Polyploidism in Deep Neural Networks: m-Parent Evolutionary Synthesis of Deep Neural Networks in Varying Population Sizes

Audrey G. Chung — University of Waterloo, ON, Canada
Paul Fieguth — University of Waterloo, ON, Canada
Alexander Wong — University of Waterloo, ON, Canada

Abstract

Evolutionary deep intelligence was recently proposed to organically produce highly efficient deep neural network architectures over successive generations. Thus far, current evolutionary synthesis processes are based on asexual reproduction, i.e., offspring neural networks are synthesized stochastically from a single parent network. In this study, we investigate the effects of m-parent sexual evolutionary synthesis (m = 1, 2, 3, 5) in combination with varying population sizes of three, five, and eight synthesized networks per generation. Experimental results were obtained using a 10% subset of the MNIST handwritten digits dataset, and show that increasing the number of parent networks results in improved architectural efficiency of the synthesized networks (approximately 150× synaptic efficiency and approximately 42–49× cluster efficiency) while resulting in only a 2–3% drop in testing accuracy.

1 Introduction

Deep learning methods, with deep neural networks [1, 2, 3, 4] in particular, have recently exploded in popularity as a result of their demonstrated ability to significantly improve the performance in various complex and challenging areas of research relative to other machine learning methods. However, this boost in performance of deep neural networks is largely attributed to increasingly large model sizes and complexity, resulting in growing storage and memory requirements. As a result, research into highly efficient deep neural networks has been conducted, and methods have been developed for significantly reducing the memory and computational requirements with minimal drop in performance.

Rather than attempting to compress existing neural networks directly, Shafiee *et al.* [5] took inspiration from nature and proposed *evolutionary deep intelligence* as a biologically-motivated approach for producing highly efficient deep neural networks by allowing networks to synthesize new networks with increasingly efficient architectures and naturally sparsify over successive generations. Instead of classical evolutionary computation approaches, Shafiee *et al.* introduced a novel probabilistic framework that models genetic encoding and environmental conditions via probability distributions. Biological evolutionary mechanisms are mimicked via: i) heredity, ii) natural selection, and iii) random mutation.

Current evolutionary deep intelligence methods [5, 6, 7], however, formulate the evolutionary synthesis process based on asexual reproduction. While effective at synthesizing increasingly efficient networks, asexual evolutionary synthesis results in limited network diversity and only explores a limited range of possible offspring networks as the structure of newly synthesized networks is highly constrained by its parent network. Relative to asexual reproduction, sexual two-parent reproduction allows for rapid adaptation to changing environments and has the potential to accelerate evolution by several orders of magnitude due to the increased diversity in the population [8]. Generalizing on the idea of sexual evolutionary synthesis, we explore polyploidism in deep neural networks and the effects of m-parent evolutionary synthesis on offspring network diversity and efficiency in the context of various population sizes where there are multiple potential parent network candidates.

2 Methods

In this study, we extend Shafiee *et al.*'s cluster-driven genetic encoding [6] to investigate the effects of m-parent evolutionary synthesis in varying population sizes. Two main concepts are explored: i) the effect of varying the number of parent networks, and ii) the effect of varying population size. The evolutionary deep intelligence scheme in [6] is generalized to use m parents during the evolutionary synthesis process (as shown in Figure 1). At each generation, m parent networks from the preceding generation are combined via a mating function to synthesize new offspring networks.

2.1 m-Parent Evolutionary Synthesis

Let the network architecture be formulated as $\mathcal{H}(N,S)$, where N denotes the set of possible neurons and S the set of possible synapses in the network. Each neuron $n_j \in N$ is connected to neuron $n_k \in N$ via a set of synapses $\bar{s} \subset S$, such that the synaptic connectivity $s_j \in S$ has an associated $w_j \in W$ to denote the connection's strength. In the seminal evolutionary deep intelligence paper [5], the synthesis probability $P(\mathcal{H}_g|\mathcal{H}_{g-1},\mathcal{R}_g)$ of a new network at generation g is approximated by the synaptic probability $P(S_g|W_{g-1},R_g)$ to emulate heredity through the generations of networks, and is also conditional on an environmental factor model $\mathcal{R}_g$ to imitate natural selection via a changing environment for successive generations of networks to adapt to. The synthesis probability is formulated as:

$$P(\mathcal{H}_g|\mathcal{H}_{g-1},\mathcal{R}_g) \simeq P(S_g|W_{g-1},R_g). \tag{1}$$

In the case of m-parent evolutionary synthesis, a newly synthesized network $\mathcal{H}_{g(i)}$ can be dependent on a subset of all previously synthesized networks $\mathcal{H}_{G_i}$, and is encoded as

$$P(\mathcal{H}_{g(i)}|\mathcal{H}_{G_i},\mathcal{R}_{g(i)}) \simeq P(S_{g(i)}|W_{G_i},R_{g(i)}) \tag{2}$$

where G_i is the set of network indices corresponding to previous networks on which $\mathcal{H}_{g(i)}$ is dependent, and $g(i)$ gives the generation number corresponding to the i^{th} network. Note that in the general case, the number of networks in subset $\mathcal{H}_{G_i}$ and the range of generational dependency $g(G_i)$ is only constrained by the number and generational range of already synthesized networks.

In this work, we propose a generalized form of the synthesis probability $P(\mathcal{H}_{g(i)}|\mathcal{H}_{G_i},\mathcal{R}_{g(i)})$ via the incorporation of a m-parent evolutionary synthesis process to drive network diversity and adaptability. In [6], the cluster synthesis probability $P(s_{g,c}|W_{g-1},\mathcal{R}^c_g)$ and the synapse synthesis probability $P(s_{g,i}|w_{g-1,i},\mathcal{R}^s_g)$ of the i^{th} synthesized network have been conditional on the network architecture and synaptic strength of a single parent network in the previous generation and the environmental factor models. To explore the effects of m-parent evolutionary synthesis in evolutionary deep intelligence, we reformulate the synthesis probability to combine the cluster and synapse probabilities of m parent networks $\mathcal{H}_{G_i}$ during the synthesis of an offspring network via some cluster-level mating function $\mathcal{M}_c(\cdot)$ and some synapse-level mating function $\mathcal{M}_s(\cdot)$:

$$\begin{aligned} P(\mathcal{H}_{g(i)}|\mathcal{H}_{G_i},\mathcal{R}_{g(i)}) = \\ \prod_{c \in C} \Big[P(s_{g(i),c}|\mathcal{M}_c(W_{\mathcal{H}_{G_i}}),\mathcal{R}^c_{g(i)}) \cdot \\ \prod_{j \in c} P(s_{g(i),j}|\mathcal{M}_s(w_{\mathcal{H}_{G_i},j}),\mathcal{R}^s_{g(i)}) \Big]. \end{aligned} \tag{3}$$

2.2 Mating Rituals of Deep Neural Networks in Varying Population Sizes

In the context of this study, we limit $\mathcal{H}_{G_i}$ to networks in the immediately preceding generation, i.e., for a newly synthesized network $\mathcal{H}_{g(i)}$ at generation $g(i)$, the m parent networks in $\mathcal{H}_{G_i}$ are from generation $g(i)-1$. As such, we propose the cluster-level and synapse-level mating functions to be as follows:

$$\mathcal{M}_c(W_{\mathcal{H}_{G_i}}) = \sum_{k=1}^{m} \alpha_{c,k} W_{\mathcal{H}_k} \tag{4}$$

$$\mathcal{M}_s(w_{\mathcal{H}_{G_i},j}) = \sum_{k=1}^{m} \alpha_{s,k} w_{\mathcal{H}_k,j} \tag{5}$$

where $W_{\mathcal{H}_k}$ represents the cluster's synaptic strength for the k^{th} parent network $\mathcal{H}_k \in \mathcal{H}_{G_i}$. Similarly, $w_{\mathcal{H}_k,j}$ represents the synaptic strength of a synapse j within cluster c for the k^{th} parent network $\mathcal{H}_k \in \mathcal{H}_{G_i}$.

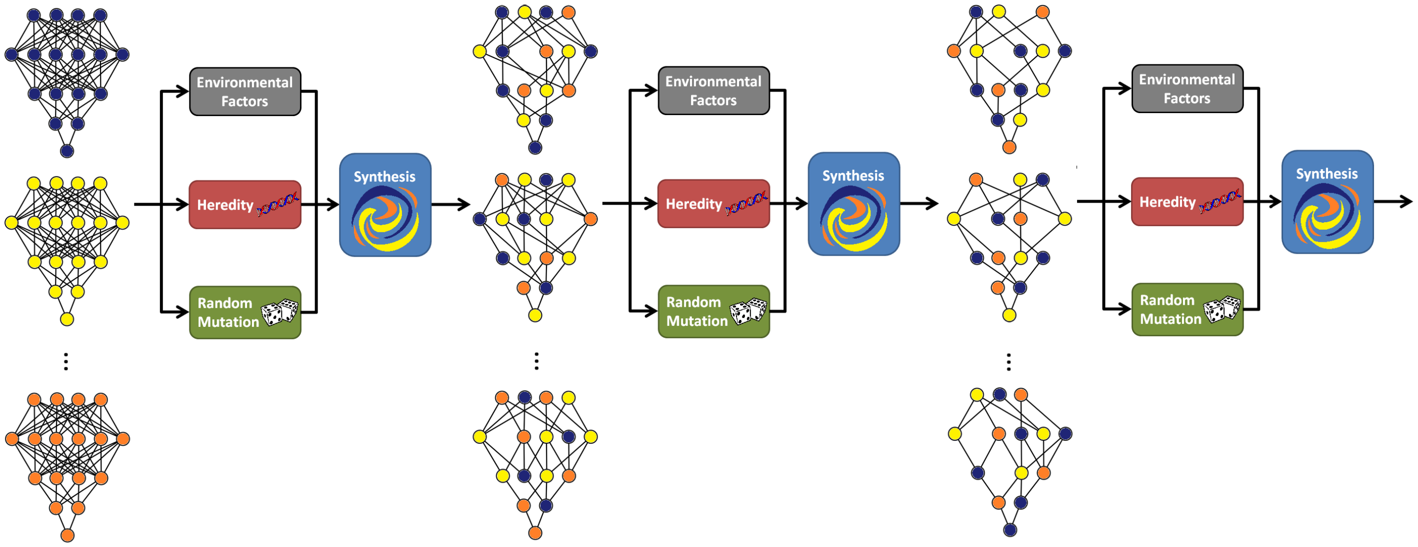


Fig. 1: The proposed evolutionary synthesis process over successive generations. The effects of m-parent evolutionary synthesis are explored via the combination of m parent networks during the synthesis of offspring networks. At each generation, m parent networks from the preceding generation are combined via a mating function to synthesize new offspring networks.

3 Results

3.1 Experimental Setup

The m-parent evolutionary synthesis of deep neural networks was performed over several generations for m = 1, 2, 3, and 5, and the effects of m-parent evolutionary synthesis in varying population sizes of three, five, and eight synthesized networks per generation were explored using a 10% subset of the MNIST [9] hand-written digits dataset with the first generation ancestor networks trained using the LeNet-5 architecture [10].

In this study, m parents were randomly selected (without replacement) from the population of networks in the immediately preceding generation and weighted equally in the mating functions $\mathcal{M}_c$ and $\mathcal{M}_s$.

Similar to Shafiee *et al.*'s work [6], we designed the environmental factor models $\mathcal{R}^c_{g(i)}$ and $\mathcal{R}^s_{g(i)}$ to enforce that an offspring deep neural network is limited to a fraction of the total number of synapses available in the previous generation, allowing for the synthesized deep neural networks to become progressively more efficient in the successive generations while minimizing any loss in accuracy. Due to the mating functions $\mathcal{M}_c$ and $\mathcal{M}_s$, newly synthesized networks inherit a network architecture structure that is the intersection of the parent network structures, and offspring networks with a higher number of parent networks result in accelerated sparsification. To simulate relatively comparable sparsification rates over generations, the environmental factor models $\mathcal{R}^c_{g(i)}$ and $\mathcal{R}^s_{g(i)}$ are modelled as functions of the number of parents m:

$$\begin{aligned}\mathcal{R}^c_{g(i)} &= (1-r_c)^m\\ \mathcal{R}^s_{g(i)} &= (1-r_s)^m.\end{aligned} \tag{6}$$

Thus, the sparsification thresholds r_c and r_s were varied such that the environmental factor models $\mathcal{R}^c_{g(i)}$ and $\mathcal{R}^s_{g(i)}$ were approximately equal across all m. Each filter (i.e., collection of kernels) was considered as a synaptic cluster in the multi-factor synapse probability model, and both the synaptic efficiency and cluster efficiency were assessed along with testing accuracy.

3.2 Experimental Results

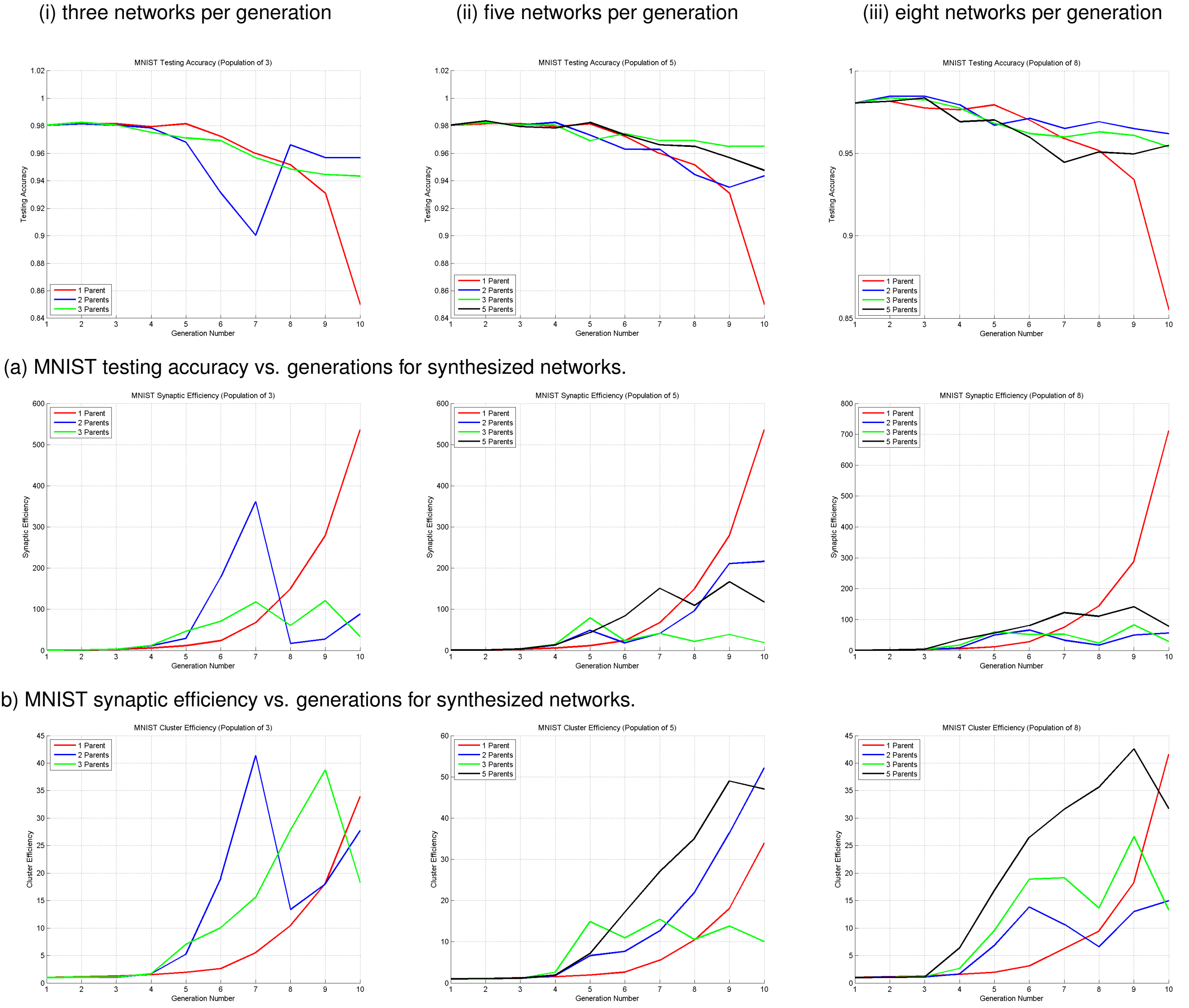
In this study, the effects of m-parent evolutionary synthesis in varying population sizes were investigated using m = 1, 2, 3, and 5, and population sizes of three, five, and eight synthesized networks per generation. At each generation, the network testing accuracy was evaluated and the corresponding architectural efficiency was assessed in terms of cluster efficiency (defined as the reduction in the total number of kernels in a network relative to the first generation ancestor network) and synaptic efficiency (defined as the reduction in the total number of synapses in a network relative to the first generation ancestor network). Figure 2 shows the testing accuracy, synaptic sparsity, and cluster sparsity of networks synthesized using m-parent evolutionary synthesis as a function of generation number, and evaluated on the subset of the MNIST dataset for populations of three, five, and eight networks per generation, respectively. Note that there is generally a trade-off between testing accuracy and architectural efficiency, i.e., testing accuracy decreases as synaptic efficiency and cluster efficiency increase. For all experiments, the original fully-trained ancestor network (generation 1) trained on 10% of the MNIST dataset had a testing accuracy of 98% with 143,136 synapses and 7,200 kernels (corresponding to a 1-channel input LeNet architecture [10]).

Note that in all cases, 1-parent evolutionary synthesis has the steepest decline in testing accuracy as generation increases, with testing accuracy decreasing approximately 13% by generation 10; in comparison, the testing accuracy of 2-parent, 3-parent, and 5-parent evolutionary synthesis remain relatively high, dropping only 2–3% by generation 10. While 1-parent evolutionary synthesis shows the most increase in synaptic sparsity (as expected due to the correspondingly steep drop in testing accuracy), 5-parent evolutionary synthesis shows an approximately 150× increase in synaptic sparsity at generation 9 for populations of five and eight networks per generation while maintaining a testing accuracy of approximately 95%, and achieved cluster efficiency of 49× and 42× for populations of five and eight networks per generation, respectively. This indicates that increasing the number of parents during evolutionary synthesis can allow for the synthesis of more efficient network architectures through increased network diversity with minimal loss in testing accuracy.

Interestingly, there appears to be noticeably more variability in the overall testing accuracy and network efficiency trends for networks synthesized using m-parent evolutionary synthesis in a population of three networks per generation, particularly in the case of the 2-parent evolutionary synthesis. Presumably anomalous, we speculate that the network structures of the parent networks were likely sufficiently diverse to allow for accelerated sparsification due to the mating functions, resulting in a rapid decrease in testing accuracy and corresponding increase in synaptic and cluster efficiency. It should be noted, then, that the combination of drastically differing network architectures could likely result in very few viable offspring networks, and that a balance between network diversity and structural consistency must be found to reach an optimal evolutionary synthesis process.

4 Discussion

In this work, we explored the effects of m-parent evolutionary synthesis in varying population sizes in an evolutionary deep intelligence approach. Overall, the use of m-parent evolutionary synthesis showed that increasing the number of parent networks results in improved architectural efficiency of the synthesized networks (approximately 150× synaptic efficiency and approximately 42–49× cluster efficiency at generation 9) while maintaining relatively high



(a) MNIST testing accuracy vs. generations for synthesized networks.

(b) MNIST synaptic efficiency vs. generations for synthesized networks.

(c) MNIST cluster efficiency vs. generations for synthesized networks.

Fig. 2: MNIST testing accuracy and network efficiency (synaptic and cluster) for a populations of three, five, and eight synthesized networks per generation using one (red), two (blue), three (green), and five (black) parents.

testing accuracy (2–3% drop by generation 10).

With the current random parent selection method, the population size at each generation does not appear to affect testing accuracy or architectural efficiency; however, the impact of increasing population size (and the resulting diversity in networks) in the case of directed parent selection is unclear, as parent networks with drastically differing architectures likely result in minimally viable offspring networks. Future work in this area includes the development of a more sophisticated parent selection method from the pool of potential parent neural networks to incorporate the notion of "survival of the fittest" [11], and the investigation of various methods for combining the parent neural networks.

Acknowledgements

This research has been supported by the Canada Research Chairs Program and Natural Sciences and Engineering Research Council of Canada (NSERC). The authors also thank Nvidia for the GPU hardware used in this study through the Nvidia Hardware Grant Program.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[3] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 6645–6649.

[4] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in *Adv ances in neural information processing systems*, 2014, pp. 1799–1807.

[5] M. J. Shafiee, A. Mishra, and A. Wong, "Deep learning with darwin: Evolutionary synthesis of deep neural networks," *arXiv preprint arXiv:1606.04393*, 2016.

[6] M. J. Shafiee and A. Wong, "Evolutionary synthesis of deep neural networks via synaptic cluster-driven genetic encoding," *NIPS Workshops*, 2016.

[7] M. J. Shafiee, E. Barshan, and A. Wong, "Evolution in Groups: A deeper look at synaptic cluster driven evolution of deep neural networks," *FTC*, 2017.

[8] J. F. Crow and M. Kimura, "Evolution in sexual and asexual populations," *American Naturalist*, pp. 439–450, 1965.

[9] Y. LeCun, C. Cortes, and C. J. Burges, "The mnist database of handwritten digits," 1998.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] G. C. Williams, *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton University Press, 2008.