Large Learning Agents: Towards Continually Aligned Robots with Scale in RL

Doctoral Consortium

Bram Grooten
Eindhoven University of Technology

ABSTRACT

In the field of deep reinforcement learning significant progress has been made, but it seems we are missing the power of the scaling laws evident in large language models. This research aims to pioneer the development of large learning agents (**LLA**s) that can take advantage of efficient scaling. We focus on creating agents that generalize strongly, quickly adapt to continuously changing environments, and integrate the reinforcements received through human feedback. We believe that this is a key step towards the long-term vision for continually aligned and intelligent agents.

KEYWORDS

Deep Reinforcement Learning, Continual Learning, AI Alignment, Scale, Efficiency, Sparsity.

ACM Reference Format:
Bram Grooten. 2024. Large Learning Agents: Towards Continually Aligned Robots with Scale in RL: Doctoral Consortium. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Throughout my PhD journey and beyond, I strive to contribute at the intersection of reinforcement learning, continual learning, and AI alignment. My research is driven by a threefold question, seeking to address fundamental challenges in the development of aligned and intelligent agents:

How can we design large learning agents that

(1) *generalize robustly in a wide variety of environments,*
(2) *adapt rapidly to continually evolving environments,*
(3) *and incorporate reinforcements provided by humans?*

By embracing the principles of efficient scaling, we aspire to develop agents that are not only proficient in their specific tasks but are also equipped to handle the unpredictable nature of real-world environments. In the subsequent sections we will delve into each of these subquestions, exploring current methodologies in the literature, identifying challenges, and proposing novel approaches.

2 REINFORCEMENT LEARNING

The core of this research lies in the area of deep reinforcement learning, where artificial neural networks are instrumental in learning a policy π, often alongside value functions V or Q. The current methods in deep RL predominantly use relatively shallow and narrow networks; the potential of large learning agents is underexplored.

2.1 Scale

The question is whether we can scale up these neural networks to be deeper and wider, to benefit from the scaling laws that we see today in large language models (LLMs). These models have demonstrated that increased scale can lead to significant improvements in learning capabilities and generalization [11, 18, 29, 31]. This research hypothesizes that similar principles of scaling can be applied to deep RL, unlocking new levels of efficiency and effectiveness.

In this context, we introduce the concept of Large Learning Agents (LLAs) - a shift towards leveraging larger neural network architectures in RL. The Bitter Lesson [28] teaches us that approaches that make use of the power of computation generally lead to more capable learning systems in the long run. Thus, LLAs are envisioned as a step towards harnessing the computational resources available today to scale up the capabilities of RL agents.

When we improve the efficiency of our neural networks, we can take better advantage of the available compute. Sparse neural networks have the potential to require less memory (less parameters) while maintaining the same representational power (number of neurons) [20]. Or for equal compute, i.e. parameters, we can train networks with more neurons! We should take advantage of the fact these larger sparser networks perform better than dense networks *for the same parameter count* [14]. On top of that, training sparse neural networks can be faster with certain hardware [7, 12, 33, 35].

Part of my work focuses on the improvement of dynamic sparse training (DST) methods [4, 13, 21] to train neural networks that are sparse from scratch. DST methods search for the optimal sparse network structure by periodically pruning and growing weights, inspired by our own brain's plasticity, which also drops and grows synapses [3, 5, 25]. We found that within DST, pruning weights based on magnitude alone is a simple yet effective mechanism [23].

2.2 Focus

Agents that can focus on the most task-relevant inputs in a certain problem generally perform better. We demonstrated this in our work on Automatic Noise Filtering [15, ANF], which uses dynamic sparse training to adjust the structure of a neural network over time. The network learns to grow more connections to input neurons that provide task-relevant information, and prune weights that are connected to irrelevant inputs. We have shown that even in environments where 99% of the input features are irrelevant to the task, ANF gains adequate performance, in contrast to dense (fully-connected) neural networks.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand.

2.3 Generalization

In our latest work [16, MaDi] we demonstrate that learning to ***ma**sk* ***di**stractions* in image-based RL can benefit an agent's generalization performance. MaDi introduces a lightweight Masker network at the front of the architecture. It learns to mask task-irrelevant pixels via the reward signal only, without the need for additional segmentation labels. MaDi shows state-of-the-art performance on challenging benchmarks such as the DeepMind Control Generalization Benchmark [17] and the Distracting Control Suite [27]. A first step in answering research question (1).

Vision Transformers [11], which have an internal attention mechanism [31], were not able to find sufficient focus by themselves in the benchmarks we tested. The addition of MaDi's Masker network significantly improved their generalization performance. Perhaps with enough pretraining, ViT-like architectures could gain state-of-the-art performance. We hope to find methods that scale efficiently, requiring the least amount of pretraining possible.

The ambition is that with the appropriate focus mechanisms, RL algorithms, and sufficient scale, we can make agents that are able to learn even faster than humans. Some works show promising directions in this regard [10, 32, 34], as model-free deep RL can nowadays learn to play most Atari games up to human level in the equivalent of just two hours of gameplay [26].

2.4 Physical AGI

The pursuit of Artificial General Intelligence (AGI) is a frontier in our field, and recent developments have further defined its trajectory. DeepMind's recent publication outlines six distinct levels of *cognitive AGI*, providing a framework for understanding and measuring progress in this area [22]. My long-term ambition extends towards algorithms that can be applied to *physical* robots, to hopefully move towards physical AGI as well. The first self-driving cars have become a reality, and my childhood dream of creating a household robot is still in the back of my mind.

In our previously mentioned MaDi paper [16], we showed that physical agents can also improve their generalization ability by masking distractions. We trained a UR5 robotic arm in a visual reaching task. The goal was to reach the webcam on the tip of its arm toward a red circle, located randomly on a white screen. Through asynchronous MaDi, the robot can learn this in real-time, approximately two hours. Furthermore, when we replaced the white background by random videos during test time, the agent did not get distracted and could still perform the task excellently.

3 CONTINUAL LEARNING

When agents or robots are deployed in the real world, it will become increasingly important to ensure they can continually adapt to new situations. The field of lifelong or continual learning investigates this [6, 24], where agents need to learn multiple tasks sequentially.

Literature has shown that our current neural networks can lose plasticity over time when trying to learn continually [1, 8, 9]. Methods that mitigate this often use a sense of resetting or reinitializing parts of the network. Dynamic sparse training methods similarly reinitialize some weights periodically, which I believe can be quite effective in maintaining plasticity. The fact that these networks are *sparse* or *incomplete* gives another advantage: when a weight is pruned, we can grow a new connection in a *different* location, instead of always having to reset parameters or neurons in-place. Sparsity allows us more "room to play with." A promising direction for research question (2).

In this regard, it seems important for networks to be able to determine which parts are forgettable. In a setting with limited compute, we will not be able to perform all tasks learned in a long continual sequence perfectly.[1] Perhaps the idea of learning to focus on the relevant parts of the network can help in continual learning too, as ANF [15] accomplished in noisy RL environments.

4 AI ALIGNMENT

In the fast-moving field of artificial intelligence, it is important to consider safety and ethical aspects in our work [2]. A natural way to integrate this into reinforcement learning algorithms is through the approach of human-in-the-loop RL [19, 30]. This methodology does not assume that a reward function is given by the environment, but that AI agents will have to learn it themselves; directly from human feedback. This is a project that I am currently working on, progressing towards research question (3).

We might provide an approximate initial reward function to the AI agents that we think is useful, as a head start, but supply human feedback along the way as it is learning continually. An agent will need to learn to update not only its policy, but also its initial reward function with the reinforcements it receives.

We need agents that are able to adapt quickly to human feedback, such that they can function in the real world. We want robots that adjust their behavior according to human preferences. Even if our culture, norms, and values evolve over time, the continually aligned agent keeps learning to adapt its reward function. Hopefully, this can be a vital tool in the creation of continually aligned AI agents.

5 CONCLUSION

This research at the intersection of deep reinforcement learning, continual learning, and AI alignment focuses on developing Large Learning Agents (LLAs) that harness the power scaling laws. Our discoveries of techniques like Automatic Noise Filtering [15] and Masking Distractions [16] demonstrate progress towards agents that can effectively generalize and adapt in challenging environments. The approach of dynamic sparse training, as part of this research, has opened new avenues for investigation in the areas of efficient scaling and continual learning. Our application to physical robotic tasks has shown encouraging results, indicating the feasibility of these techniques in real-world scenarios. Significantly, incorporating human feedback into the learning loop emerges as a critical aspect in aligning AI with ethical standards and human preferences. We believe this can be a promising direction to ensure that our agents operate safely and responsibly.

ACKNOWLEDGMENTS

I wish to thank my supervisors Decebal Mocanu and Mykola Pechenizkiy, as well as Matt Taylor, for their continued guidance and support in this research. This work is part of the AMADeuS project of the Open Technology Programme (project number 18489), which is partly financed by the Dutch Research Council (NWO).

[1] Definitely not if the sequence is infinitely long, while the compute budget is not.

REFERENCES

[1] Zaheer Abbas, Rosie Zhao, Joseph Modayil, Adam White, and Marlos C Machado. 2023. Loss of plasticity in continual deep reinforcement learning. *arXiv preprint arXiv:2303.07507* (2023). URL: https://arxiv.org/abs/2303.07507.

[2] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete Problems in AI Safety. *arXiv preprint arXiv:1606.06565* (2016). URL: https://arxiv.org/abs/1606.06565.

[3] Paul Bach-y Rita, Carter C Collins, Frank A Saunders, Benjamin White, and Lawrence Scadden. 1969. Vision Substitution by Tactile Image Projection. *Nature* 221, 5184 (1969), 963–964. URL: https://www.nature.com/articles/221963a0.

[4] Guillaume Bellec, David Kappel, Wolfgang Maass, and Robert Legenstein. 2018. Deep Rewiring: Training very sparse deep networks. *International Conference on Learning Representations* (2018). URL: https://arxiv.org/abs/1711.05136.

[5] Elisa Castaldi, Claudia Lunghi, and Maria Concetta Morrone. 2020. Neuroplasticity in adult human visual cortex. *Neuroscience & Biobehavioral Reviews* 112 (2020), 542–552. URL: https://www.sciencedirect.com/science/article/pii/S0149763419303288.

[6] Zhiyuan Chen and Bing Liu. 2018. *Lifelong Machine Learning*. Vol. 1. Springer. URL: https://link.springer.com/book/10.1007/978-3-031-01581-6.

[7] Selima Curci, Decebal Constantin Mocanu, and Mykola Pechenizkiyi. 2021. Truly Sparse Neural Networks at Scale. *arXiv preprint arXiv:2102.01732* (2021). URL: https://arxiv.org/abs/2102.01732.

[8] Shibhansh Dohare, Juan Hernandez-Garcia, Parash Rahman, Richard Sutton, and A Rupam Mahmood. 2023. Loss of Plasticity in Deep Continual Learning. *arXiv preprint arXiv:2306.13812* (2023). URL: https://arxiv.org/abs/2306.13812.

[9] Shibhansh Dohare, Richard S Sutton, and A Rupam Mahmood. 2021. Continual Backprop: Stochastic Gradient Descent with Persistent Randomness. *arXiv preprint arXiv:2108.06325* (2021). URL: https://arxiv.org/abs/2108.06325.

[10] Pierluca D'Oro, Max Schwarzer, Evgenii Nikishin, Pierre-Luc Bacon, Marc G Bellemare, and Aaron Courville. 2022. Sample-Efficient Reinforcement Learning by Breaking the Replay Ratio Barrier. In *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=OpC-9aBBVJe.

[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*. URL: https://arxiv.org/abs/2010.11929.

[12] Erich Elsen, Marat Dukhan, Trevor Gale, and Karen Simonyan. 2020. Fast Sparse ConvNets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14629–14638. URL: https://arxiv.org/abs/1911.09723.

[13] Utku Evci, Trevor Gale, Jacob Menick, Pablo Samuel Castro, and Erich Elsen. 2020. Rigging the Lottery: Making All Tickets Winners. In *International Conference on Machine Learning*. PMLR, 2943–2952. URL: https://arxiv.org/abs/1911.11134.

[14] Laura Graesser, Utku Evci, Erich Elsen, and Pablo Samuel Castro. 2022. The State of Sparse Training in Deep Reinforcement Learning. In *International Conference on Machine Learning*. PMLR, 7766–7792. URL: https://arxiv.org/abs/2206.10369.

[15] Bram Grooten, Ghada Sokar, Shibhansh Dohare, Elena Mocanu, Matthew E. Taylor, Mykola Pechenizkiy, and Decebal Constantin Mocanu. 2023. Automatic Noise Filtering with Dynamic Sparse Training in Deep Reinforcement Learning. *The 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (2023). URL: https://arxiv.org/abs/2302.06548.

[16] Bram Grooten, Tristan Tomilin, Gautham Vasan, Matthew E. Taylor, A. Rupam Mahmood, Meng Fang, Mykola Pechenizkiy, and Decebal Constantin Mocanu. 2024. MaDi: Learning to Mask Distractions for Generalization in Visual Deep Reinforcement Learning. *The 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (2024). URL: https://arxiv.org/abs/2312.15339.

[17] Nicklas Hansen and Xiaolong Wang. 2021. Generalization in Reinforcement Learning by Soft Data Augmentation. In *International Conference on Robotics and Automation*. URL: https://arxiv.org/abs/2011.13389.

[18] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling Laws for Neural Language Models. *arXiv preprint arXiv:2001.08361* (2020). URL: https://arxiv.org/abs/2001.08361.

[19] Kimin Lee, Laura M Smith, and Pieter Abbeel. 2021. PEBBLE: Feedback-Efficient Interactive Reinforcement Learning via Relabeling Experience and Unsupervised Pre-training. In *International Conference on Machine Learning*. PMLR, 6152–6163. URL: https://arxiv.org/abs/2106.05091.

[20] Decebal Constantin Mocanu, Elena Mocanu, Tiago Pinto, Selima Curci, Phuong H Nguyen, Madeleine Gibescu, Damien Ernst, and Zita A Vale. 2021. Sparse Training Theory for Scalable and Efficient Agents. *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems* (2021). URL: https://arxiv.org/abs/2103.01636.

[21] Decebal Constantin Mocanu, Elena Mocanu, Peter Stone, Phuong H Nguyen, Madeleine Gibescu, and Antonio Liotta. 2018. Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. *Nature communications* 9, 1 (2018), 1–12. URL: https://arxiv.org/abs/1707.04780.

[22] Meredith Ringel Morris, Jascha Sohl-dickstein, Noah Fiedel, Tris Warkentin, Allan Dafoe, Aleksandra Faust, Clement Farabet, and Shane Legg. 2023. *"Levels of AGI": Operationalizing Progress on the Path to AGI*. Technical Report. URL: https://arxiv.org/abs/2311.02462.

[23] Aleksandra I Nowak, Bram Grooten, Decebal Constantin Mocanu, and Jacek Tabor. 2023. Fantastic Weights and How to Find Them: Where to Prune in Dynamic Sparse Training. *Advances in Neural Information Processing Systems* (2023). URL: https://arxiv.org/abs/2306.12230.

[24] German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. 2019. Continual lifelong learning with neural networks: A review. *Neural Networks* 113 (2019), 54–71. URL: https://doi.org/10.1016/j.neunet.2019.01.012.

[25] Alberto E Pereda. 2014. Electrical synapses and their functional interactions with chemical synapses. *Nature Reviews Neuroscience* 15, 4 (2014), 250–263. URL: https://www.nature.com/articles/nrn3708.

[26] Max Schwarzer, Johan Samir Obando Ceron, Aaron Courville, Marc G Bellemare, Rishabh Agarwal, and Pablo Samuel Castro. 2023. Bigger, Better, Faster: Human-level Atari with human-level efficiency. In *International Conference on Machine Learning*. PMLR, 30365–30380. URL: https://arxiv.org/abs/2305.19452.

[27] Austin Stone, Oscar Ramirez, Kurt Konolige, and Rico Jonschkowski. 2021. The Distracting Control Suite – A Challenging Benchmark for Reinforcement Learning from Pixels. *arXiv preprint arXiv:2101.02722* (2021). URL: https://arxiv.org/abs/2101.02722.

[28] Richard Sutton. 2019. The Bitter Lesson. *Incomplete Ideas (blog)* (2019). URL: http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

[29] Mingxing Tan and Quoc Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning*. PMLR, 6105–6114. URL: https://arxiv.org/abs/1905.11946.

[30] Matthew E Taylor. 2023. Reinforcement Learning Requires Human-in-the-Loop Framing and Approaches. In *HHAI*. 351–360. URL: https://irll.ca/files/publications/23HHAI.pdf.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *Advances in Neural Information Processing Systems* 30 (2017). URL: https://arxiv.org/abs/1706.03762.

[32] Ziyu Wang, Victor Bapst, Nicolas Heess, Volodymyr Mnih, Remi Munos, Koray Kavukcuoglu, and Nando de Freitas. 2016. Sample Efficient Actor-Critic with Experience Replay. In *International Conference on Learning Representations*. URL: https://arxiv.org/abs/1611.01224.

[33] Wieger Wesselink, Bram Grooten, Qiao Xiao, Cassio de Campos, and Mykola Pechenizkiy. 2023. Nerva: a Truly Sparse Implementation of Neural Networks. *Sparse Neural Networks workshop at ICLR* (2023).

[34] Weirui Ye, Shaohuai Liu, Thanard Kurutach, Pieter Abbeel, and Yang Gao. 2021. Mastering Atari Games with Limited Data. *Advances in Neural Information Processing Systems* 34 (2021), 25476–25488. URL: https://arxiv.org/abs/2111.00210.

[35] Aojun Zhou, Yukun Ma, Junnan Zhu, Jianbo Liu, Zhijie Zhang, Kun Yuan, Wenxiu Sun, and Hongsheng Li. 2020. Learning N:M Fine-grained Structured Sparse Neural Networks From Scratch. *International Conference on Learning Representations* (2020). URL: https://arxiv.org/abs/2102.04010.