



Universiteit
Leiden

The Netherlands

Enhancing autonomy and efficiency in goal-conditioned reinforcement learning

Yang, Z.

Citation

Yang, Z. (2025, February 26). *Enhancing autonomy and efficiency in goal-conditioned reinforcement learning*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/4196074>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4196074>

Note: To cite this publication please use the final published version (if applicable).

References

- Akkaya, I., Andrychowicz, M., Chociej, M., Litwin, M., McGrew, B., Petron, A., Paino, A., Plappert, M., Powell, G., Ribas, R., et al. (2019). Solving rubik’s cube with a robot hand. *arXiv preprint arXiv:1910.07113*.
- Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O., & Zaremba, W. (2017). Hindsight experience replay. *Advances in neural information processing systems*, 30.
- Anthropic. (2024, September). Claude.
- Baranes, A., & Oudeyer, P.-Y. (2013). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, 61(1), 49–73.
- Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47, 253–279.
- Berkman, E. T. (2018). The neuroscience of goals and behavior change. *Consulting Psychology Journal: Practice and Research*, 70(1), 28.
- Blundell, C., Uria, B., Pritzel, A., Li, Y., Ruderman, A., Leibo, J. Z., Rae, J., Wierstra, D., & Hassabis, D. (2016). Model-free episodic control. *arXiv preprint arXiv:1606.04460*.
- Blundell, C., Uria, B., Pritzel, A., Li, Y., Ruderman, A., Leibo, J. Z., Rae, J. W., Wierstra, D., & Hassabis, D. (2016). Model-free episodic control. *CoRR*, abs/1606.04460.
- Burda, Y., Edwards, H., Storkey, A., & Klimov, O. (2018). Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*.
- Campero, A., Raileanu, R., Kuttler, H., Tenenbaum, J. B., Rocktäschel, T., & Grefenstette, E. (2021). Learning with amigo: Adversarially motivated intrinsic goals. *International Conference on Learning Representations*.

References

- Chevalier-Boisvert, M., Dai, B., Towers, M., Perez-Vicente, R., Willems, L., Lahlou, S., Pal, S., Castro, P. S., & Terry, J. (2024). Minigrid & miniworld: Modular & customizable reinforcement learning environments for goal-oriented tasks. *Advances in Neural Information Processing Systems*, 36.
- Chevalier-Boisvert, M., Willems, L., & Pal, S. (2018). Minimalistic gridworld environment for openai gym.
- Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, 395(6699), 272–274.
- Colas, C., Karch, T., Sigaud, O., & Oudeyer, P. (n.d.). Intrinsically motivated goal-conditioned reinforcement learning: A short survey. arxiv 2020. *arXiv preprint arXiv:2012.09830*.
- Colas, C., Fournier, P., Chetouani, M., Sigaud, O., & Oudeyer, P.-Y. (2019). Curious: Intrinsically motivated modular multi-goal reinforcement learning. *International conference on machine learning*, 1331–1340.
- Colas, C., Karch, T., Sigaud, O., & Oudeyer, P.-Y. (2022). Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: A short survey. *Journal of Artificial Intelligence Research*, 74, 1159–1199.
- Degrave, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., Ewalds, T., Hafner, R., Abdolmaleki, A., de Las Casas, D., et al. (2022). Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602(7897), 414–419.
- Deng, F., Park, J., & Ahn, S. (2023). Facing off world model backbones: Rnns, transformers, and s4. *arXiv preprint arXiv:2307.02064*.
- Ding, Y., Florensa, C., Abbeel, P., & Phielipp, M. (2019). Goal-conditioned imitation learning. *Advances in neural information processing systems*, 32.
- Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O., & Clune, J. (2021). First return, then explore. *Nature*, 590(7847), 580–586.
- Escontrela, A., Adeniji, A., Yan, W., Jain, A., Peng, X. B., Goldberg, K., Lee, Y., Hafner, D., & Abbeel, P. (2024). Video prediction models as rewards for reinforcement learning. *Advances in Neural Information Processing Systems*, 36.
- Eysenbach, B., Gu, S., Ibarz, J., & Levine, S. (2017). Leave no trace: Learning to reset for safe and autonomous reinforcement learning. *arXiv preprint arXiv:1711.06782*.
- Eysenbach, B., Gupta, A., Ibarz, J., & Levine, S. (2018). Diversity is all you need: Learning diverse skills without a reward function.

- Eysenbach, B., Zhang, T., Levine, S., & Salakhutdinov, R. R. (2022). Contrastive learning as goal-conditioned reinforcement learning. *Advances in Neural Information Processing Systems*, 35, 35603–35620.
- Fan, L., Wang, G., Jiang, Y., Mandlekar, A., Yang, Y., Zhu, H., Tang, A., Huang, D.-A., Zhu, Y., & Anandkumar, A. (2022). Minedojo: Building open-ended embodied agents with internet-scale knowledge. *Advances in Neural Information Processing Systems*, 35, 18343–18362.
- Fang, M., Zhou, T., Du, Y., Han, L., & Zhang, Z. (2019). Curriculum-guided hindsight experience replay. *Advances in neural information processing systems*, 32.
- Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatain, M., Novikov, A., Ruiz, F. J., Schrittwieser, J., Swirszcz, G., et al. (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930), 47–53.
- Florensa, C., Held, D., Geng, X., & Abbeel, P. (2018). Automatic goal generation for reinforcement learning agents. *International conference on machine learning*, 1515–1528.
- Fu, J., Korattikara, A., Levine, S., & Guadarrama, S. (2019). From language to goals: Inverse reinforcement learning for vision-based instruction following. *arXiv preprint arXiv:1902.07742*.
- Google. (2024, September). Gemini.
- Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.
- Gupta, A., Yu, J., Zhao, T. Z., Kumar, V., Rovinsky, A., Xu, K., Devlin, T., & Levine, S. (2021). Reset-free reinforcement learning via multi-task learning: Learning dexterous manipulation behaviors without human intervention. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 6664–6671.
- Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In J. G. Dy & A. Krause (Eds.), *Proceedings of the 35th international conference on machine learning, ICML 2018, stockholmsmässan, stockholm, sweden, july 10-15, 2018* (pp. 1856–1865, Vol. 80). PMLR.
- Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., et al. (2018). Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*.
- Hafner, D., Lillicrap, T., Ba, J., & Norouzi, M. (2019). Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*.

References

- Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2020). Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*.
- Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*.
- Haldar, S., Pari, J., Rai, A., & Pinto, L. (2023). Teach a robot to fish: Versatile imitation from one minute of demonstrations. *arXiv preprint arXiv:2303.01497*.
- Hansen, N., Su, H., & Wang, X. (2023). Td-mpc2: Scalable, robust world models for continuous control. *arXiv preprint arXiv:2310.16828*.
- Hansen, N., Wang, X., & Su, H. (2022). Temporal difference learning for model predictive control. *arXiv preprint arXiv:2203.04955*.
- Hartikainen, K., Geng, X., Haarnoja, T., & Levine, S. (2019). Dynamical distance learning for semi-supervised and unsupervised skill discovery. *arXiv preprint arXiv:1907.08225*.
- He, Q., Feng, Z., Fang, H., Wang, X., Zhao, L., Yao, Y., & Yu, K. (2023). A blockchain-based scheme for secure data offloading in healthcare with deep reinforcement learning. *IEEE/ACM Transactions on Networking*, 32(1), 65–80.
- Hu, E. S., Chang, R., Rybkin, O., & Jayaraman, D. (2023). Planning goals for exploration. *The Eleventh International Conference on Learning Representations*.
- Hu, H., Ye, J., Zhu, G., Ren, Z., & Zhang, C. (2021). Generalizable episodic memory for deep reinforcement learning. *arXiv preprint arXiv:2103.06469*.
- Jiang, Y., Gupta, A., Zhang, Z., Wang, G., Dou, Y., Chen, Y., Fei-Fei, L., Anandkumar, A., Zhu, Y., & Fan, L. (2022). Vima: General robot manipulation with multimodal prompts. *NeurIPS 2022 Foundation Models for Decision Making Workshop*.
- Kanagawa, Y. (2021). Mujoco-maze.
- Kim, J., Cho, D., & Kim, H. J. (2023). Demonstration-free autonomous reinforcement learning via implicit and bidirectional curriculum. *International Conference on Machine Learning*.
- Kim, J., hyeon Park, J., Cho, D., & Kim, H. J. (2022). Automating reinforcement learning with example-based resets. *IEEE Robotics and Automation Letters*, 7(3), 6606–6613.
- Kim, M., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., Rafailov, R., Foster, E., Lam, G., Sanketi, P., Vuong, Q., Kollar, T., Burchfiel, B., Tedrake, R., Sadigh, D., Levine, S., Liang, P., & Finn, C. (2024). Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*.

- Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Klissarov, M., D’Oro, P., Sodhani, S., Raileanu, R., Bacon, P.-L., Vincent, P., Zhang, A., & Henaff, M. (2023). Motif: Intrinsic motivation from artificial intelligence feedback. *arXiv preprint arXiv:2310.00166*.
- Kompella, V., Walsh, T. J., Barrett, S., Wurman, P., & Stone, P. (2022). Event tables for efficient experience replay. *arXiv preprint arXiv:2211.00576*.
- Kuznetsov, I., & Filchenkov, A. (2021). Solving continuous control with episodic memory. *arXiv preprint arXiv:2106.08832*.
- Le, H., Karimpanal, T. G., Abdolshah, M., Tran, T., & Venkatesh, S. (2021). Model-based episodic memory induces dynamic hybrid controls. In M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems 34: Annual conference on neural information processing systems 2021, neurips 2021, december 6-14, 2021, virtual* (pp. 30313–30325).
- Lee, S. Y., Sungik, C., & Chung, S.-Y. (2019). Sample-efficient deep reinforcement learning via episodic backward update. *Advances in Neural Information Processing Systems*, 32.
- Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1), 1334–1373.
- Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.
- Lin, Z., Zhao, T., Yang, G., & Zhang, L. (2018). Episodic memory deep q-networks. *arXiv preprint arXiv:1805.07603*.
- Liu, M., Zhu, M., & Zhang, W. (2022). Goal-conditioned reinforcement learning: Problems and solutions. *arXiv preprint arXiv:2201.08299*.
- Lu, C., Hu, S., & Clune, J. (2024). Intelligent go-explore: Standing on the shoulders of giant foundation models. *arXiv preprint arXiv:2405.15143*.
- Lu, K., Grover, A., Abbeel, P., & Mordatch, I. (2020). Reset-free lifelong learning with skill-space planning. *arXiv preprint arXiv:2012.03548*.
- Lu, K., Mordatch, I., & Abbeel, P. (2020). Adaptive online planning for continual lifelong learning.

References

- Ma, Y. J., Liang, W., Wang, G., Huang, D.-A., Bastani, O., Jayaraman, D., Zhu, Y., Fan, L., & Anandkumar, A. (2023). Eureka: Human-level reward design via coding large language models. *arXiv preprint arXiv:2310.12931*.
- Mendonca, R., Rybkin, O., Daniilidis, K., Hafner, D., & Pathak, D. (2021). Discovering and achieving goals via world models.
- Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. *International conference on machine learning*, 1928–1937.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540), 529–533.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M. A., Fidjeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nat.*, 518(7540), 529–533.
- Mu, Y., Zhang, Q., Hu, M., Wang, W., Ding, M., Jin, J., Wang, B., Dai, J., Qiao, Y., & Luo, P. (2024). Embodiedgpt: Vision-language pre-training via embodied chain of thought. *Advances in Neural Information Processing Systems*, 36.
- Nagabandi, A., Konolige, K., Levine, S., & Kumar, V. (2020). Deep dynamics models for learning dexterous manipulation. *Conference on Robot Learning*, 1101–1112.
- OpenAI. (2024, September). Chatgpt.
- Padalkar, A., Pooley, A., Jain, A., Bewley, A., Herzog, A., Irpan, A., Khazatsky, A., Rai, A., Singh, A., Brohan, A., et al. (2023). Open x-embodiment: Robotic learning datasets and rt-x models. *arXiv preprint arXiv:2310.08864*.
- Parisotto, E., & Salakhutdinov, R. (2018). Neural map: Structured memory for deep reinforcement learning. *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.
- Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. *ICML*.

- Péré, A., Forestier, S., Sigaud, O., & Oudeyer, P.-Y. (2018). Unsupervised learning of goal spaces for intrinsically motivated goal exploration. *arXiv preprint arXiv:1803.00781*.
- Pitis, S., Chan, H., Zhao, S., Stadie, B., & Ba, J. (2020). Maximum entropy gain exploration for long horizon multi-goal reinforcement learning. *International Conference on Machine Learning*, 7750–7761.
- Plappert, M., Andrychowicz, M., Ray, A., McGrew, B., Baker, B., Powell, G., Schneider, J., Tobin, J., Chociej, M., Welinder, P., Kumar, V., & Zaremba, W. (2018). Multi-goal reinforcement learning: Challenging robotics environments and request for research.
- Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, 41(3), 245–251.
- Pong, V. H., Dalal, M., Lin, S., Nair, A., Bahl, S., & Levine, S. (2019). Skew-fit: State-covering self-supervised reinforcement learning. *arXiv preprint arXiv:1903.03698*.
- Portelas, R., Colas, C., Hofmann, K., & Oudeyer, P.-Y. (2020). Teacher algorithms for curriculum learning of deep rl in continuously parameterized environments. *Conference on Robot Learning*, 835–853.
- Poudel, R. P., Pandya, H., Liwicki, S., & Cipolla, R. (2024). Recore: Regularized contrastive representation learning of world model. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22904–22913.
- Pritzel, A., Uria, B., Srinivasan, S., Badia, A. P., Vinyals, O., Hassabis, D., Wierstra, D., & Blundell, C. (2017a). Neural episodic control. *International Conference on Machine Learning*, 2827–2836.
- Pritzel, A., Uria, B., Srinivasan, S., Badia, A. P., Vinyals, O., Hassabis, D., Wierstra, D., & Blundell, C. (2017b, June). Neural episodic control. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning* (pp. 2827–2836, Vol. 70). PMLR.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. *International conference on machine learning*, 8748–8763.
- Rahman, A. A., Agarwal, P., Michalski, V., Noumeir, R., & Kahou, S. (2023). Empowering clinicians with medt: A framework for sepsis treatment. *NeurIPS 2023 Workshop on Goal-Conditioned Reinforcement Learning*.

References

- Reuss, M., Li, M., Jia, X., & Lioutikov, R. (2023). Goal-conditioned imitation learning using score-based diffusion policies. *arXiv preprint arXiv:2304.02532*.
- Salimans, T., Ho, J., Chen, X., Sidor, S., & Sutskever, I. (2017). Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*.
- Sancaktar, C., Blaes, S., & Martius, G. (2022). Curious exploration via structured world models yields zero-shot object manipulation. *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*.
- Schaul, T., Horgan, D., Gregor, K., & Silver, D. (2015). Universal value function approximators. *International conference on machine learning*, 1312–1320.
- Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. *arXiv preprint arXiv:1511.05952*.
- Schott, B. H., Henson, R. N., Richardson-Klavehn, A., Becker, C., Thoma, V., Heinze, H.-J., & Düzel, E. (2005). Redefining implicit and explicit memory: The functional neuroanatomy of priming, remembering, and control of retrieval. *Proceedings of the National Academy of Sciences*, 102(4), 1257–1262.
- Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., et al. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839), 604–609.
- Schwarzer, M., Anand, A., Goel, R., Hjelm, R. D., Courville, A., & Bachman, P. (2020). Data-efficient reinforcement learning with self-predictive representations. *arXiv preprint arXiv:2007.05929*.
- Sekar, R., Rybkin, O., Daniilidis, K., Abbeel, P., Hafner, D., & Pathak, D. (2020). Planning to explore via self-supervised world models. *ICML*.
- Sharma, A., Ahmad, R., & Finn, C. (2022). A state-distribution matching approach to non-episodic reinforcement learning. *arXiv preprint arXiv:2205.05212*.
- Sharma, A., Ahmed, A. M., Ahmad, R., & Finn, C. (2023). Self-improving robots: End-to-end autonomous visuomotor reinforcement learning. *arXiv preprint arXiv:2303.01488*.
- Sharma, A., Gupta, A., Levine, S., Hausman, K., & Finn, C. (2021). Autonomous reinforcement learning via subgoal curricula. *Advances in Neural Information Processing Systems*, 34, 18474–18486.
- Sharma, A., Xu, K., Sardana, N., Gupta, A., Hausman, K., Levine, S., & Finn, C. (2021). Autonomous reinforcement learning: Formalism and benchmarking. *arXiv preprint arXiv:2112.09605*.

- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T. P., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nat.*, 529(7587), 484–489.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *nature*, 550(7676), 354–359.
- Smith, L., Dhawan, N., Zhang, M., Abbeel, P., & Levine, S. (2019). Avid: Learning multi-stage tasks via pixel-level translation of human videos. *arXiv preprint arXiv:1912.04443*.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.
- Tulving, E. (1983). Elements of episodic memory.
- Tulving, E. (1972). Episodic and semantic memory. *Organization of memory*.
- Veeriah, V., Oh, J., & Singh, S. (2018). Many-goals reinforcement learning. *ArXiv, abs/1806.09605*.
- Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., et al. (2019). Grand-master level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782), 350–354.
- Warde-Farley, D., de Wiele, T. V., Kulkarni, T., Ionescu, C., Hansen, S., & Mnih, V. (2019). Unsupervised control through non-parametric discriminative rewards. *International Conference on Learning Representations*.
- Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, 8(3), 279–292.
- Watkins, C. J. C. H. (1989). Learning from delayed rewards.
- Wurman, P. R., Barrett, S., Kawamoto, K., MacGlashan, J., Subramanian, K., Walsh, T. J., Capobianco, R., Devlic, A., Eckert, F., Fuchs, F., et al. (2022). Outracing champion gran turismo drivers with deep reinforcement learning. *Nature*, 602(7896), 223–228.

References

- Xu, K., Verma, S., Finn, C., & Levine, S. (2020). Continual learning of control primitives: Skill discovery via reset-games. *Advances in Neural Information Processing Systems*, 33, 4999–5010.
- Yahya, A., Li, A., Kalakrishnan, M., Chebotar, Y., & Levine, S. (2017). Collective robot reinforcement learning with distributed asynchronous guided policy search. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 79–86.
- Yang, J., Gao, S., Qiu, Y., Chen, L., Li, T., Dai, B., Chitta, K., Wu, P., Zeng, J., Luo, P., et al. (2024). Generalized predictive model for autonomous driving. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14662–14672.
- Yang, X., Ji, Z., Wu, J., & Lai, Y.-K. (2021). An open-source multi-goal reinforcement learning environment for robotic manipulation with pybullet. *Annual Conference Towards Autonomous Robotic Systems*, 14–24.
- Yang, Z., Moerland, T. M., Preuss, M., & Plaat, A. (2022). Continuous episodic control. *arXiv preprint arXiv:2211.15183*.
- Yi, Z., Luo, Y., Westover, T., Katikaneni, S., Ponkiya, B., Sah, S., Mahmud, S., Raker, D., Javaid, A., Heben, M. J., et al. (2022). Deep reinforcement learning based optimization for a tightly coupled nuclear renewable integrated energy system. *Applied Energy*, 328, 120113.
- Young, K., & Tian, T. (2019). Minatar: An atari-inspired testbed for thorough and reproducible reinforcement learning experiments. *arXiv preprint arXiv:1903.03176*.
- Yu, T., Quillen, D., He, Z., Julian, R., Hausman, K., Finn, C., & Levine, S. (2019). Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. *Conference on Robot Learning (CoRL)*.
- Zhang, Y., Abbeel, P., & Pinto, L. (2020). Automatic curriculum learning through value disagreement. *Advances in Neural Information Processing Systems*, 33, 7648–7659.
- Zhang, Z., Li, Y., Bastani, O., Gupta, A., Jayaraman, D., Ma, Y. J., & Weihs, L. (2023). Universal visual decomposer: Long-horizon manipulation made easy. *arXiv preprint arXiv:2310.08581*.
- Zheng, S., Trott, A., Srinivasa, S., Parkes, D. C., & Socher, R. (2022). The ai economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science advances*, 8(18), eabk2607.
- Zhu, H., Gupta, A., Rajeswaran, A., Levine, S., & Kumar, V. (2019). Dexterous manipulation with deep reinforcement learning: Efficient, general, and low-cost.

- 2019 International Conference on Robotics and Automation (ICRA)*, 3651–3657.
- Zhu, H., Yu, J., Gupta, A., Shah, D., Hartikainen, K., Singh, A., Kumar, V., & Levine, S. (2020). The ingredients of real-world robotic reinforcement learning. *arXiv preprint arXiv:2004.12570*.

References

Acknowledgements

I would like to express my deepest gratitude to my supervisors and promoters, Thomas, Mike, and Aske. Thomas, I am really grateful you joined my supervision team after the first year of my PhD study—it has been an incredible journey. Your consistent support has greatly shaped my research, and I am truly thankful for the expertise you have shared with me. Thank you for challenging me to think critically and I appreciate everything you have done. Mike, thank you so much for always being there when I needed you. Your enthusiasm for research has inspired me, and your constant belief in my abilities has helped me navigate challenges with greater confidence. I am deeply grateful for your kindness and the personal support you have provided throughout this journey. Aske, I sincerely appreciate your encouragement and guidance during my study, as well as your swift responses and inspiring discussions. Your willingness to explore new ideas and guide me toward meaningful solutions has always been motivating. Working with you has been an absolute privilege.

I would also like to thank my collaborator Edward Hu from the University of Pennsylvania. Thank you for the wonderful collaborations and insightful discussions. Additionally, I extend my thanks to all my colleagues, whose collaboration and friendship made this journey enjoyable. Special thanks go to all members of RLG—especially Alan, Alvaro, Andrius, Annie, Andreas, Bram, Felix, Koen, Matthias, Michiel, Mike, and Tom—for their assistance with various aspects of my research and for always being willing to share their expertise and time. I would also like to thank all colleagues and friends at LIACS, especially Dr. Xueqin Chen, Dr. Wenjing Chu, Dr. Hui Feng, Dr. Xuhan Liu, Dr. Nan Pu, Dr. Hui Wang, Dr. Furong Ye, Qi Huang, Sengim Karayalcin, Shuaiqun Pan, Serban Vadineanu, Yumeng Wang, Xiaohan Wei, Weikang Wen, and Yuxuan Zhao. Thank you for the coffee breaks and the discussions we shared.

I am grateful to my surfing buddy Lixuan Zhang and my bouldering buddies, especially Dr. Yiming Dong. Thanks as well to Dr. Hao Wang, Zhaoxing Wang, and

Acknowledgements

Dr. Bing Xu for their guidance on both life and research. I am also deeply grateful to all friends from ACSSNL, especially Dr. Wei Chen, Dr. Yu Dong, Dr. Yupeng He, Dr. Wenyi Wang, Dr. Xiaoyang Zhong, Wei Fang, Nasi Liu, Tiantian Tang, Wenyu Wan, Jiayi Yang, Ying Yin, and Yuxuan Zhu. Thank you all—your continued friendship and support mean the world to me.

To my friends in China, thank you for your unwavering support. Yuan Liang and Chen Wang, your encouragement was especially meaningful during the early stages of my PhD, amidst the challenges of COVID-19. Zhao Yan, thank you for always being there. Kaiqi Fu, your research and life ambitions have been inspiring. Yuxuan Jiang, your kindness has been a constant source of comfort. To my long-time friend, Minkang Tan, and along the way, I was fortunate to meet F3—thank you Jiayang, Zhong, and Lincen for all the coffee and beers we shared.

Finally, I would like to express my deepest appreciation to my family. My heartfelt thanks go to my grandma, parents, my brother and my sister-in-law, and all my cousins, nephews, and nieces. My sincere gratitude also goes to my girlfriend for accompanying me to the end of this journey. Without your support and love, I would never have made it this far.

I am also grateful to the China Scholarship Council for their financial support, which made this endeavor possible.

This thesis would not have been possible without the support, guidance, and encouragement of all these remarkable individuals. Thank you all!