Mitigating Architectural Mismatch During the Evolutionary Synthesis of Deep Neural Networks

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 has recently shown great promise for producing small, powerful deep neural network models via the organic synthesis of increasingly efficient architectures over successive generations. Existing evolutionary synthesis processes, however, have allowed the mating of parent networks independent of architectural alignment, resulting in a mismatch of network structures. We present a preliminary study into the effects of architectural alignment during evolutionary synthesis using a gene tagging system. Surprisingly, the network architectures synthesized using the gene tagging approach resulted in slower decreases in performance accuracy and storage size; however, the resultant networks were comparable in size and performance accuracy to the non-gene tagging networks. Furthermore, we speculate that there is a noticeable decrease in network variability for networks synthesized with gene tagging, indicating that enforcing a like-with-like mating policy potentially restricts the exploration of the search space of possible network architectures.

1   Introduction

Deep neural networks (DNNs) [3, 1, 2, 4] have recently garnered widespread interest due to their demonstrated ability to improve state-of-the-art performance in many challenging areas of research. This boost in performance, however, has largely been attributed to increasingly large and complex network architectures, resulting in storage and memory requirements beyond the resources available in practical scenarios. As such, methods for producing highly efficient DNNs have been developed to reduce the memory and computational needs while maintaining performance accuracy.

Inspired by nature, Shafiee *et al.* [5] introduced *evolutionary deep intelligence* as a biologically-motivated alternative to compressing existing DNNs directly by allowing networks to organically synthesize new and increasingly efficient network architectures over successive generations. Shafiee *et al.* proposed a probabilistic framework that models the genetic encoding of networks and simulated environmental factors as probability distributions, and mimicked evolutionary mechanisms using: i) heredity, ii) natural selection, and iii) random mutation.

While existing evolutionary deep intelligence methods have leveraged asexual evolutionary synthesis [5, 6, 7, 8, 9] and m-parent sexual evolutionary synthesis [10, 11, 12], the evolutionary synthesis process combines architectural structures sequentially regardless of their relative positions in each parent network. As such, current evolutionary deep intelligence methods have employed synthesis processes that mate parent networks independent of architectural alignment, resulting in a mismatch of network structures.

In this paper, we present a preliminary study into the effects of architectural alignment during evolutionary synthesis via the introduction of a gene tagging system. Gene tagging is explored within the context of m-parent sexual evolutionary synthesis and evaluated over a range of environmental resource models. The gene tagging system allows for the proper alignment of architectural structures that originated from the same location in the ancestor network, and enforces a like-with-like mating policy.

2   Methods

In this study, we modify the existing m-parent evolutionary synthesis approach [11] to investigate the efficacy of structural alignment and its effect on the synthesis of new network architectures across a range of simulated environmental resource models. This is explored via the incorporation (i.e., structurally aligned) or omission (i.e., structurally mismatched) of gene tagging during the evolutionary synthesis process. At each generation, m parent networks from the preceding generation are structurally aligned using gene tagging and combined via a like-with-like 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 is the set of possible neurons and S denotes 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. $P(\mathcal{H}_g|\mathcal{H}_{g-1},\mathcal{R}_g)$ is also conditional on an environmental factor model $\mathcal{R}_g$ to imitate natural selection via simulated environmental resources. The synthesis probability is formulated as:

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

Generalizing to m-parent evolutionary synthesis [11], 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 formulated 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 the set of network indices G_i corresponds to the set of previous networks on which $\mathcal{H}_{g(i)}$ is dependent, and $g(i)$ represents the generation number corresponding to the ith 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.

The synthesis probability to combine the cluster and synapse probabilities of m parent networks $\mathcal{H}_{G_i}$ is represented by some cluster-level mating function $\mathcal{M}_c(\cdot)$ and some synapse-level mating function $\mathcal{M}_s(\cdot)$:

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

These mating functions are formulated using an intersection-based mating policy [11], i.e.,:

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

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

where $W_{\mathcal{H}_k}$ represents the cluster's synaptic strength for the kth 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 kth parent network $\mathcal{H}_k \in \mathcal{H}_{G_i}$.

2.2   Mitigating Architecture Mismatch via Gene Tagging

To encourage like-with-like mating during evolutionary synthesis, this study introduces a gene tagging system to enforce structural alignment, i.e., only mating architectural clusters originating from the same location in the ancestor network. As such, the cluster-level and synapse-level mating functions proposed in [11] are reformulated as follows:

$$\mathcal{M}_c(W_{\mathcal{H}_{G_i}}) = \prod_{k=1}^{\mathcal{K}_c} \alpha_{c,k} W_{\mathcal{H}_k} \tag{6}$$

$$\mathcal{M}_s(w_{\mathcal{H}_{G_i},j}) = \prod_{k=1}^{\mathcal{K}_c} \alpha_{s,k} w_{\mathcal{H}_k,j} \tag{7}$$
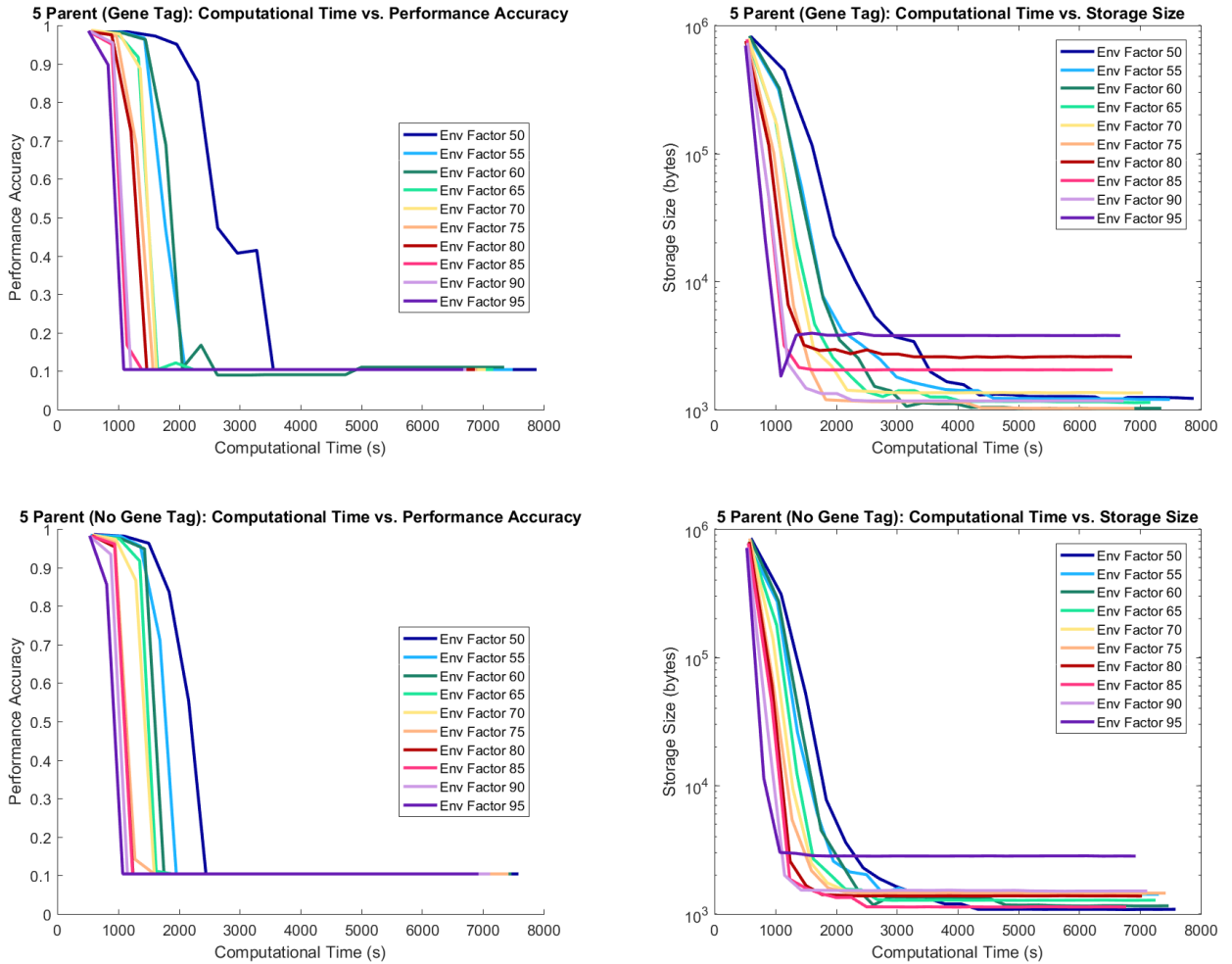
Fig. 1: Performance accuracy (left) and storage size (right) for 5-parent evolutionary synthesis with gene tagging (top row) and no gene tagging (bottom row) using various environmental factor models. Plots best viewed in colour.

where $\mathcal{K}_c$ is the subset of parent networks with existing architectural clusters corresponding to a single gene tagged cluster $c \in C$, and C is the set of clusters that exists in $\mathcal{H}_{g(i)}$.

This study also introduces a parameter-based lower bound on the subset of parent networks with existing architectural clusters $\mathcal{K}_c$. Prior studies [10, 11, 12] employed an intersection-based mating policy where a cluster was synthesized only if the cluster existed within all m parent networks (i.e., $\mathcal{K}_c = m$). To allow for more flexibility within the m-parent evolutionary synthesis, we propose a *%POP* parameter to control the proportion of parent network architectures that must contain any given cluster during evolutionary synthesis. As such, only clusters that exist within the specified proportion of parent networks are synthesized (i.e., $m\%POP \leq \mathcal{K}_c \leq m$), and the intersection-based policy can be achieved using $\%POP = 1$.

3 Results

3.1 Experimental Setup

The effect of structural alignment during the m-parent evolutionary synthesis was evaluated using 10% of the MNIST [13] handwritten digits dataset (Figure 2) to increase the training speed of the synthesized network architectures and increase the inherently low intra-class variation within the MNIST dataset. For this preliminary study, 5-parent evolutionary synthesis and a *%POP* of 60% was used, and $\mathcal{H}_{G_i}$ was restricted to networks in the immediately preceding generation, i.e., for a newly synthesized network $\mathcal{H}_{g(i)}$, the $m = 5$ parent networks in $\mathcal{H}_{G_i}$ are from generation $g(i) - 1$.

The first generation ancestor network was trained using the LeNet-5 architecture [14], and the synthesized offspring networks were assessed using performance accuracy on the MNIST dataset and storage size (representative of the architectural efficiency of a network) of the networks with respect to the computational time
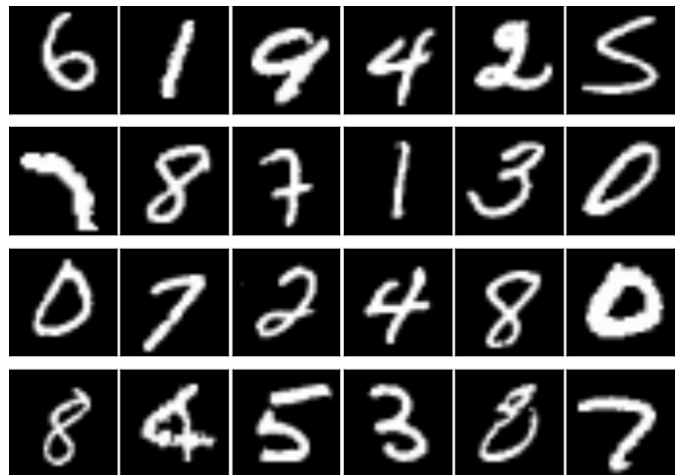

Fig. 2: Sample images from the MNIST hand-written digits dataset [13].

required. Similar to [12], each filter was considered as a synaptic cluster in the multi-factor synapse probability model, and the cluster-level environmental factor model $\mathcal{R}^c_{g(i)}$ and the synapse-level environmental factor model $\mathcal{R}^s_{g(i)}$ were varied together from 50% to 95% at 5% increments, i.e.,:

$$\mathcal{R}^c_{g(i)}, \mathcal{R}^s_{g(i)} = \{50, 55, 60, 65, 70, 75, 80, 85, 90, 95\}\% \qquad (8)$$

The synthesized networks were assessed using performance accuracy on the MNIST dataset and storage size (representative of the architectural efficiency of a network) of the networks with respect to the computational time required.
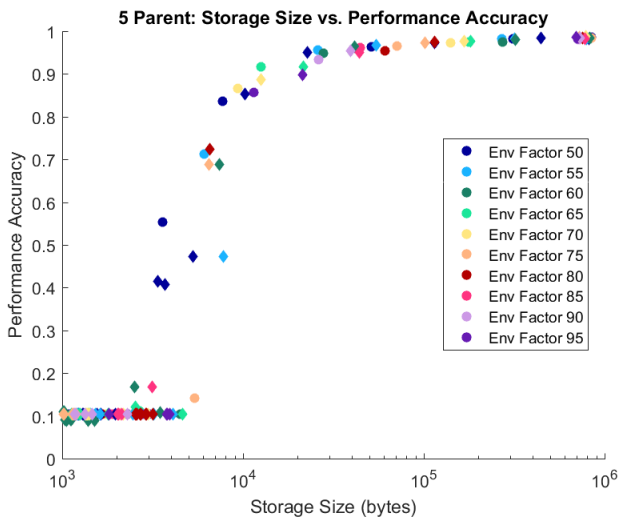
Fig. 3: Performance accuracy as a function of storage size for 5-parent sexual evolutionary synthesis using various environmental factor models. Networks synthesized using gene tagging (diamond) show minimal to no noticeable difference relative to networks synthesized without gene tagging (round) in terms of maintaining performance accuracy while decreasing storage size. Plots best viewed in colour.

3.2 Experimental Results

Figure 1 shows the performance accuracy and storage size for 5-parent evolutionary synthesis with gene tagging (top row) and no gene tagging (bottom row) given various cluster-level and synapse-level environmental factor models. The incorporation of gene tagging resulted in a more gradual decrease in both performance accuracy and storage size relative to computational time, while the omission of gene tagging produced network architectures with more rapidly decreasing performance accuracy and storage size. Note that the bottom plateau in performance accuracy is at 10% akin to random guessing, as the MNIST dataset consists of 10 classes of handwritten digits.

While there is no inherent limit in network reduction when using the evolutionary deep intelligence approach, Figure 1 shows that there are natural plateaus in network storage size, particularly with the higher environmental factor models, i.e., environmental factor models of 80%, 85%, and 95%. In addition, note that these plateaus in network storage size tend to occur more when synthesizing networks using gene tagging.

Figure 3 shows performance accuracy as a function of storage size for 5-parent sexual evolutionary synthesis using various cluster-level and synapse-level environmental factor models, where the best synthesized networks are closest to the top left corner, i.e., high performance accuracy and low storage size. Networks synthesized using gene tagging (diamond points) appear to be minimally worse relative to networks synthesized without gene tagging (round points) in terms of maintaining performance accuracy while decreasing storage size.

4 Discussion

In this work, we presented a preliminary study into the effects of architectural alignment during evolutionary synthesis via the introduction of a gene tagging system to enforce a like-with-like mating policy. Surprisingly, the network architectures synthesized using the gene tagging approach resulted in slower decreases in performance accuracy and storage size. However, the resultant networks were comparable in storage size and performance accuracy to the networks synthesized without gene tagging.

We speculate that the use of gene tagging inherently decreases overall network variability across the synthesized network architectures, indicating that enforcing a like-with-like mating policy via the use of gene tagging potentially restricts the exploration of the search space of possible network architectures. As a result, synthesized networks may fail to achieve the same increases in architectural efficiency as architectures synthesized with more random approaches. Future work includes further investigation into the ef-

fects of architectural alignment via the use of different $\%POP$s and varying the number of parents networks m, and the exploration of how gene tagging potentially affects the architectural variability in synthesized networks.

References

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

[2] 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.

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

[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," *Neural Processing Letters*, 2017.

[6] M. J. Shafiee and A. Wong, "Evolutionary synthesis of deep neural networks via synaptic cluster-driven genetic encoding," in *Advances in Neural Information Processing Systems*, 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," *arXiv preprint arXiv:1704.02081*, 2017.

[8] M. J. Shafiee, F. Li, and A. Wong, "Exploring the imposition of synaptic precision restrictions for evolutionary synthesis of deep neural networks," *arXiv preprint arXiv:1707.00095*, 2017.

[9] M. J. Shafiee, E. Barshan, F. Li, B. Chwyl, M. Karg, C. Scharfenberger, and A. Wong, "Learning efficient deep feature representations via transgenerational genetic transmission of environmental information during evolutionary synthesis of deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 979–986.

[10] A. G. Chung, M. Javad Shafiee, P. Fieguth, and A. Wong, "The mating rituals of deep neural networks: Learning compact feature representations through sexual evolutionary synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1220–1227.

[11] A. G. Chung, P. Fieguth, and A. Wong, "Polyploidism in deep neural networks: m-parent evolutionary synthesis of deep neural networks in varying population sizes," *Journal of Computational Vision and Imaging Systems*, vol. 3, no. 1, 2017.

[12] A. G. Chung, P. Fieguth, and A. Wong, "Nature vs. nurture: The role of environmental resources in evolutionary deep intelligence," *Computer and Robot Vision (CRV), 2018 15th Conference on*, 2018.

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

[14] 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.