusai.github.io/>, 2020. [Online]. Available: https://campusai.github.io/ml/variational_inference
- [55] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN 1441412697
- [56] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, “Openai gym,” 2016.
- [57] A. Juliani, V.-P. Berges, E. Vckay, Y. Gao, H. Henry, M. Mattar, and D. Lange, “Unity: A General Platform for Intelligent Agents,” sep 2018. [Online]. Available: <http://arxiv.org/abs/1809.02627>
- [58] “The Shortest Path Through a Maze - Edward F. Moore - Google Books.” [Online]. Available: https://books.google.com/books/about/The_Shortest_Path_Through_a_Maze.html?id=IVZBHAAACAAJ
- [59] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba, “Hindsight experience replay,” 2018.
- [60] B. van Niekerk, S. James, A. Earle, and B. Rosman, “Will it blend? composing value functions in reinforcement learning,” 2018.
- [61] J. Hunt, A. Barreto, T. Lillicrap, and N. Heess, “Composing entropic policies using divergence correction,” in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 2911–2920. [Online]. Available: <http://proceedings.mlr.press/v97/hunt19a.html>

- [62] T. Haarnoja, V. Pong, A. Zhou, M. Dalal, P. Abbeel, and S. Levine, “Composable deep reinforcement learning for robotic manipulation,” 2018.
- [63] H. Hasselt, “Double Q-learning,” in *Advances in Neural Information Processing Systems*, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds., vol. 23. Curran Associates, Inc., 2010. [Online]. Available: <https://proceedings.neurips.cc/paper/2010/file/091d584fced301b442654dd8c23b3fc9-Paper.pdf>
- [64] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, “Prioritized experience replay,” 2016.
- [65] M. Hessel, J. Modayil, H. van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver, “Rainbow: Combining improvements in deep reinforcement learning,” 2017.
- [66] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems* 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [67] R. Munos, T. Stepleton, A. Harutyunyan, and M. G. Bellemare, “Safe and efficient off-policy reinforcement learning,” 2016.
- [68] Z. Wang, V. Bapst, N. Heess, V. Mnih, R. Munos, K. Kavukcuoglu, and N. de Freitas, “Sample efficient actor-critic with experience replay,” 2017.
- [69] F. Taschin, “Acer-torch implementation,” 2021. [Online]. Available: <https://github.com/fedetask/ACER-torch/>
- [70] S. Kapturowski, G. Ostrovski, J. Quan, R. Munos, and W. Dabney, “Recurrent experience replay in distributed reinforcement learning,” in *7th International Conference on Learning Representations, ICLR 2019*, 2019.

- [71] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.
- [72] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," 2020.
- [73] "Deepmind ai reduces google data centre cooling bill by 40%," <https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40>, accessed: 2021-05-28.

For DIVA

```
{
  "Author1": {
    "Last name": "Canal Anton",
    "First name": "Oleguer",
    "Local User Id": "oleguer@kth.se",
    "E-mail": "oleguer@kth.se",
    "organisation": {"L1": "School of Electrical Engineering and Computer Science ",
                    }
  },
  "Degree": {"Educational program": "Master's Programme, Machine Learning, 120 credits"},
  "Title": {
    "Main title": "Automatic game-testing with personality",
    "Subtitle": "Multi-task reinforcement learning for automatic game-testing",
    "Language": "eng" },
  "Alternative title": {
    "Main title": "Automatisk speltestning med personlighet",
    "Subtitle": "Multi-task förstärkning lärande för automatisk speltestning",
    "Language": "swe"
  },
  "Supervisor1": {
    "Last name": "Ghadirzadeh",
    "First name": "Ali",
    "E-mail": "algh@kth.se",
    "organisation": {"L1": "School of Electrical Engineering and Computer Science ",
                    "L2": "DIVISION OF ROBOTICS, PERCEPTION AND LEARNING" }
  },
  "Supervisor2": {
    "Last name": "Gordillo",
    "First name": "Camilo",
    "E-mail": "cgordillo@ea.com",
  },
  "Supervisor2": {
    "Last name": "Bergdahl",
    "First name": "Joakim",
    "E-mail": "jbergdahl@ea.com",
  },
  "Examiner1": {
    "Last name": "Wahlberg",
    "First name": "Bo",
    "E-mail": "bo@kth.se",
  },
  "Cooperation": { "Partner_name": "SEED, Electronic Arts"},
  "Other information": {
    "Year": "2021", "Number of pages": "xii,59"
  }
}
```