Biology Inspired Growth in Meta-Learning

ABSTRACT

Unlike biological systems most current neural networks models do not change their structure during learning. In most machine learning problems, it is important to choose an appropriate network structure because simple networks are likely to under-fit while complex networks are less plastic and more computational expensive to train. To address this challenge, we introduce a dynamically growing network that starts from a small network and adds layers while training. We demonstrate an acceleration of learning using a simple dynamically growing network compared to over baseline models with fixed network structures on a meta-learning task.

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**.

KEYWORDS

meta-learning, developmental neural networks, bio-inspired, maml

1 INTRODUCTION

Artificial Neural Networks (ANNs) have accomplished many amazing feats including surpassing human performance in certain narrow domains. While ANNs are inspired by the brain, they are merely a loose abstraction and are still trailing human performance in many ways. One path towards achieving more human like intelligence is to look biology for futher inspiration [3]. In particular, the ability to change structure over a lifetime is extremely underutilized in ANNs. ANNs are generally trained by modifying the weights of a large fixed structure, however biological brains start from a small size, grow new structures and prune old structures adaptively. This is one potential mechanism that might be necessary to allow ANNs to overcome some of the challenges we discuss next.

Despite ANNs' exceptional performance under ideal conditions, they still face challenges. Network performance can be sensitive to structure, and the correct structure for a problem is difficult to find. Another challenge is that, typically, after learning one task, a network that learns an additional new task will impair or erase its ability to perform the originally learned tasks. This is commonly known as Catastrophic Forgetting (CF). One approach to performing multiple tasks is to use multiple networks, but this requires more resources and redundancy and is not easily scalable. It is also noteworthy that separated networks fail to employ the principle of neural reuse[1], which is vital to biological learning, and allows for the application of knowledge to span across tasks. Meta-Learning, or learning how to learn, attempts to train a network to be trained rapidly. Meta-learning evaluates a network based on how accurately it can solve a task within only a few gradient steps. Most meta-learning algorithms only train the weights of a network and not the structure, despite its importance in the learning process. In dynamic environments such as online meta-learning, which much more closely resemble the real world scenarios, the optimal structure of the network may change along with the task.

2 RELATED WORK

Neural Architecture Search is a field which searches the space of networks structures, and neuroevolution identifies good structures through evolutionary algorithms [6] [2]. However, this involves training a population of networks and individual networks do not have any structure changes during their operation. Cross validation is used to determine network structure in offline multitasking, where all tasks are introduced at the beginning of training. However, this strategy becomes much more difficult in an online setting. All of these methods attempt to find a single network structure that is suitable for a specific problem instead of a universal strategy.

Limited efforts have been made to use networks which develop structure during their lifetime. One major approach is Cartesian Genetic Programming, in which a network maintains neurons and dendrites which can move, die, and reproduce [5]. Other works have explored changing network structure in more traditional networks. Branching networks which grow when new tasks come have been explored in the continual learning environment, but suffer from oversimplified assumptions and computation expense when many tasks are present [4] [8] [7].

There is little work in structural network development in a Meta-Learning setting, and we aim in this paper to provide the first investigation of applying network structure development to an online Meta-Learning environment. Based on the problems encountered in selecting proper structures for ANNs and biological inspiration of the best example of general intelligence we have, we hypothesize that by beginning with a smaller structure and continuing to remove unnecessary structure may reduce computation requirements.

3 METHODOLOGY

We attempt to perform meta-learning with a developing network structure. We selected Model-Agnostic Meta-Learning (MAML) as the algorithm of tackling CF and use a growing convolutional network as the backbone for image classification. Our network starts with a basic convolutional network structure and simply adds additional convolutional layers at spaced intervals growing from 4 convolutional layers to 8. New layers are initialized with predetermined special weights so that it can preserve the output from previous layer. This is to minimize the disruption from adding a layer to the training process, which proves to be extremely disruptive to the training process and hard to recover.

4 RESULTS

Our first experiment (Figure 1) was to see the impact of adding convolutional layers in simple image classification task. We compared network growth in non-disruptive ways and disruptive ways. Both of them dynamically add 5 layers when training for 50 epochs. The disruptive growth seriously impacts the network performance and it struggles to recover the accuracy before being disrupted. This has a huge negative impact on the training accuracy. The non-disruptive way has a much better accuracy than the non-disruptive one, with negligible disruption to the performance. Although the

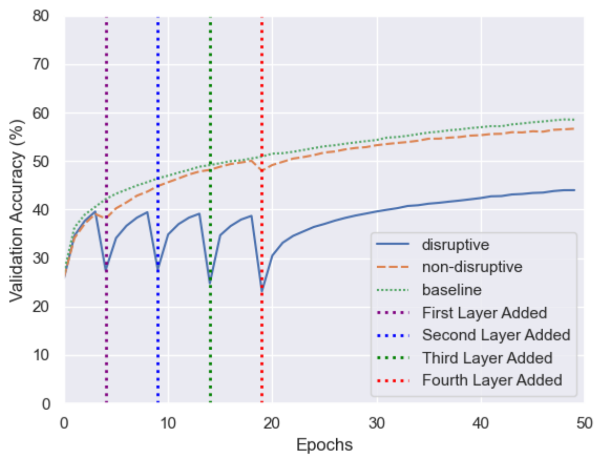

Figure 1: Effect of Network Growth on Supervised Learning (CIFAR-10). Adding convolutional layers in simple image classification task with randomized weights (disruptive), and with an identity matrix (non-disruptive). The non-disruptively added layers still reduce accuracy, but it is minimized. There is no improvement over the baselines for this simple task.

performance at the end does not display any advantage over the baseline, it is within our expectation: supervised learning in image recognition is a relatively mature field and there is not much room for improvement by network growth. However, we hypothesized that the network growth can perform better for multitasking compared to non-growing networks.

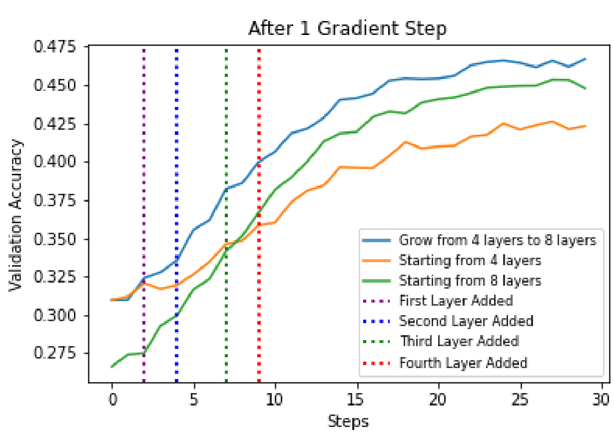

Figure 2: Effect of Network Growth on Meta-Learning (Mini-ImageNet). This is the mean performance of 10 experiment's validation accuracy of growing network trained with MAML on miniimagenet. The meta-learning task is 5 ways and one shot. The results are representative of gradient steps above 0.

This experiment demonstrates the advantage of growing network when encountering a more complex task. In this experiment, the task in 5 ways and the growing network grows from 4 layers to 8 layers and each growth is marked by a vertical line. This growing network's performance is compared to a 4-layer and a 8-layer convolutional network without any growth. When starting from 8 layers, the initial performance is not as good as starting from 4 layers because there are more parameters to train. It is plausible that the network can learn faster when starting from a smaller size. However, a small network can easily underfit due to lack of training parameters. Network growth can potentially solve that problem. After the first graph (marked by the first vertical line) the growing network grows from 4 layers to 5 layers and it immediately starts outperforming the 4-layer baseline. After that, continuous growth of the growing network continues to outperform both baselines.

5 CONCLUSION

Small networks may be trained faster but tend to under-fit, while large networks are good at difficult tasks, but computationally expensive to train. Therefore, growth from simple networks to complex ones could help to combine both advantages. Our simplistic model of including growth in the network has demonstrated potential to improve meta-learning. Layers can be added to complex networks during training with minimal disruption to the training process, and improve the final the results. More sophisticated growth rules and in coordination with pruning rules may be able to further improve meta-learning by determining not only effective weights for a given structure and task, but an effective structure and weights for a given task.

REFERENCES

[1] 2020. *Neural reuse in multifunctional neural networks for control tasks*. ALIFE 2021: The 2021 Conference on Artificial Life, Vol. ALIFE 2020: The 2020 Conference on Artificial Life. https://doi.org/10.1162/isal_a_00319 arXiv:https://direct.mit.edu/isal/proceedings-pdf/isal2020/32/210/1908600/isal_a_00319.pdf

[2] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. 2019. Neural architecture search: A survey. *The Journal of Machine Learning Research* 20, 1 (2019), 1997–2017.

[3] Dhireesha Kudithipudi, Mario Aguilar-Simon, Jonathan Babb, Maxim Bazhenov, Douglas Blackiston, Josh Bongard, Andrew P. Brna, Suraj Chakravarthi Raja, Nick Cheney, Jeff Clune, Anurag Daram, Stefano Fusi, Peter Helfer, Leslie Kay, Nicholas Ketz, Zsolt Kira, Soheil Kolouri, Jeffrey L. Krichmar, Sam Kriegman, Michael Levin, Sandeep Madireddy, Santosh Manicka, Ali Marjaninejad, Bruce McNaughton, Risto Miikkulainen, Zaneta Navratilova, Tej Pandit, Alice Parker, Praveen K. Pilly, Sebastian Risi, Terrence J. Sejnowski, Andrea Soltoggio, Nicholas Soures, Andreas S. Tolias, Darío Urbina-Meléndez, Francisco J. Valero-Cuevas, Gido M. van de Ven, Joshua T. Vogelstein, Felix Wang, Ron Weiss, Angel Yanguas-Gil, Xinyun Zou, and Hava Siegelmann. 2022. Biological underpinnings for lifelong learning machines. *Nature Machine Intelligence* 4, 3 (March 2022), 196–210. https://doi.org/10.1038/s42256-022-00452-0

[4] Yongxi Lu, Abhishek Kumar, Shuangfei Zhai, Yu Cheng, Tara Javidi, and Rogerio Feris. 2016. Fully-adaptive Feature Sharing in Multi-Task Networks with Applications in Person Attribute Classification. https://doi.org/10.48550/ARXIV.1611.05377

[5] J.F. Miller and S.L. Smith. 2006. Redundancy and computational efficiency in Cartesian genetic programming. *IEEE Transactions on Evolutionary Computation* 10, 2 (2006), 167–174. https://doi.org/10.1109/TEVC.2006.871253

[6] Kenneth O Stanley, Jeff Clune, Joel Lehman, and Risto Miikkulainen. 2019. Designing neural networks through neuroevolution. *Nature Machine Intelligence* 1, 1 (2019), 24–35.

[7] Ju Xu and Zhanxing Zhu. 2018. Reinforced Continual Learning. https://doi.org/10.48550/ARXIV.1805.12369

[8] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. 2017. Lifelong Learning with Dynamically Expandable Networks. https://doi.org/10.48550/ARXIV.1708.01547