SAND2017-7531C

The Center for Cyber Defenders

Expanding computer security knowledge

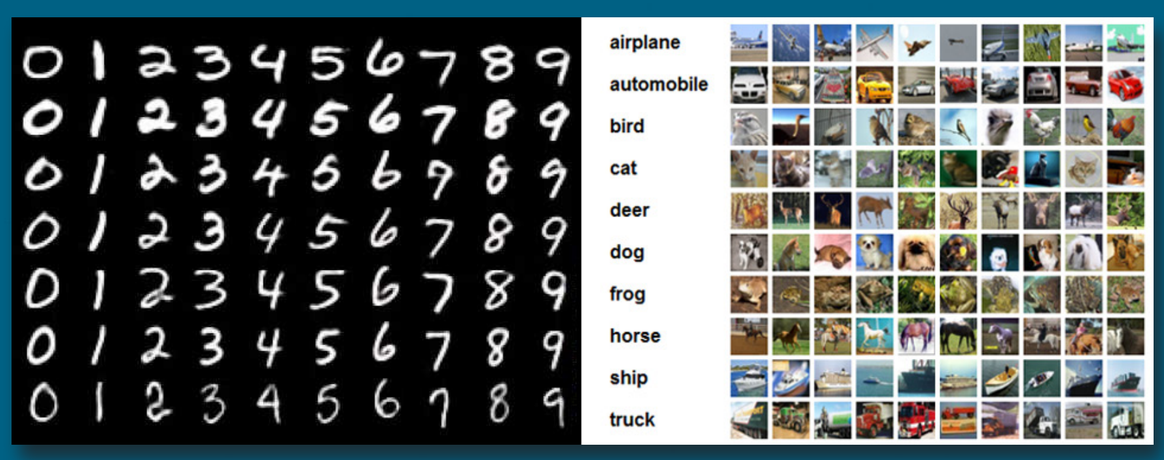
Applying Neurogenesis Deep Learning to Convolutional Neural Networks

Brendan Curran, Missouri University of Science and Technology
Min Je Jung, Georgia Institute of Technology

Project Mentor: Timothy Draelos, Org. 6362

Problem Statement:

It is difficult to utilize a trained deep neural network to recognize data that has never been encountered. Training on new data is likely to erase the understanding of the original data. The brain utilizes a mechanism known as neurogenesis to create new neurons capable of adapting to novel inputs. Designing a similar process for deep neural networks could provide the ability to expand their utility.



MNIST Dataset Cifar-10 Dataset

Objectives and Approach:

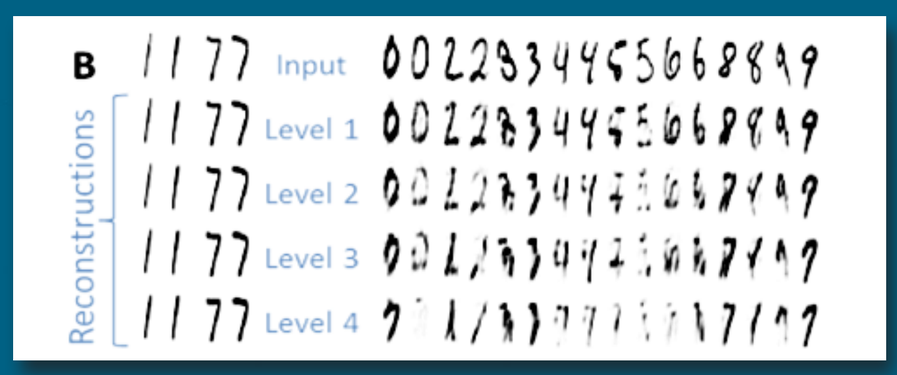
Train a classifier on a dataset such as MNIST or CIFAR-10 to recognize new classes such as letters in a NIST dataset or new objects in CIFAR-100.

To determine when and where to add new filters to a convolutional neural network (CNN):

- Transform classifier into autoencoder
- Reconstruct images with autoencoder on a layer by layer basis to quantify how well network represents input
- If reconstruction error is too high, add new filter at that layer
- Train only new weights on the novel data
- Transfer encoder weights into classifier and retrain output layer to include additional classes

Results:

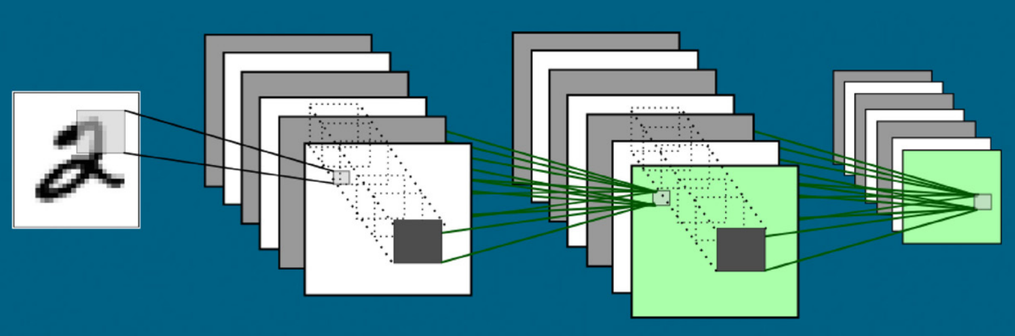
Transforming a classifier into an autoencoder has shown the ability to differentiate learned and unlearned data. Reconstruction errors are significantly larger for new classes of data and increase as feature detectors become more specific. Neurogenesis Deep Learning (NDL) has already been successfully applied to autoencoders and needs to be ported to these convolutional autoencoders. We expect to maintain high performance while incorporating new data.



Reconstructions of all MNIST classes on a layer by layer basis using a network trained just on 1's and 7's

Impact and Benefits:

Training large networks is costly in terms of time and computing resources. Neurogenesis Deep Learning could reduce the need to train entirely new networks and instead utilize the information contained in existing static networks. NDL could be used as a continuous learning method, resulting in large networks expanding their knowledge and performance with changing data.



Convolution architecture highlighting new filters and weights to be trained in green

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2017-XXXXC.