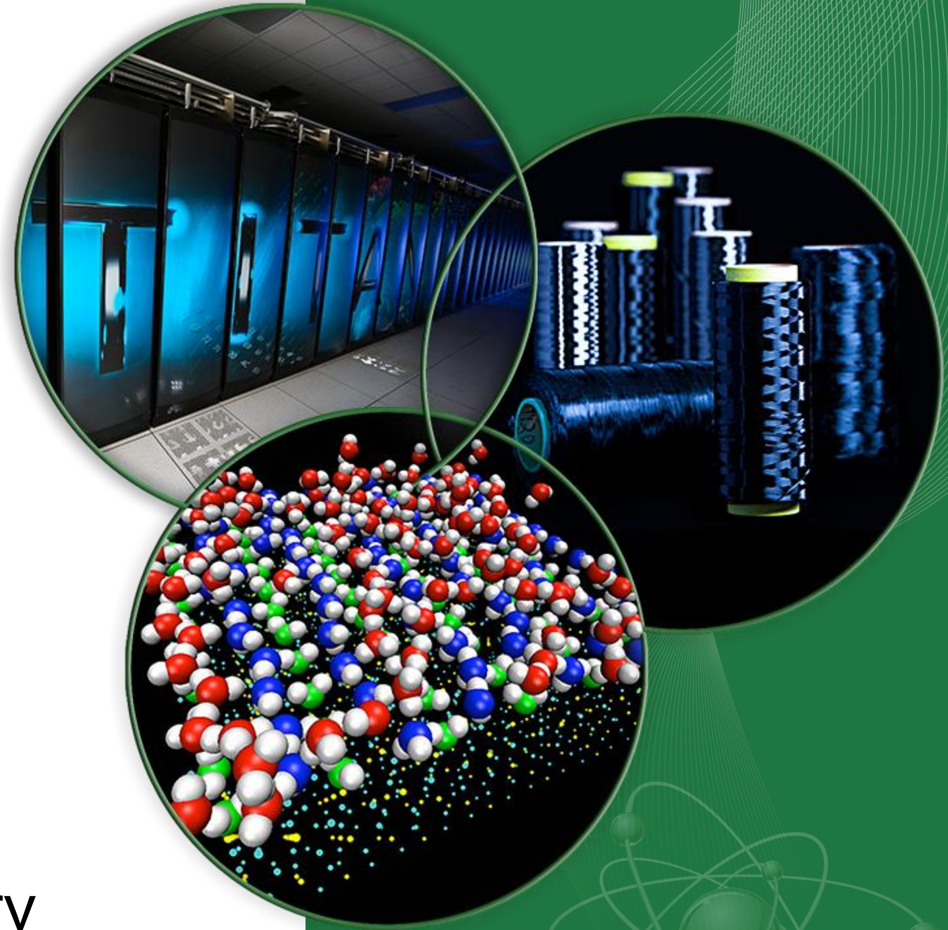
Deep Learning



Steven R. Young

Computational Data Analytics

Oak Ridge National Laboratory

ORNL is managed by UT-Battelle
for the US Department of Energy

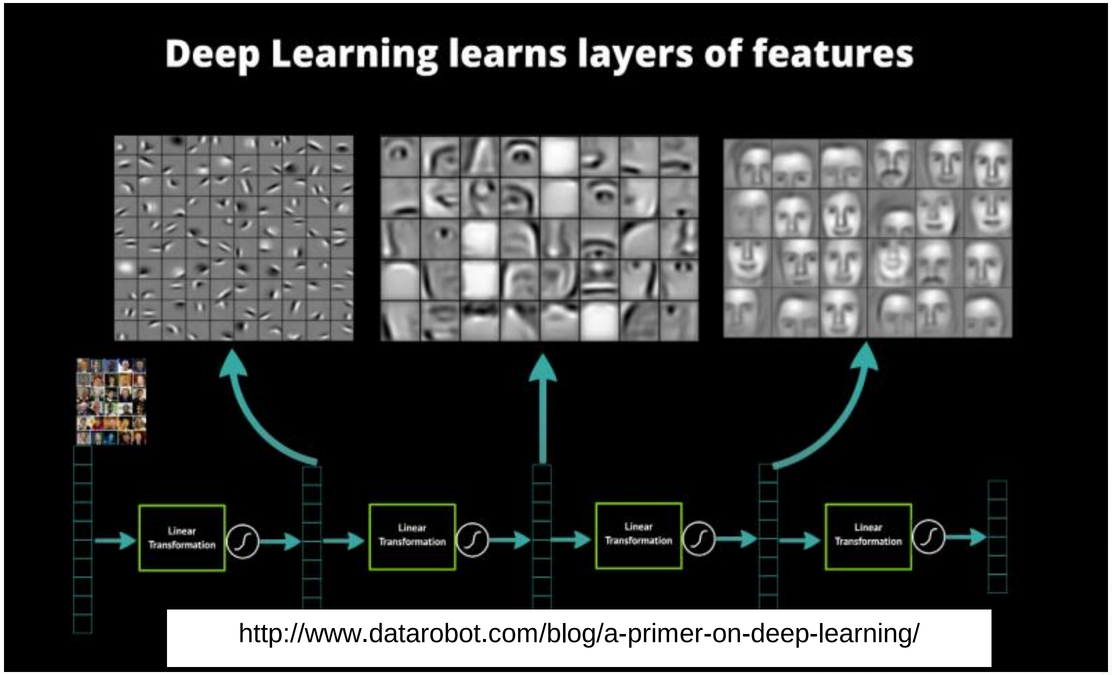
Oak Ridge National Laboratory

Outline

- Overview of Deep Learning
- Deep Learning and Neutrinos
- Network Design
- Deep Learning and Scientific Data

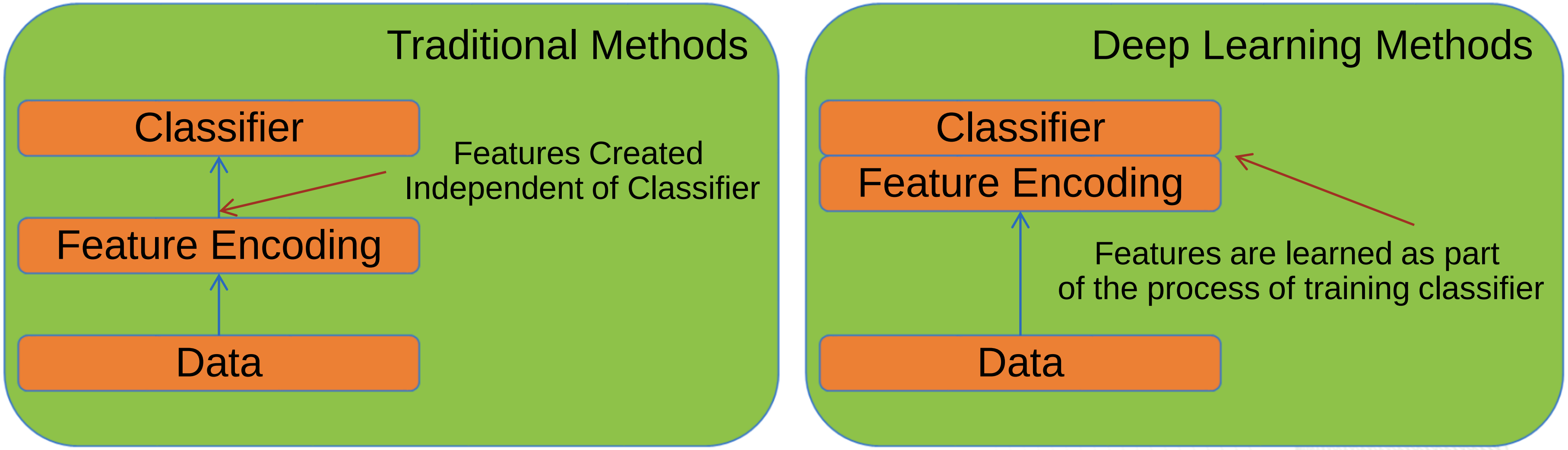
Deep Learning Shows Promise for Large Datasets

- Deep learning is data driven feature extraction supported by hierarchy of neuron layers
  - Lower layers learn local detail
  - Higher layers learn global concepts



What are the Goals of Deep Learning?

- Remove/reduce the need for domain level experts to determine what are important features of the data.
- The model learns what is important.
- The model works “directly” with the data.



Where is Deep Learning successful?

Challenging Problems

New State of the Art Results

Object Classification

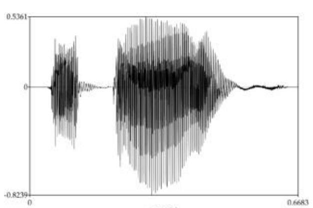
- 3.6% error on ImageNet competition
- **5x** decrease in error over results prior to first DL submission

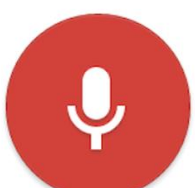
Face Recognition

http://vis-www.cs.umass.edu/lfw/

- 99.97% accuracy on LFW dataset.
- Only **14 errors out of 6000 pairs**.
- 5 were mislabeled in the dataset
- Human Level: 97.53%

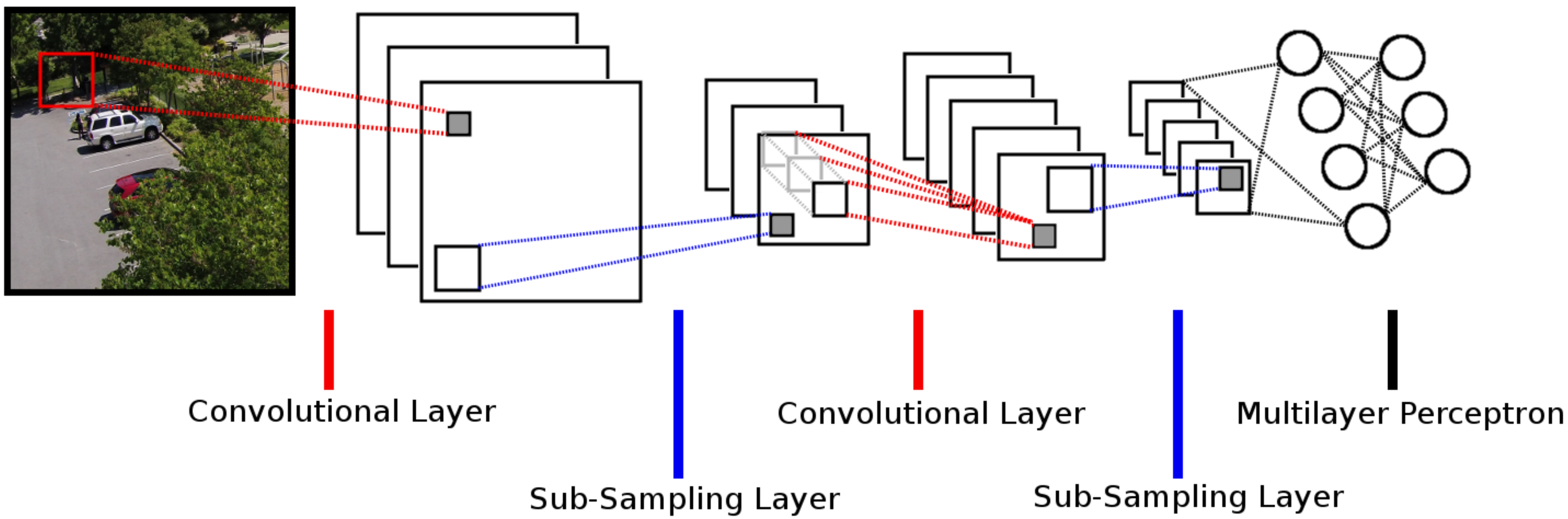
Speech Recognition





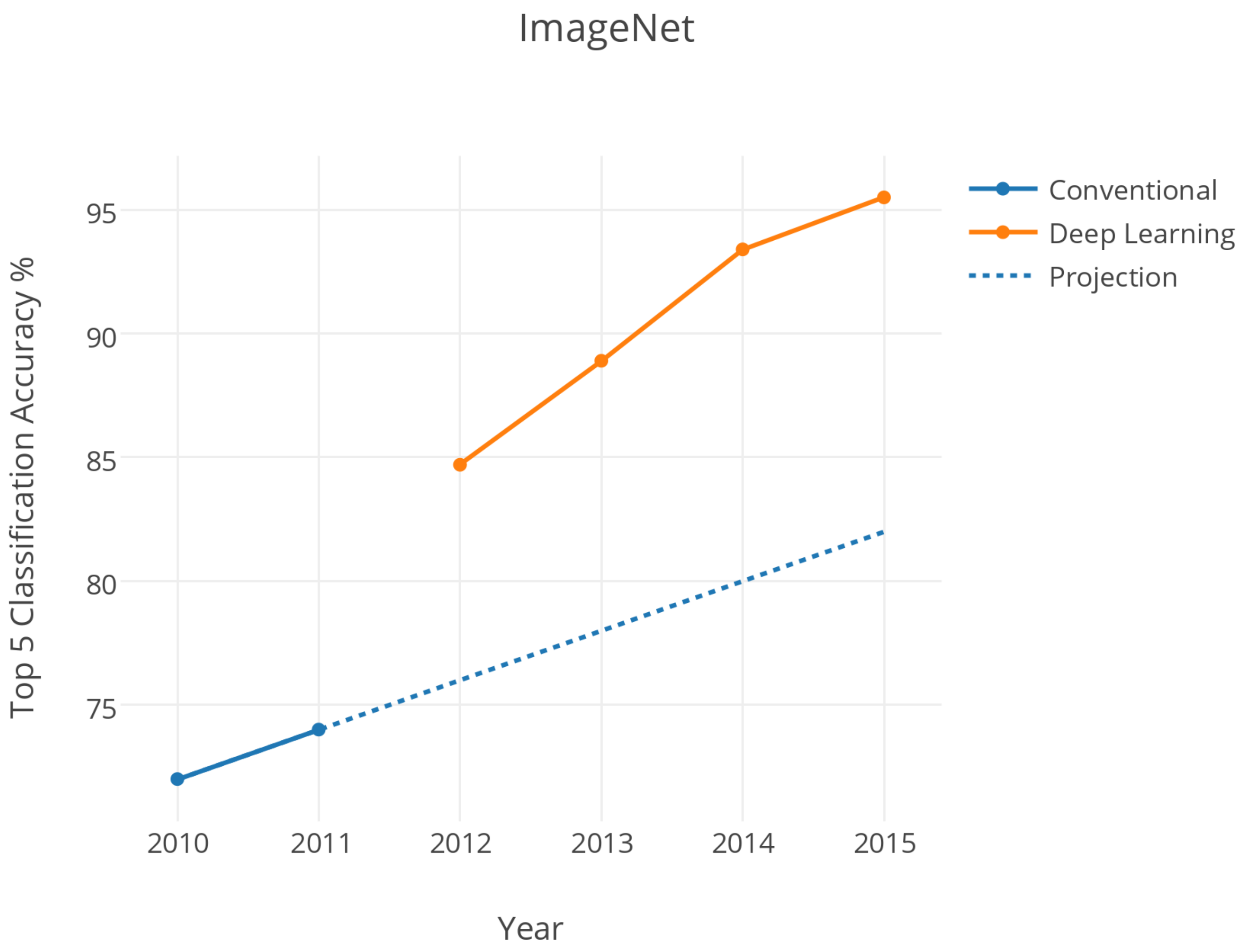
- Used in Google's production speech recognition software.
- Provides significant improvement on many standard benchmarks over previous methods.

Convolutional Neural Network (CNN)



- Input image is convolved with hidden units in the first convolutional layer
- The resulting feature maps is then sub-sampled using max pooling.
- Process is continued until the output of the final averaging layer is provided to a multilayer perceptron.

ImageNet Competition

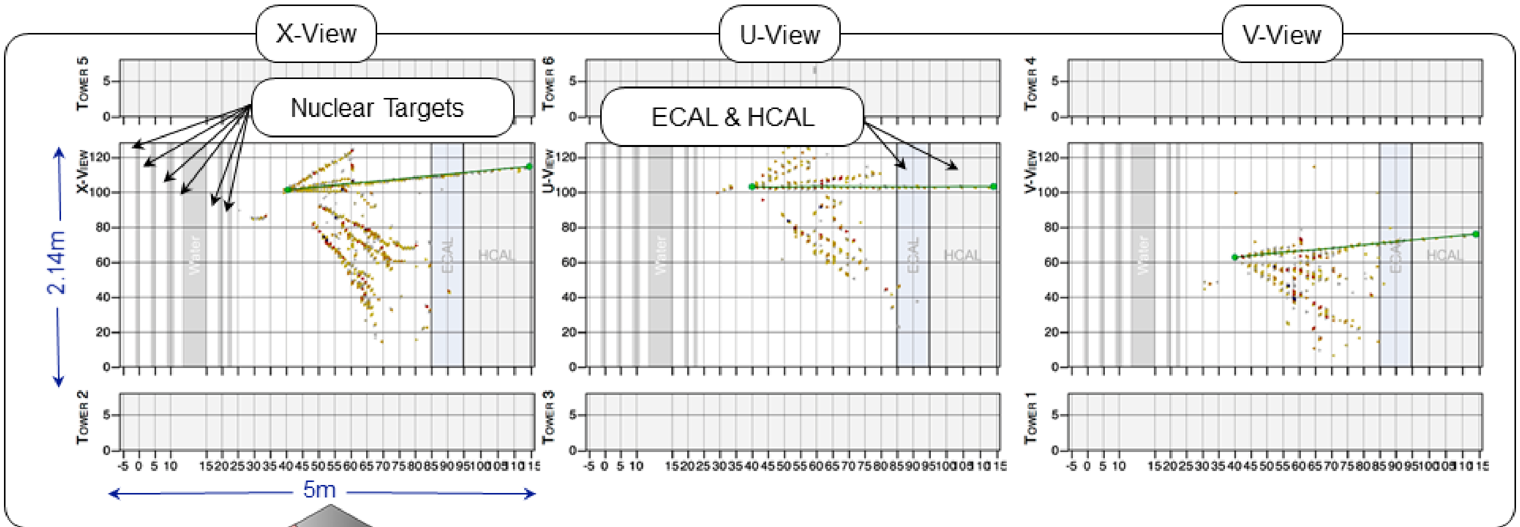


Neutrino Deep Learning Work

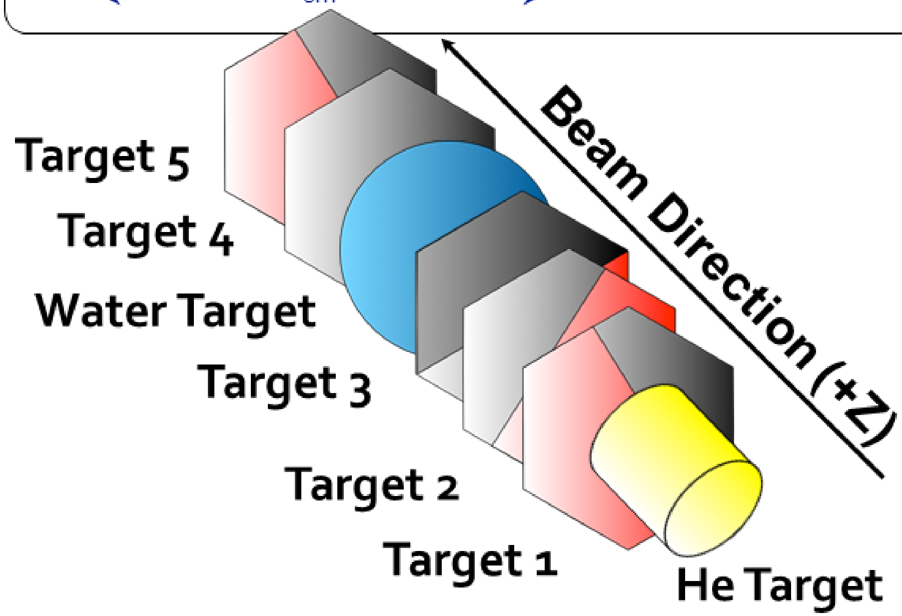
Neutrino Deep Learning Work

- At least two experiments investigating DL for a variety of tasks
- NOvA
  - Classifying event interaction type
    - ν_μ CC, ν_e CC, ν_τ CC, ν NC
  - https://arxiv.org/abs/1604.01444
- MINERvA
  - Vertex reconstruction

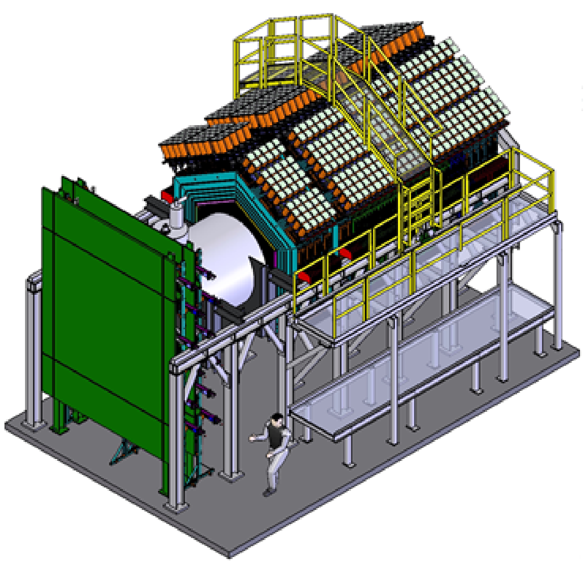
MINERvA



X-View
U-View
V-View
Nuclear Targets
ECAL & HCAL
2.14m
5m
Water
ECAL
HCAL

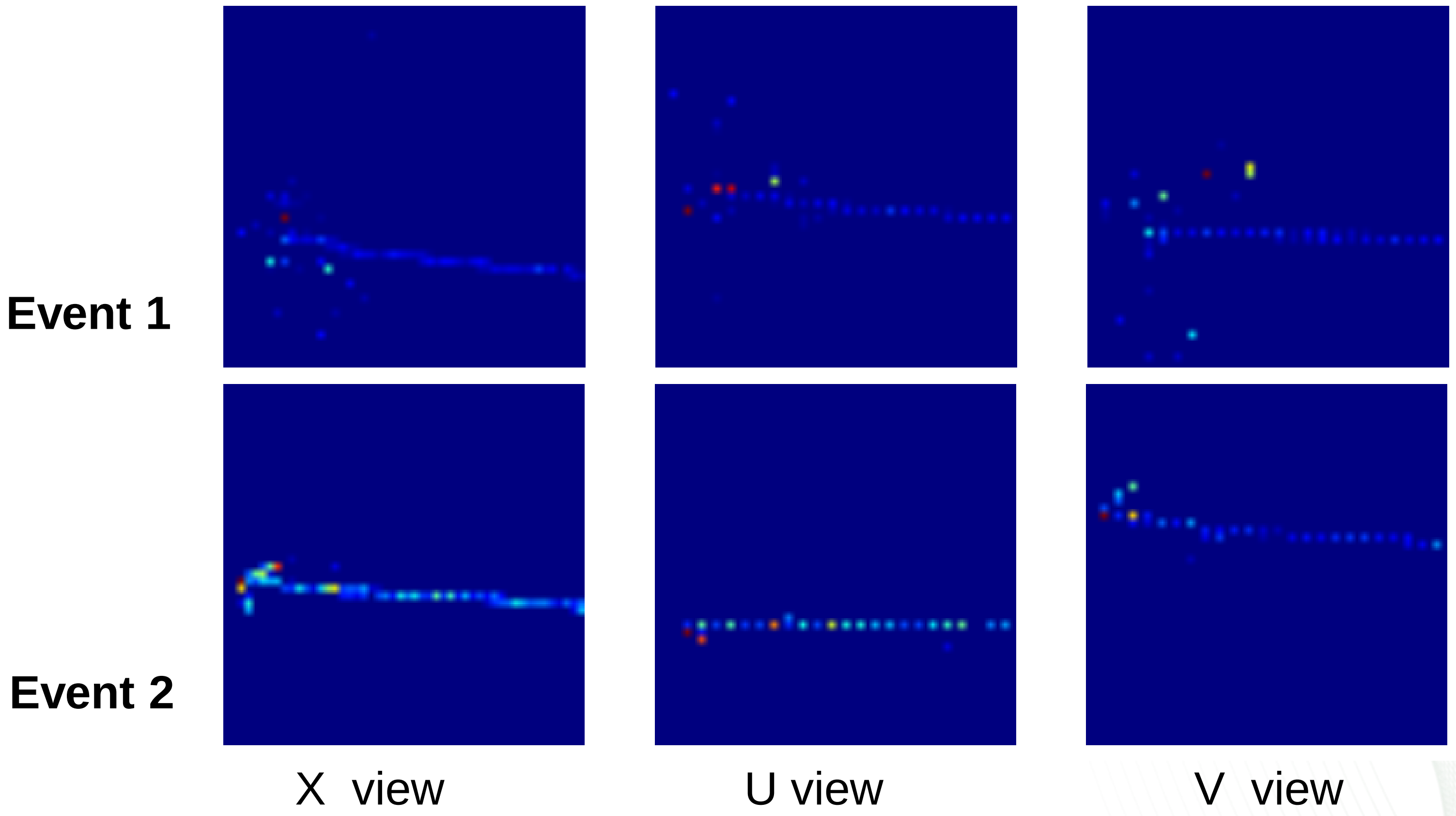


Target 5
Target 4
Water Target
Target 3
Target 2
Target 1
He Target
Beam Direction (+Z)



MINERvA Vertex Reconstruction

- Data: Simulation data. Energy lattice provided for each event.
- Goal: Find location of neutrino interaction.



MINERvA Vertex Segment Classification

Goal: Classify which segment the vertex is located in.

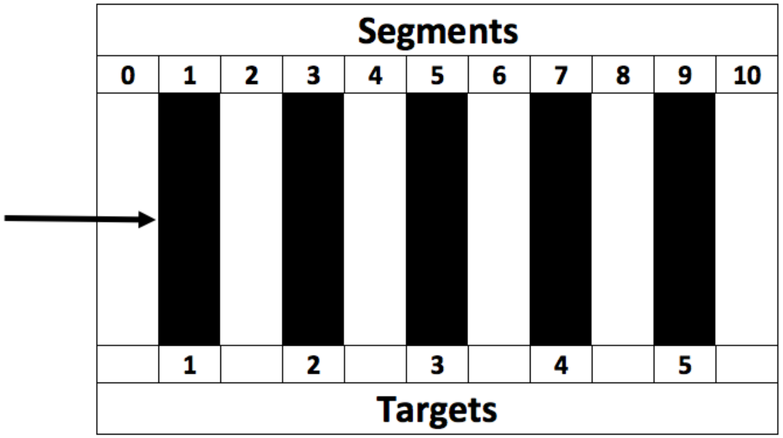
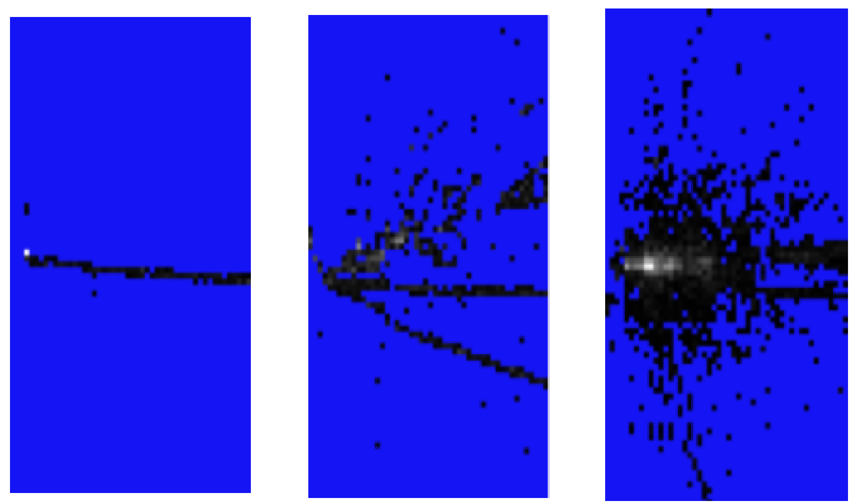


Table 1: Class distribution

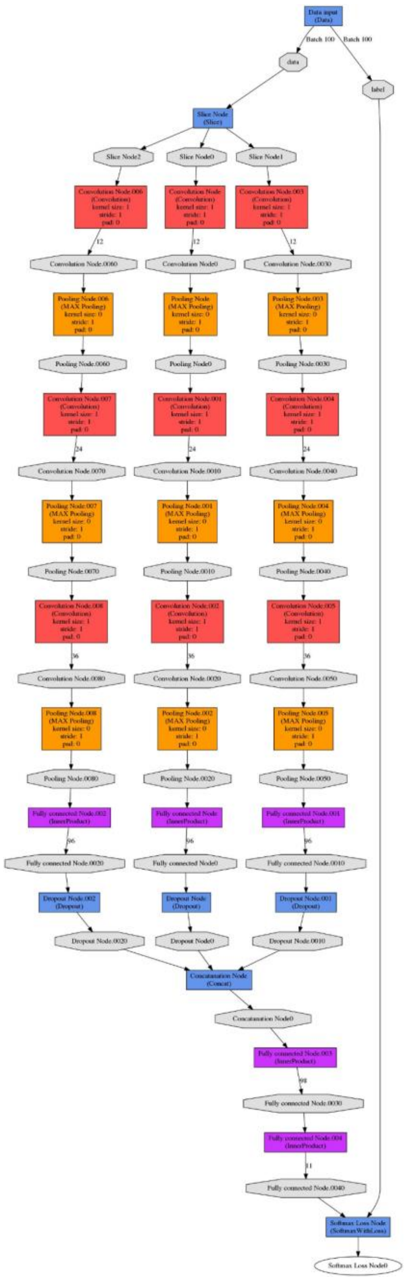
Target	1	2	3	4	5
Distribution	12.9%	13.8%	11.4%	8.4%	10.8%
Segment	1	3	5	7	9

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



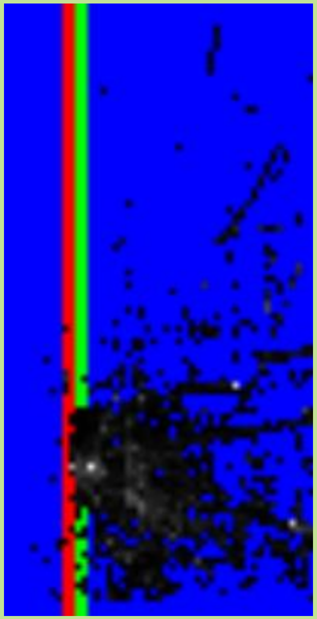
Challenge: Events can have very different characteristics.

How to combine data from 3 views?

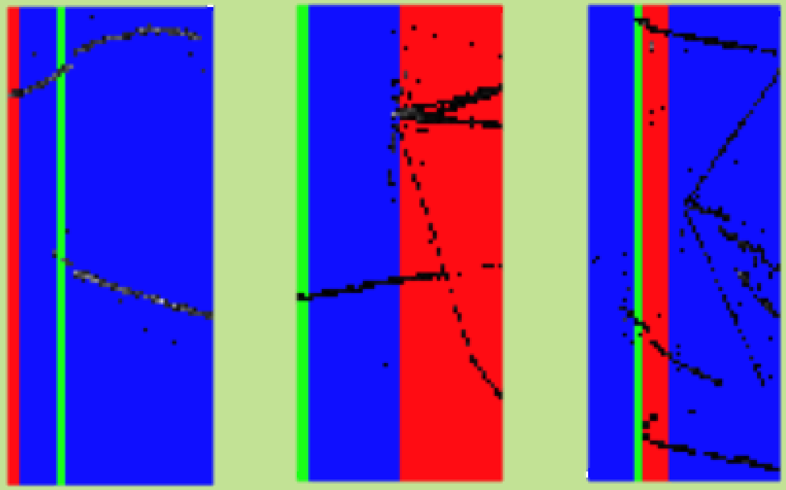


Model	Test Set Accuracy
Previous Methods	91.90%
CNN – 1 View	80.42%
CNN – 3 Views + 1 Column	88.71%
CNN – 3 Views + 3 Columns	**93.58%**

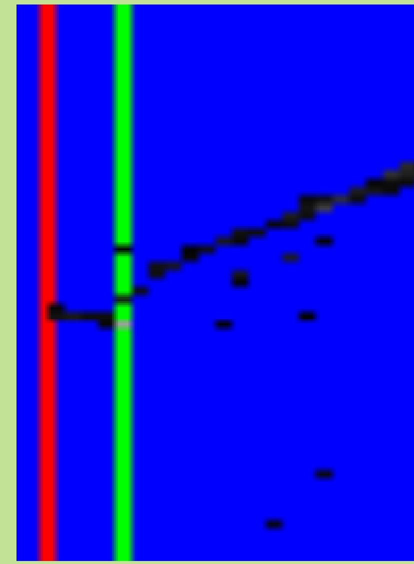
Misclassified Events



Boundary



Multiple Interactions



Backward Tracks

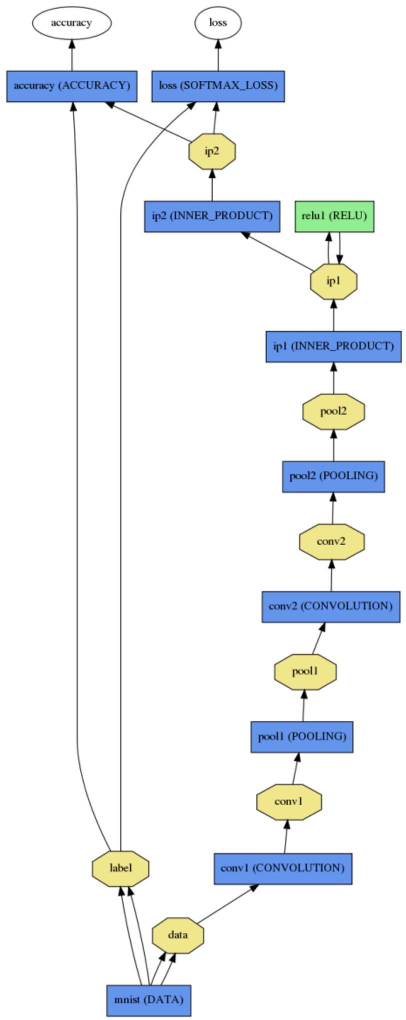
Why that network design/hyper-parameters?

# convLayers	Kernel Sizes ($\{h\}\times w$)	Accuracy
Three	$\{6, 6, 3\} \times 3$	93.58%
Four	$\{\mathbf{8}, \mathbf{8}, \mathbf{7}, \mathbf{6}\} \times \mathbf{3}$	**94.09%**
Five	$\{8, 7, 7, 3, 3\} \times 3$	93.55%

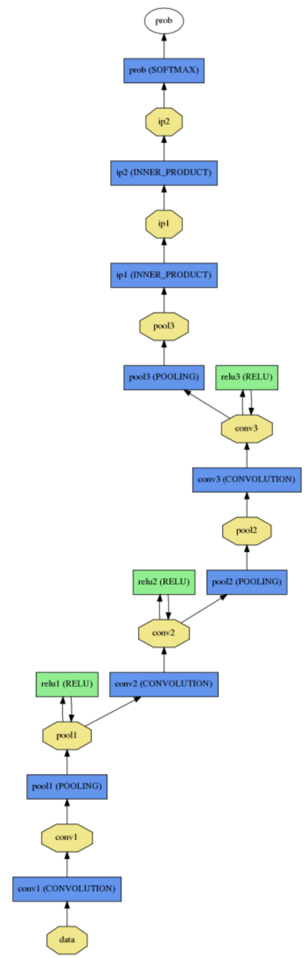
- Is that the best accuracy possible?
  - Better network hyper-parameter choices?
- Other problems to solve?
  - Different networks for those?
- Leverage ORNL's Titan supercomputer to improve performance & expand to other problems

Deep Learning Network Design

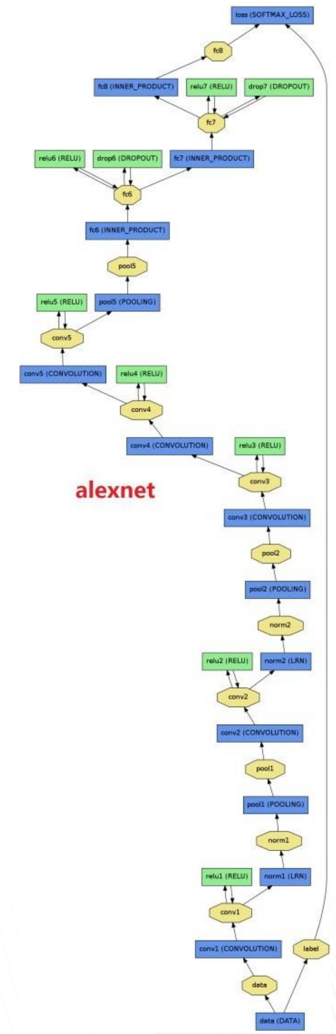
Network Design for Different Datasets



MNIST



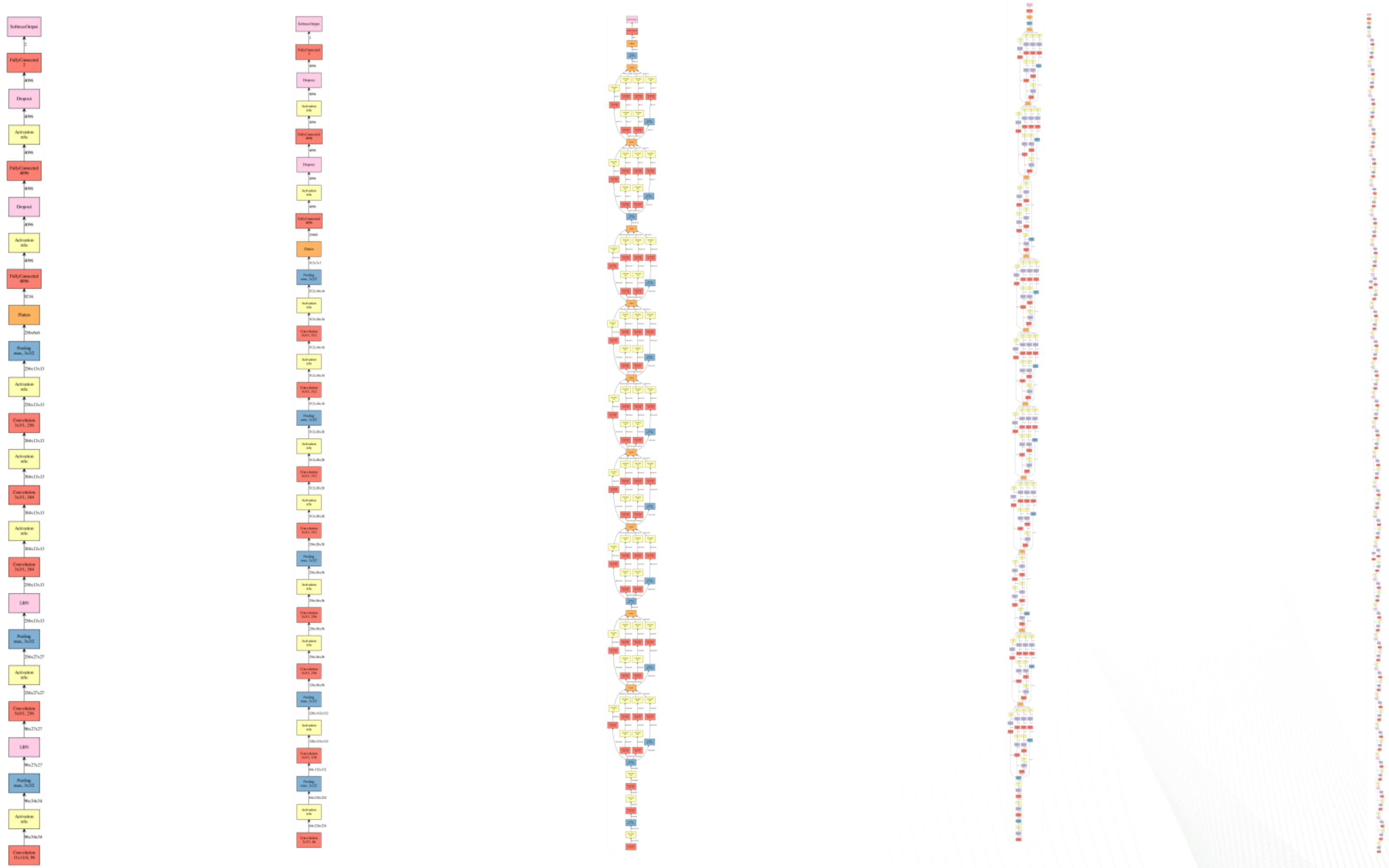
CIFAR-10



ImageNet

OAK RIDGE National Laboratory

Progression of ImageNet Network Design

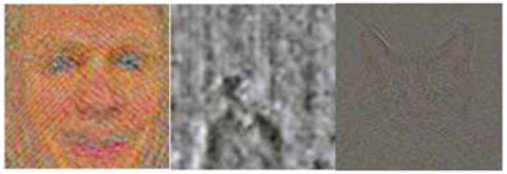


AlexNet VGG GoogLeNet InceptionNet Resnet

Not quite there, so what's needed?

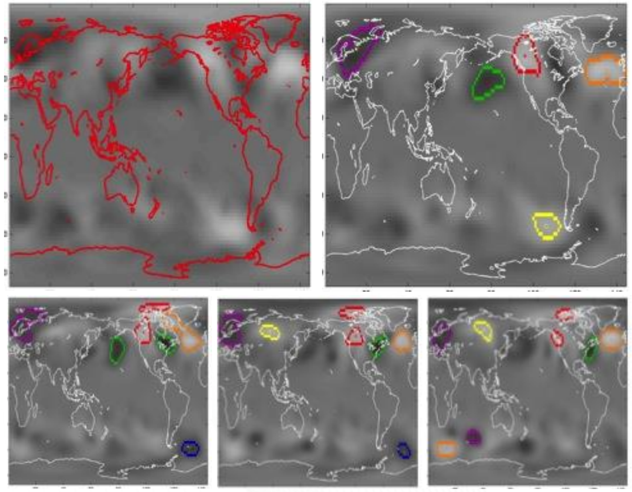
- Current research involving toy problems / data sets; Real applications driven by commercial interests
- Domain expertise and computational training costs limit adaptability to new data sets

Improve Adaptability of Deep Learning



(a) Face (b) Body (c) Cat

*Reference: A. Coates, B. Huval, T. Wang, D. J. Wu, and A. Y. Ng. "Deep learning with COTS HPC systems." In International Conference on Machine Learning, 2013.

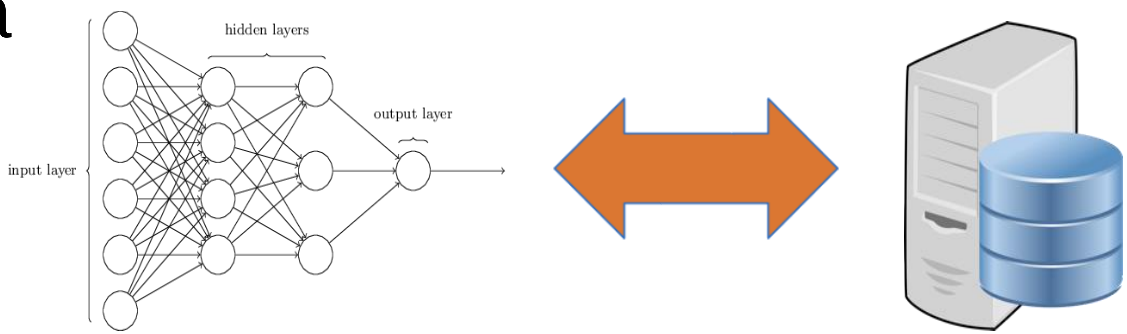


From simple & small data sets…

…to more complex & bigger data sets

Problem: Adaptability Challenge

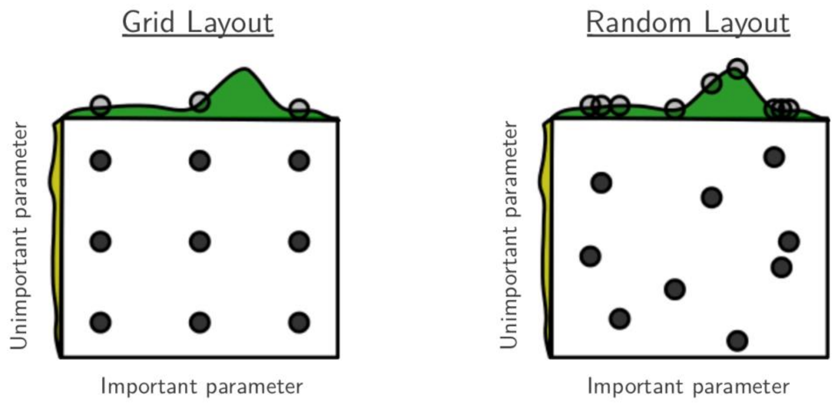
- ***Premise:*** For every data set, there exists a corresponding neural network that performs ideally with that data



- What's the ideal neural network architecture (i.e., hyper-parameters) for a particular data set ?
- Current approach: educated guessing
  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

Hyper-parameter Selection



- Manual search, guess and check
  - Requires domain knowledge
- Grid search
  - Exponential growth with high-dimensional hyper-parameter space
  - Doesn’t exploit low effective dimension for discovery
- Random search
  - By itself, not adaptive (no use of prior information)

OAK RIDGE National Laboratory

MENNDL: Multi-node Evolutionary Neural Networks for Deep Learning

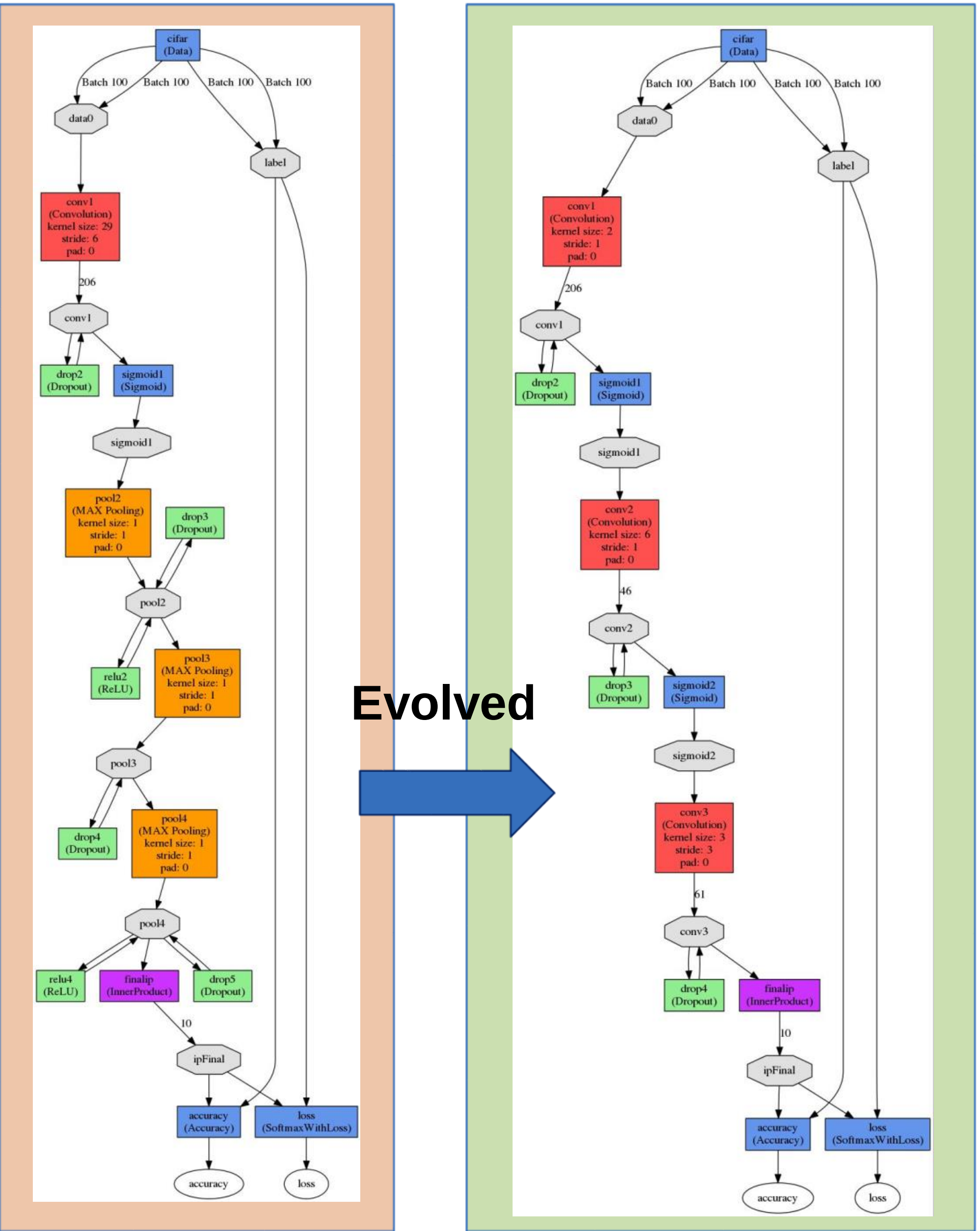
- Evolutionary algorithm as a solution for searching hyper-parameter space for deep learning
  - Focus on Convolutional Neural Networks
  - Evolve *only* the topology with EA; typical training process
  - Generally: Provide *scalability* and *adaptability* for many data sets and compute platforms
- Leverage more GPUs; ORNL's Titan has 18k GPUs
  - Next generation, Summit, will have more
- Provide the ability to analyze hierarchical patterns from large data sets
  - Often high dimensional, thousands of variables
  - Climate science, material science, physics, etc.

Proof of Concept Using CIFAR-10 Data

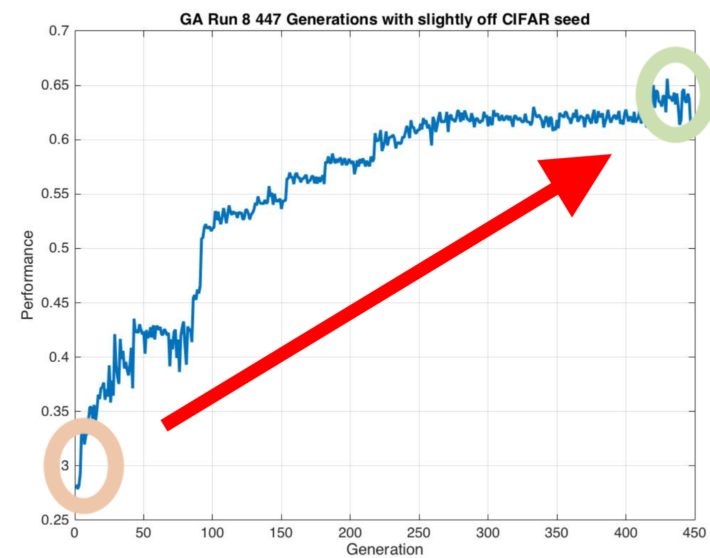
- CIFAR-10 data: Images of 10 classes of objects
- Using MENNDL, can we evolve the topology of a poorly performing CNN to perform well on CIFAR-10?

OAK RIDGE National Laboratory

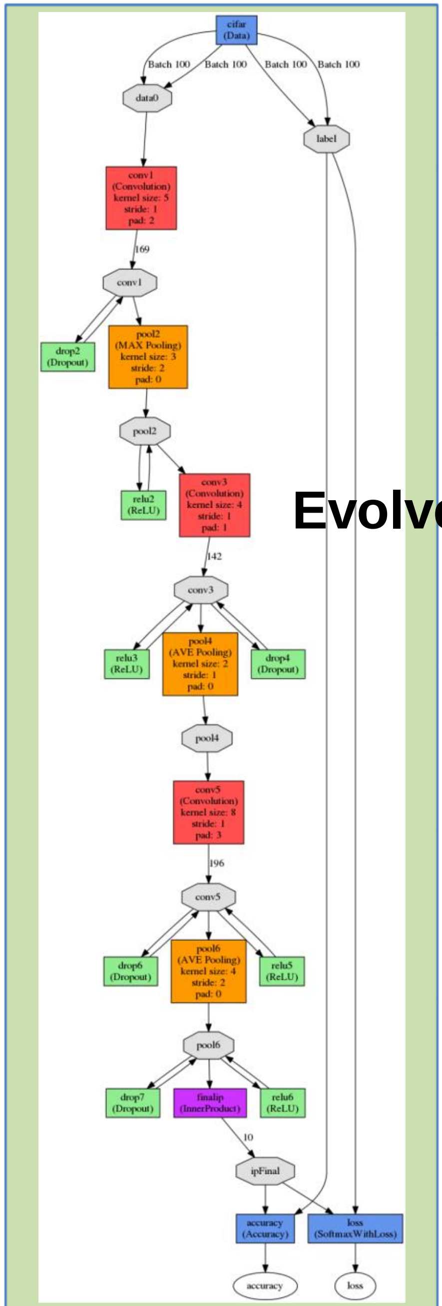
Hyper-parameter Values vs Performance



- 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**

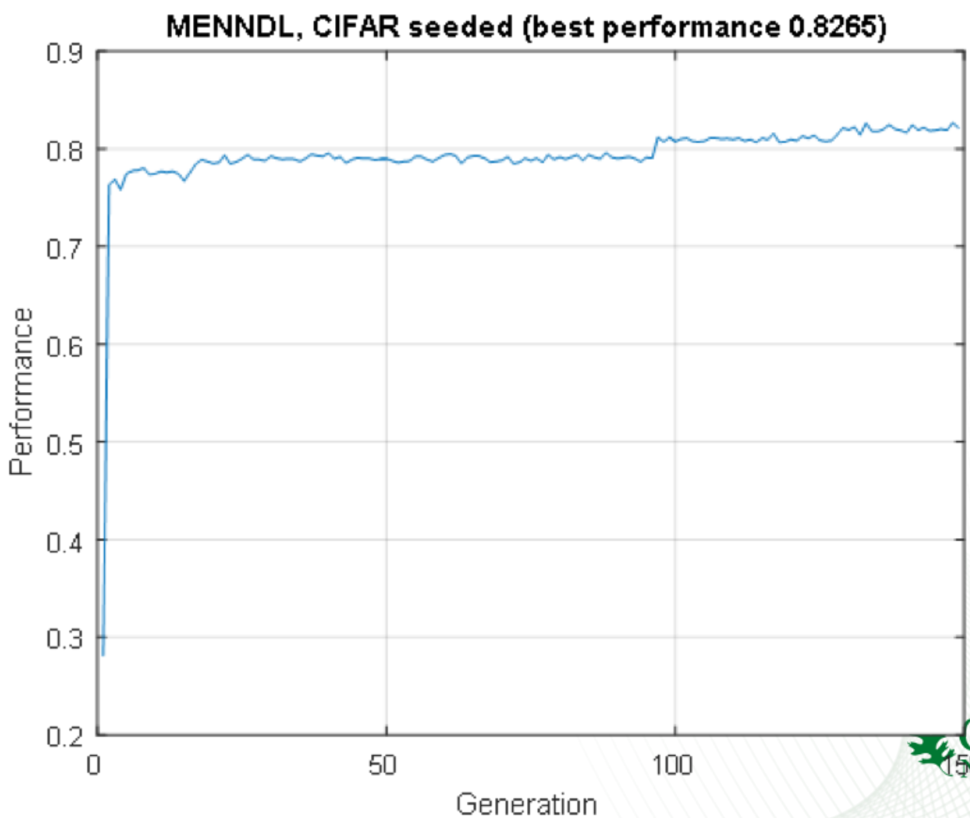


Hyper-parameter Values vs Performance



Evolved

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



OAK RIDGE
National Laboratory

Deep Learning and Scientific Data

Applying DL for Scientific Use

- Challenges
  – Scientists don't want a black box, they want to the system to explain how the system arrived at a result.
    - Good news! It's not a black box.
    - Bad news! It's not clear how to break down thousands/millions/billions of parameters/calculations into an easily conveyed explanation of a result.
  – Scientists want better ways to quantify the uncertainty of results.
    - Small changes can cause dramatic changes in results.
    - How to measure similarity of training data to testing data?
- Opportunity
  – Many science fields have high quality simulations that can be leveraged for training data.

Simulation Data for Training

Acknowledgements

- Gabriel Perdue (FNAL) and the entire MINERvA collaboration
- Robert Patton (ORNL) and the “Scalable Deep Learning Algorithms for Exascale Data Analytics” LDRD team
- Adam Terwilliger (Grand Valley State University)
- David Isele (University of Pennsylvania)

Backup Slides

Per Segment Accuracy

Target	Segment	Previous	DL
-	0	*78.9%*	*78.1%*
1	1	92.2%	**96.4%**
-	2	*88.4%*	*88.5%*
2	3	91.5%	**96.4%**
-	4	*88.6%*	*89.2%*
3	5	91.2%	**95.4%**
-	6	*95.1%*	*95.1%*
4	7	89.1%	**93.4%**
-	8	**73.7%**	61.3%
5	9	88.8%	**94.9%**
-	10	**98.0%**	96.8%