Biology Inspired Growth in Meta-Learning

Cullen A LaKemper*
Cehong Wang*
Jason A Yoder
lakempca@rose-hulman.edu
Rose-Hulman Institute of Technology
Terre Haute, Indiana, USA

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 computationally 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 other 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

ACM Reference Format:
Cullen A LaKemper, Cehong Wang, and Jason A Yoder. 2022. Biology Inspired Growth in Meta-Learning. In *Genetic and Evolutionary Computation Conference Companion (GECCO '22 Companion), July 9–13, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3520304.3533945

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 to 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 grow new structures and prune old structures adaptively. Dynamically adjusting structure during training might be necessary to allow ANNs to overcome certain challenges.

One such ongoing challenge is that ANN 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 (NE) identifies good structures through evolutionary algorithms [6] [2]. However, NE utilizes a population and individual networks do not have any structural changes during their operation. Cross validation can be 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 ANNs which develop structure during their lifetime. One major approach is Cartesian Genetic Programming, in which a network maintains soma and dendrites which can move, die, and reproduce [5]. Other works have explored developing traditional network structures. Branching networks which grow when new tasks come have been explored in the continual learning environment, but have oversimplified assumptions and high computational expense [4] [8] [7].

In this paper we aim to provide the first investigation of applying network structure development to a 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 growing from small to large and continually removing unnecessary structure we may reduce computation requirements.

*Both authors contributed equally to this research.

GECCO '22 Companion, July 9–13, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9268-6/22/07.
https://doi.org/10.1145/3520304.3533945

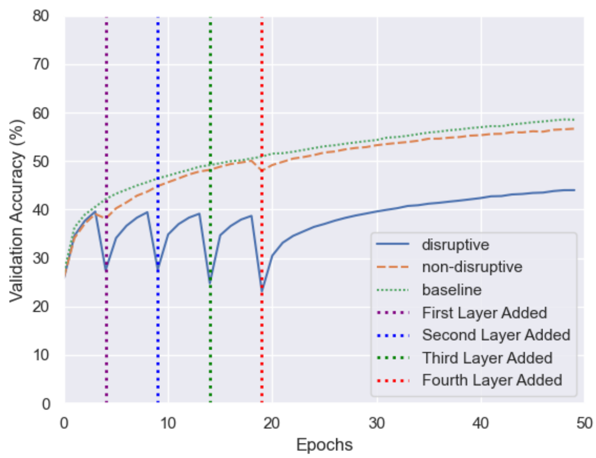

Figure 1: Effect of Network Growth on Supervised Learning (CIFAR-10). Adding convolutional layers in a simple image classification task with randomized weights (disruptive) vs. an identity matrix (non-disruptive). The non-disruptively added layers still reduce accuracy, but it is minimized. There is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

3 METHODOLOGY

We perform meta-learning with a developing network structure. We selected the Model-Agnostic Meta-Learning (MAML) algorithm to train a growing convolutional network 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 to preserve the output from the 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 for recovery.

4 RESULTS

Our first experiment (Figure 1) evaluated adding convolutional layers in a simple image classification task. We compared network growth in non-disruptive and disruptive ways. They both dynamically add 5 layers and train for 50 epochs. The disruptive growth drastically reduces network accuracy and recovery after growth is slow. The non-disruptive growth has negligible disruption to the performance and much better overall accuracy. Although the 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. However, we hypothesized network growth to be more beneficial in multitasking environments.

Our next experiment demonstrates the potential advantage of a growing network when encountering a more complex task. In particular, a growing network's (from 4 to 8 layers) performance is compared to that of 4-layer and 8-layer convolutional networks with no growth. The intial performance of the larger network (8-layer) is worse than that of the smaller network (4-layer), possibly because

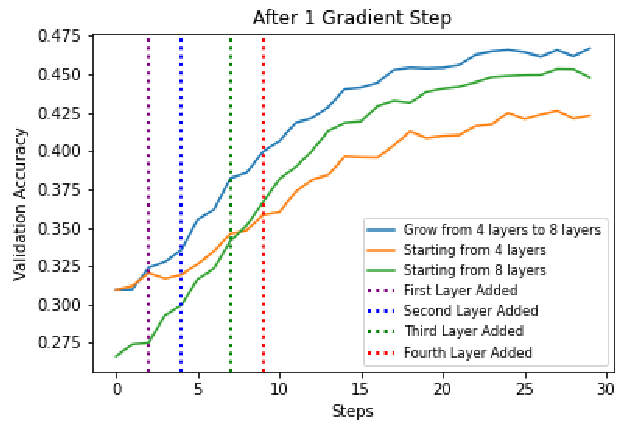

Figure 2: Effect of Network Growth on Meta-Learning (Mini-ImageNet). This is the mean performance of 10 experiment's validation accuracy of growing networks trained with MAML on miniimagenet. The meta-learning task is 5 ways and one shot. The results are representative of gradient steps above 0.

there are more parameters to train. However, a small network can easily underfit due to lack of training parameters. After the first growth from 4 to 5 layers the growing network immediately starts outperforming the 4-layer baseline and continues to outperform both baselines for the remainder of the experiment.

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. Growth from small networks to larger ones could help to combine both advantages. Our simple growth model of adding layers while minimizing disruption to the training process has demonstrated the potential to improve meta-learning. More sophisticated growth rules and 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

[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 and et al. 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 and et al. 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 and et al. 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 and et al. 2017. Lifelong Learning with Dynamically Expandable Networks. https://doi.org/10.48550/ARXIV.1708.01547