Assessing Architectural Similarity in Populations of Deep Neural Networks

Audrey G. Chung*, Paul Fieguth, and Alexander Wong

Vision and Image Processing Research Group, University of Waterloo, Waterloo, ON, Canada

Waterloo Artificial Intelligence Institute, University of Waterloo, Waterloo, ON, Canada

*agchung@uwaterloo.ca

Abstract

Evolutionary deep intelligence has recently shown great promise for producing small, powerful deep neural network models via the synthesis of increasingly efficient architectures over successive generations. Despite recent research showing the efficacy of multi-parent evolutionary synthesis, little has been done to directly assess architectural similarity between networks during the synthesis process for improved parent network selection. In this work, we present a preliminary study into quantifying architectural similarity via the percentage overlap of architectural clusters. Results show that networks synthesized using architectural alignment (via gene tagging) maintain higher architectural similarities within each generation, potentially restricting the search space of highly efficient network architectures.

1. Introduction

The use of deep neural networks (DNNs) [1, 7] has become ubiquitous over the last few years due to their demonstrated efficacy in many challenging application areas, including image classification [6, 12], pose estimation [13, 9], and speech recognition [5, 14]. However, the modelling accuracy of high-performance DNNs is a result of increased model size and complexity, rendering them impractical for real-world scenarios with limited computational and memory resources. As a result, methods for reducing the computational requirements of DNNs while maintaining performance accuracy are highly desirable.

One such method is *evolutionary deep intelligence* [10]. Inspired by nature, Shafiee *et al.* proposed a biologically-motivated method for synthesizing increasingly efficient and compact network architectures over successive generations from existing high-performance DNNs. While the seminal papers in evolutionary deep intelligence [10, 11] formulated the synthesis process as asexual evolutionary synthesis, recent work [4, 2] has investigated the use of sexual evolutionary synthesis to produce populations of increasingly compact DNNs at each generations.

Most recently, Chung *et al.* [3] conducted an initial study into mitigating architectural mismatch during sexual evolutionary synthesis via a gene tagging system. While results showed no notable difference in performance accuracy, it raises an interesting question: how can we assess the architectural similarity of DNNs in a meaningful and useful way?

In this work, we present a preliminary study exploring the quantification of network architectural similarity in populations of evolutionary synthesized neural networks via percentage overlap of architectural clusters. Architectural similarity is explored within the context of multi-parent sexual evolutionary synthesis, and will allow for the development of improved similarity-based mating policies during the evolutionary synthesis of highly efficient networks.

2. Methods

In this paper, we investigate the quantification of architectural similarity using generations of networks synthesized via multi-parent evolutionary synthesis with and without gene tagging [3].

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 [11], 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.

Extending on [10, 11], Chung *et al.* [4] generalized the synthesis process multi-parent (m-parent) evolutionary synthesis where a newly synthesized network $\mathcal{H}_{g(i)}$ can be dependent on a subset of all previously synthesized networks $\mathcal{H}_{G_i}$, with G_i corresponding to the set of previous networks

on which $\mathcal{H}_{g(i)}$ is dependent and $g(i)$ representing the generation number corresponding to the ith network.

The synthesis probability combining the 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 $\mathcal{M}_s(\cdot)$:

$$P(\mathcal{H}_{g(i)}|\mathcal{H}_{G_i}, \mathcal{R}_{g(i)}) = \prod_{C \in \mathcal{C}} \Big[P(s_{g(i),C}|\mathcal{M}_c(W_{\mathcal{H}_{G_i}}), \mathcal{R}^c_{g(i)}) \cdot$$
$$\prod_{j \in C} P(s_{g(i),j}|\mathcal{M}_s(w_{\mathcal{H}_{G_i},j}), \mathcal{R}^s_{g(i)}) \Big].$$
(1)

2.2. Architecture Alignment via Gene Tagging

To encourage like-with-like mating during evolutionary synthesis, Chung *et al*. [3] recently introduced 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 are formulated as follows:

$$\mathcal{M}_c(\overline{W}_{\mathcal{H}_{G_i}}) = \prod_{k \in \mathcal{K}_c} \alpha_{c,k} \overline{W}_{\mathcal{H}_k} \qquad (2)$$

$$\mathcal{M}_s(\overline{w}_{\mathcal{H}_{G_i},j}) = \prod_{k \in \mathcal{K}_c} \alpha_{s,k} \overline{w}_{\mathcal{H}_k,j} \qquad (3)$$

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

2.3. Architectural Cluster Overlap

To investigate the quantification of architectural similarity in the context of multi-parent sexual evolutionary synthesis, the percentage overlap of architectural clusters between two networks is formulated as the proportion of intersecting clusters:

$$\%overlap_{AB} = \frac{C_A \cap C_B}{C_A}, \qquad (4)$$

where C_A and C_B represent the sets of architectural clusters that exist in the two networks being compared.

Percentage overlap of architectural clusters is an intuitive representation of network architecture similarity made viable in the context of multi-parent evolutionary synthesis by leveraging the gene tagging system [3]. As such, gene tagging (which allows for architectural alignment during evolutionary synthesis) can similarly be used to calculate percentage overlap of existing architectural clusters originating from the same location in the ancestor network. Percentage overlap is indicative of network population diversity within a generation, *e.g.*, relatively low average percentage overlap would indicate a generation of synthesized networks with comparatively higher architectural variability.

3. Results

3.1. Experimental Setup

In this study, we used the network architectures synthesized in [3] with the least aggressive environmental factor model ($R^c_{g(i)}, R^s_{g(i)} = 50$) and trained on the MNIST dataset [8]. Architectural similarity was assessed on the first seven generations of networks (after which the performance accuracy degraded to random guessing) synthesized with and without gene tagging.

3.2. Experimental Results

Figure 1 shows the performance accuracy as a function of storage size for the populations of synthesized networks in the first seven generations, where the best synthesized networks are closest to the top left, *i.e.*, high performance accuracy and low storage size. Networks synthesized using gene tagging show a slightly slower progression in maintaining performance accuracy while decreasing storage size relative to networks synthesized without gene tagging.

Synthesizing networks with gene tagging and without gene tagging both produced architectures that increase in variability over successive generations; however, networks synthesized with gene tagging diversify more slowly than those without gene tagging (as shown in Table 1). Figure 1 and Table 1 also suggest that generations of networks approaching an optimal tradeoff between performance accuracy and storage size tend to also have the highest architectural variability, *e.g.*, in generations 3 and 4.

Lastly, it is worth noting that the increasing percentage overlap in generations 6 and 7 of networks synthesized without gene tagging is a result of sparse, low-variability architectures that can no longer represent the problem space (*i.e.*, performance accuracy of 10% on the 10-class MNIST dataset, equivalent to random guessing). Similarly, the percentage overlap in generations 6 and 7 of networks synthesized with gene tagging increases as the performance accuracy begins to rapidly decrease.

4. Discussion

We presented a preliminary study in assessing architectural similarity between deep neural networks to improve the sexual evolutionary synthesis process. Results show that networks synthesized using gene tagging have less architectural variability than networks synthesized without gene tagging, as quantified by relatively higher overlap percentages of architectural clusters. This indicates that the use of gene tagging is potentially restricting the exploration of highly efficient network architectures in the search space. Future work includes further investigation into quantities of information, *e.g.*, mutual information, as well as the development of a custom similarity metric for optimal architectural similarity during sexual evolutionary synthesis.

Table 1. Average percentage overlap of architectural clusters in network models for the first seven generations of 5-parent sexual evolutionary synthesis. Note that the increasing percentage overlap in generations 6 and 7 of networks synthesized without gene tagging is a result of sparse, low-variability architectures that can no longer represent the problem space, while the unpredictability of percentage overlap in generations 6 and 7 of networks synthesized with gene tagging may be a result of some (but not all) networks having sparse, low-variability architectures.

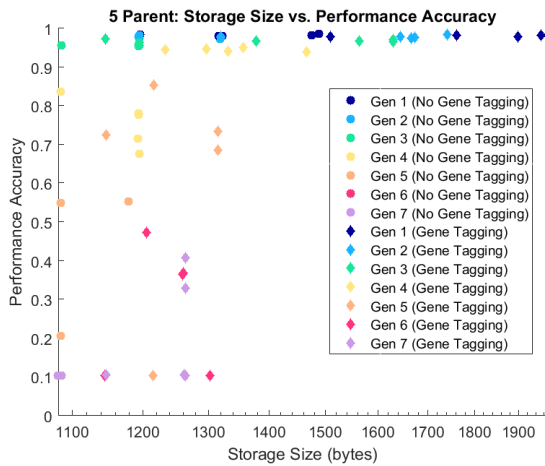| Gen No.	Gene Tagging	No Gene Tagging
1	93.75%	93.71%
2	87.59%	78.11%
3	83.49%	68.84%
4	71.81%	66.64%
5	73.17%	68.44%
6	69.09%	82.74%
7	73.48%	91.05%

Figure 1. Performance accuracy as a function of storage size for the first seven generations of 5-parent sexual evolutionary synthesis for networks synthesized with gene tagging (diamond) and without gene tagging (round). Plots best viewed in colour.

References

[1] Y. Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[2] 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*, 3(1), 2017.

[3] A. G. Chung, P. Fieguth, and A. Wong. Mitigating architectural mismatch during the evolutionary synthesis of deep neural networks. *arXiv preprint arXiv:1811.07966*, 2018.

[4] 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 International Conference on Computer Vision*, pages 1220–1227, 2017.

[5] 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*, pages 6645–6649. IEEE, 2013.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[7] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[8] Y. LeCun, C. Cortes, and C. Burges. MNIST handwritten digit database. 1998.

[9] A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision*, pages 483–499. Springer, 2016.

[10] M. J. Shafiee, A. Mishra, and A. Wong. Deep learning with darwin: evolutionary synthesis of deep neural networks. *Neural Processing Letters*, 48(1):603–613, 2018.

[11] M. J. Shafiee and A. Wong. Evolutionary synthesis of deep neural networks via synaptic cluster-driven genetic encoding. *Advances in Neural Information Processing Systems*, 2016.

[12] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[13] 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 *Advances in neural information processing systems*, pages 1799–1807, 2014.

[14] W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, and A. Stolcke. The microsoft 2017 conversational speech recognition system. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5934–5938. IEEE, 2018.