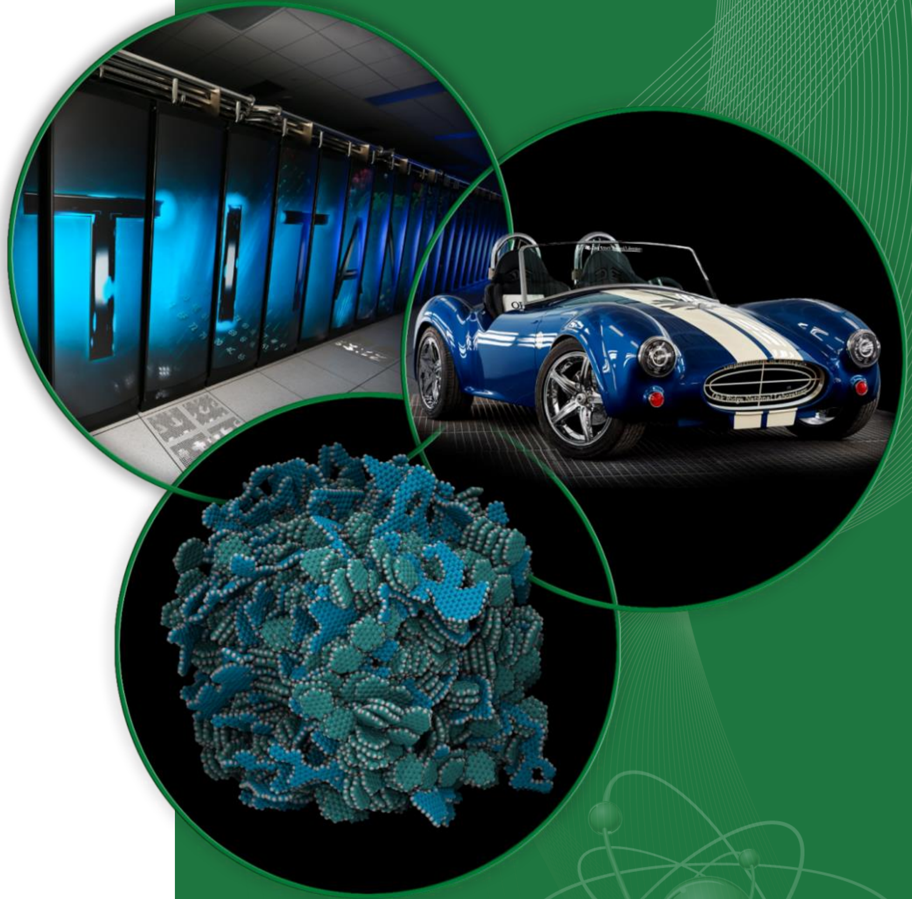
Evolving Deep Networks Using HPC

Steven R. Young, Derek C. Rose,
Travis Johnston, William T. Heller,
Thomas P. Karnowski, Thomas E. Potok
and Robert M. Patton
Oak Ridge National Laboratory

Gabriel Perdue
Fermi National Accelerator Laboratory

Jonathan Miller
Universidad Tecnica Federico Santa Maria



Overview

- Deep Learning for Scientific Data
- Challenges
- Tools
- Next Steps

Deep Learning for Scientific Data

Commercial Data

State of the Art Results

Object Recognition

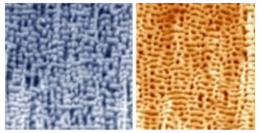
Face Recognition

Characteristics

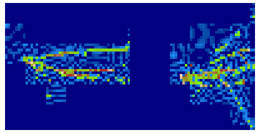
- Data is easy to collect
- Inexpensive labels

Scientific Data

Challenging New Domains



Materials Science



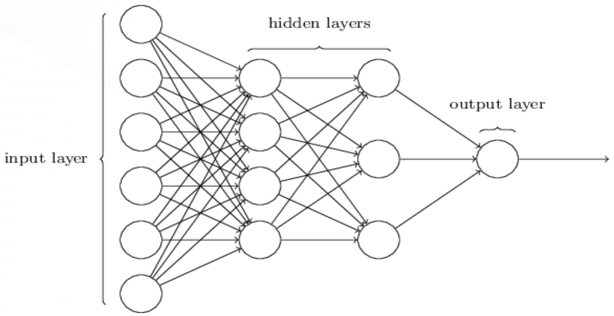
High Energy Physics

Characteristics

- Data is difficult to collect
- Few labels available

Problem: Adaptability Challenge

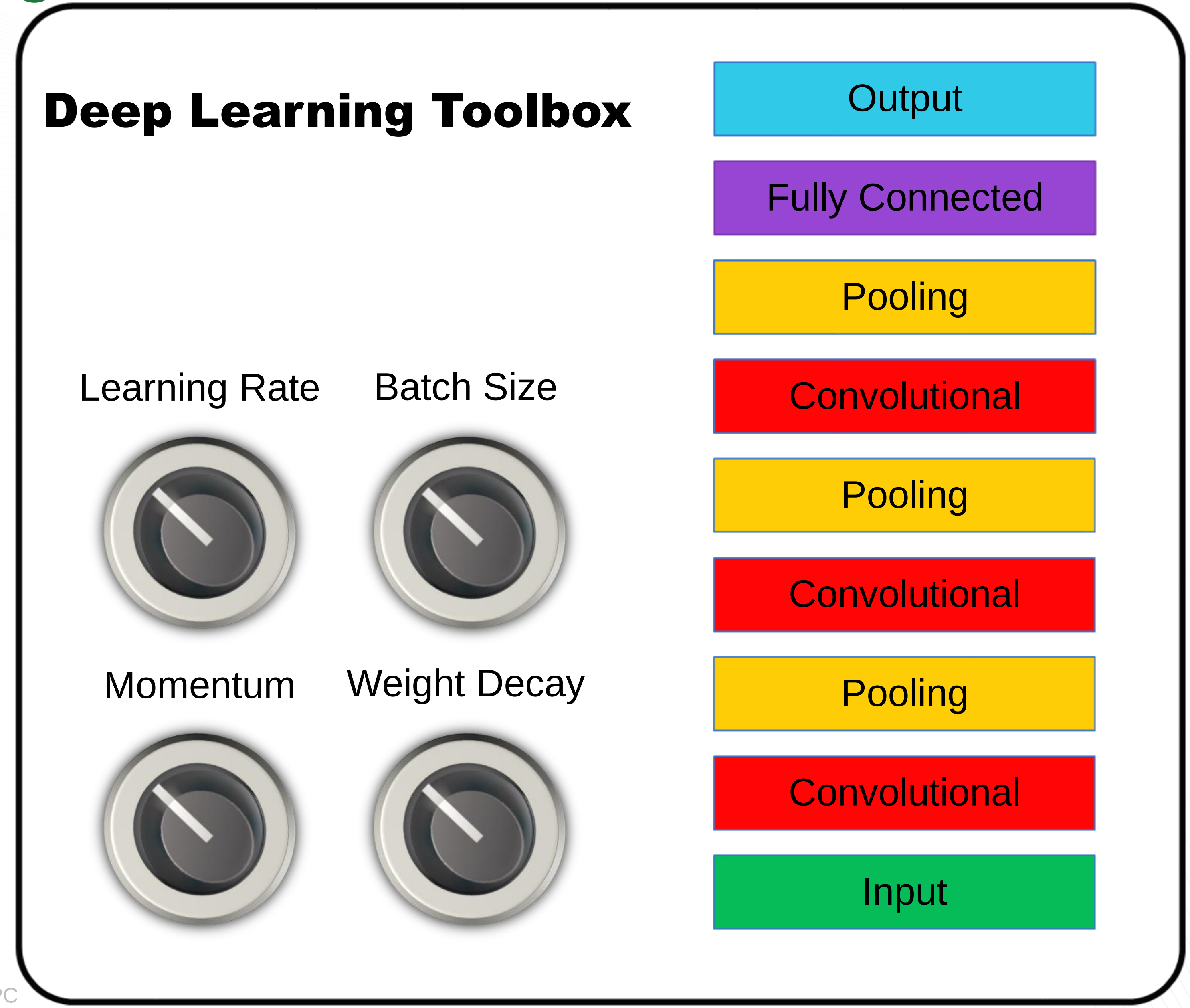
- ***Premise:*** For every data set, there exists a corresponding neural network that performs ideally with that data



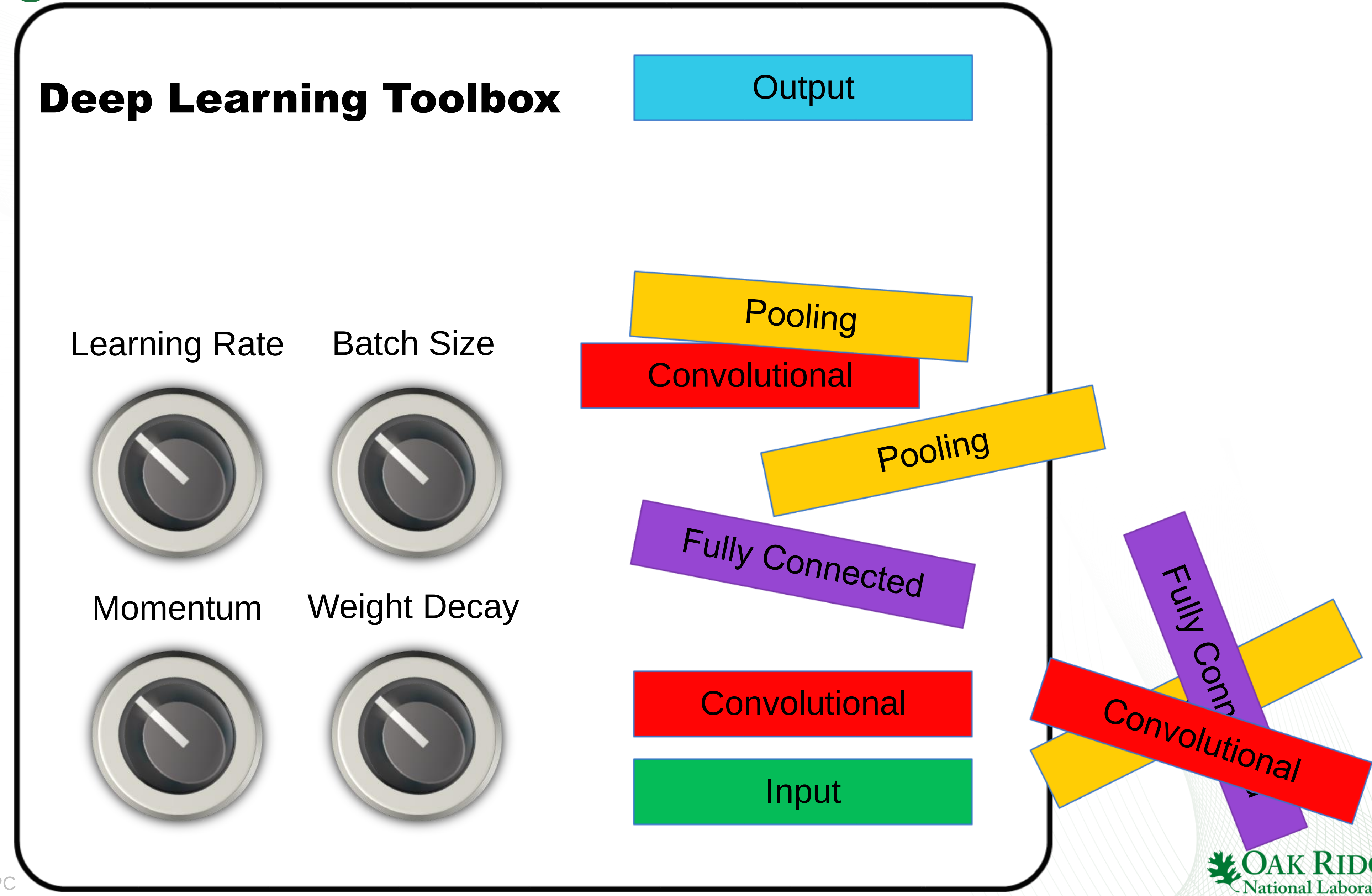
- What's the ideal neural network architecture (i.e., hyper-parameters) for a particular data set ?
- Widely-used approach: intuition
  1. Pick some deep learning software (Caffe, Torch, Theano, etc)
  2. Design a set of parameters that defines your deep learning network
  3. Try it on your data
  4. If it doesn't work as well as you want, go back to step 2 and try again.

OAK RIDGE National Laboratory

The Challenge

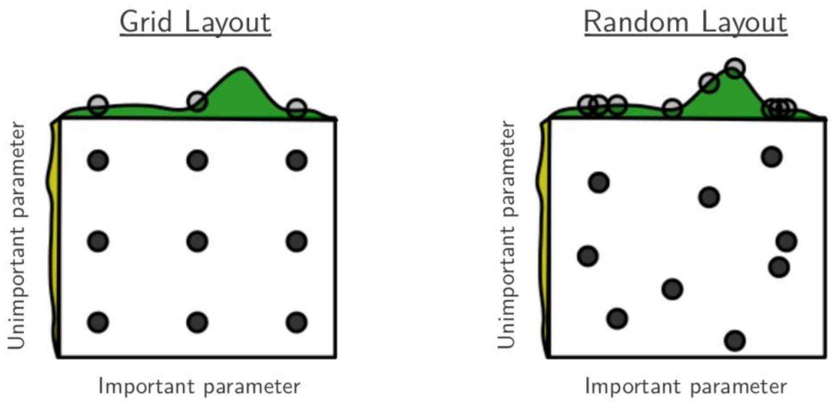


The Challenge



Current Approaches to Hyper-parameter Optimization

- Use *out-of-the-box* network
  - Why spend time trying to create your own network when there are already so many good ones available? Surely, one of those networks will also solve your problem.



Bergstra, J, and Bengio, Y. Random Search for Hyperparameter Optimization, Journal of Machine Learning Research, Feb. 2012.

- *Tune* an out-of-the-box network
  - **Hyper-parameter sweeps**
    Assumes independence of hyper-parameters
  - **Grid search**
    Requires training an exponential number of networks (infeasible)
  - **Random search**
    Significant improvement over grid search, but doesn't make use of information learned during training.

OAK RIDGE National Laboratory

What can we do with Titan?

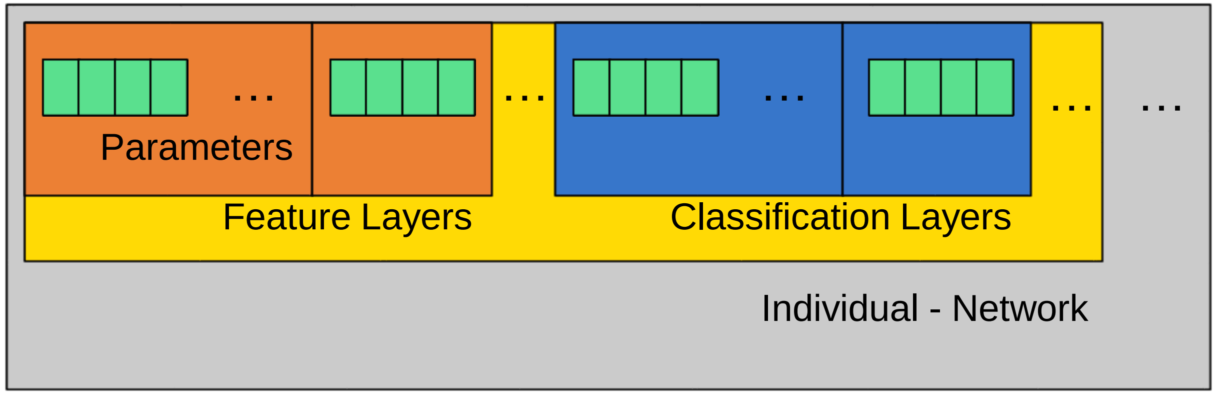
18,688 GPUs

OAK RIDGE National Laboratory

MENNDL: Multi-node Evolutionary Neural Networks for Deep Learning

- Evolutionary algorithm as a solution for searching hyper-parameter space for deep learning
  - Focus on Convolutional Neural Networks
  - Evolve *only* the topology with EA; typical SGD training process
  - Generally: Provide *scalability* and *adaptability* for many data sets and compute platforms
- Leverage more GPUs; ORNL's Titan has 18k GPUs
  - Next generation, Summit, will have increased GPU capability
- Provide the ability to apply DL to new datasets quickly
  - Climate science, material science, physics, etc.

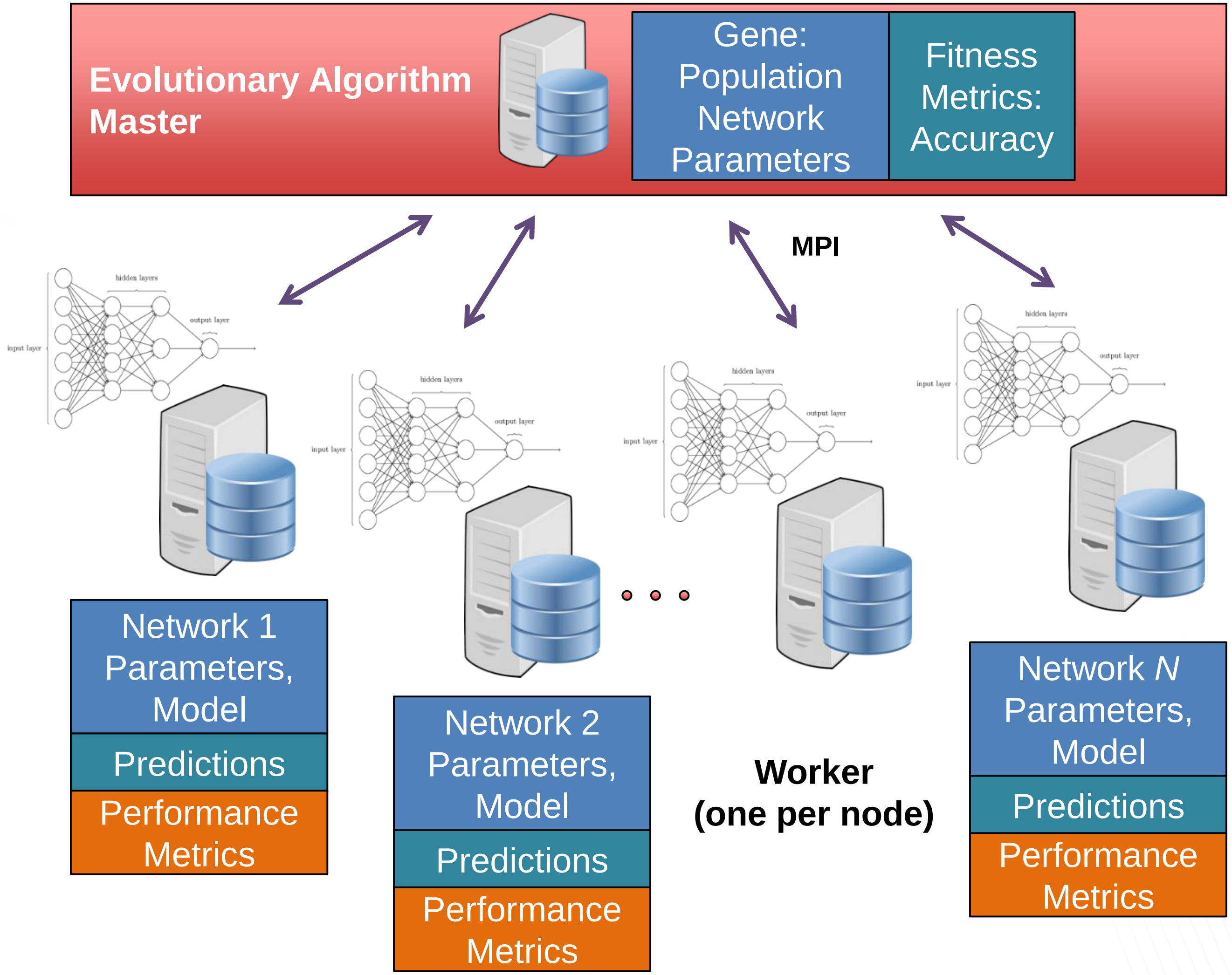
Designing the Genetic Code



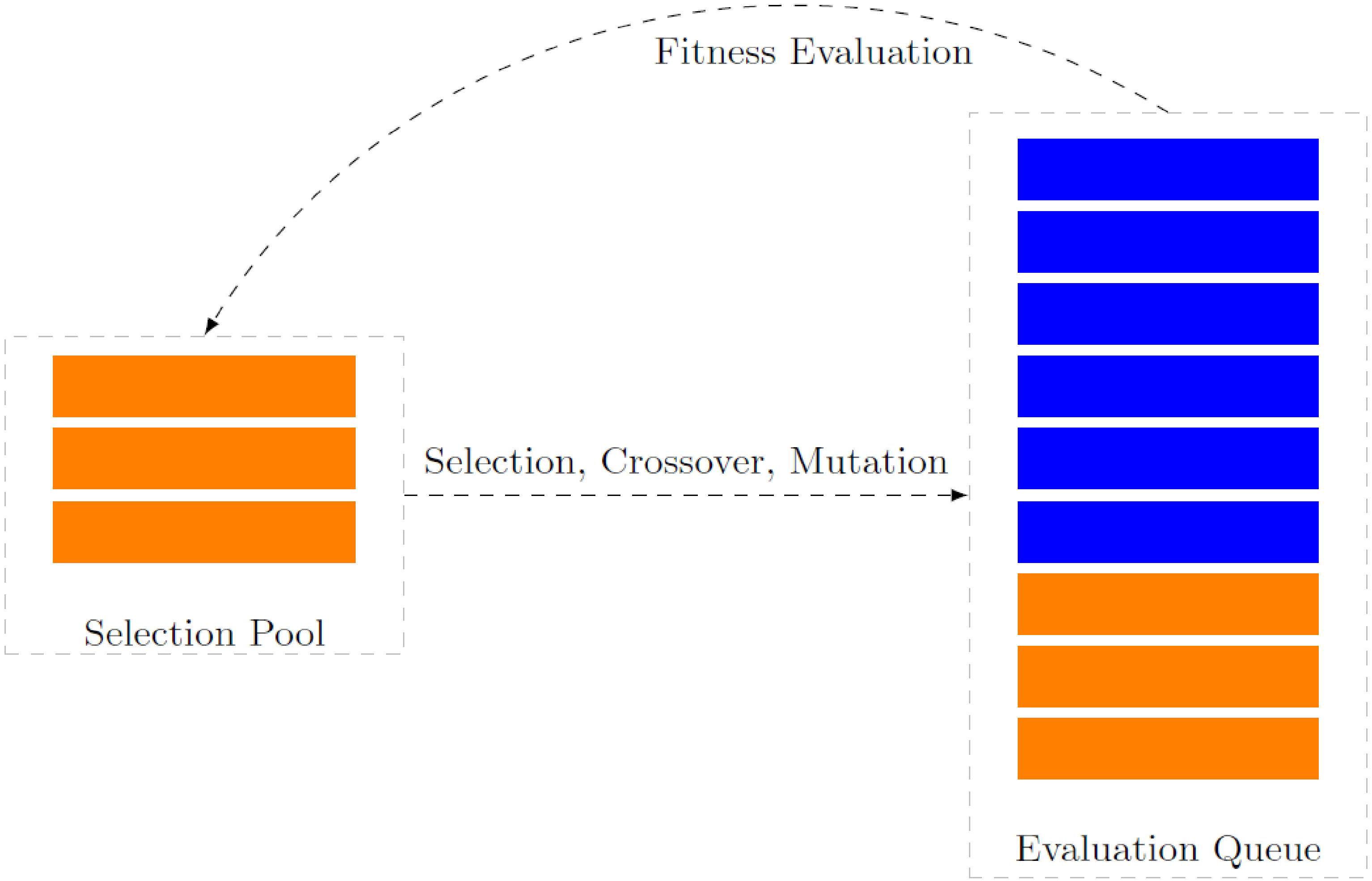
Population – Group of Networks

- Goal: facilitate complete network definition exploration
- Each population member is a network which has a genome with sets of genes
  - Fixed width set of genes corresponds to a layer
    - Layers contain multiple distinct parameters
  - Restrict layer types based on section
    - Feature extraction and classification
    - Minor guided design in network, otherwise we attempt to fully encompass all layer types

MENNDL: Communication

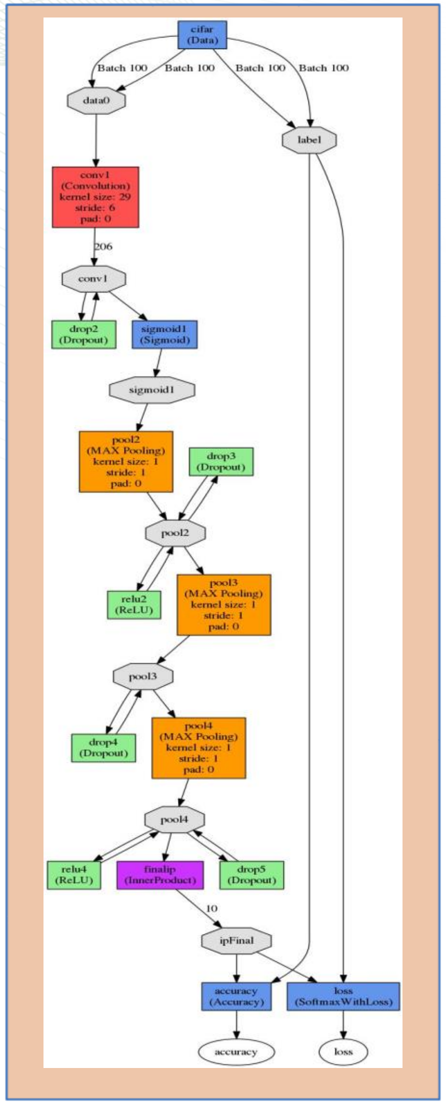


Asynchronous Evolutionary Algorithm

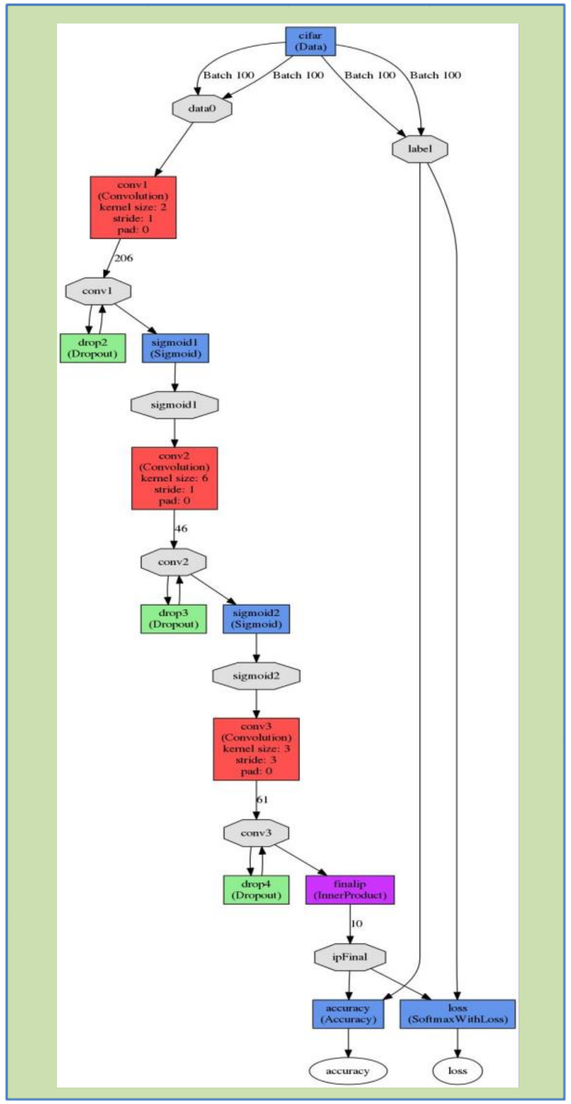


OAK RIDGE National Laboratory

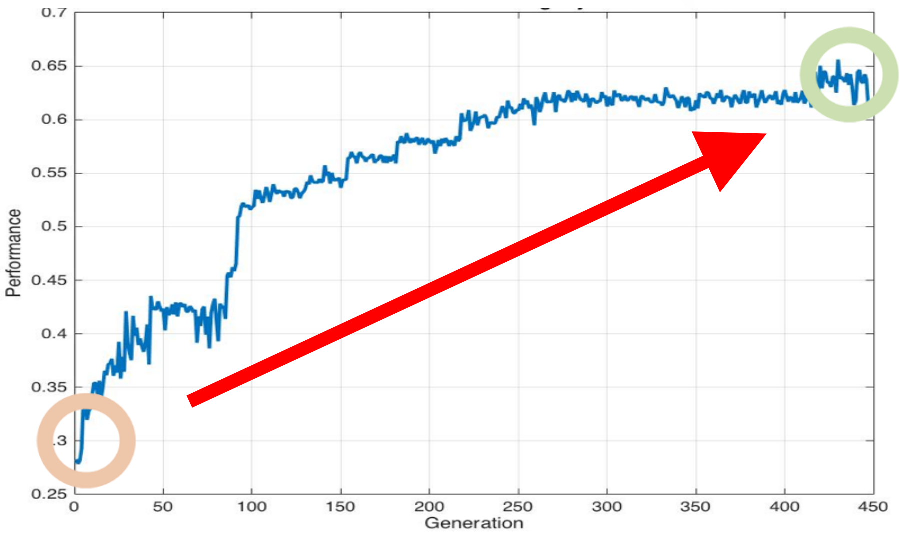
Hyper-parameter Values and Improved Performance



Evolved



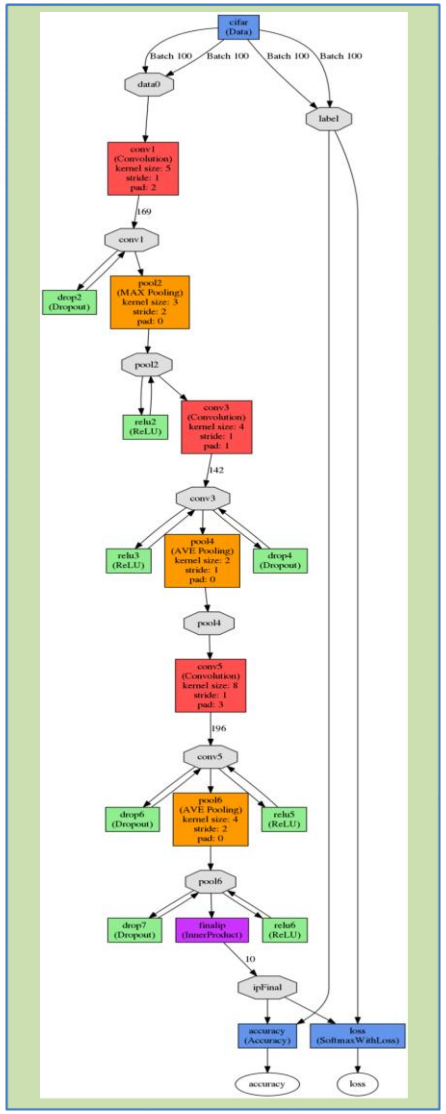
- Currently T&E of latest code that changes all possible parameters (e.g., # of layers, layer types, etc)
- Using just 4 nodes
- **From 27% to 65% Accuracy**



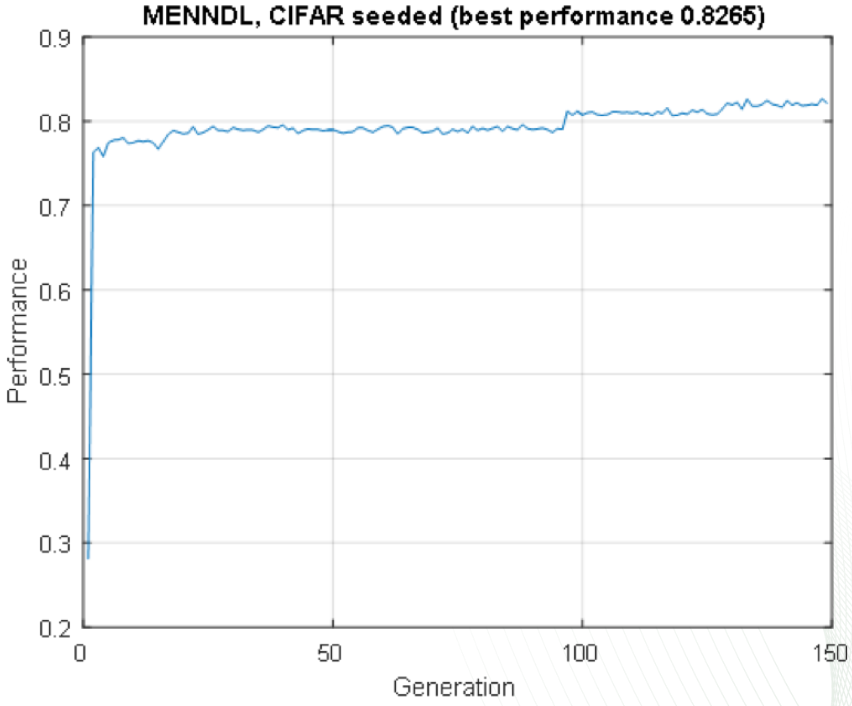
OAK RIDGE National Laboratory

Hyper-parameter Values and Improved Performance

Evolved

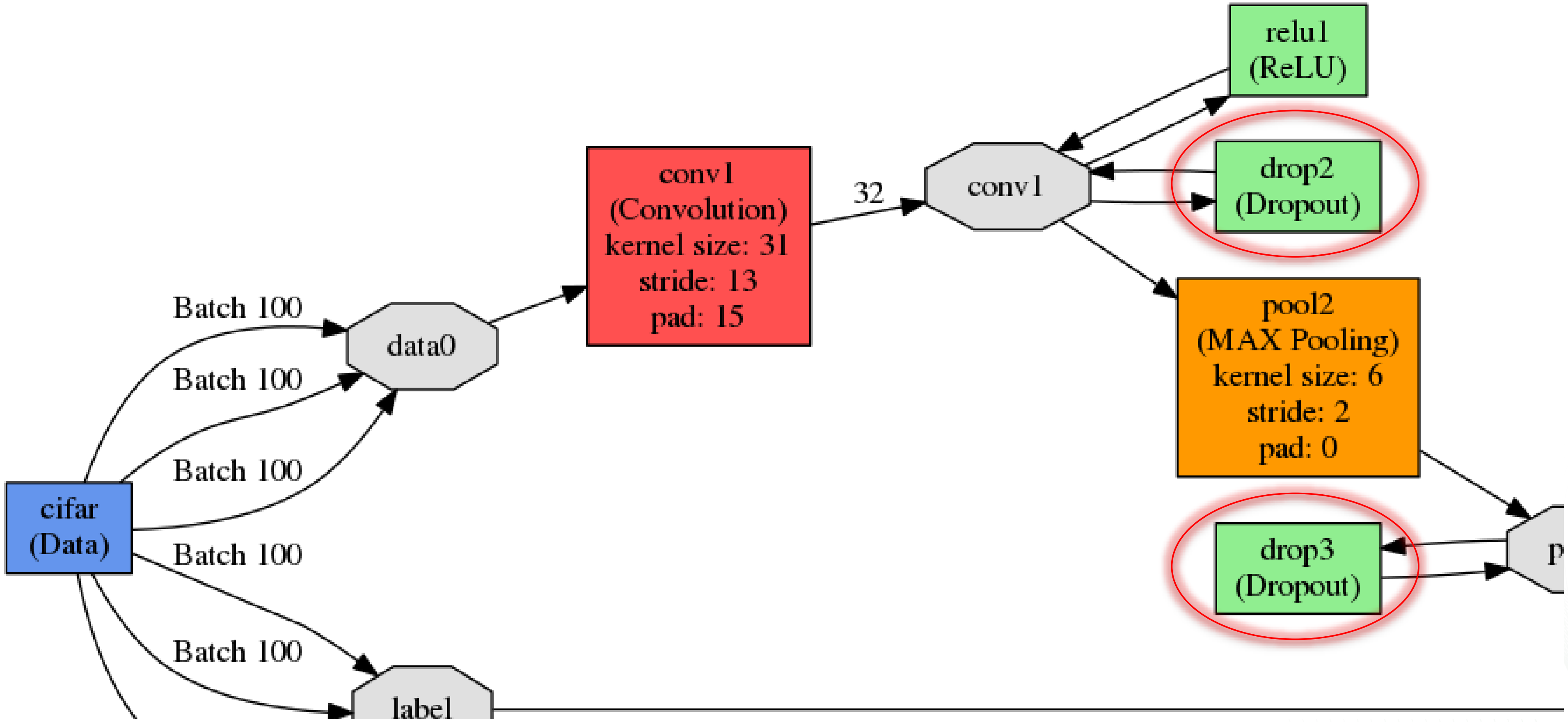


- Improved performance over known good network
- Using just 4 nodes
- **From 75% to 82%**



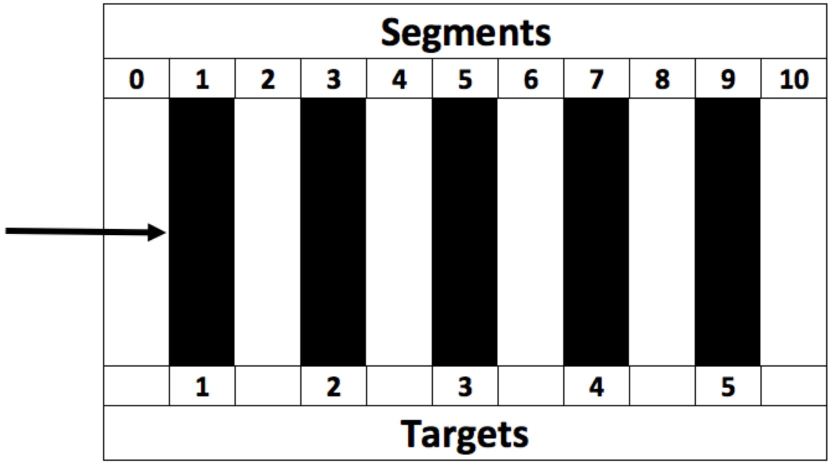
OAK RIDGE National Laboratory

Unusual Layers (limited training examples)



OAK RIDGE
National Laboratory

Application: MINERvA Detector Vertex Reconstruction



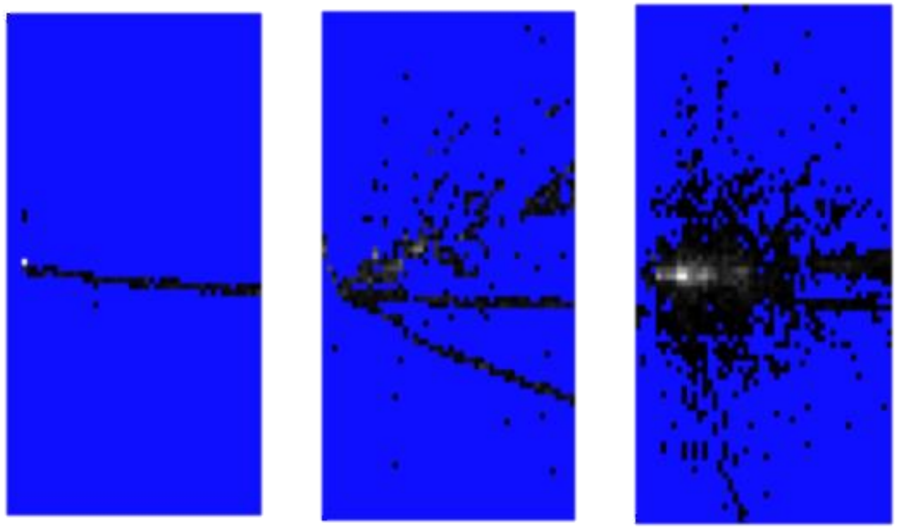


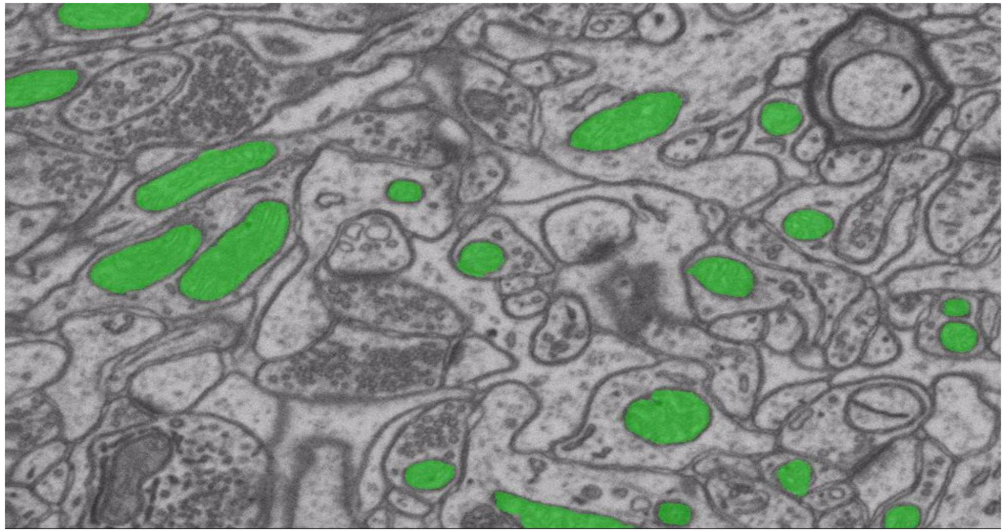
Table 1: Class distribution

Target	1	2	3	4	5
Distribution	12.9%	13.8%	11.4%	8.4%	10.8%
Segment	1	3	5	7	9

Segment	0	2	4	6	8	10
Distribution	2.4%	4.7%	4.8%	13.5%	1.2%	16.0%

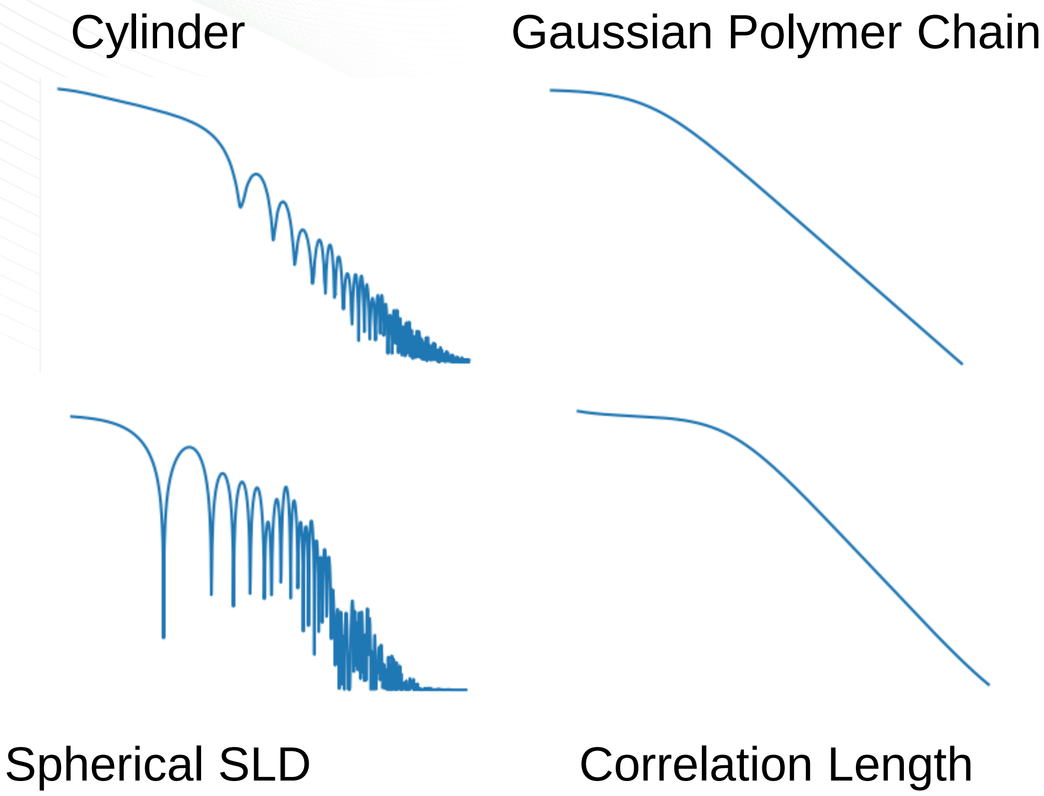
- Goal: Classify which segment the vertex is located in.
- Challenge: Events can have very different characteristics.
- Increased accuracy from 78% using seeded network to 82%

Application: 3D Electron Microscopy



- St. Jude Children's Research Hospital is interested in developing tools which will aid biologists in labeling and analyzing new image volumes for the location, density, shape, and other characteristics of sub-cellular structures such as mitochondria.
- Segmentation of 3D electron microscopy (EM) imagery is an important initial characterization task as mitochondria are relatively distinct but occur in a variety of locations, shapes, and sizes.
- MENNDL evaluated nearly 900k convolutional networks on +18k of Titan's nodes for 24 consecutive hours.
- Achieved a classification accuracy of 93.8%, representing a 30% reduction in error vs. a human expert defined network configuration.

Application: Small Angle Neutron Scattering Model Identification



- Goal is to identify model that corresponds to scattering data.
- Provide line plot of intensity vs. scattering length to convolutional neural network.
- Increased accuracy from 68.5% using a human designed network to 76.0%

MENNDL Current Status

- Scaled to 18,000 nodes of Titan
- 460,000 Networks evaluated in 24 hours
- Expanding to more complex topologies
- Evaluating on a wide range of science datasets
- Preparing for Summit (6 Volta GPUs per node, 4,600 nodes)

Questions