



Influence of random topology in artificial neural networks: A survey

Sara Kaviani, Insoo Sohn*

Division of Electronics & Electrical Engineering, Dongguk University, Seoul, Republic of Korea

Received 12 December 2019; accepted 15 January 2020

Available online xxxx

Abstract

Due to the fully-connected complex structure of Artificial Neural Networks (ANNs), systems based on ANN may consume much computational time, energy and space. Therefore, intense research has been recently centered on changing the topology and design of ANNs to obtain high performance. To explore the influence of network structure on ANNs complex systems topologies have been applied in these networks to have more efficient and less complex structures while they are more similar to biological systems at the same time. In this paper, the methodology and results of some recent papers are summarized and discussed in which the authors investigated the efficacy of random complex networks on the performance of Hopfield associative memory and multi-layer ANNs compared with ANNs with small-world, scale-free and regular structures.

© 2020 The Korean Institute of Communications and Information Sciences (KICS). Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Complex systems; Artificial neural networks; Random networks

Contents

1. Introduction.....	1
2. Basic concepts.....	2
2.1. Regular graphs.....	2
2.2. Random networks.....	2
2.3. Small-world networks.....	2
2.4. Scale-free networks.....	3
3. Artificial neural networks with complex structures.....	3
3.1. Hopfield associative memory.....	3
3.2. Multi-layer artificial neural networks.....	4
4. Discussion and conclusion.....	4
Declaration of competing interest.....	5
Acknowledgments.....	5
References.....	5

1. Introduction

Artificial Neural Networks (ANNs), inspired from biological brain networks, have obtained a fair amount of success in various domains such as speech and pattern recognition [1],

climate forecasting [2] and disease diagnosis [3]. The backbone of the ANNs development was a neural network model introduced by Hopfield [4] which has been used for associative memory and has produced significant results about improving their memory capacity and pattern stability and retrieval through extensive investigations [5,6]. Another type of ANNs, imitating the biological brain structure, called the Multi-layer perceptrons was invented to understand the information flow through the neural network and the learning process.

* Corresponding author.

E-mail addresses: s.kaviani@dongguk.edu (S. Kaviani), isohn@dongguk.edu (I. Sohn).

Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

<https://doi.org/10.1016/j.ict.2020.01.002>

2405-9595/© 2020 The Korean Institute of Communications and Information Sciences (KICS). Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

One major problem of conventional ANNs is that they usually consist of large number of neurons which are fully-connected to each other. As a result, ANNs structure would be very complex while they do not guarantee the maximum performance. However, biological investigations demonstrated that, natural neural networks have a sparse topology rather than fully-connected topology as in ANNs. In multi-layer perceptrons, as an example, “deep compression” is studied and it is shown that efficiently compressing deep neural networks using pruning methods can reduce the storage requirement of neural networks without affecting their accuracy [7].

Therefore, topology always affect the behavior and performance of the networks. Complex systems along with their graph theoretical basis [8] have attracted much attentions in various fields [9,10]. Based on the complex system topologies that are similar to biological systems network structure, application of different types of complex topologies in ANNs has been investigated in recent years [11–14] solving the complexity problem of ANNs. Although many real-world complex systems are not completely random in researches, randomly connected structures with various connectivity rules have shown outstanding results in improving functionality and performance of different types of ANNs [15–18].

In this paper, we review and explore articles which have studied the influence of topology on the performance of ANNs based on complex systems structures. We specifically concentrate on the effect of random structures which has been defined and used in complex systems theory. In these investigations after applying various complex structures in Hopfield associative memory and multi-layer perceptrons as two important types of ANNs and comparing the networks performance with each other, the random networks was shown to be the most efficient structures among fully-connected, small-world and scale-free networks.

The paper is organized as follows. First in Section 2 we briefly explain some basic concepts about complex systems structures and topologies. Then in Section 3 we summarize and discuss about some recent results of applying complex structures in two important types of ANNs. Finally in Section 4 we discuss some conclusions.

2. Basic concepts

In this section, we briefly discuss some preliminaries and the most important concepts of graph theory and basic realizations of complex networks and their characteristic.

2.1. Regular graphs

A network is simply defined as a series of interconnected nodes as its basic elements. The characteristics of nodes and their connections depend on the nature of the corresponding network. The degree k of a node is the number of links between the node and its neighbors. Nodes with a degree significantly more than the average degree of the network per node (z) are denoted as “hubs”. The probability of each node to have k links is called the “degree distribution” p_k of the

network. The most simple network is the regular (Reg.) graph that is a highly ordered lattice in a sense that each node is connected to all of its nearest neighbors. Therefore, all the nodes have the same degree k . These graphs do not offer any random and high connectivity. The conventional form of ANNs are usually regular graphs which are fully-connected (FC) to their neighbors.

2.2. Random networks

The simplest types of complex networks are random graphs. A specific type of random graphs can be constructed by randomly choosing two nodes and connect them together by a link and continue until the number of links become $\frac{Nz}{2}$ where N is the total number of nodes. This graph is called “Erdős–Renyie” (ER) random graph [8,19]. The degree distribution for ER random graphs is:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \quad (1)$$

where p is the probability of a link connection. For $N \gg k$ the degree distribution can be approximated by

$$p_k \simeq e^{-z} \frac{z^k}{k!} \quad (2)$$

which shows a Poisson distribution with the mean $\langle k \rangle = z$. A generalization of ER random graph is to assume that the probability of making links between node i and node j is p_{ij} which is independent of the other connections. These graphs are called “Kovalenko” random graphs [20] and will turn to ER if all p_{ij} ’s are equal.

2.3. Small-world networks

The following are two important definitions in small-world networks:

- *The small-world effect :*
In all kinds of networks the “small-world effect” occurs which indicates that the average path length between two nodes of a network can be orders of magnitude smaller than the total number of nodes.
- *Clustering Coefficient (C) :*
The clustering coefficient is the probability that two neighbors of a particular node are connected together.

Opposed to the real-world networks, in the thermodynamic limit ($N \rightarrow \infty$), the clustering coefficient C disappears in random graphs. Therefore, the small-world networks, which are neither regular nor random and interpolate between these two, are designed to exhibit a finite clustering coefficient while they show the small-world effect. The small-world networks have a fixed number of nodes and growth does not exist in their structure. Two types of small-world networks have been proposed:

1. *Watts–Strogatz model (WS):* In this model [21], starting from a regular lattice, one rewires all the links one by one from one end with probability p and connect that

end to a randomly chosen node from the rest of the graph. The average degree z of the lattice will remain unchanged, however the degree of each node can be smaller or larger than z .

2. *Newman–Watts model (NM)*: In this model [22], instead of rewiring the existed links, some links will be added to the network called “shortcuts”. Two nodes that are chosen at random will be connected together if they are not already connected to each other.

2.4. Scale-free networks

In contrast with random and small-world graphs, scale-free (SF) networks are open, as the most real networks, in a sense that they are constructed by adding a finite number of nodes to the network. As the time passes, the number of nodes will be greater which is the same for world wide web that grows exponentially by addition of new web pages. In random graphs it is assumed that the probability of two nodes being connected is uniform and random but in scale-free networks the “preferential attachment” or “rich-get-richer” arises. This is when the probability of getting connected to another node differs for each node. The constructed scale-free network include some “hubs” which, in addition to high number of links, have the growing probability of incoming connections. The most well-known type of these networks is the Barabasi–Albert (BA) model [8,23] with preferential attachment. The procedure of forming these networks is as follows:

1. Start with n disconnected nodes.
2. At every time step t a new node and $m \leq n$ links are added to the network.
3. The m links are connected to the existing nodes with probability

$$P(k_i) = \frac{k_i}{\sum_j k_j}. \quad (3)$$

Hence, $P(k_i)$ shows that the probability of being attached to a node depends on the number of links it has which indicates the “rich-get-richer” rule. After t time steps we have a network with $N = t + m$ nodes and mt links. The constructed scale-free networks degree distribution is

$$p_k \sim k^{-\gamma} \quad \text{for } \gamma > 1. \quad (4)$$

In real-world networks this scaling exists only for large k .

3. Artificial neural networks with complex structures

In this section we review the methodology and the results of some recent articles in which the complex network topologies have been applied on ANNs and compared to each other.

3.1. Hopfield associative memory

Hopfield (HF) networks are recurrent neural networks made up of binary neurons that evolve based on spin glass models in physics in order to hit a local minimum of energy. These neural networks can be used as associative memory

(i.e., also called Hopfield-type attractor neural network) in a sense that the Hopfield model can be assumed as a memory which can be addressed through its contents. In the Hopfield associative memory, the local minimums are called “stored patterns” [5,24]. An input pattern overlap with a stored pattern, even if it is perturbed by an amount of noise, is checked and the pattern will be recognized. The performance of these memories are evaluated by the stability of the stored patterns which increases as the induced noise to the pattern decrease, ability to recognize the stored patterns from an error-induced state and the networks storage capacity. After realizing that the Hopfield model [4] was a generalization of the infinite range spin glass [25] with the same theoretical background, researchers tried to investigate the model in its fully connected and randomly diluted structures [26]. In [27], randomly diluted networks shown to be much more simple and more useful than fully connected networks in which loss of connections does not only cause in loss of storage capacity but also increase it per connection.

One of the first attempts to apply random networks on ANNs was presented by McGraw et al. in 2003 [17]. They studied the computational performance of Hopfield-type attractor neural networks with asynchronous updating in random [5] and Hebbian [28] connection strength with different types of topologies such as regular lattice, ER random, WS small-world, and BA scale-free networks. It was shown that the network as a whole with random structure outperforms other topologies with the same number of nodes and connections in storage, stability and retrieval of the patterns. To prove their claim, they computed the overlap parameter by which the overall pattern retrieval can be recognized:

$$m^\mu \equiv \frac{1}{N} \sum_{i=1}^N x_i \xi_i^\mu \quad (5)$$

where the output of the i th node is shown by $x_i = \pm 1$ and ξ_i^μ is the i th memory state of the μ th pattern. In their simulations, it can be seen that by increasing the number of stored patterns p the overlap, which shows stability (when $m_{init} = 1$) and retrievability of the network (when $m_{init} = 0.5$), decreases. On the other hand, while the regular lattice degrades more rapidly than the other topologies, the random network shows the most slow reduction and the small-world network performance is between these two. The scale-free network degrades in a little bit more rapid manner than the random structure, but scale-free networks were shown to be more efficient in partial recognition of the patterns due to their highly connected neurons (hubs). Computing the overlap parameter for the scale-free networks without taking into account the low degree nodes confirms that eliminating these nodes would not affect the networks performance. Furthermore, due to recent researches, the connections between real neurons in the biological brain of higher animals as well as *C.elegans* [29] are not completely random and show small-world behavior. Hence, the authors suggested that small-world networks with moderate shortcuts could be almost efficient with less wiring costs than random networks. However, for the mean connectivity used in [17] ($\langle k \rangle = 50$), the network does not have the scale-free property, so there

is still lack of experiments about scale-free topology which seems to be one of the most important and the most similar structure to the brain neural network.

In 2006, Lu et al. [18], explored the effects of topology on the Hopfield-type associative memory neural networks as well. They also compared between different topologies with fixed number of nodes and connections but they added more details and some new results about the scale-free structure and its potential to improve the performance. They obtained the same conclusions as [17]. The authors generally claimed that more random topologies with less locality (lower clustering coefficient), would improve the pattern stability and retrievability of associative memory. Hence, adding a random component or some shortcuts to the network can largely enhance the network's performance. According to their simulations with the same procedure and metrics as in [17], ER random and BA scale-free networks compete with each other in having the most slow degradation of stability and retrievability due to the increase of stored patterns. The better structure between ER random and BA scale-free in each network can be recognized by a constant c that can be obtained numerically. If p is the random connection probability and $\langle k \rangle$ is the average degree of the network, then ER random network outperforms BA scale-free structure when $\frac{p}{\langle k \rangle} \leq c$ and opposite when $\frac{p}{\langle k \rangle} > c$. Moreover, according to McGraw's article, where it has been shown that larger nodes are less prone to errors, they claimed that asynchronous updating from larger nodes to smaller nodes shows better results than asynchronous updating in random and proved it through simulations.

3.2. Multi-layer artificial neural networks

Inspired by the brain's neuronal structure, these networks are designed by placing neurons in different layers. Conventionally, multi-layer ANNs consist of one input layer, one or more hidden layers and one output layer. The signal (information) propagates through the network and pass one layer to another in a way that the output of each layer will be the input for the next layer. The ANNs with multiple hidden layers are called "Deep neural networks" (DNNs) that are mostly used in image and voice recognition [30].

In 2012, in their neuroscience paper, Hill et al. [31] after accumulating various data from living brain of Wistar rats, found an interesting result about the rats brains functional connectivity. Their analysis showed that local neural microcircuits can be remarkably modeled by random synaptic formations. In 2016 Shafiee et al. [15], highly motivated by the findings of Hill et al. about the random formation of brain map, they introduced the concept of StochasticNet (Stoch. Net) in which the connectivity between neurons in DNNs is defined to be stochastic. This stochastic synaptic connections, taking advantage of Kovalenko random graph model [20], seems to be more efficient for some specific tasks. They calculated train and test errors for different image datasets (CIFAR-10 [32], MNIST [33], SVHN [34], and STL-10 [35]) to evaluate their claim. Results showed that with less than half (39%) the number of neural connections of a conventional DNN, comparable

accuracy in CIFAR-10, MNIST, and SHVN datasets can be obtained and higher accuracy of $\sim 6\%$ for STL-10 dataset observed compared with the conventional DNN. The results of the STL-10 dataset proves that when there is a low number of training samples, the StochasticNet can perform very efficiently. The StochasticNets also operate faster and show reduction in overfitting in comparison with the conventional DNN due to the smaller gap between training and test errors.

In 2019, for the first time Adjodah et al. [16], explored how to optimize the topology of communication between agents in deep reinforcement learning (DRL) (i.e., a DNN which learns by trial and error), built upon evolution strategies (ES) algorithm [36], to obtain better performance (i.e., the average total reward each agent obtain). They introduce a networked decentralized variant of ES, as NetES and investigated both experimentally and theoretically the influence of changing the topology of the neural network to ER random, scale-free, small-world, and fully connected networks while running ES on DRL. The authors claimed that ER random networks not only outperform all other structures but also perform better than the fully-connected networks by only $\frac{1}{3}$ number of neurons in the fully-connected network. After observing the best performance among the other structures, NetES was evaluated on five different DRL benchmarks selected from Open source Roboschool benchmark [37], MuJoCo framework [38], using ER structure. The experiments indicated that the performance is largely enhanced and ER networks with $\frac{1}{3}$ number of neurons outperforms fully-connected networks on all benchmark tasks from 9.8% on MuJoCo Ant-v1 to 798% on MuJoCo Humanoid-v1. Based on the experiments they concluded that ER can perform better in smaller networks.

4. Discussion and conclusion

Based on the results of the recent research works on applying complex systems topologies as discussed in the previous section, also shown in Table 1, we can conclude that ANNs with alternative complex topologies rather than being fully-connected show high performance with less complexity and reduced computational time and energy.

According to their investigations about Hopfield associative memory as an ANN, it has been found that randomly connected Hopfield networks perform better than regular and fully connected networks in general. Indeed randomness and regularity play the role of two opposite structures in terms of pattern retrieval. In other words, as the randomness increases, the capacity and memory retrieval by the network is improved. Moreover, it has been perceived that adding a certain number of shortcuts to a regular network, in order to see small-world behavior in the network, would enhance the pattern stability and retrievability. These small-world ANNs have been shown to perform better than regular ANNs and have a comparable performance with random networks but with less long-distance connections which result in less energy and computation time in associative memories. On the other hand, implementing scale-free topology in an associative memory network demonstrated that it can perform as well as random networks and

Table 1

The summary of the results of applying different complex structures in Hopfield associative memory (HF) and two types of multi-layer ANNs.

Article	ANN	Compared structures	Metrics	Best Structure	Dataset\Framework	Details
[17]	HF	ER, WS, BA, Reg. (asynchronous updating in random)	● stability \ retrieval ◇ partial retrieval ★ time \ wiring cost	● ER > BA > WS > Reg. ◇ BA ★ WS	–	Addition of shortcuts to a Reg. lattice enhance the stability and retrieval
[18]	HF	ER, WS, BA, Reg. (asynchronous updating from larger to smaller nodes)	stability \ retrieval	ER ~ BA > WS > Reg.	–	(1) if $\frac{p}{(k)} \leq c \rightarrow \text{ER} > \text{BA}$ if $\frac{p}{(k)} > c \rightarrow \text{BA} > \text{ER}$ (2) More randomness and less clustering increase performance.
[15]	DNN	Stoch.Net, DNN	● accuracy ◇ accuracy ★ speed \ overfitting	● Stoch.Net ~ DNN ◇ Stoch.Net > DNN ★ Stoch.Net > DNN	● CIFAR-10, MNIST, SVHN ◇ STL-10 ★ all datasets	Stoch.Net with $\frac{1}{2}$ connections perform the same or better than DNN.
[16]	DRL	FC, ER, SF, SW	performance	ER > SF > SW > FC	Humanoid-v1 (Roboschool and MuJoCo), HalfCheetah-v1, Hopper-v1 and ANT-v1 (MuJoCo)	ER with $\frac{1}{3}$ num. of neurons outperform FC from 9.8% on MuJoCo Ant-v1 to 798% on MuJoCo Humanoid-v1.

in some cases they will even outperform random structures depending on the mean degree of the lattice and the number of stored patterns [18]. Scale-free ANNs have been found to perform significantly better in partial image retrieval because of their highly connected hubs. This leads us to an interesting result that scale-free networks are very efficient in improving storage capacity.

Furthermore, as one of the most important ANNs, in DNNs, the stochasticNet structure with less than half connections of conventional DNN can improve overfitting while keeping the same or better learning accuracy (it differs on different datasets) and always increase the operational speed. In DRLs, using ES algorithm with different structures (BA scale-free, WS small-world, fully-connected and ER random networks) lead to the ER random structure show the best performance among the others with $\frac{1}{3}$ neurons of a fully-connected ANNs.

Last but not least, although natural complex systems have shown small-world and scale-free behavior, in the recent reviewed articles the role of these crucial networks in improving the ANNs performance has been neglected to some extent and more numerical and analytic computations and simulations seem to be needed specially in multi-layer ANNs. In our future work we will review various articles which have been published in recent years and explore the role of other topologies rather than random networks in more detail and with different methodologies.

Declaration of competing interest

The authors declare that there is no conflict of interest in this paper.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea

(NRFK) funded by the Ministry of Education (2018R1D1A1B07041981).

References

- [1] G.E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, *IEEE Trans. Audio Speech Lang. Process.* 20 (1) (2011) 30–42.
- [2] M.C.V. Ramirez, H.F. de Campos Velho, N.J. Ferreira, Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region, *J. Hydrol.* 301 (1–4) (2005) 146–162.
- [3] O. Erkaymaz, O. Mahmut, M. Perc, Performance of small-world feedforward neural networks for the diagnosis of diabetes, *Appl. Math. Comput.* 311 (2017) 22–28.
- [4] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Nat. Acad. Sci.* 79 (8) (1982) 2554–2558.
- [5] D.J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks*, Cambridge University Press, 1992.
- [6] *An Introduction To the Modeling of Neural Networks*, Vol. 2, Cambridge University Press, 1992.
- [7] S. Han, H. Mao, W.J. Dally, Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding, 2015, arXiv preprint [arXiv:1510.00149](https://arxiv.org/abs/1510.00149).
- [8] C. Gros, *Complex and Adaptive Dynamical Systems*, Springer, 2010.
- [9] I. Sohn, Small-world and scale-free network models for IoT systems, *Mob. Inf. Syst.* 2017 (2017).
- [10] Y. Lee, I. Sohn, Reconstructing damaged complex networks based on neural networks, *Symmetry* 9 (12) (2017) 310.
- [11] D. Stauffer, A. Aharoni, L. da Fontoura Costa, J. Adler, Efficient Hopfield pattern recognition on a scale-free neural network, *Eur. Phys. J. B* 32 (3) (2003) 395–399.
- [12] D. Simard, L. Nadeau, H. Kröger, Fastest learning in small-world neural networks, *Phys. Lett. A* 336 (1) (2005) 8–15.
- [13] J.W. Bohland, A.A. Minai, Efficient associative memory using small-world architecture, *Neurocomputing* 38 (2001) 489–496.
- [14] J.I. Perotti, F.A. Tamarit, S.A. Cannas, A scale-free neural network for modelling neurogenesis, *Physica A* 371 (1) (2006) 71–75.

- [15] M.J. Shafiee, P. Siva, A. Wong, Stochasticnet: Forming deep neural networks via stochastic connectivity, *IEEE Access* 4 (2016) 1915–1924.
- [16] D. Adjodah, D. Calacci, A. Dubey, A. Goyal, P. Krafft, E. Moro, A. Pentland, Communication topologies between learning agents in deep reinforcement learning, 2019, arXiv preprint [arXiv:1902.06740](https://arxiv.org/abs/1902.06740).
- [17] P.N. McGraw, M. Menzinger, Topology and computational performance of attractor neural networks, *Phys. Rev. E* 68 (4) (2003) 047102.
- [18] J. Lu, J. He, J. Cao, Z. Gao, Topology influences performance in the associative memory neural networks, *Phys. Lett. A* 354 (5–6) (2006) 335–343.
- [19] E. Renyi, On random graph, *Publ. Mat.* 6 (1959) 290–297.
- [20] I.N. Kovalenko, Theory of random graphs, *Cybernet. Systems Anal.* 7 (4) (1971) 575–579.
- [21] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, *Nature* 393 (6684) (1998) 440.
- [22] M.E. Newman, D.J. Watts, Renormalization group analysis of the small-world network model, *Phys. Lett. A* 263 (4–6) (1999) 341–346.
- [23] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512.
- [24] J. Hertz, A. Krogh, R.G. Palmer, H. Horner, Introduction to the theory of neural computation, *Phys. Today* 44 (1991) 70.
- [25] D. Sherrington, S. Kirkpatrick, Solvable model of a spin-glass, *Phys. Rev. Lett.* 35 (26) (1975) 1792.
- [26] I. Kanter, H. Sompolinsky, Associative recall of memory without errors, *Phys. Rev. A* 35 (1) (1987) 380.
- [27] B. Derrida, E. Gardner, A. Zippelius, An exactly solvable asymmetric neural network model, *Europhys. Lett.* 4 (2) (1987) 167.
- [28] D.O. Hebb, *The Organization of Behavior*, Vol. 65, Wiley, New York, 1949, Cambridge University Press, 1992.
- [29] N.A. Dunn, S.R. Lockery, J.T. Pierce-Shimomura, J.S. Conery, A neural network model of chemotaxis predicts functions of synaptic connections in the nematode *Caenorhabditis elegans*, *J. Comput. Neurosci.* 17 (2) (2004) 137–147.
- [30] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Netw.* 61 (2015) 85–117, [arXiv:1404.7828](https://arxiv.org/abs/1404.7828).
- [31] S.L. Hill, Y. Wang, I. Riachi, F. Schürmann, H. Markram, Statistical connectivity provides a sufficient foundation for specific functional connectivity in neocortical neural microcircuits, *Proc. Natl. Acad. Sci.* 109 (42) (2012) E2885–E2894.
- [32] A. Krizhevsky, G. Hinton, Learning Multiple Layers of Features from Tiny Images, Vol. 1, no. 4, Technical report, University of Toronto, 2009.
- [33] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [34] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A.Y. Ng, Reading Digits in Natural Images with Unsupervised Feature Learning, 2011.
- [35] A. Coates, A.Y. Ng, H. Lee, An analysis of single-layer networks in unsupervised feature learning, In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 215–223.
- [36] T. Salimans, J. Ho, X. Chen, S. Sidor, I. Sutskever, Evolution strategies as a scalable alternative to reinforcement learning, 2017, [arXiv:1703.03864](https://arxiv.org/abs/1703.03864).
- [37] Open AI, Roboschool, 2017. <https://github.com/openai/roboschool>. (Accessed 30 September 2017).
- [38] E. Todorov, T. Erez, Y. Tassa, Mujoco: A physics engine for model-based control, in: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2012, pp. 5026–5033.