

Accelerating the Evolution of Convolutional Neural Networks with Node-Level Mutations and Epigenetic Weight Initialization

Travis Desell
University of North Dakota
tdesell@cs.und.edu

ABSTRACT

This paper examines three generic strategies for improving the performance of neuro-evolving convolutional neural networks (CNNs): node-level mutation operations, epigenetic weight initialization and pooling connections. These were implemented using the Evolutionary eXploration of Augmenting Convolutional Topologies (EXACT) algorithm. Results were gathered over the period of a month using a volunteer computing project, where over 225,000 CNNs were trained and evaluated across 16 different EXACT searches. The node mutation operations were shown to dramatically improve evolution rates over traditional edge mutation operations (as used by the NEAT algorithm), and epigenetic weight initialization was shown to further increase the accuracy and generalizability of the trained CNNs. As a negative but interesting result, allowing for pooling connections was shown to degrade the evolution progress. The best trained CNNs reached 99.46% accuracy on the MNIST test data in under 13,500 CNN evaluations – accuracy comparable with some of the best human designed CNNs.

CCS CONCEPTS

• **Computing methodologies** → **Massively parallel algorithms; Neural networks; Genetic algorithms;**

KEYWORDS

Neuroevolution, Convolutional Neural Networks, Image Classification, Distributed Evolutionary Algorithms

ACM Reference Format:

Travis Desell. 2018. Accelerating the Evolution of Convolutional Neural Networks with Node-Level Mutations and Epigenetic Weight Initialization. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan*. ACM, New York, NY, USA, Article 4, 2 pages. <https://doi.org/10.1145/3205651.3205792>

1 INTRODUCTION

By investigating strategies for accelerating the neuro-evolution process, this work provides advances towards answering a big open challenge in machine learning: *What is the optimal architecture for a convolutional neural network?* In particular, this work investigates: *i) node-level mutation operators* which allow for faster growth of the evolved CNNs; *ii) epigenetic weight initialization* where child CNNs

retain weights from their already trained parent(s); and *iii) fractional max pooling* [2], which allows for the use of pooling connections between arbitrary sized input and output feature maps.

In "*epigenetic*" *weight initialization*, the initial weights of a child genome are inherited from their parent genomes, instead of using randomly initialized parameters. Graham's *fractional max pooling* [2] was utilized to allow pooling operations between arbitrarily sized feature maps – edges could be either convolutional or pooling. Lastly, five new mutation operators were investigated, in addition to the original NEAT-like operations initially implemented in EXACT. These were *i) enable/disable node*, which enable/disable a node along with its incoming and outgoing edges; *ii) add node*, which divides the graph by depth and randomly selects input/output nodes for a new node at the depth of the partition; *iii) split node* which disables a node and creates two new ones, which randomly create edges to the inputs/outputs of the split node; and *iv) merge node* which disables two nodes and creates a new one with edges to all the inputs/outputs of the merged nodes. Due to space limitations the reader is referred to Desell [1] for further implementation details.

2 RESULTS

Four different types of EXACT searches were run for a period of a month on the Citizen Science Grid¹ volunteer computing project: *i) node and edge operations, without pooling*; *ii) node and edge operations, with pooling*; *iii) only edge operations, without pooling*; *iv) only edge operations, with pooling*. All 16 searches were run simultaneously to minimize external influences such as the availability of volunteered hosts and network outages. Each search evaluated between 13,000 and 14,500 different trained CNNs. Table 1 presents the best, average, and worst validation and testing error and prediction rates for each of these searches.

Effects of Node Mutations and Pooling. The searches with node mutations and no pooling performed the best, with significant improvements over those without node mutations and with pooling. In terms of error and prediction rates, the searches with node mutations and pooling performed comparably to the searches without pooling and without node mutations. The searches with pooling and without node operations performed the worst.

Epigenetic Weight Initialization. Table 3 provides statistics for the validation and testing error and predictions of the best evolved CNNs from each of the four search types. Each of these CNNs were retrained with weights initialized by the standard strategy recommended by He *et al.* [3]. Table 2 shows the min, average and maximum of these values, having retrained each of the best CNNs five times.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5764-7/18/07.

<https://doi.org/10.1145/3205651.3205792>

¹<https://csgrid.org>

Search	Validation Error			Testing Error			Validation Predictions			Testing Predictions		
	Worst	Avg	Best	Worst	Avg	Best	Worst	Avg	Best	Worst	Avg	Best
Node+Edge 1	259.49	234.71	224.37	245.48	220.99	188.62	99.20%	99.28%	99.36%	99.14%	99.26%	99.46%
Node+Edge 2	300.00	276.16	258.75	296.74	267.02	243.16	99.07%	99.17%	99.27%	99.03%	99.14%	99.30%
Node+Edge 3	262.86	240.13	232.08	249.33	222.21	189.20	99.20%	99.27%	99.37%	99.13%	99.25%	99.39%
Node+Edge 4	281.66	256.58	247.89	273.01	235.60	210.81	99.08%	99.19%	99.28%	99.04%	99.21%	99.36%
Node+Edge, Pooling 1	346.27	320.52	302.79	359.28	326.21	274.14	98.89%	99.02%	99.12%	98.82%	99.00%	99.20%
Node+Edge, Pooling 2	335.53	296.80	283.96	350.83	299.32	271.21	98.95%	99.07%	99.18%	98.82%	99.03%	99.25%
Node+Edge, Pooling 3	349.16	335.53	316.00	441.07	358.78	312.25	98.83%	98.96%	99.11%	98.74%	98.92%	99.15%
Node+Edge, Pooling 4	357.96	333.84	323.10	372.10	320.63	284.59	98.85%	98.96%	99.12%	98.79%	98.96%	99.17%
Edge 1	324.57	294.86	278.76	329.86	294.91	254.20	99.01%	99.09%	99.18%	98.86%	99.03%	99.18%
Edge 2	321.82	308.53	298.78	363.19	319.48	280.03	98.97%	99.07%	99.17%	98.79%	98.98%	99.18%
Edge 3	292.04	286.42	271.55	339.46	282.48	225.64	99.07%	99.13%	99.22%	98.95%	99.07%	99.26%
Edge 4	327.44	293.25	277.72	327.98	292.53	254.06	98.97%	99.09%	99.22%	98.89%	99.03%	99.19%
Edge, Pooling 1	390.09	368.79	350.83	417.36	376.38	331.14	98.76%	98.87%	99.02%	98.65%	98.79%	99.07%
Edge, Pooling 2	378.39	359.22	338.24	449.45	394.67	329.94	98.79%	98.86%	98.95%	98.49%	98.72%	98.98%
Edge, Pooling 3	389.28	354.73	327.78	421.46	359.62	314.51	98.79%	98.92%	99.02%	98.67%	98.82%	98.99%
Edge, Pooling 4	392.37	382.92	342.80	478.48	383.83	310.83	98.67%	98.79%	98.95%	98.56%	98.79%	99.08%

Table 1: EXACT Search Error and Prediction Rates

Search	Validation Error			Testing Error			Validation Predictions			Testing Predictions		
	Worst	Avg	Best	Worst	Avg	Best	Worst	Avg	Best	Worst	Avg	Best
Node+Edge	313.83	291.71	268.00	260.92	240.57	223.70	99.09%	99.17%	99.26%	99.19%	99.26%	99.30%
Node+Edge, Pooling	469.48	392.53	357.46	344.27	321.45	300.77	98.92%	98.97%	99.00%	98.93%	98.99%	99.06%
Edge	380.41	355.73	331.26	317.61	298.38	290.50	98.91%	98.99%	99.08%	98.85%	99.03%	99.15%
Edge, Pooling	569.96	539.27	470.25	571.04	501.45	444.72	98.43%	98.51%	98.64%	98.20%	98.46%	98.64%

Table 2: Retraining Error and Prediction Rates

Search Type	Weights	Val. Err.	Test Err.	Val. Pred.	Test Pred.
Node+Edge	93813	259.49	188.62	99.20%	99.46%
Node+Edge, Pooling	50387	335.53	271.21	98.99%	99.25%
Edge	50285	292.04	225.64	99.12%	99.26%
Edge, Pooling	30792	342.80	310.83	98.95%	99.08%

Table 3: Best Evolved Neural Networks

3 DISCUSSION AND FUTURE WORK

Epigenetic weight initialization was shown to provide a significant improvement to test error rates, which were reduced by 0.11% to 0.44%, and on average by 0.20% to 0.62% when compared to weights initialized by the standard randomized method. Node mutations reduced test error by 0.21% for the best found genomes and on average 0.175% for the searches without pooling and by 0.18% for the best found genomes and by 0.1575% for the searches with pooling. These results are significant as error rates were already below 1%.

Allowing edges to be either convolutional or fractional max pooling edges, was shown to degrade performance. Overall, it seems that the additional complexity of the search space caused by allowing for pooling edges did not overcome any benefit provided by allowing pooling. Further this does reflect on the sentiment of some members of the machine learning community, e.g. Geoffrey Hinton describing the pooling operation as a “big mistake” [4] and other work showing well performing CNNs which do not require pooling [5].

Current work is focusing on reproducing the same results on other data sets such as the CIFAR and TinyImage datasets. Additionally, as EXACT can track insertion rates from the various recombination operators, additional benefits may be gained by online adaptation of how frequently the operations are used. For example, it may improve performance to initial focus on operations which grow the CNN size (adding and splitting nodes and edges) and then later focus on operations which refine the CNN (disabling nodes and edges).

REFERENCES

- [1] Travis Desell. 2017. Developing a Volunteer Computing Project to Evolve Convolutional Neural Networks and Their Hyperparameters. In *The 13th IEEE International Conference on eScience (eScience 2017)*. 19–28. <https://doi.org/10.1109/eScience.2017.14>
- [2] Benjamin Graham. 2014. Fractional max-pooling. *arXiv preprint arXiv:1412.6071* (2014).
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.
- [4] Geoffrey Hinton. 2014. Machine Learning SubReddit - AMA Geoffrey Hinton. (2014). [Accessed Online 2018] https://www.reddit.com/r/MachineLearning/comments/2lmo0l/ama_geoffrey_hinton/clyj4jv/.
- [5] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. 2014. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806* (2014).