# Are skip connections necessary for biologically plausible learning rules?

Daniel Jiwoong Im, Rutuja Patil, Kristin Branson, {imd,patilr,bransonk}@hhmi.janelia.org - HHMI Janelia Research Campus

## Motivations

- Backpropagation (BP) is the workhorse of deep learning.
- To understand learning in real neural networks, we must consider biologically motivated credit assignment methods such as random feedback alignment (FA) [1] and difference target propgation (DTP) [2].
- The problems with FA and DTP:
  1. Does not scale to deep networks
  2. Highly sensitive to hyper-parameter choice.
- Skip connections are shortcuts that jumps over layers. They are common in real neural networks and improve performance in artificial neural networks.

## Related Work

Machine Learning

The concept of skip or dense connection in deep learning was introduced with residual networks[3], highway networks[4], and densely networks[5].

Neuroscience

1. The brain contains of many examples of both skip connections. For example, neocortex has a similar structure to residual nets[6].
2. In mouse brain, the connectivity strength follows log-normally distribution, which implies sparse long-range connections[7].

## Methods

We compare the performance of

1. Standard neural network (NN) versus densely-connected neural network (DN, all forward skip connections),
2. Convolutional neural network (ConvNet) versus densely connected ConvNet (DenseConvNet).

trained using BP, FA, and DTP.

## Performance

Comparing the performance of BP, FA, DTP on MNIST dataset for different network depths (using best hyper-parameters).

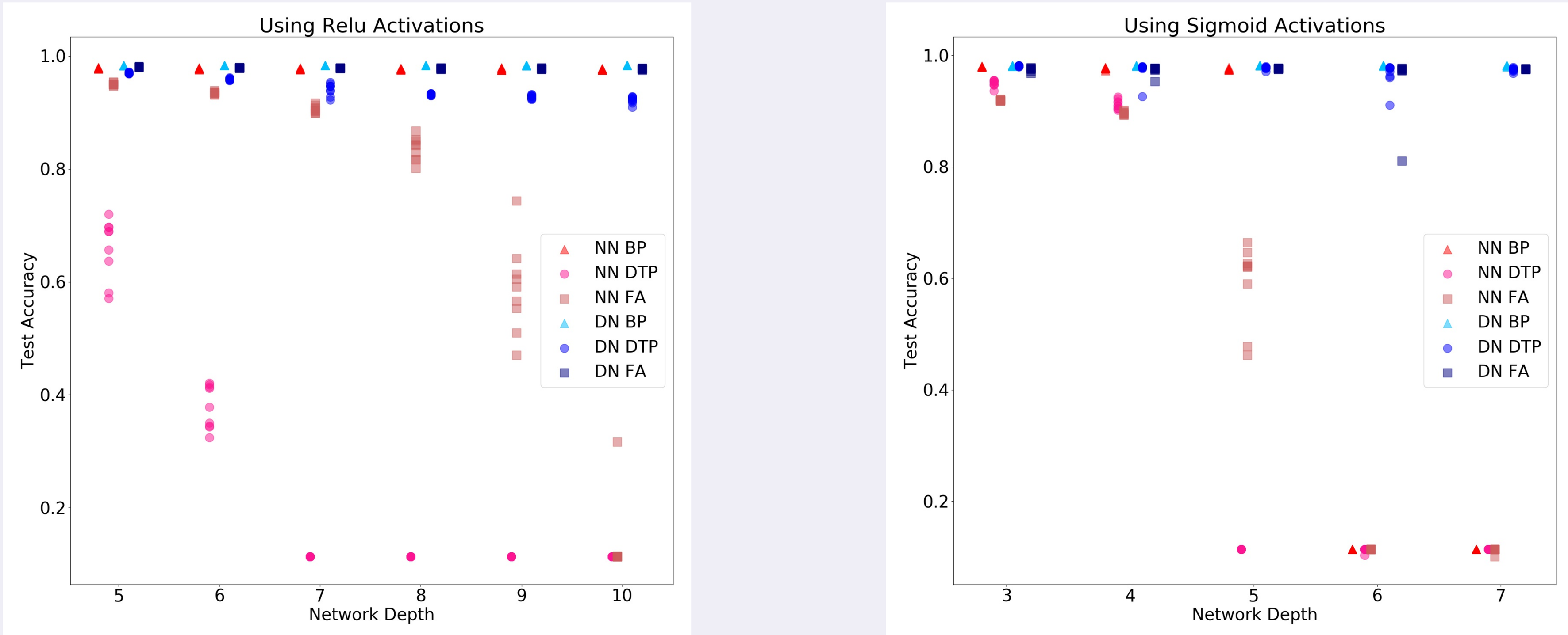


Figure 1: NN vs DN

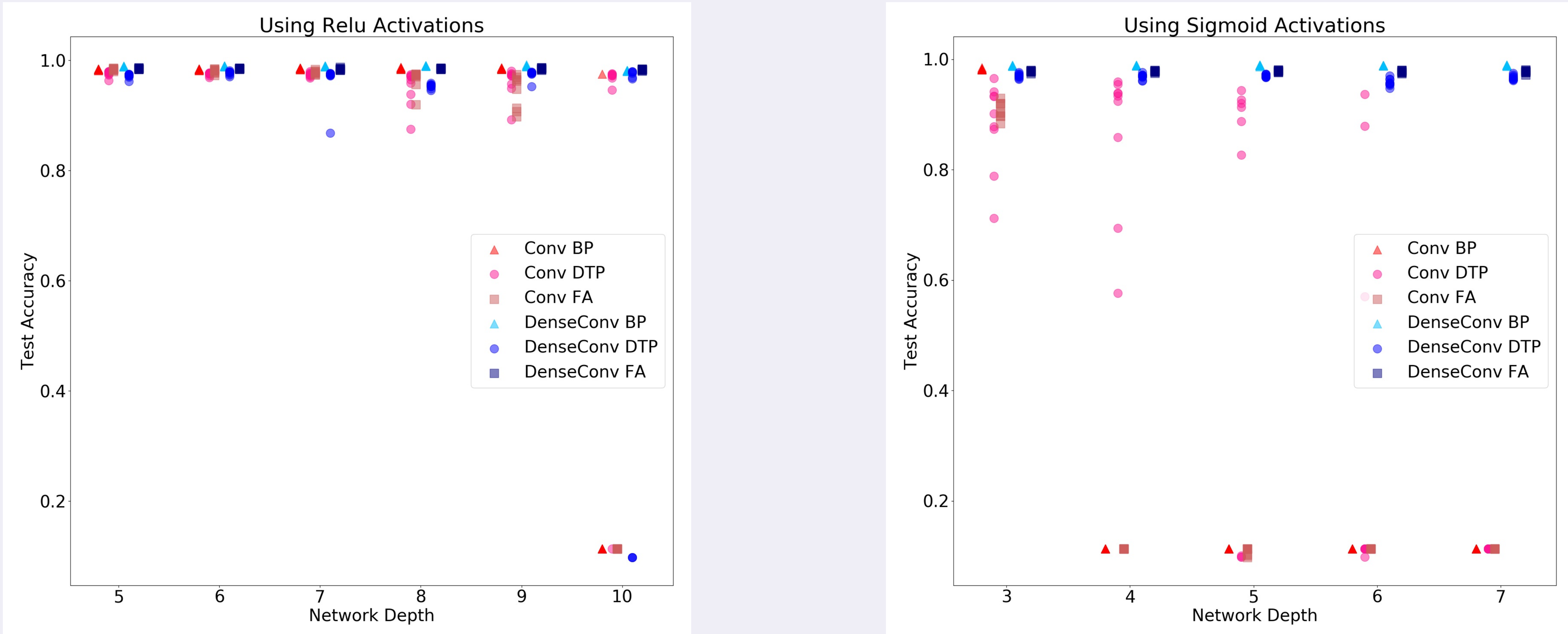


Figure 2: ConvNet vs DenseConv

- The test accuracy for FA and DTP approaches chance as network depth increases for NN and ConvNet, but remains high for DN and DenseConvNet.
- DTP gains the most from having dense connections, because it is local learning rule.

## Hyper-parameters

The table provides a range of hyper-parameters explored for various learning rules. We explored three sets of learning rates and five sets of the early stopping starting points.

| | Sigmoid | | | Relu | | |
|---|---|---|---|---|---|---|
| | Learning Rate | Early Stop | Depth | Learning Rate | Early Stop | Depth |
| BP | (0.1-0.001) | (200k-800k) | (3-7) | (0.001-0.00001) | (100k-300k) | (5-10) |
| FA | (0.01-0.0001) | (200k-800k) | (3-7) | (0.001-0.00001) | (600k-1000k) | (5-10) |
| DTP | (0.01-0.0001) | (200k-800k) | (3-7) | (0.001-0.00001) | (50k-150k)[1] | (5-10) |

## Discussion

- (Past) Biologically-inspired learning rules have been studied in isolation.
- (Future) Biologically-inspired learning rules should be considered in the types of architectures found in real neural networks.

## Sensitivity Analysis

The sensitivity of test accuracy over different hyper-parameter choices. X-axis corresponds to different hyper-parameter choices sorted from best to worst performances.

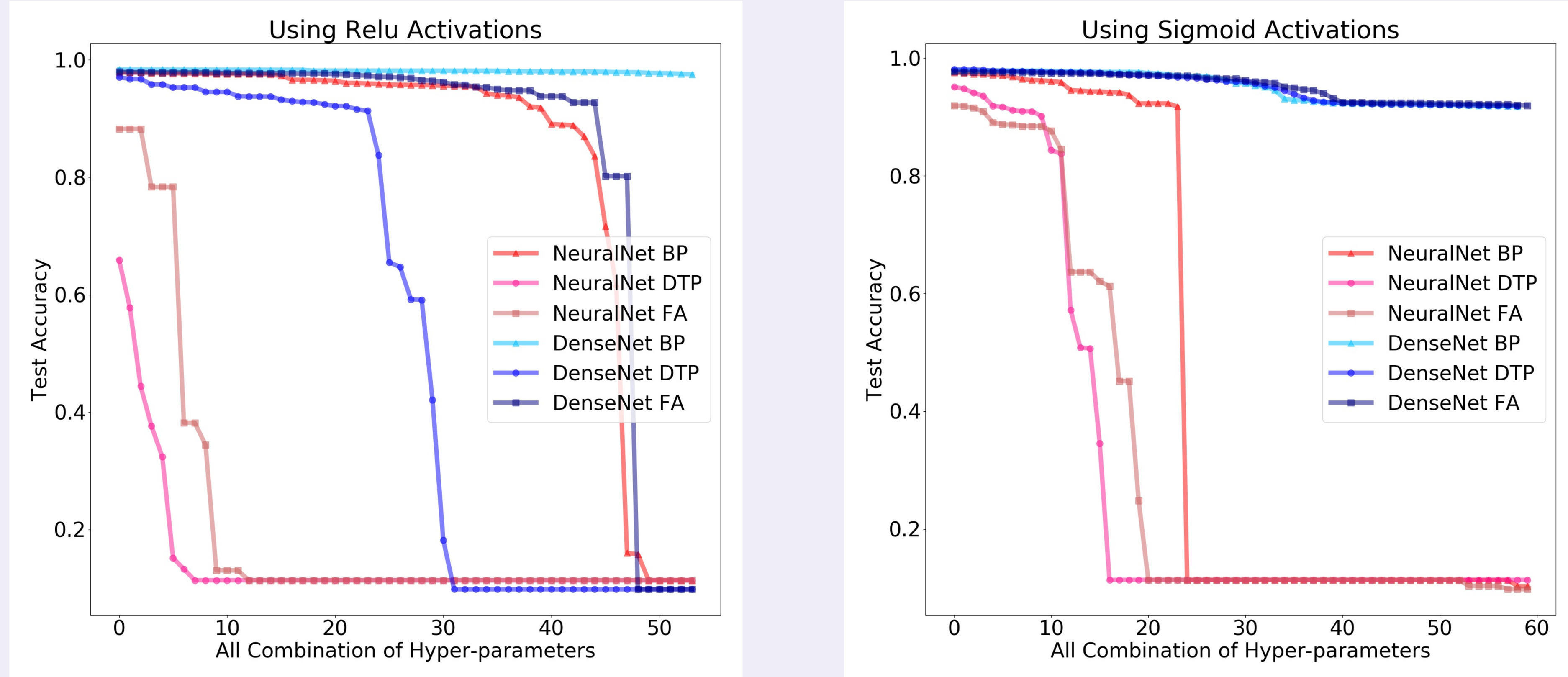


Figure 3: NN vs DN

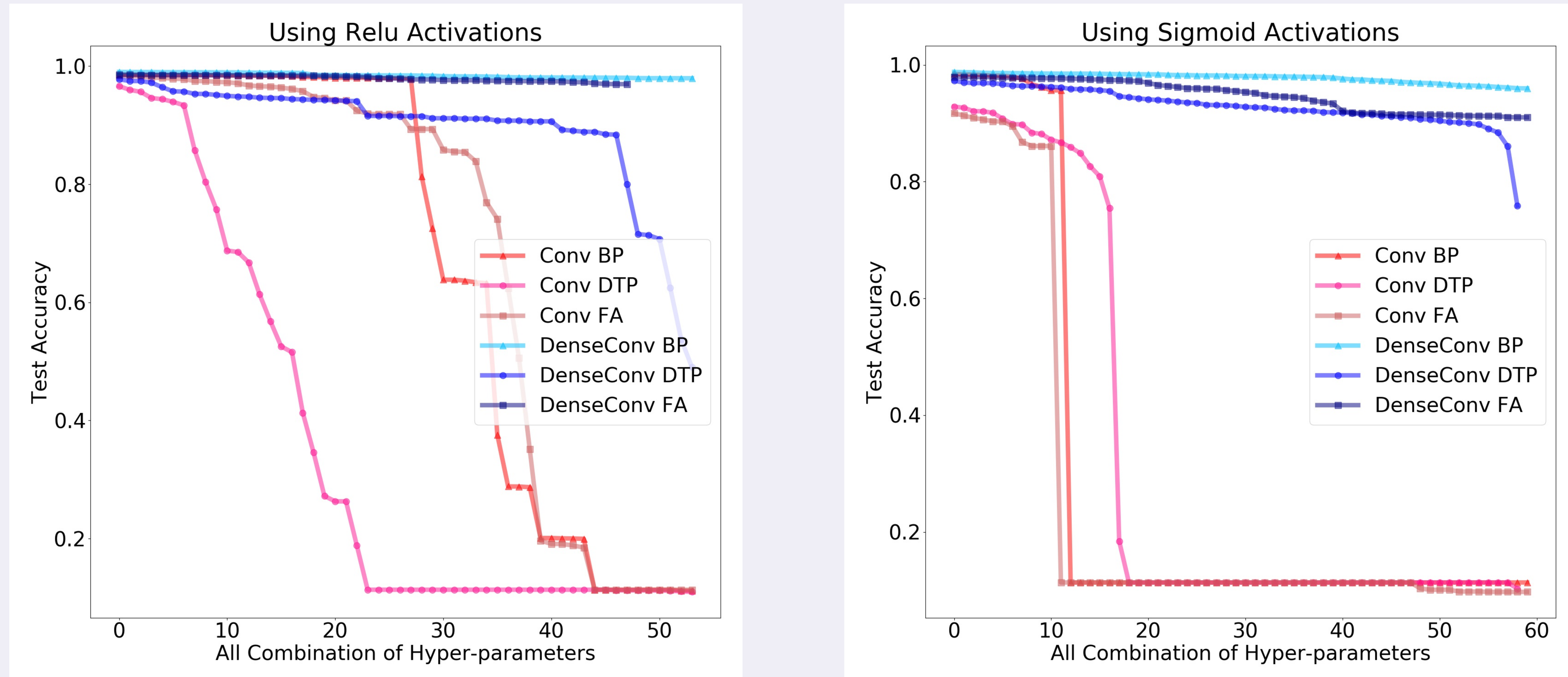


Figure 4: ConvNet vs DenseConv

- The performance of FA and DTP for NN and ConvNet (red lines) is highly dependent on hyper-parameter choice, but not for DN and DenseConvNet (blue lines)

**Having dense connections makes the model more robust to differnet hyper-parameters.**

## Reference

1.Timothy P. Lillicrap, Daniel Cownden, Douglas B. Tweed, and Colin J. Akerman. Random synaptic feedback weights support error backpropagation for deep learning. In Nature Communication, pages 207—214. 2016.

2.Dong-Hyun Lee, Saizheng Zhang, Asja Fischer, and Yoshua Bengio. Difference target propagation. European conference of machine learning, 2015. 3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. Computer Vision and Pattern Recognition (CVPR), 2015

4. Rupesh Kumar Srivastava, Klaus Greff, Jürgen Schmidhuber. Highway Networks, International Conference on Machine Learning Workshop 2015

5. Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. (CVPR), 2017

6. Alex M. Thomson. Neocortical layer 6, a review. Front. Neuroanat, 2010

7. Seung Wook Oh, Julie A. Harris, and Hongkui Zeng. A mesoscale connectome of the mouse brain. In Nature, pages 207—214. 2014