References

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018. ISBN 0262039249

[2] K. O. Stanley and R. Miikkulainen, "Evolving neural networks through augmenting topologies," *Evol. Comput.*, vol. 10, no. 2, p. 99–127, Jun. 2002. doi: 10.1162/106365602320169811. [Online]. Available: https://doi.org/10.1162/106365602320169811

[3] F. Vannella, G. Iakovidis, E. A. Hakim, E. Aumayr, and S. Feghhi, "Remote electrical tilt optimization via safe reinforcement learning," 2021.

[4] H. Proshansky and G. Murphy, "The effects of reward and punishment on perception." *The Journal of Psychology: Interdisciplinary and Applied*, vol. 13, pp. 295–305, 1942. doi: 10.1080/00223980.1942.9917097. [Online]. Available: https://doi.org/10.1080/00223980.1942.9917097

[5] K. Barr, *ASIC Design in the Silicon Sandbox: A Complete Guide to Building Mixed-Signal Integrated Circuits*. New York: McGraw-Hill Education, 2007. ISBN 9780071481618. [Online]. Available: https://www.accessengineeringlibrary.com/content/book/9780071481618

[6] A. Nikou, A. Mujumdar, M. Orlic, and A. V. Feljan, "Symbolic reinforcement larning for safe ran control," *Internaltinal Conference of Autonomous Agents and Multi Agent systems (AAMAS) (to appear)*, 2021.

[7] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019. doi: 10.1109/COMST.2019.2916583

[8] V. Yajnanarayana, H. Ryden, and L. Hevizi, "5g handover using reinforcement learning," *2020 IEEE 3rd 5G World Forum (5GWF)*,

Sep 2020. doi: 10.1109/5gwf49715.2020.9221072. [Online]. Available: http://dx.doi.org/10.1109/5GWF49715.2020.9221072

[9] M. Kavalerov, Y. Likhacheva, and Y. Shilova, "A reinforcement learning approach to network routing based on adaptive learning rates and route memory," in *SoutheastCon 2017*, 2017. doi: 10.1109/SECON.2017.7925316 pp. 1–6.

[10] D. Silver, A. Huang, C. Maddison, A. Guez, L. Sifre, G. 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, pp. 484–489, 01 2016. doi: 10.1038/nature16961

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness, M. Bellemare, A. Graves, M. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–33, 02 2015. doi: 10.1038/nature14236

[12] M. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, "The arcade learning environment: An evaluation platform for general agents," *Journal of Artificial Intelligence Research*, vol. 47, 07 2012. doi: 10.1613/jair.3912

[13] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," 2016.

[14] D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke, and S. Levine, "Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation," 2018.

[15] P. Abbeel, A. Coates, M. Quigley, and A. Ng, "An application of reinforcement learning to aerobatic helicopter flight." 01 2006, pp. 1–8.

[16] G. Zheng, F. Zhang, Z. Zheng, Y. Xiang, N. Yuan, X. Xie, and Z. Li, "Drn: A deep reinforcement learning framework for news recommendation," 04 2018. ISBN 978-1-4503-5639-8 pp. 167–176.

[17] I. Munemasa, Y. Tomomatsu, K. Hayashi, and T. Takagi, "Deep reinforcement learning for recommender systems," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018. doi: 10.1109/ICOIACT.2018.8350761 pp. 226–233.

[18] 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.

[19] T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *CoRR*, 09 2015.

[20] H. v. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, ser. AAAI'16. AAAI Press, 2016, p. 2094–2100.

[21] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, "Dueling network architectures for deep reinforcement learning," in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ser. ICML'16. JMLR.org, 2016, p. 1995–2003.

[22] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1928–1937. [Online]. Available: http://proceedings.mlr.press/v48/mniha16.html

[23] H. Mania, A. Guy, and B. Recht, "Simple random search provides a competitive approach to reinforcement learning," 2018.

[24] Y. Yu, "Towards sample efficient reinforcement learning," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, ser. IJCAI'18. AAAI Press, 2018. ISBN 9780999241127 p. 5739–5743.

[25] T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever, "Evolution strategies as a scalable alternative to reinforcement learning," 2017.

[26] F. P. Such, V. Madhavan, E. Conti, J. Lehman, K. O. Stanley, and J. Clune, "Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning," 2018.

[27] E. Conti, V. Madhavan, F. P. Such, J. Lehman, K. O. Stanley, and J. Clune, "Improving exploration in evolution strategies for deep reinforcement learning via a population of novelty-seeking agents," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS'18. Red Hook, NY, USA: Curran Associates Inc., 2018, p. 5032–5043.

[28] H. Liu, K. Simonyan, O. Vinyals, C. Fernando, and K. Kavukcuoglu, "Hierarchical representations for efficient architecture search," 2018.

[29] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014.

[30] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," 2016.

[31] E. Strubell, A. Ganesh, and A. Mccallum, "Energy and policy considerations for deep learning in nlp," 01 2019. doi: 10.18653/v1/P19-1355 pp. 3645–3650.

[32] N. Jouppi, A. Borchers, R. Boyle, P.-l. Cantin, C. Chao, C. Clark, J. Coriell, M. Daley, M. Dau, J. Dean, B. Gelb, C. Young, T. Ghaemmaghami, R. Gottipati, W. Gulland, R. Hagmann, C. Ho, D. Hogberg, J. Hu, and N. Boden, "In-datacenter performance analysis of a tensor processing unit," *ACM SIGARCH Computer Architecture News*, vol. 45, pp. 1–12, 06 2017. doi: 10.1145/3140659.3080246

[33] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed. USA: John Wiley ; Sons, Inc., 1994. ISBN 0471619779

[34] C. Boutilier, T. Dean, and S. Hanks, "Decision-theoretic planning: Structural assumptions and computational leverage," *The Journal of Artificial Intelligence Research (JAIR)*, vol. 11, 07 1999. doi: 10.1613/jair.575

[35] A. A. Markov, “Rasprostranenie zakona bol’shih chisel na velichiny, zavisyaschie drug ot druga,” *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, vol. 15, no. 135-156, p. 18, 1906.

[36] ——, “An example of statistical investigation of the text eugene onegin concerning the connection of samples in chains,” *Science in Context*, vol. 19, no. 4, p. 591–600, 2006. doi: 10.1017/S0269889706001074

[37] R. Bellman, “Dynamic programming and stochastic control processes,” *Information and Control*, vol. 1, no. 3, pp. 228–239, 1958. doi: https://doi.org/10.1016/S0019-9958(58)80003-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019995858800030

[38] R. E. Bellman, *The Theory of Dynamic Programming*. Santa Monica, CA: RAND Corporation, 1954.

[39] L. Bottou, “Online learning and stochastic approximations,” 1998.

[40] F. B. Fitch, “Warren s. mcculloch and walter pitts. a logical calculus of the ideas immanent in nervous activity. bulletin of mathematical biophysics, vol. 5 (1943), pp. 115–133.” *Journal of Symbolic Logic*, vol. 9, no. 2, p. 49–50, 1944. doi: 10.2307/2268029

[41] S. Herculano-Houzel, “The human brain in numbers: a linearly scaled-up primate brain,” *Frontiers in human neuroscience*, vol. 3, pp. 31–31, Nov 2009. doi: 10.3389/neuro.09.031.2009 19915731[pmid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19915731

[42] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[43] D. Rumelhart, G. Hinton, and J. Mcclelland, “A general framework for parallel distributed processing,” *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, 01 1986.

[44] L. B. Rall, “Automatic differentiation: Techniques and applications,” in *Lecture Notes in Computer Science*, 1981.

[45] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[46] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314, 1989.

[47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. Red Hook, NY, USA: Curran Associates Inc., 2012, p. 1097–1105.

[48] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wang, "Image and video compression with neural networks: A review," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, p. 1683–1698, Jun 2020. doi: 10.1109/tcsvt.2019.2910119. [Online]. Available: http://dx.doi.org/10.1109/TCSVT.2019.2910119

[49] H. Purwins, B. Li, T. Virtanen, J. Schluter, S.-Y. Chang, and T. Sainath, "Deep learning for audio signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, p. 206–219, May 2019. doi: 10.1109/jstsp.2019.2908700. [Online]. Available: http://dx.doi.org/10.1109/JSTSP.2019.2908700

[50] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," 2019.

[51] C. Sousa, "An overview on weight initialization methods for feedforward neural networks," 07 2016. doi: 10.1109/IJCNN.2016.7727180

[52] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Journal of Machine Learning Research - Proceedings Track*, vol. 9, pp. 249–256, 01 2010.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *IEEE International Conference on Computer Vision (ICCV 2015)*, vol. 1502, 02 2015. doi: 10.1109/ICCV.2015.123

[54] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," 2020.

[55] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. The MIT Press, 04 1992. ISBN 9780262275552. [Online]. Available: https://doi.org/10.7551/mitpress/1090.001.0001

[56] S. Haupt and R. Haupt, "1.1 a genetic algorithms and their applications in environmental sciences," 01 2003.

[57] S. Baluja and R. Caruana, "Removing the genetics from the standard genetic algorithm," USA, Tech. Rep., 1995.

[58] P. Larranaga, C. Kuijpers, R. Murga, I. Inza, and S. Dizdarevic, "Genetic algorithms for the travelling salesman problem: A review of representations and operators," *Artificial Intelligence Review*, vol. 13, pp. 129–170, 01 1999. doi: 10.1023/A:1006529012972

[59] P. Larranaga, C. Kuijpers, R. Murga, and Y. Yurramendi, "Learning bayesian network structures by searching for the best ordering with genetic algorithms," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 26, no. 4, pp. 487–493, 1996. doi: 10.1109/3468.508827

[60] C. C. Da Ronco and E. Benini, *A Simplex-Crossover-Based Multi-Objective Evolutionary Algorithm*. Dordrecht: Springer Netherlands, 2014, pp. 583–598. ISBN 978-94-007-6818-5. [Online]. Available: https://doi.org/10.1007/978-94-007-6818-5_41

[61] N. Hansen and A. Auger, "Cma-es: Evolution strategies and covariance matrix adaptation," in *Proceedings of the 13th Annual Conference Companion on Genetic and Evolutionary Computation*, ser. GECCO '11. New York, NY, USA: Association for Computing Machinery, 2011. doi: 10.1145/2001858.2002123. ISBN 9781450306904 p. 991–1010. [Online]. Available: https://doi.org/10.1145/2001858.2002123

[62] D. Wierstra, T. Schaul, J. Peters, and J. Schmidhuber, "Natural evolution strategies," in *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, 2008. doi: 10.1109/CEC.2008.4631255 pp. 3381–3387.

[63] K. Stanley, "Compositional pattern producing networks: A novel abstraction of development," *Genetic Programming and Evolvable Machines*, vol. 8, pp. 131–162, 06 2007. doi: 10.1007/s10710-007-9028-8

[64] K. Stanley, D. D'Ambrosio, and J. Gauci, "A hypercube-based encoding for evolving large-scale neural networks," *Artificial life*, vol. 15, pp. 185–212, 02 2009. doi: 10.1162/artl.2009.15.2.15202

[65] F. S. Melo, S. P. Meyn, and M. I. Ribeiro, "An analysis of reinforcement learning with function approximation," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New

York, NY, USA: Association for Computing Machinery, 2008. doi: 10.1145/1390156.1390240. ISBN 9781605582054 p. 664–671. [Online]. Available: https://doi.org/10.1145/1390156.1390240

[66] H. Van Hasselt, "Reinforcement learning in continuous state and action spaces," 08 2013. doi: 10.1007/978-3-642-27645-37

[67] 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

[68] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," 11 2016.

[69] Y. Dong and X. Zou, "Mobile robot path planning based on improved ddpg reinforcement learning algorithm," in *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*, 2020. doi: 10.1109/ICSESS49938.2020.9237641 pp. 52–56.

[70] A. Nair, P. Srinivasan, S. Blackwell, C. Alcicek, R. Fearon, A. De Maria, V. Panneershelvam, M. Suleyman, C. Beattie, S. Petersen, S. Legg, V. Mnih, K. Kavukcuoglu, and D. Silver, "Massively parallel methods for deep reinforcement learning," 07 2015.

[71] Y. Tassa, T. Erez, and E. Todorov, "Synthesis and stabilization of complex behaviors through online trajectory optimization," 10 2012. doi: 10.1109/IROS.2012.6386025. ISBN 978-1-4673-1737-5 pp. 4906–4913.

[72] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, pp. 503–507, 05 2015. doi: 10.1038/nature14422

[73] J. Pugh, L. Soros, and K. Stanley, "Quality diversity: A new frontier for evolutionary computation," *Frontiers in Robotics and AI*, vol. 3, 07 2016. doi: 10.3389/frobt.2016.00040

[74] J. Lehman and K. Stanley, "Exploiting open-endedness to solve problems through the search for novelty," in *ALIFE*, 2008.

[75] M. Jaderberg, V. Dalibard, S. Osindero, W. M. Czarnecki, J. Donahue, A. Razavi, O. Vinyals, T. Green, I. Dunning, K. Simonyan, C. Fernando,

and K. Kavukcuoglu, "Population based training of neural networks," 2017.

[76] H. Liu, K. Simonyan, O. Vinyals, C. Fernando, and K. Kavukcuoglu, "Hierarchical Representations for Efficient Architecture Search," *arXiv e-prints*, p. arXiv:1711.00436, Nov. 2017.

[77] T. Gangwani and J. Peng, "Policy optimization by genetic distillation," 2018.

[78] S. Khadka and K. Tumer, "Evolution-guided policy gradient in reinforcement learning," 2018.

[79] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," 06 2016.

[80] M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, "The arcade learning environment: An evaluation platform for general agents," *J. Artif. Int. Res.*, vol. 47, no. 1, p. 253–279, May 2013.

[81] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities," in *Proceedings of the April 18-20, 1967, Spring Joint Computer Conference*, ser. AFIPS '67 (Spring). New York, NY, USA: Association for Computing Machinery, 1967. doi: 10.1145/1465482.1465560. ISBN 9781450378956 p. 483–485. [Online]. Available: https://doi.org/10.1145/1465482.1465560

[82] M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, "The arcade learning environment: An evaluation platform for general agents," *Journal of Artificial Intelligence Research*, vol. 47, p. 253–279, Jun 2013. doi: 10.1613/jair.3912. [Online]. Available: http://dx.doi.org/10.1613/jair.3912

[83] G. Neglia, G. Calbi, D. Towsley, and G. Vardoyan, "The role of network topology for distributed machine learning," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019. doi: 10.1109/INFOCOM.2019.8737602 pp. 2350–2358.

[84] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–73, 2021/05/18/ 1992, full publication date: JULY 1992. [Online]. Available: http://www.jstor.org/stable/24939139

[85] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[86] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, Nov. 1997. doi: 10.1162/neco.1997.9.8.1735. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[87] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014.

[88] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[89] H. Robbins and S. Monro, "A Stochastic Approximation Method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400 – 407, 1951. doi: 10.1214/aoms/1177729586. [Online]. Available: https://doi.org/10.1214/aoms/1177729586

Appendix A

Supplementary Information

A.1 Additional Technical Information

This section contains some additional technical information about the thesis. Table A.1 contains the technical specifications of the Atari 2600 games utilized in the thesis. Table A.2 provides the hyperparameter used for training DQN on RET simulator. Table A.3 gives the technical details of Experiment 4.

Game	Size per frame (state space)	Number of actions (action space)	Reward range
amidar	(250, 160, 3)	10	$(-\infty, \infty)$
assault	(250, 160, 3)	7	$(-\infty, \infty)$
asterix	(210, 160, 3)	9	$(-\infty, \infty)$
asteroids	(210, 160, 3)	14	$(-\infty, \infty)$
atlantis	(210, 160, 3)	4	$(-\infty, \infty)$
enduro	(210, 160, 3)	9	$(-\infty, \infty)$
frostbite	(210, 160, 3)	18	$(-\infty, \infty)$
gravitar	(210, 160, 3)	18	$(-\infty, \infty)$
kangaroo	(210, 160, 3)	18	$(-\infty, \infty)$
seaquest	(210, 160, 3)	18	$(-\infty, \infty)$
skiing	(250, 160, 3)	3	$(-\infty, \infty)$
venture	(210, 160, 3)	18	$(-\infty, \infty)$
zaxxon	(210, 160, 3)	18	$(-\infty, \infty)$
beamrider	(210, 160, 3)	9	$(-\infty, \infty)$
qbert	(210, 160, 3)	6	$(-\infty, \infty)$
spaceinvaders	(210, 160, 3)	6	$(-\infty, \infty)$

Table A.1 – Specification of Atari 2600 games used for the experiments.