This ICCV workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

The Mating Rituals of Deep Neural Networks: Learning Compact Feature Representations through Sexual Evolutionary Synthesis

Audrey G. Chung Mohammad Javad Shafiee Paul Fieguth Alexander Wong
University of Waterloo
200 University Ave W, Waterloo, ON, N2L 3G1
agchung@uwaterloo.ca

Abstract

Evolutionary deep intelligence was recently proposed as a method for achieving highly efficient deep neural network architectures over successive generations. Drawing inspiration from nature, we propose the incorporation of sexual evolutionary synthesis. Rather than the current asexual synthesis of networks, we aim to produce more compact feature representations by synthesizing more diverse and generalizable offspring networks in subsequent generations via the combination of two parent networks. Experimental results were obtained using the MNIST and CIFAR-10 datasets, and showed improved architectural efficiency and comparable testing accuracy relative to the baseline asexual evolutionary neural networks. In particular, the network synthesized via sexual evolutionary synthesis for MNIST had approximately double the architectural efficiency (cluster efficiency of 34.29× and synaptic efficiency of 258.37×) in comparison to the network synthesized via asexual evolutionary synthesis, with both networks achieving a testing accuracy of ∼97%.

1. Introduction

Deep learning methods, especially deep neural networks [2, 9, 15, 27], have recently exploded in popularity due to their demonstrated ability to significantly improve the performance over other machine learning methods in various challenging areas of research. However, this boost in performance of deep neural networks is largely attributed to increasingly large model sizes, resulting in growing storage and memory requirements.

These computational requirements make high-performance deep neural networks infeasible for devices without access to cloud computing. For many practical situations such as self-driving cars and smartphone applications, the available computing resources are limited to low-power, embedded GPUs and CPUs; with such limited computational power and storage, smaller and more compact versions of deep neural networks are highly desirable. As such, research into compact feature representations via 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.

One of the first approaches for adapting the size of a neural network was optimal brain damage [18]. The method removed unimportant weights (as determined using the second derivative of the objective function as a saliency approximation of a parameter) from the network to improve network generalizability, increase the speed of learning, and reduce the number of training sampled required.

Gong *et al.* [8] proposed a network compression framework where vector quantization was leveraged to shrink the storage requirements of deep neural networks trained for computer vision tasks. Gong *et al.* noted that vector quantization has clear advantages over existing matrix factorization methods, and found a good balance between model size and accuracy could be achieved via the application of k-means clustering to the weights or product quantization.

Han *et al.* [11] introduced deep compression to address the limitations of computational power and memory that comes with embedded systems via a three stage pipeline: pruning, trained quantization, and Huffman coding. The method reduced the storage requirements of a neural network by 35x to 49x with no loss in accuracy. Han *et al.* [10] also reduced the storage and computational requirements of neural networks with no drop in accuracy by training a network to learn which weights are important, pruning the unimportant connections, and retraining the network to fine tune the remaining weights.

Another method for deep compression is hashing [3], which uses a low-cost hash function to group network weights into hash buckets with a single shared parameter value. Exploiting the redundancy in both network layers, the HashedNet architecture leverages the idea of weight-sharing and allows for considerable savings in terms of

memory and storage.

Other methods for reducing the computational requirements of neural networks include low rank approximations [12, 13]. Jaderberg *et al.* [13] used low-rank expansions to speed up the computation of deep neural networks (specifically the convolutional layers of convolutional neural networks) by exploiting cross-channel or filter redundancy to construct a low rank basis of filters. Similarly, Ioannou *et al.* [12] created computationally efficient networks using low rank representations of convolutional filters by learning a set of small basis filters that are then combined into more complex filters.

Sparsity learning [5, 19, 28] is another strategy used to sparsify deep neural networks. Feng and Darrell [5] demonstrated a novel method for learning components of the structure of a neural network by incorporating the Indian Buffet Process prior; especially effective when there is limited labelled training data, this method captures complex data distributions in an unsupervised generative manner. Liu *et al.* [19] showed how to reduce the number of parameters in neural networks via sparse decomposition by exploiting both intra-channel and inter-channel redundancy. Lastly, Wen *et al.* [28] recently proposed a Structured Sparsity Learning (SSL) method to regularize the structures within deep neural networks (e.g., filters shapes, channels, layer depth).

Rather than attempting to compress existing deep neural networks into smaller and more compact representations directly, Shafiee *et al.* [21] proposed an entirely novel concept: *Can deep neural networks naturally evolve to be highly efficient?* Inspired by biological evolution, Shafiee *et al.* developed an *evolutionary deep intelligence* approach to produce highly efficient and compact deep neural networks by allowing these networks to synthesize new networks with increasingly compact representations and naturally sparsify over successive generations. Biological evolutionary mechanisms are mimicked via three computational constructs: i) heredity, ii) natural selection, and iii) random mutation.

While previous studies [1, 7, 24, 25, 26] have been conducted that leverage the idea of using evolutionary techniques to generate and train neural networks, there are key differences between these and the evolutionary deep intelligence method proposed by Shafiee *et al.* [21]. Past works have primarily focused on improving a network's training and accuracy, while evolutionary deep intelligence shifts the focus to organically synthesizing networks with high architectural efficiency. In addition, these previous studies use classical evolutionary computation approaches such as genetic algorithms and evolutionary programming, while Shafiee *et al.* introduced a novel probabilistic framework that models genetic encoding and environmental conditions via probability distributions.

More recently, Shafiee *et al.* proposed a modification of the original evolutionary deep intelligence approach via synaptic cluster-driven genetic encoding [22]. Further investigating the genetic encoding scheme used to mimic heredity, Shafiee *et al.* proposed the incorporation of synaptic clustering into the genetic encoding scheme, and introduced a multi-factor synapse probability model. Modelling the synaptic probability as a product of the probability of synthesis of a particular cluster of synapses and the probability of synthesis of a particular synapse within the cluster, this new genetic encoding scheme demonstrated state-of-the-art performance while producing significantly more efficient network architectures and compact feature representations specifically tailored for GPU-accelerated applications.

The current work in evolutionary deep intelligence [21, 22], however, formulates the evolutionary synthesis process based on asexual reproduction; that is, offspring neural networks are synthesized by stochastically sparsifying a clone of their parent network. While effective at synthesizing efficient networks with comparable testing accuracies, asexual evolutionary synthesis results in a limited range of possible offspring networks as the offspring network structure is highly constrained by the parent network. Motivated by the aim of promoting diversity in evolutionary deep neural networks, we explore the use of sexual reproduction when synthesizing offspring network architectures in this study.

Evolutionarily speaking, sexual reproduction is thought to have developed in living organisms due to the fact that it favours the survival of groups rather than individuals by allowing for accelerated adaptation to changing environments via the combination of mutations occurring in distinct individuals in a single descendant [6, 20]. Relative to asexual reproduction, sexual reproduction has the potential to accelerate evolution by several orders of magnitude [4], with its effects most prominent in a large population with a high frequency of beneficial mutations. This motivates the idea that the use of sexual reproduction in evolutionary synthesis can accelerate the generation-by-generation development of useful compact feature representations.

To evaluate the validity of sexual reproduction in evolutionary deep intelligence, we propose an extension of Shafiee *et al.*'s cluster-driven genetic encoding approach [22] via the incorporation of a second parent network during the synthesis of an offspring network at each generation. The methodology for the proposed model is described in Section 2. The experimental setup and results are presented in Section 3. Lastly, conclusions and future work are discussed in Section 4.

2. Methods

In this work, we propose an extension of Shafiee *et al.*'s cluster-driven genetic encoding [22] via an adaptation of

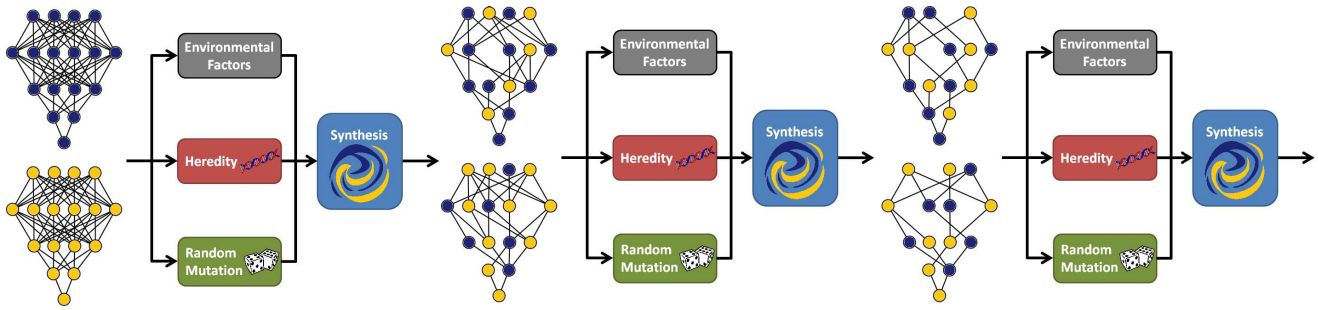


Figure 1: The proposed evolutionary synthesis process over successive generations as an extension of cluster-driven genetic encoding. The effects of sexual evolutionary synthesis are explored via the incorporation of a second parent network during the synthesis of offspring networks. At each generation, two parent networks from the preceding generation are combined via a mating function to synthesize new offspring networks.

the network synthesis process towards more compact feature representations using sexual evolutionary synthesis. To explore the effects of sexual evolutionary synthesis when synthesizing offspring network architectures, we augment the evolutionary deep intelligence scheme in [22] to incorporate a second parent network during the process of synthesizing new offspring networks as shown in Figure 1. At each generation, two parent networks from the preceding generation are combined via a mating function to synthesize new offspring networks containing information from both parent networks.

2.1. Sexual 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 paper on evolutionary deep intelligence [21], 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. That is, the synthesis probability can be formulated as follows:

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

More recently, Shafiee *et al.* [22] reformulated the synthesis probability to incorporate the multi-factor synaptic probability model and different quantitative environmental factor models at the synapse and cluster levels:

$$P(\mathcal{H}_g|\mathcal{H}_{g-1}, \mathcal{R}_g) = \prod_{c \in C} \left[P(s_{g,c}|W_{g-1}, \mathcal{R}_g^c) \cdot \prod_{j \in c} P(s_{g,j}|w_{g-1,j}, \mathcal{R}_g^s) \right] \tag{2}$$

where $\mathcal{R}_g^c$ and $\mathcal{R}_g^s$ represent the environmental factor models enforced during the synaptic cluster synthesis and the synapse synthesis, respectively. $P(s_{g,c}|W_{g-1}, \mathcal{R}_g^c)$ represents the probability of synthesis for a given cluster of synapses $s_{g,c}$; that is, $P(s_{g,c}|W_{g-1}, \mathcal{R}_g^c)$ denotes the likelihood that a synaptic cluster $s_{g,c}$ (for all clusters $c \in C$) will exist in the network architecture in generation g given the cluster's synaptic strength in generation $g-1$ and the cluster-level environmental factor model. Comparably, $P(s_{g,j}|w_{g-1,j}, \mathcal{R}_g^s)$ represents the likelihood of the existence of synapse j within the synaptic cluster c in generation g given the synaptic strength in the previous generation $g-1$ and synapse-level environmental factor model. This multi-factor probability model encourages both the persistence of strong synaptic clusters and the persistence of strong synaptic connectivity over successive generations [22].

With asexual evolutionary synthesis, however, a limited range of possible offspring networks is explored as the structure of each network is constrained by its parent network. Taking inspiration from nature, we aim to increase the diversity and compactness of evolutionary deep neural networks by incorporating information from multiple parent networks when synthesizing a new offspring network. In addition to increasing diversity, previous work in the field of evolutionary biology has concluded that sexual reproduction will accelerate adaptation to a new environment given that the genetic variance arises due to a changing environment [23]. This motivates the idea that more efficient and diverse offspring networks with increasingly compact feature representations can be synthesized in fewer generations using sexual evolutionary synthesis, particularly in the case

of non-stationary environmental factor models.

In this work, we propose a further modification of the synthesis probability $P(\mathcal{H}_g|\mathcal{H}_{g-1}, \mathcal{R}_g)$ via the incorporation of a two-parent synthesis process to drive network diversity and adaptability by mimicking sexual reproduction. Thus far for the ith synthesized network, the cluster synthesis probability $P(s_{g,c}|W_{g-1}, \mathcal{R}_g^c)$ and the synapse synthesis probability $P(s_{g,i}|w_{g-1,i}, \mathcal{R}_g^s)$ 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 sexual evolutionary synthesis in evolutionary deep intelligence, we reformulate the synthesis probability to combine the cluster and synapse probabilities of two parent networks, e.g., $\mathcal{H}_A$ and $\mathcal{H}_B$, during the synthesis of an offspring network via some cluster-level mating function $\mathcal{M}_c(\cdot)$ and some synapse-level mating function $\mathcal{M}_s(\cdot)$:

$$P(\mathcal{H}_{g,i}|\mathcal{H}_A, \mathcal{H}_B, \mathcal{R}_g) = \prod_{c \in C} \Big[P(s_{g,c}|\mathcal{M}_c(W_{\mathcal{H}_A}, W_{\mathcal{H}_B}), \mathcal{R}_{g(i)}^c) \cdot \prod_{j \in c} P(s_{g,j}|\mathcal{M}_s(w_{\mathcal{H}_A,j}, w_{\mathcal{H}_B,j}), \mathcal{R}_{g(i)}^s) \Big]. \quad (3)$$

2.2. Mating Rituals of Deep Neural Networks

In this work, we restrict the the parent networks, $\mathcal{H}_A$ and $\mathcal{H}_B$, to the immediately preceding generation; that is, for an offspring $\mathcal{H}_{g,i}$ at generation g, the parent networks $\mathcal{H}_A$ and $\mathcal{H}_B$ are from generation $g-1$. We propose the cluster-level and synapse-level mating functions to be as follows:

$$\mathcal{M}_c(W_{\mathcal{H}_A}, W_{\mathcal{H}_B}) = \alpha_c W_{\mathcal{H}_A} + \beta_c W_{\mathcal{H}_B} \quad (4)$$

$$\mathcal{M}_s(w_{\mathcal{H}_A,j}, w_{\mathcal{H}_B,j}) = \alpha_s w_{\mathcal{H}_A,j} + \beta_s w_{\mathcal{H}_B,j} \quad (5)$$

where $W_{\mathcal{H}_A}$ and $W_{\mathcal{H}_B}$ represent the cluster's synaptic strength for parent networks $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Similarly, $w_{\mathcal{H}_A,j}$ and $w_{\mathcal{H}_B,j}$ represent the synaptic strength of a synapse j within cluster c for parent networks $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.

2.3. Realization of Genetic Encoding

In this study, we employ the simple realization of cluster-driven genetic encoding proposed in [22]. The probability of synthesis for a given synapse cluster $s_{g,c}$ is realized as:

$$P(s_{g,c}|W_{g-1}) = \exp\Big(\frac{\Sigma_{i \in c} \lfloor w_{g-1,i} \rfloor}{Z} - 1\Big) \quad (6)$$

where $\lfloor \cdot \rfloor$ encodes a synaptic weight truncation and Z is the normalization factor required to construct a probability distribution, i.e., $P(s_{g,c}|W_{g-1}) \in [0, 1]$. The truncation of synaptic weights lessens the impact of weak synapses in a synaptic cluster.

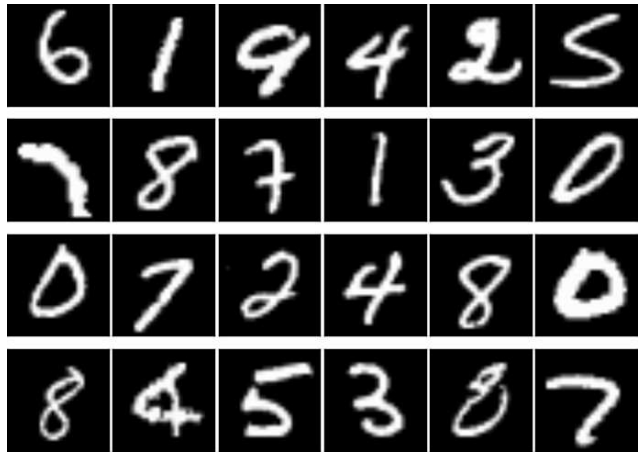


Figure 2: Sample images from the MNIST hand-written digits dataset [17].

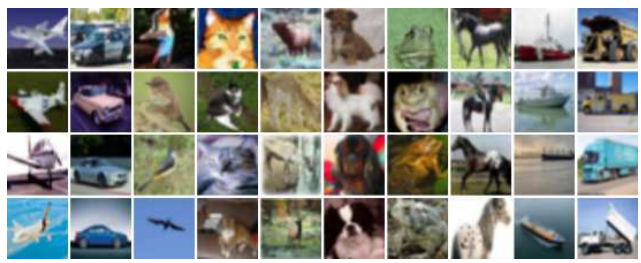


Figure 3: Sample images from the CIFAR-10 object classification dataset [14].

Similarly, the probability of synthesis for a particular synapse $s_{g,i}$ within a synaptic cluster c is realized as:

$$P(s_{g,i}|w_{g-1,i}) = \exp\Big(\frac{w_{g-1,i}}{a} - 1\Big) \quad (7)$$

where z is a layer-wise normalization factor. This genetic encoding scheme allows for the simultaneous consideration of both inter-synapse relationships and individual synapse strength [22].

3. Results

3.1. Experimental Setup

The asexual and sexual evolutionary synthesis of deep neural networks were performed over multiple generations, and the effects of sexual evolutionary synthesis relative to asexual evolutionary synthesis were explored using the MNIST [17] hand-written digits and CIFAR-10 [14] object classification datasets with the first generation ancestor networks trained using the LeNet-5 architecture [16]. Figure 2

and Figure 3 shows sample images from the MNIST and CIFAR-10 datasets, respectively.

Similar to Shafiee *et al.*'s work [22], 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 70% of the total number of synapses in its parent network in the previous generation; this allows increasingly more compact feature representations and for the synthesized deep neural networks to become progressively more efficient in the successive generations while minimizing any loss in accuracy. In addition, 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

An extension of Shafiee *et al.*'s cluster-driven genetic encoding scheme [22] for asexual evolutionary synthesis, the use of sexual evolutionary synthesis during the synthesis of offspring networks for evolutionary deep intelligence was explored in this study. 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 4 and Figure 5 show the testing accuracy, synaptic sparsity, and cluster sparsity of networks synthesized using asexual and sexual evolutionary synthesis as a function of generation number, and evaluated on the MNIST and CIFAR-10 datasets, respectively. Note that in both the asexual and sexual evolutionary synthesis cases, there is a trade-off between testing accuracy and architectural efficiency, i.e., testing accuracy decreases as synaptic efficiency and cluster efficiency increase.

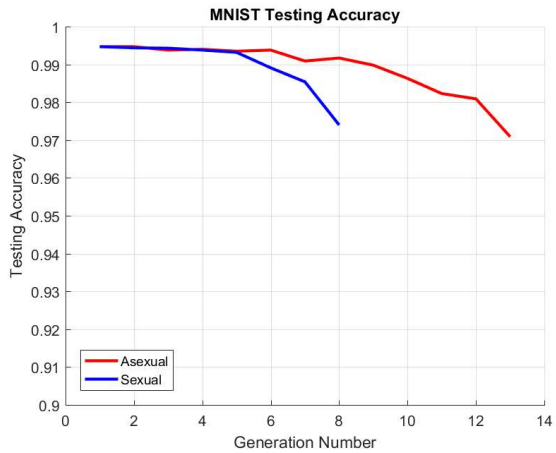
Figure 4 shows the MNIST testing accuracy and network efficiency (synaptic and cluster) for networks synthesized using asexual (red) and sexual (blue) evolutionary synthesis. Figure 4 (a) shows the testing accuracy for networks synthesized using asexual and sexual evolutionary synthesis. For this experiment, the original fully-trained ancestor network (generation 1) had a testing accuracy of 99.5% with 143,136 synapses and 7,200 kernels (corresponding to a 1-channel input LeNet architecture [16]). As expected, both asexual and sexual evolutionary synthesis produced networks with slight decreases in testing accuracy (approximately 3%); however, note that sexual evolutionary synthesis produced this network by generation 8 with a testing accuracy of 97.4% while asexual evolutionary produced a corresponding network at generation 13 with a testing accuracy of 97.1%.

Figure 4 (b) shows the synaptic sparsity for networks synthesized using asexual and sexual evolutionary synthesis and evaluated using the MNIST dataset. While networks synthesized via sexual evolutionary synthesis and asexual evolutionary synthesis produced similar testing accuracies, the network synthesized using sexual evolutionary synthesis at generation 8 has a synaptic efficiency of $258\times$ and the network synthesized using asexual evolutionary synthesis at generation 13 has a synaptic efficiency of $139\times$. Figure 4 (b) shows that the synaptic efficiency of networks synthesized using sexual evolutionary synthesis increases more steeply over generations than the synaptic efficiency of networks synthesized using asexual evolutionary synthesis. With almost double the synaptic efficiency for comparable testing accuracy, this indicates that networks synthesized via sexual evolutionary synthesis produce notably more efficient and compact feature representations.

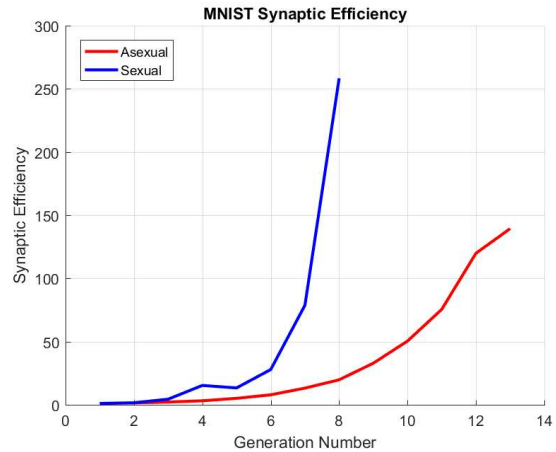
Figure 4 (c) shows the cluster sparsity for networks synthesized using asexual and sexual evolutionary synthesis and evaluated using the MNIST dataset. Like synaptic efficiency, the cluster efficiency of networks synthesized using sexual evolutionary synthesis is noticeably higher than the cluster efficiency of networks synthesized using asexual evolutionary synthesis, and Figure 4 (c) shows a similar trend of increasing cluster efficiency over generations for networks synthesized via sexual evolutionary synthesis relative to asexual evolutionary synthesis. Specifically, the cluster efficiency of the network synthesized using sexual evolutionary synthesis at generation 8 is $34.3\times$ while the cluster efficiency of the network synthesized using asexual evolutionary synthesis at generation 13 is $14.1\times$, further indicating the potential of sexual evolutionary synthesis to produce more efficient and compact feature representations.

Figure 5 shows the CIFAR-10 testing accuracy and network efficiency (synaptic and cluster) for networks synthesized using asexual (red) and sexual (blue) evolutionary synthesis. Figure 5 (a) shows the testing accuracy for networks synthesized using asexual and sexual evolutionary synthesis. In this experiment, the original fully-trained ancestor network (generation 1) had a testing accuracy of 77.1% with 144,736 synapses and 7,264 kernels (corresponding to a 3-channel input LeNet architecture [16]). While both asexual and sexual evolutionary synthesis produced networks with decreased testing accuracy (approximately $3\% - 4\%$), sexual evolutionary synthesis produced this network by generation 4 with a testing accuracy of 74.1% while asexual evolutionary produced a corresponding network at generation 7 with a testing accuracy of 73.4%. Similarly for a testing accuracy decrease of approximately 10%, sexual evolutionary synthesis produced this network by generation 5 with a testing accuracy of 67.9% while asexual evolutionary produced a corresponding network at generation 10 with a testing accuracy of 67.9%.

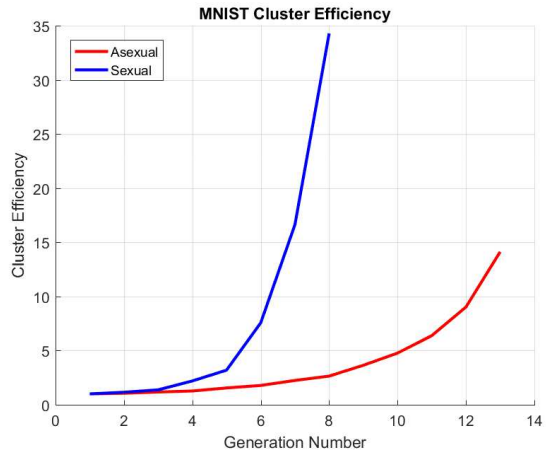
Figure 5 (b) shows the synaptic sparsity for networks



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

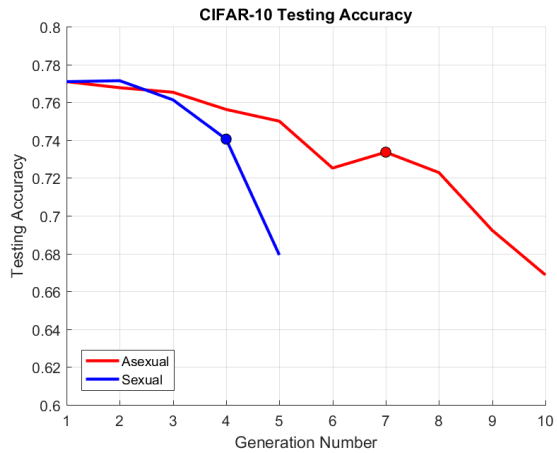


(b) MNIST synaptic efficiency vs. generations for synthesized networks.

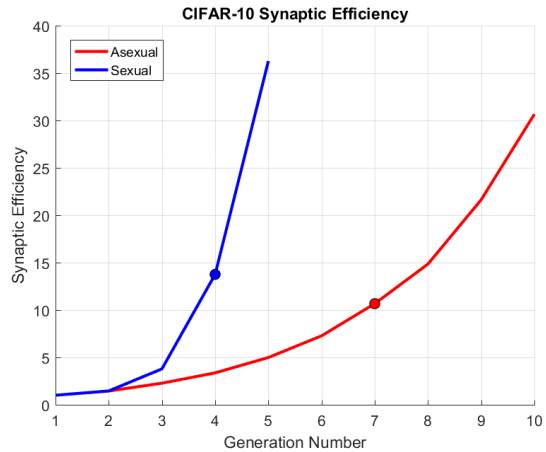


(c) MNIST cluster efficiency vs. generations for synthesized networks.

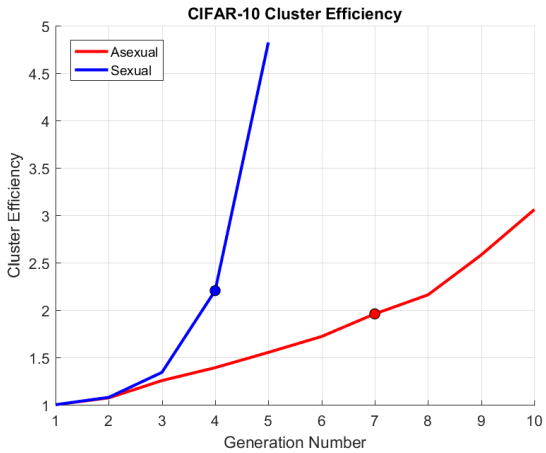
Figure 4: MNIST testing accuracy and network efficiency (synaptic and cluster) for networks synthesized using asexual (red) and sexual (blue) evolutionary synthesis.



(a) CIFAR-10 testing accuracy vs. generations for synthesized networks; point at 3 – 4% accuracy drop marked.



(b) CIFAR-10 synaptic efficiency vs. generations for synthesized networks; point at 3 – 4% accuracy drop marked.



(c) CIFAR-10 cluster efficiency vs. generations for synthesized networks; point at 3 – 4% accuracy drop marked.

Figure 5: CIFAR-10 testing accuracy and network efficiency (synaptic and cluster) for networks synthesized using asexual (red) and sexual (blue) evolutionary synthesis.

synthesized using asexual and sexual evolutionary synthesis and evaluated using the CIFAR-10 dataset. With the $3\% - 4\%$ drop in testing accuracy, the networks synthesized via sexual and asexual evolutionary synthesis have similar synaptic efficiencies of $10.7\times$ and $13.8\times$, respectively. At the 10% testing accuracy drop, the network synthesized using sexual evolutionary synthesis at generation 5 has a synaptic efficiency of $36.3\times$ and the network synthesized using asexual evolutionary synthesis at generation 10 has a synaptic efficiency of $30.7\times$. While only a slight increase, this increase in synaptic efficiency still allows for somewhat more compact feature representations, and networks synthesized using sexual evolutionary synthesis exhibit a more exponential increase in synaptic efficiency relative to asexual evolutionary synthesis.

Figure 5 (c) shows the cluster sparsity for networks synthesized using asexual and sexual evolutionary synthesis and evaluated using the CIFAR-10 dataset. Unlike the synaptic efficiency where the increase at the 10% testing accuracy drop is marginal, however, the cluster efficiency of networks synthesized using sexual evolutionary synthesis is notably higher than the cluster efficiency of networks synthesized using asexual evolutionary synthesis. Quantitatively, there is marginal improvement in cluster efficiency when comparing the networks with $3\% - 4\%$ drop in testing accuracy; the cluster efficiency of the network synthesized using sexual evolutionary synthesis is $2.20\times$ at generation 4, and the cluster efficiency of the network synthesized using asexual evolutionary synthesis is $1.96\times$ at generation 7. At the 10% testing accuracy drop, the difference in cluster efficiency becomes more pronounced. The cluster efficiency of the network synthesized using sexual evolutionary synthesis at generation 5 is $4.82\times$ while the cluster efficiency of the network synthesized using asexual evolutionary synthesis at generation 10 is $3.06\times$. A more obvious increase relative to synaptic efficiency both in terms of increase in cluster efficiency over generations and quantitatively, the increase in cluster efficiency of networks synthesized via sexual evolutionary synthesis allows for more compact networks and feature representations.

Notice that in both the experiments with MNIST and CIFAR-10 datasets, fewer generations were required to reach similar levels of network performance (in this study, the expected drop in testing accuracy) using sexual evolutionary synthesis. It is worth noting that achieving a more efficient deep neural network at earlier generations is beneficial as it reduces the number of training steps required, which is the most computationally complex aspect of the evolutionary process. While more obvious in the experiment with the MNIST dataset, it can also be seen that sexual evolutionary synthesis produced networks that are more efficient on both the synaptic and cluster levels. This is particularly noticeable when comparing the trends of synaptic and cluster efficiency increase over generations, as the increases in efficiency for networks synthesized via sexual evolutionary synthesis follow a steeper exponential trend than networks synthesized via asexual evolutionary synthesis. Lastly, it is worth noting that the MNIST dataset allows for considerably more increases in architectural efficiency relative to the CIFAR-10 dataset; this is likely due to the simplicity of the MNIST dataset (1-channel images of handwritten digits) relative to the CIFAR-10 dataset (3-channel natural images of objects).

4. Conclusion

In this work, we explored the effects of sexual evolutionary synthesis when synthesizing offspring deep neural networks in an evolutionary deep intelligence approach. An extension of the cluster-driven genetic encoding scheme proposed by Shafiee *et al.* [22], we incorporated a second parent network into the offspring network synthesis process at each generation.

Overall, the use of sexual evolutionary synthesis showed noticeable improvement in the architectural efficiency of the synthesized networks with maintaining comparable testing accuracy. In the case of the MNIST dataset, the offspring network synthesized via asexual evolutionary synthesis at generation 13 had a cluster efficiency of only $14.1\times$ and a synaptic efficiency of $139\times$ while the offspring network synthesized via sexual evolutionary synthesis at generation 8 had approximately double the architectural efficiency (cluster efficiency of $34.3\times$ and synaptic efficiency of $258\times$); both networks had a testing accuracy of around 97%. Similarly for the CIFAR-10 dataset, the offspring network synthesized via asexual evolutionary at generation 10 had a cluster efficiency of only $3.06\times$ and a synaptic efficiency of $30.7\times$ while the offspring network synthesized via sexual evolutionary synthesis at generation 5 had a cluster efficiency of $4.82\times$ and a synaptic efficiency of $36.3\times$; both networks had a testing accuracy of around 68%.

This suggests that sexual evolutionary synthesis in evolutionary deep intelligence via the synthesis of offspring neural networks using two parent networks can produce more efficient network architectures and increasingly compact feature representations, allow for higher levels of generalizability and adaptability in synthesized networks, and encourage more diversity in the genetic encoding. As such, further investigation into the effects of sexual versus asexual evolutionary synthesis for network synthesis in evolutionary deep intelligence would be beneficial, particularly with the deep neural networks in a changing environment, i.e., non-stationary environmental factor models.

Future work in this area includes a more thorough investigation into various methods for combining the parent neural networks, e.g., determining whether favouring the genetics of the parent with a better fitness score (such as testing

accuracy) would allow for a higher probability of passing on beneficial traits to the offspring network and subsequent generations. Other potential areas of future work include associating strong task performance with specific synapses or synaptic clusters in a deep neural network, and favouring these specific sections of different parent networks during the offspring synthesis process. Lastly, it would be beneficial to extend the mating and synthesis process to include mate selection from a pool of potential neural network mates to incorporate the notion of "survival of the fittest" [29] into the evolutionary deep intelligence approach to allow for a stronger overall population of networks.

References

[1] P. J. Angeline, G. M. Saunders, and J. B. Pollack. An evolutionary algorithm that constructs recurrent neural networks. *IEEE transactions on Neural Networks*, 5(1):54–65, 1994. 2

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

[3] W. Chen, J. T. Wilson, S. Tyree, K. Q. Weinberger, and Y. Chen. Compressing neural networks with the hashing trick. *CoRR, abs/1504.04788*, 2015. 1

[4] J. F. Crow and M. Kimura. Evolution in sexual and asexual populations. *American Naturalist*, pages 439–450, 1965. 2

[5] J. Feng and T. Darrell. Learning the structure of deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2749–2757, 2015. 2

[6] R. A. Fisher. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1930. 2

[7] J. Gauci and K. Stanley. Generating large-scale neural networks through discovering geometric regularities. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pages 997–1004. ACM, 2007. 2

[8] Y. Gong, L. Liu, M. Yang, and L. Bourdev. Compressing deep convolutional networks using vector quantization. *arXiv preprint arXiv:1412.6115*, 2014. 1

[9] 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. 1

[10] S. Han, H. Mao, and W. J. Dally. Deep compression: Compressing deep neural network with pruning, trained quantization and huffman coding. *CoRR, abs/1510.00149*, 2, 2015. 1

[11] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In *Advances in Neural Information Processing Systems*, pages 1135–1143, 2015. 1

[12] Y. Ioannou, D. Robertson, J. Shotton, R. Cipolla, and A. Criminisi. Training cnns with low-rank filters for efficient image classification. *arXiv preprint arXiv:1511.06744*, 2015. 2

[13] M. Jaderberg, A. Vedaldi, and A. Zisserman. Speeding up convolutional neural networks with low rank expansions. *arXiv preprint arXiv:1405.3866*, 2014. 2

[14] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images. 2009. 4

[15] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 1

[16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 4, 5

[17] Y. LeCun, C. Cortes, and C. J. Burges. The mnist database of handwritten digits, 1998. 4

[18] Y. LeCun, J. S. Denker, S. A. Solla, R. E. Howard, and L. D. Jackel. Optimal brain damage. In *NIPs*, volume 2, pages 598–605, 1989. 1

[19] B. Liu, M. Wang, H. Foroosh, M. Tappen, and M. Pensky. Sparse convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 806–814, 2015. 2

[20] H. J. Muller. Some genetic aspects of sex. *The American Naturalist*, 66(703):118–138, 1932. 2

[21] M. J. Shafiee, A. Mishra, and A. Wong. Deep learning with darwin: Evolutionary synthesis of deep neural networks. *arXiv preprint arXiv:1606.04393*, 2016. 2, 3

[22] M. J. Shafiee and A. Wong. Evolutionary synthesis of deep neural networks via synaptic cluster-driven genetic encoding. *arXiv preprint arXiv:1609.01360*, 2016. 2, 3, 4, 5, 7

[23] J. M. Smith. Evolution in sexual and asexual populations. *American Naturalist*, pages 469–473, 1968. 3

[24] K. O. Stanley, B. D. Bryant, and R. Miikkulainen. Real-time neuroevolution in the nero video game. *IEEE transactions on evolutionary computation*, 9(6):653–668, 2005. 2

[25] K. O. Stanley and R. Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127, 2002. 2

[26] S. S. Tirumala, S. Ali, and C. P. Ramesh. Evolving deep neural networks: A new prospect. In *Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on*, pages 69–74. IEEE, 2016. 2

[27] 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*, pages 1799–1807, 2014. 1

[28] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li. Learning structured sparsity in deep neural networks. In *Advances In Neural Information Processing Systems*, pages 2074–2082, 2016. 2

[29] G. C. Williams. *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton University Press, 2008. 8