

Deep Learning with Darwin: Evolutionary Synthesis of Deep Neural Networks

Mohammad Javad Shafiee¹  · Akshaya Mishra¹ · Alexander Wong¹

Published online: 20 November 2017
© Springer Science+Business Media, LLC 2017

Abstract Taking inspiration from biological evolution, we explore the idea of “Can deep neural networks *evolve* naturally over successive generations into highly efficient deep neural networks?” by introducing the notion of synthesizing new highly efficient, yet powerful deep neural networks over successive generations via an evolutionary process from ancestor deep neural networks. The architectural traits of ancestor deep neural networks are encoded using synaptic probability models, which can be viewed as the ‘DNA’ of these networks. New descendant networks with differing network architectures are synthesized based on these synaptic probability models from the ancestor networks and computational environmental factor models, in a random manner to mimic heredity, natural selection, and random mutation. These offspring networks are then trained into fully functional networks, like one would train a newborn, and have more efficient, more diverse network architectures than their ancestor networks, while achieving powerful modeling capabilities. Experimental results for the task of visual saliency demonstrated that the synthesized ‘evolved’ offspring networks can achieve state-of-the-art performance while having network architectures that are significantly more efficient (with a staggering ~ 48 -fold decrease in synapses by the fourth generation) compared to the original ancestor network.

Keywords Deep neural networks · Evolutionary · EvoNet · Deep learning · Saliency detection · Convolutional neural networks

1 Introduction

Deep learning, especially deep neural networks [1–4] have shown considerable promise through tremendous results in recent years, significantly improving the accuracy of a variety of challenging problems when compared to other machine learning methods [5–10].

✉ Mohammad Javad Shafiee
mjshafiee@uwaterloo.ca

¹ Systems Design Engineering Department, University of Waterloo, Waterloo, ON, Canada

However, deep neural networks require high performance computing systems due to the tremendous quantity of computational layers they possess, leading to a massive quantity of parameters to learn and compute. This issue of architectural complexity has increased greatly in recent years [7, 11, 12], driven by the demand for increasingly deeper and larger deep neural networks to boost modeling accuracy. As such, it has become increasingly more difficult to take advantage of such complex deep neural networks in scenarios where computational and energy resources are scarce.

To enable the widespread use of deep learning, there has been a recent drive towards obtaining highly-efficient deep neural networks with strong modeling power. Much of the work in obtaining efficient deep neural networks have focused on deterministically compressing trained deep neural networks [13] to tackle the memory issue of deep neural networks. In the pioneering work by Lecun et al. [13], an optimal brain damage method was proposed where synapses are pruned based on their strengths. This deterministic pruning approach was further extended upon by Han et al. [15]. Gong et al. [14] proposed the use of traditional lossy compression techniques such as vector quantization, which was further leveraged by Han et al. [16] along with traditional lossless compression techniques such as Huffman coding to further reduce the memory requirements of deep neural networks. Chen et al. [17] further introduced the use of hashing of weights in deep neural networks to reduce the memory requirements of deep neural networks.

Structural sparsity techniques have been recently proposed for addressing the memory, computational complexity and processing issues of deep neural networks. For example, low rank matrix approximation methods [28–30] have been proposed in which the convolution kernels are factorized by smaller kernel sizes. Furthermore, sparsity learning approaches [31–33] have been proposed that involved the optimization of sparse kernels during the training of the deep neural networks.

Rather than attempting to take an existing deep neural network and compress it into a smaller representation heuristically, the following idea is considered here: *Can deep neural networks evolve naturally over successive generations into highly efficient deep neural networks?* Using an example of evolutionary progress towards efficiency from nature, a recent study by Moran et al. [18] proposed that the eyeless Mexican cavefish evolved to lose its vision system over generations due to the high metabolic cost of vision. Therefore, by evolving naturally over generations in a way where the cavefish lost its vision system, the amount of energy expended is significantly reduced and thus improves survivability in subterranean habitats where food availability is low. The ability to mimic the biological evolutionary process for the task of producing highly-efficient deep neural networks over successive generations can have considerable benefits.

In this study, we entertain a different notion for producing highly-efficient deep neural networks by introducing the evolutionary synthesis of deep neural networks over successive generations based on ancestor deep neural networks. While the idea of leveraging evolutionary computation concepts for training and generating deep neural networks have been previously explored in literature [19–23], there are significant key differences between these previous studies and this study:

- While previous studies have focused on improving the accuracy and training of deep neural networks, to the best of the authors' knowledge this study is the first to explore and focus on the notion of evolutionary synthesis of deep neural networks with high network architectural efficiency over successive generations.
- While the evolutionary computational approaches leveraged by these previous studies are classical approaches such as genetic algorithms and evolutionary programming, this

study introduces a new probabilistic framework where evolution mechanisms such as genetic encoding and environmental conditions are modeled via probability distributions, and the stochastic synthesis process leverages these probability models to produce deep neural networks at successive generations. To the best of the authors' knowledge, this study is the first to leverage a probabilistic approach to evolutionary synthesis of deep neural networks.

- To the best of the authors' knowledge, the new approach introduced in this study is the first to achieve evolution and synthesis of deep neural networks with very deep, large neural network architectures that have been demonstrated to provide great performance in recent years [7, 11, 12]. Previous studies have focused on deep neural networks with smaller and shallower network architectures, as the approaches used in such studies are more difficult to scale to very deep, large network architectures.

It is also important to note that while previous methods such as Dropout [26] and Drop-Connect [27] have explored the notion of probabilistically deactivating and reactivating synapses, this is only performed temporarily during training for the purpose of regularization and as such the topology of the resulting deep neural network does not change and remains dense at the end. This is significantly different than the approach proposed in this study, where the synapse topology changes for deep neural networks at each generation of evolution, with the resulting deep neural networks evolving to possess significantly more efficient network architectures than its ancestors.

2 Methodology

The proposed evolutionary synthesis of deep neural networks is primarily inspired by real biological evolutionary mechanisms. In nature, traits that are passed down from generation to generation through DNA may change over successive generations due to factors such as natural selection and random mutation, giving rise to diverse and enhanced traits in later generations. To realize the idea of evolutionary synthesis for producing deep neural networks, we introduce a number of computational constructs to mimic the following mechanisms of biological evolution: (i) *Heredity*, (ii) *Natural Selection*, and (iii) *Random Mutation*.

2.1 Heredity

Here, the idea of heredity is modeled by encoding the architectural traits of deep neural networks in the form of synaptic probability models, which are used to pass down traits from generation to generation. One can view these synaptic probability models as the 'DNA' of the networks. Let $\mathcal{H} = (\mathcal{N}, \mathcal{S})$ denote the possible architecture of a deep neural network, with \mathcal{N} denoting the set of possible neurons and \mathcal{S} denoting the set of possible synapses, with $s_k \in \mathcal{S}$ denoting a synapse between two neurons $(n_i, n_j) \in \mathcal{N}$. One can encode the architectural traits of a deep neural network as $P(\mathcal{H}_g | \mathcal{H}_{g-1})$, which denotes the conditional probability of the architecture of a network in generation g (denoted by \mathcal{H}_g), given the architecture of its ancestor network in generation $g - 1$ (denoted by \mathcal{H}_{g-1}).

If we were to treat areas of strong synapses in an ancestor network in generation g as desirable traits to be inherited by descendant networks at generation g , where descendant networks have a higher probability of having similar areas of strong synapses as its ancestor network, one can instead encode the architectural traits of a deep neural network as the synaptic probability $P(S_g | \mathcal{W}_{g-1})$, where $w_{g-1,k} \in \mathcal{W}_{g-1}$ encodes the synaptic strength of

each synapse $s_{g-1,k}$. Modeling $P(S_g|\mathcal{W}_{g-1})$ as an exponential distribution, with the probability of each synapse in the network assumed to be independently distributed, one arrives at

$$P(S_g|\mathcal{W}_{g-1}) = \prod_i \exp\left(\frac{w_{g-1,i}}{Z} - 1\right), \quad (1)$$

where Z is a normalization constant.

2.2 Natural Selection and Random Mutation

The ideas of natural selection and random mutation are mimicked through the introduction of a network synthesis process for synthesizing descendant networks, which takes into account not only the synaptic probability model encoding the architectural traits of the ancestor network, but also an environmental factor model to mimic the environmental conditions that help drive natural selection, in a random manner that drives random mutation. More specifically, a synapse is synthesized randomly between two possible neurons in a descendant network based on $P(S_g|\mathcal{W}_{g-1})$ and an environmental factor model $\mathcal{F}(\mathcal{E})$, with the neurons in the descendant network synthesized subsequently based on the set of synthesized synapses. As such, the architecture of a descendant network at generation g can be synthesized randomly via synthesis probability $P(\mathcal{H}_g)$, which can be expressed by

$$P(\mathcal{H}_g) = \mathcal{F}(\mathcal{E}) \cdot P(S_g|\mathcal{W}_{g-1}). \quad (2)$$

The environmental factor model $\mathcal{F}(\mathcal{E})$ can be the combination of quantitative environmental conditions that are imposed upon the descendant networks that they must adapt to.

To have a better intuitive understanding, let us examine an illustrative example of how one can impose environmental conditions using $\mathcal{F}(\mathcal{E})$ to promote the evolution of highly efficient deep neural networks.

2.3 Efficiency-Driven Evolutionary Synthesis

One of the main environmental factors in encouraging energy efficiency during evolution is to restrict the resources available. For example, in a study by Moran et al. [18], it was proposed that the eyeless Mexican cavefish lost its vision system over generations due to the high energetic cost of neural tissue and low food availability in subterranean habitats. Their study demonstrated that the cost of vision is about 15% of resting metabolism for a 1-g eyed phenotype, thus losing their vision system through evolution has significant energy savings and thus improves survivability. As such, inspired by that available resources are computationally restricted to descendant networks to encourage the evolution of highly-efficient deep neural networks.

Considering the aforementioned example, the descendant networks must take on network architectures with more efficient energy consumption than this original ancestor network to be able to survive. The main factor in energy consumption is the quantity of synapses and neurons in the network. Therefore, to mimic environmental constraints that encourage the evolution of highly-efficient deep neural networks, an environmental constraint $\mathcal{F}(\mathcal{E}) = C$ is introduced that probabilistically constrains the quantity of synapses that can be synthesized in the descendant network (which in effect also constrains the quantity of neurons that can be synthesized), such that descendant networks are forced to evolve more efficient network architectures than their ancestor networks.

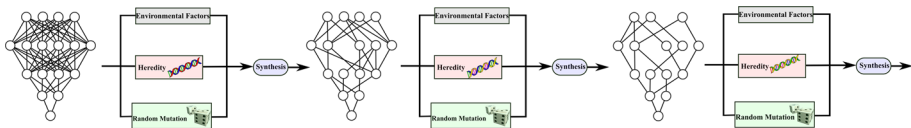


Fig. 1 Evolutionary synthesis process of highly-efficient deep neural networks. The architectural traits of ancestor networks are modeled with a probabilistic ‘DNA’ in the form of a synaptic probability model. At each generation, a new offspring deep neural network is synthesized in a stochastic manner (mimicking random mutation) based on the environmental factors (mimicking natural selection) and the probabilistic DNA of its ancestor (mimicking heredity). This process is repeated generation after generation to synthesize progressively more efficient yet powerful deep neural networks

Therefore, given $P(S_g|\mathcal{W}_{g-1})$ and $\mathcal{F}(\mathcal{E}) = C$, the synthesis probability $P(\mathcal{H}_g)$ can be formulated as

$$P(\mathcal{H}_g) = C \cdot P(S_g|\mathcal{W}_{g-1}), \quad (3)$$

where C is the highest percentage of synapses desired in the descendant network. The random element of the network synthesis process mimics the random mutation process and promotes network architectural diversity.

Given the probabilistic framework introduced above, the proposed evolutionary synthesis of highly-efficient deep neural networks can be described as follows. As seen in Fig. 1, given an ancestor deep neural network at generation $g - 1$, the architectural traits of the ancestor network is encapsulated in the form of a synaptic probability model $P(S_g|\mathcal{W}_{g-1})$ according to (1), which can be viewed as its probabilistic ‘DNA’ and mimicks the heredity process. Based on this probabilistic ‘DNA’ $P(S_g|\mathcal{W}_{g-1})$ and environmental constraint factors $\mathcal{F}(\mathcal{E})$ mimicking the natural selection process, a synthesis probability $P(\mathcal{H}_g)$ is then constructed according to (3). Finally, given synthesis probability $P(\mathcal{H}_g)$, a descendant offspring deep neural network is synthesized in a stochastic manner that mimics the process of random mutation. The offspring deep neural networks are then trained and the above processes are repeated generation after generation to produce more efficient, more diverse offspring deep neural networks that also retain powerful modeling capabilities:

$$s_{g,k} \text{ exists in } \mathcal{H}_g \text{ if } P(s_{g,k}) \geq U(0; 1), \quad (4)$$

where $U(0; 1)$ is a uniformly distributed random number from a uniform distribution between 0 and 1. The synthesized descendant networks at generation g are then trained into fully-functional networks, like one would train a newborn, and the evolutionary synthesis process is repeated for producing successive generations of descendant networks.

3 Experimental Results

To investigate the efficacy of the proposed evolutionary synthesis of highly-efficient deep neural networks, experiments were performed using the MSRA-B [24] and HKU-IS datasets [25] for the task of visual saliency. This task was chosen given the importance for biological beings to detect objects of interest (e.g., prey, food, predators) for survival in complex visual environments, and can provide interesting insights into the evolution of networks. Three generations of descendant deep neural networks (second, third, and fourth generations) were synthesized within an artificially constrained environment beyond the original, first-generation ancestor network. The environmental constraint imposed during synthesis in this study is that the

descendant networks should not have more than 40% of the total number of synapses that its direct ancestor network possesses (i.e., $C = 0.4$ was determined empirically as it was found that this parameter setting provided strong network complexity reduction performance through comprehensive testing.), thus encouraging the evolution of highly-efficient deep neural networks. The network architecture of the original, first generation ancestor network used in this study, and details on the tested datasets and performance metrics are as follow.

3.1 Network Architecture

The network architecture of the original, first generation ancestor network used in this study builds upon the VGG16 very deep convolutional neural network architecture [7] for the purpose of image segmentation as follows. The outputs of the c3, c4, and c5 stacks from the VGG16 architecture are fed into newly added c6, c7, c8 stacks, respectively. The output of the c7 and c8 stacks are then fed into d1 and d2 stacks. The concatenated outputs of the c6, d1, and d2 stacks are then fed into the c9 stack. The output of the c5 stack is fed into c10 and c11 stacks. Finally, the combined output of the c9, c10 and c11 stacks are fed into a softmax layer to produce final segmentation result. The details of different stacks are as follows: c1: 2 convolutional layers of 64, 3×3 local receptive fields, c2: 2 convolutional layers of 128, 3×3 local receptive fields, c3: 3 convolutional layers of 256, 3×3 local receptive fields, c4: 3 convolutional layers of 512, 3×3 local receptive fields, c5: 3 convolutional layers of 512, 3×3 local receptive fields, c6: 1 convolutional layers of 256, 3×3 local receptive fields, c7 and c8: 1 convolutional layers of 512, 3×3 local receptive fields, c9: 1 convolutional layers of 384, 1×1 local receptive fields, c10 and c11: 2 convolutional layers of 512, 11×11 local receptive fields and 384, 1×1 local receptive fields, d1 and d2 are deconvolutional layers.

3.2 Datasets

The MSRA-B dataset [24] consists of 5000 natural images and their corresponding ground truth maps where the salient objects in the images are segmented with pixel-wise annotation. The dataset is divided into training, validation and testing groups containing 2500, 500 and 2000 images, respectively. Figure 2 the HKU-IS dataset [25] consists of 4447 natural images and their corresponding ground truth maps where the salient objects in the images are segmented with pixel-wise annotation. The entire dataset is used as a testing group for the descendant networks trained on the training group of the MSRA-B dataset. Figure 3 illustrates some of the example images from the dataset with their corresponding ground truths.

3.3 Performance Metrics

To evaluate the performance of the evolved descendant deep neural networks at different generations two metrics were computed for each of the descendant deep neural networks across the 2000 test images of the MSRA-B dataset that were not used for training;

- *Mean absolute error (MAE)*, is a measure which evaluates how close the generated saliency result is to the ground truth:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

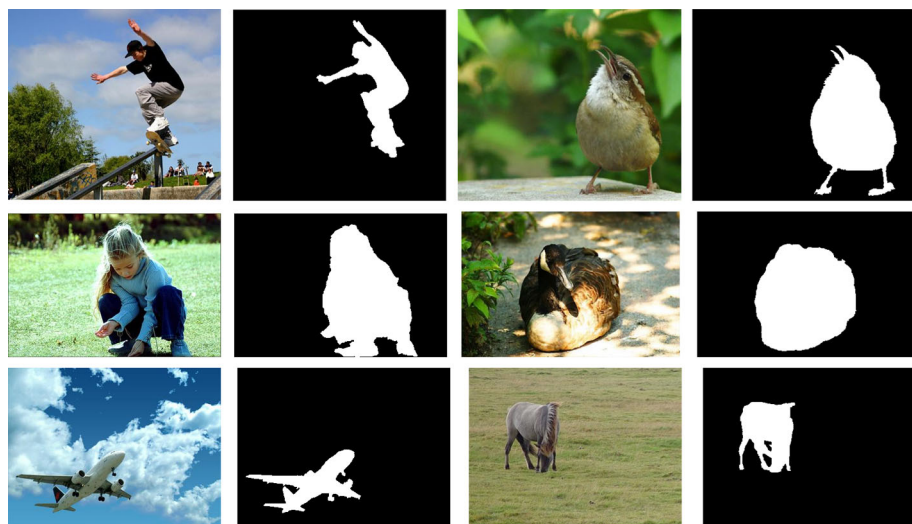


Fig. 2 MSRA-B image dataset: this dataset contains 5000 natural images divided into 2500, 500 and 2000 images as training, validation and test samples, respectively. The ground truth maps are provided with pixel-wise annotation. Examples of images and their corresponding ground truth maps in the MSRA-B image dataset are shown here

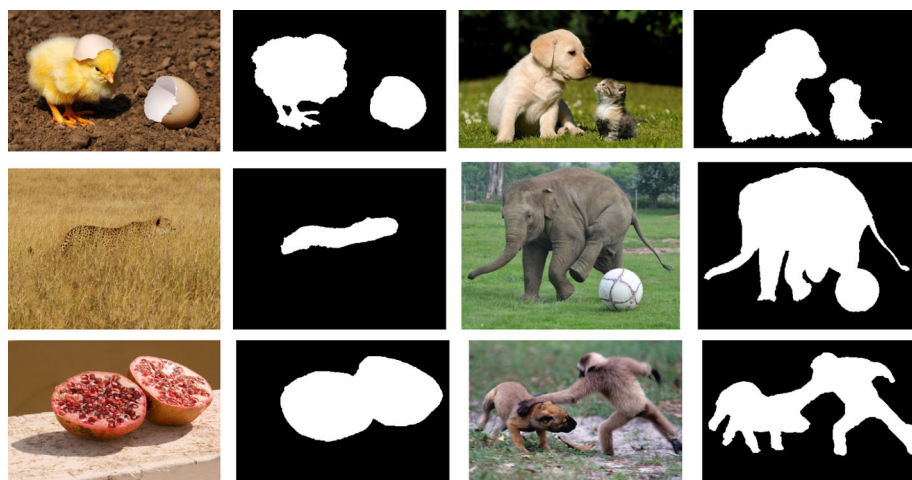


Fig. 3 HKU-IS image dataset: this dataset contains 4447 natural images, and the entire dataset is used as a testing group for the descendant deep neural networks trained on the training group of the MSRA-B dataset

- F_β score, is a general evaluation metric for binary classification which considers both positive and negative misclassification at the same time where here is the wrong pixel annotation as salient pixels or background:

$$F_\beta = (1 + \beta^2) \cdot \frac{Pr \cdot Re}{\beta^2 \cdot Pr + Re} \quad (6)$$

where Pr and Re are the computed precision and recall based on the available ground truth and $\beta^2 = 0.3$ [25].

Table 1 Performance metrics for different generations of synthesized offspring networks for MSRA-B dataset and comparison to the proposed method by Li et al. [25]

Generation	Number of synapses	Architectural efficiency	F_β score	MAE
1	63,767,232	$1\times$	0.875	0.0743
2	15,471,797	$4.12\times$	0.876	0.0739
3	3,603,007	$17.69\times$	0.861	0.0813
4	1,333,010	$47.83\times$	0.850	0.0863
Li et al. [25]	–	–	0.865	~ 0.11

Table 2 Performance metrics for different generations of synthesized offspring networks for HKU-IS dataset compared to the proposed method by Li et al. [25]

Generation	Number of synapses	Architectural efficiency	F_β score	MAE
1	63,767,232	$1\times$	0.830	0.0914
2	15,471,797	$4.12\times$	0.826	0.0911
3	3,603,007	$17.69\times$	0.775	0.1087
4	1,333,010	$47.83\times$	0.753	0.1190
Li et al. [25]	–	–	0.80	~ 0.08

The same performance metrics was also computed for the original, first generation ancestor deep neural network.

3.4 Architectural Efficiency Over Successive Generations

The detailed experimental results describing the number of synapses, architectural efficiency (defined here as the reduction of synapses in the network compared to the original, ancestor deep neural network in the first generation), F_β score, and mean absolute error (MAE) presented in Tables 1 and 2 for the MSRA-B and HKU-IS datasets, respectively. A number of insightful observations can be made with respect to change in the architectural efficiency over successive generations of descendant deep neural networks.

First, it can be observed that the performance differences from one generation of descendant networks to the next generation are small for MSRA-B ($< 3\%$ between first generation and the fourth generation), while the performance differences are small for HKU-IS between the first two generations ($< 0.5\%$) before larger performance differences in the third and fourth generations ($< 8\%$ between the first and fourth generations). These results indicate that the modeling power of the ancestor network are well-preserved in the descendant networks.

Second, it can be observed that the descendant networks in the second and third generations can achieve state-of-the-art F_β scores for MSRA-B (0.876 at second generation and 0.861 at third generation, compared to 0.865 as reported by Li et al. [25] for their state-of-the-art visual saliency method), while having network architectures that are significantly more efficient compared to the first generation ancestor network (~ 18 -fold decrease in synapses). A similar trend was observed for HKU-IS, though persisting only in the second generation (0.826 compared to 0.8 reported in [25], while achieving a ~ 4 -fold decrease in synapses over ancestor network). What is more remarkable is that the descendant network at the fourth

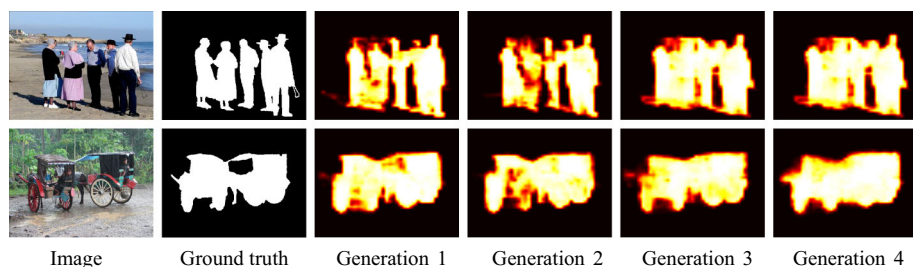


Fig. 4 Example test images from the tested datasets, and the corresponding visual saliency maps generated by the descendant deep neural networks at different generations

generation maintains strong F_β scores (0.850 for MSRA-B and 0.753 for HKU-IS), while having network architectures that are incredibly efficient (~ 48 -fold decrease in synapses) compared to the first generation ancestor network. This ~ 48 -fold increase in architectural efficiency while maintaining modeling power clearly show the efficacy of producing highly-efficient deep neural networks over successive generations via the proposed evolutionary synthesis.

3.5 Visual Saliency Variations Over Successive Generations

To gain additional insights, Fig. 4 demonstrate example test images from the MSRA-B dataset and the HKU-IS dataset, respectively, along with the corresponding visual saliency maps generated by the descendant networks at different generations. It can be observed that the descendant networks at all generations consistently identified the objects of interest in the scene as visually salient. It is also interesting to observe that by the fourth generation, with a ~ 48 -fold decrease in synapses compared to the first generation ancestor network, the ability to distinguish fine-grained visual saliency starts to diminish. These observations are interesting in that, similar to biological evolution, they show that the descendant networks evolved over successive generations in such a way that important traits (e.g., general ability to identify salient objects) are retained from its ancestors while less important traits (e.g., ability to distinguish fine-grained saliency) diminish in favor of adapting to environmental constraints (e.g., growing highly-efficient architectures due to imposed constraints).

These experimental results show that, by taking inspiration from biological evolution, the proposed evolutionary synthesis of deep neural networks can lead to the natural evolution of deep neural networks over successive generations into highly efficient, yet powerful deep neural networks, and thus a promising direction for future exploration in deep learning.

4 Conclusion

The idea of evolving deep neural networks naturally over successive generations to synthesize highly efficient deep neural networks was explored in this study. Inspired by biological evolution, a new probabilistic framework for evolutionary synthesis of highly efficient deep neural networks was proposed that avoids the high computational complexity faced when leveraging traditional evolutionary approaches for evolving deep neural networks. The architecture traits of ancestor deep neural networks was passed to offspring networks via a probabilistic ‘DNA’ in the form of synaptic probability models to mimic heredity. The environmental fac-

tors are enforced via computational modelling and, along with the probabilistic ‘DNA’ are used to stochastically synthesize offspring deep neural networks to mimic the natural selection and random mutation processes. The new synthesized offspring networks were then trained and the evolutionary process is repeated over successive generations. Experimental results showed that the synthesized off-spring networks can perform as well as state-of-the-art approaches while providing highly efficient network architectures. In particular, a highly-efficient deep neural network with a ~ 48 -fold decrease in number of synapses within a VGG-based network architecture was synthesized for the object saliency that was able to achieve comparable results to state-of-the-art methods, thus demonstrating the efficacy of the proposed framework.

References

1. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
2. Graves A, Mohamed A-R, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: *IEEE international conference on acoustics, speech and signal processing*, pp 6645–6649
3. Bengio Y (2009) Learning deep architectures for AI. *Found Trends Mach Learn* 2(1):1–127
4. Tompson J, Jain A, LeCun Y, Bregler C (2014) Joint training of a convolutional network and a graphical model for human pose estimation. In: *Proceedings of advances in neural information processing systems (NIPS)*, pp 1799–1807
5. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Proceedings of advances in neural information processing systems (NIPS)*, pp 1097–1105
6. Farabet C, Couprie C, Najman L, LeCun Y (2013) Learning hierarchical features for scene labeling. In: *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*
7. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556
8. Hinton G, Deng L, Yu D, Dahl GE, Mohamed A-R, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN et al (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. In: *IEEE signal processing magazine*
9. Hannun A, Case C, Casper J, Catanzaro B, Diamos G, Elsen E, Prenger R, Satheesh S, Sengupta S, Coates A et al (2014) Deep speech: scaling up end-to-end speech recognition. *CoRR*, abs/1412.5567
10. Amodei D, Anubhai R, Battenberg E, Case C, Casper J, Catanzaro B, Chen J, Chrzanowski M, Coates A, Diamos G et al (2015) Deep speech 2: end-to-end speech recognition in English and Mandarin. *CoRR*, abs/1512.02595
11. Srivastava RK, Greff K, Schmidhuber J (2015) Training very deep networks. In: *Proceedings of advances in neural information processing systems (NIPS)*, pp 2377–2385
12. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, pp 1–9
13. LeCun Y, Denker JS, Solla SA, Howard RE, Jackel LD (1989) Optimal brain damage. In: *Advances in neural information processing systems (NIPS)*
14. Gong Y, Liu L, Yang M, Bourdev L (2014) Compressing deep convolutional networks using vector quantization. *CoRR*, abs/1412.6115
15. Han S, Mao H, Dally WJ (2015) Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. *CoRR*, abs/1510.00149
16. Han S, Pool J, Tran J, Dally W (2015) Learning both weights and connections for efficient neural network. In: *Advances in neural information processing systems (NIPS)*
17. Chen W, Wilson JT, Tyree S, Weinberger KQ, Chen Y (2015) Compressing neural networks with the hashing trick. *CoRR*, abs/1504.04788
18. Moran D, Softley R, Warrant EJ (2015) The energetic cost of vision and the evolution of eyeless Mexican cavefish. *Sci Adv* 1:e1500363
19. Peter GMS, Angeline J, Pollack JB (1994) An evolutionary algorithm that constructs recurrent neural networks. In: *IEEE transactions on neural networks*
20. Stanley KO, Bryant BD, Miikkulainen R (2005) Real-time neuroevolution in the NERO video game. In: *IEEE transactions on evolutionary computation*

21. Stanley KO, Miikkulainen R (2002) Evolving neural networks through augmenting topologies. *Evol Comput* 10:99–127
22. Gauci J, Stanley KO (2007) Generating large-scale neural networks through discovering geometric regularities. In: *Proceedings of the 9th annual conference on Genetic and evolutionary computation*. pp 997–1004
23. Tirumala SS, Ali S, Ramesh CP (2016) Evolving deep neural networks: a new prospect. In: *12th International conference on natural computation, Fuzzy systems and knowledge discovery (ICNC-FSKD)*. pp 69–74
24. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H-Y (2011) Learning to detect a salient object. In: *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*. pp 353–367
25. Li G, Yu Y (2015) Visual saliency based on multiscale deep features. In: *IEEE conference on computer vision and pattern recognition (CVPR)*
26. Nitish S, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res JMLR* 15:1929–1958
27. Wan L, Zeiler M, Zhang S, LeCun Y, Fergus R (2013) Regularization of neural networks using dropconnect. In: *International conference on machine learning (ICML)*
28. Ioannou Y, Robertson D, Shotton J, Cipolla R, Criminisi A (2015) Training CNNs with low-rank filters for efficient image classification. *arXiv preprint [arXiv:1511.06744](https://arxiv.org/abs/1511.06744)*
29. Jaderberg M, Vedaldi A, Zisserman A (2014) Speeding up convolutional neural networks with low rank expansions. *arXiv preprint [arXiv:1405.3866](https://arxiv.org/abs/1405.3866)*,
30. Denton E, Zaremba W, Bruna J, LeCun Y, Fergus R (2014) Exploiting linear structure within convolutional networks for efficient evaluation. In: *Proceedings of advances in neural information processing systems (NIPS)*. pp 1269–1277
31. Feng J, Darrell T (2015) Learning the structure of deep convolutional networks. In: *Proceedings of the IEEE international conference on computer vision*, pp 2749–2757
32. Liu B, Wang M, Foroosh H, Tappen M, Pensky M (2015) Sparse convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 806–814
33. Wen W, Wu C, Wang Y, Chen Y, Li H (2016) Learning structured sparsity in deep neural networks. *arXiv preprint [arXiv:1608.03665](https://arxiv.org/abs/1608.03665)*