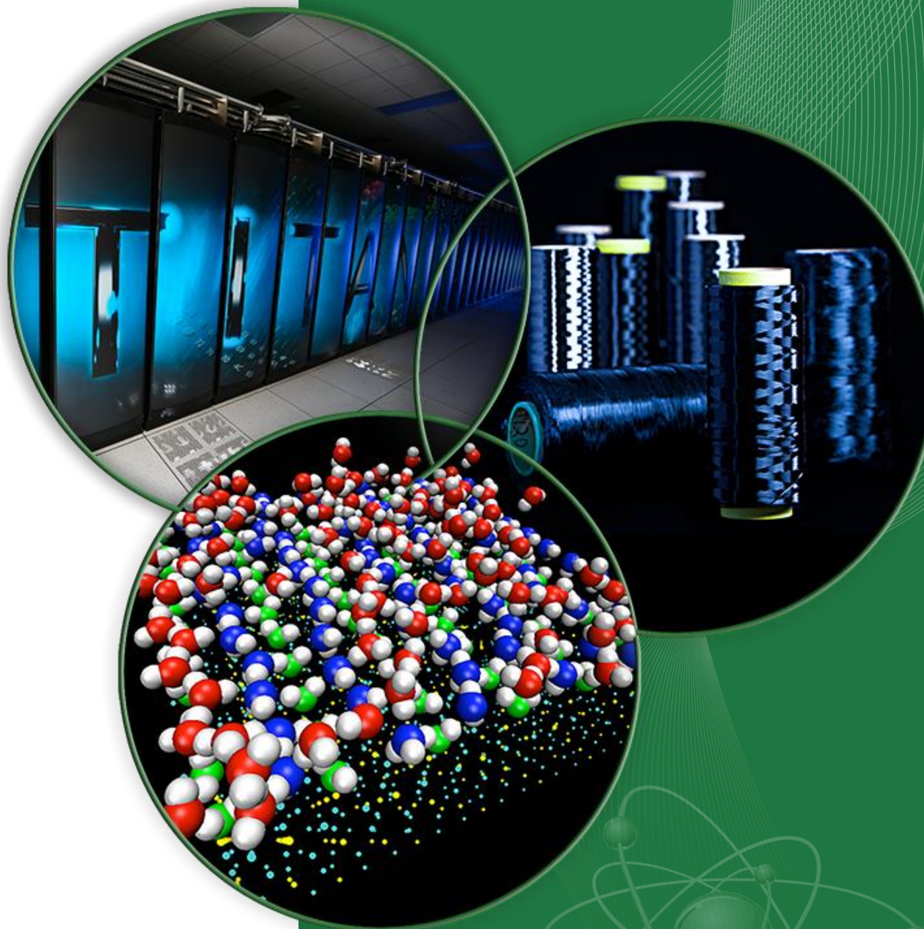
Optimizing Deep Learning Hyper-Parameters Through an Evolutionary Algorithm



Steven R. Young, Ph.D.

Computational Data Analytics

Oak Ridge National Laboratory

Where is Deep Learning successful?

Challenging Problems

Object Classification

Face Recognition

Speech Recognition

New State of the Art

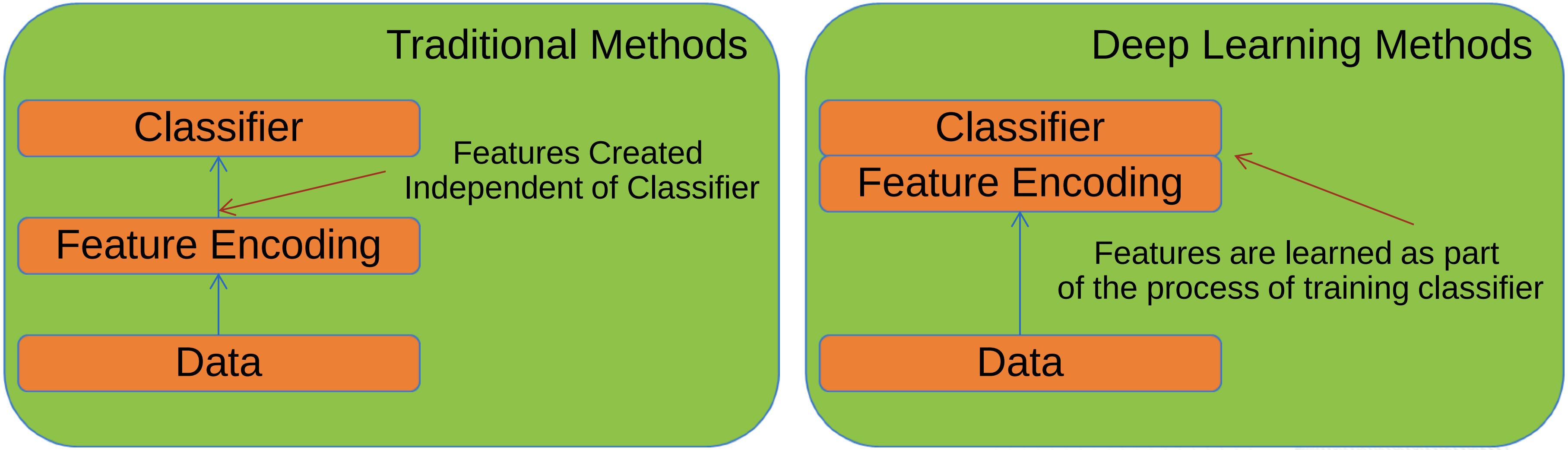
- 5% error on ImageNet competition
- **5x** decrease in error over results prior to first DL submission

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

- Used in production speech recognition software for major internet companies.
- Provides significant improvement on many standard benchmarks over previous methods.

What are the Goals of Deep Learning?

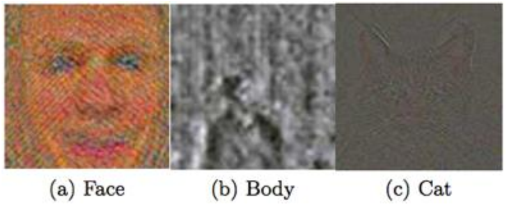
- Remove/reduce the need for domain level experts to determine what are important features of the data.
- The model learns what is important.
- The model works directly with the data.



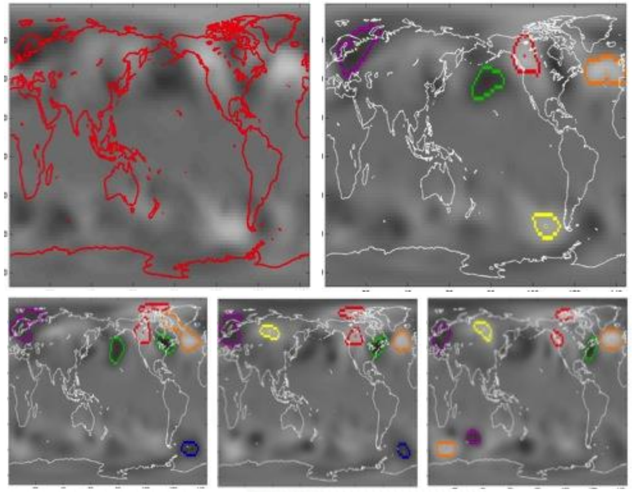
What's Needed?

- 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



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



From datasets of commercial interest…

…to datasets of scientific interest.

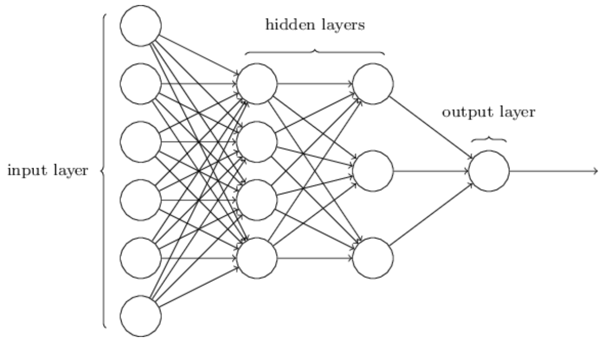
Deep Learning Challenges

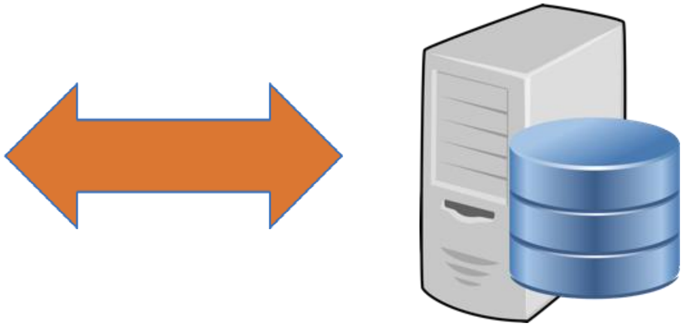
- Scale
  - How big should the network be?
  - How to leverage distributed computing / training?
- Computation time
  - How to cut training time / costs?
  - How to improve training performance?
- Network Specification
  - How to apply DL to new, unknown data sets?
  - What hyper-parameters to use for a new data set?

Addressing these challenges will empower scientists to tackle new problems using DL methods.

Network Specification Challenge

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



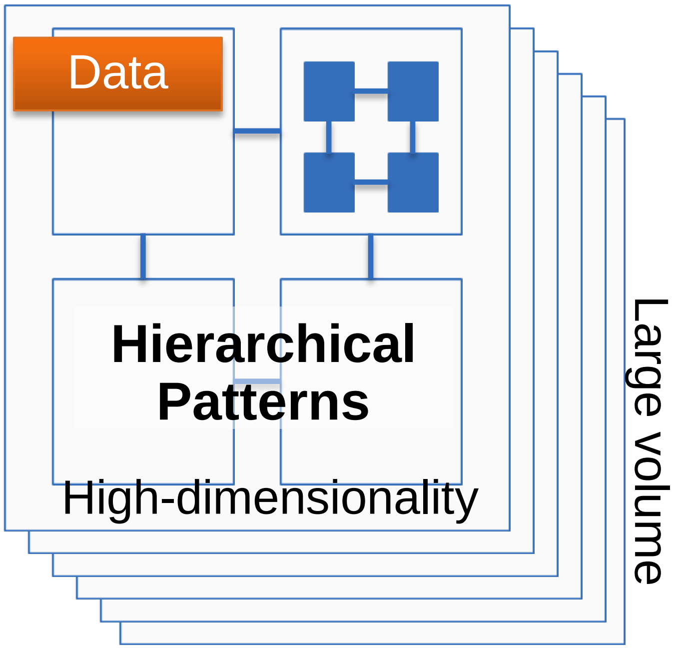


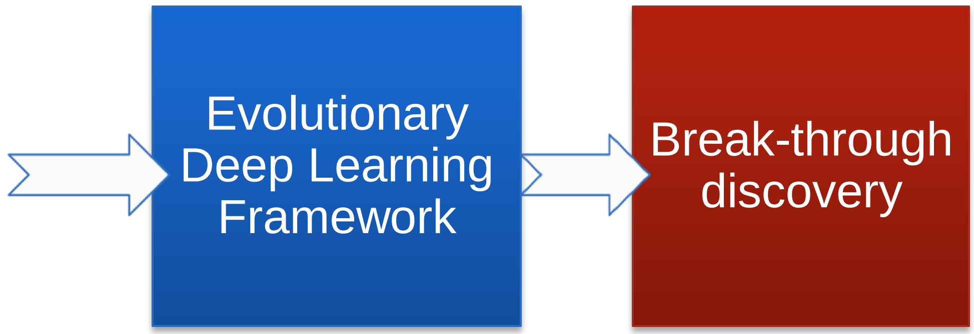
- What's the ideal neural network architecture (i.e., hyper-parameters) for a particular data set ?
- Widely-used approach: intuition + random search

Oak Ridge National Laboratory

MENNDL: Multi-node Evolutionary Neural Networks for Deep Learning

Provide a capability to analyze hierarchical patterns from extremely large data sets consisting of thousands of variables



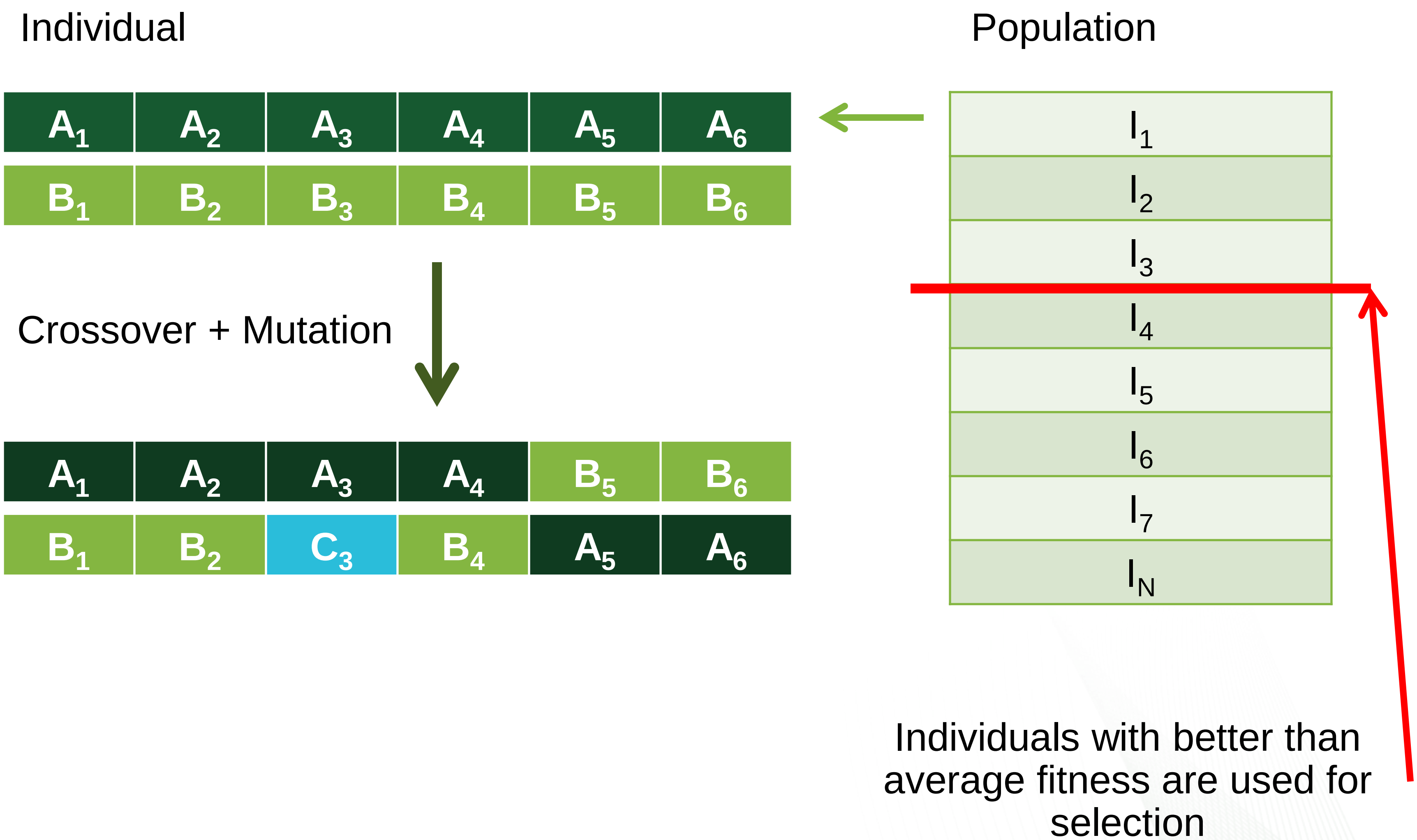


- Evolve *only* the topology with EA; typical training process for learning weights
- Convolutional Neural Networks evolved by Evolutionary Algorithms (essentially a memetic algorithm) to provide *scalability* and *adaptability* to both data sets and compute platforms
- Leverage more GPUs; ORNL's Titan has 18k GPUs; next generation machine will have more
- Apply to extremely large scientific data sets in areas such as climate science, material science, etc.

MENNDL Code (C / C++, CUDA)

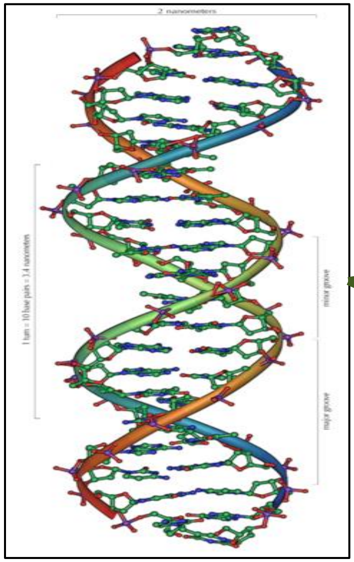
- ORNL developed Evolutionary Algorithm based on C++ code previously developed on ORNL's Jaguar; tested to 5,000 nodes on Jaguar (27% of Jaguar)
  - MPI is used to distribute population fitness evaluation across nodes
- Convolutional Neural Network leveraging open source deep learning library, Caffe
  - Written in C++; GPU enabled using CUDA
  - Developed and under active development at UC Berkeley
  - Supported by NVIDIA's cuDNN library
  - Network topology defined in plain text file format

MENNDL Evolutionary Algorithm



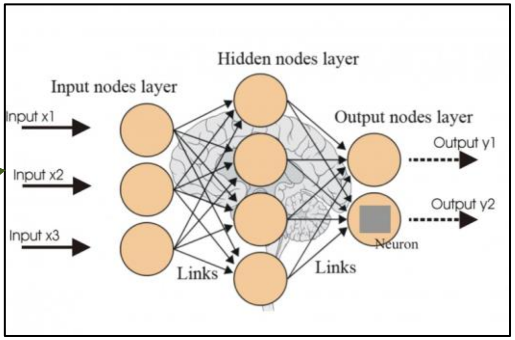
MENNDL: Fitness Evaluation

Node 1



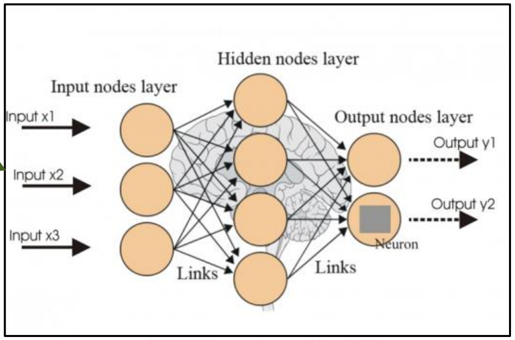
MPI

Node 2



…

Node N



Selection, Mutation, and Crossover are handled by a single node

Fitness of the population is evaluated on the remainder of available nodes

MENNDL Workflow

MENNDL

Visualize parameters from all networks

EDEN
Exploratory Data analysis ENvironment

Gain Insight

Identify best networks for new, challenging problems

Design new experiments

Proof of Concept using CIFAR-10 data

- CIFAR-10 data: Images of 10 classes of objects
- Using MENNDL, can we evolve 6 hyper-parameters of a known CNN that performs well on CIFAR-10?

CIFAR Deep Learning Network

airplane … truck

Inner Product Layer (10 outputs)

Inner Product Layer (64 outputs)

Pool3 Layer (Avg pooling, K=3)

Conv3 Layer (64 outputs, K=5) (ReLU)

Pool2 Layer (Avg pooling, K=3)

Conv2 Layer (32 outputs, K=5) (ReLU)

Pool1 Layer (Max pooling, K=3)

Conv1 Layer (32 outputs, K=5)

CIFAR Training Data

Hyper-parameter Values vs Performance

- As performance of the CNN improves, hyper-parameter values start trending toward specific values
- 500 nodes of Titan for 24 hours, evolving for 70 generations and creating 35,000 hyper-parameter sets
- Clearly defined trends in performance versus kernel size, number of outputs, and runtime

CIFAR

airplane … truck

Inner Product Layer (10 outputs)

Inner Product Layer (64 outputs)

Pool3 Layer (Avg pooling, K=3)

Conv3 Layer (64 outputs, K=5) (ReLU)

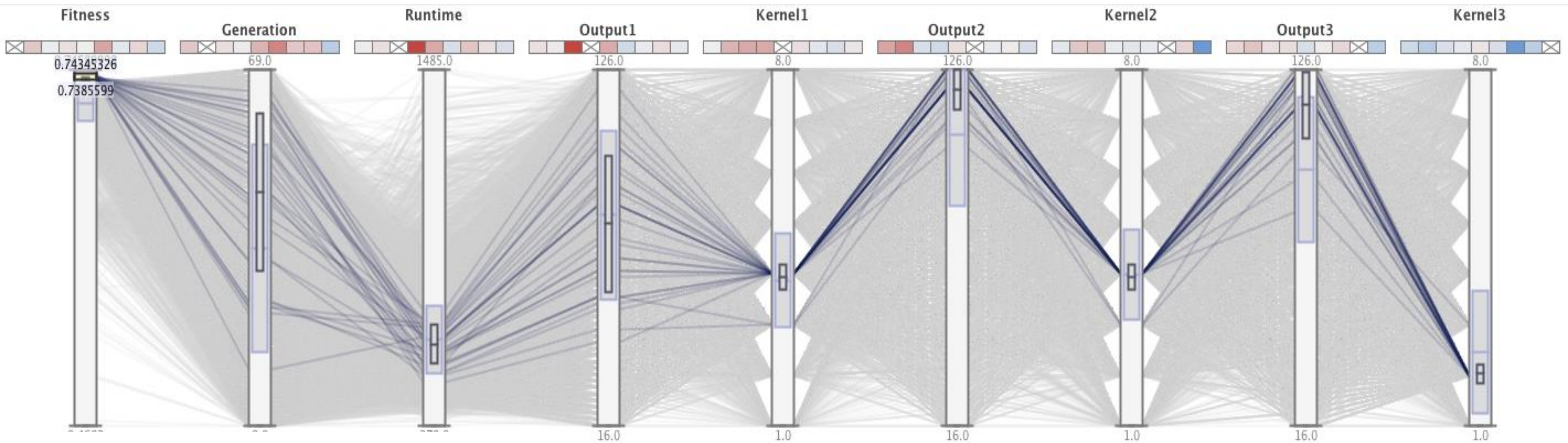
Pool2 Layer (Avg pooling, K=3)

Conv2 Layer (32 outputs, K=5) (ReLU)

Pool1 Layer (Max pooling, K=3)

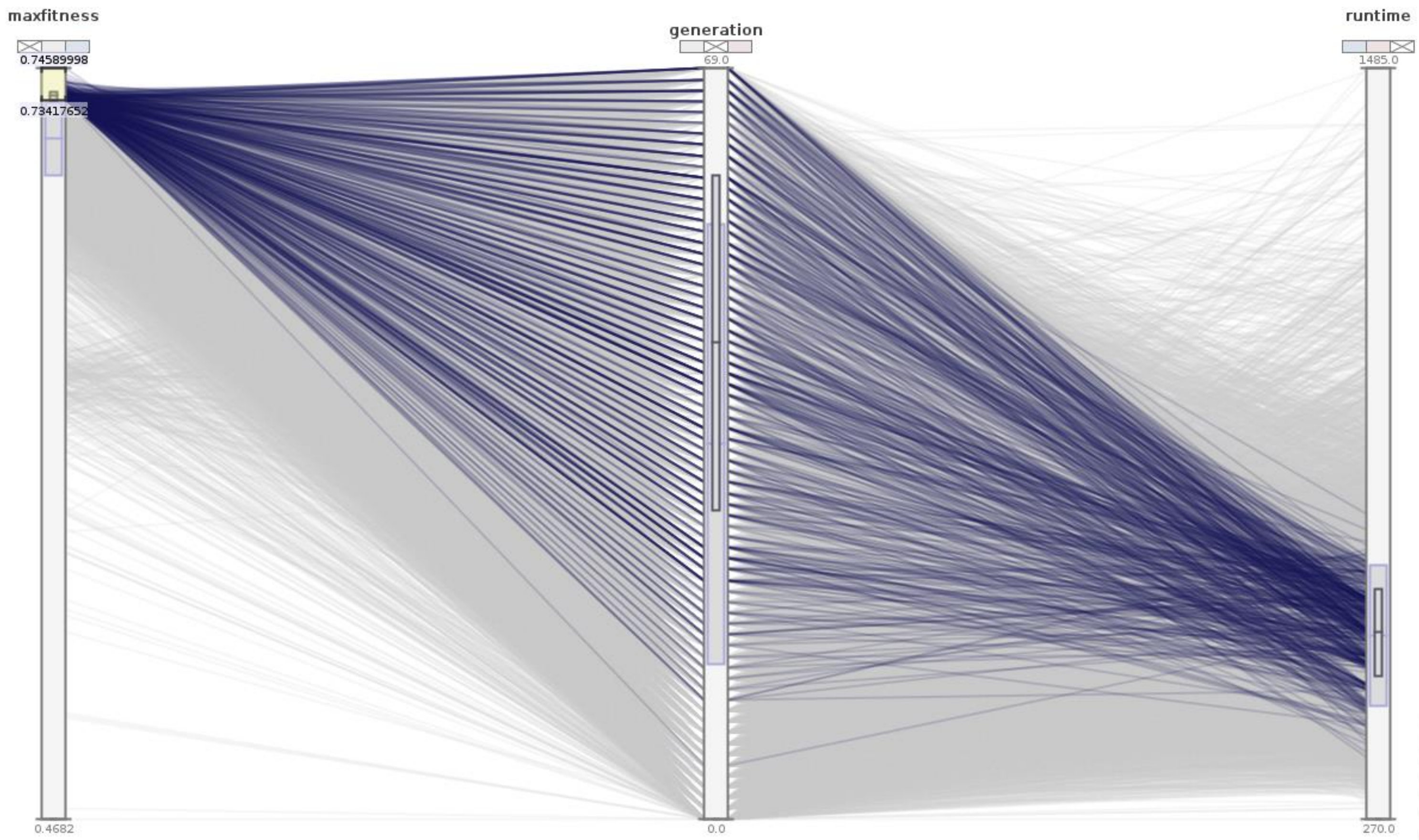
Conv1 Layer (32 outputs, K=5)

CIFAR Training Data



EDEN: http://cda.ornl.gov/projects/eden/

Improved Fitness as Hyper-parameters Evolve



Future Steps: Proof of Adaptability

MNIST

0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

Inner Product Layer (10 outputs)

Rectified Linear Unit Layer (500 outputs)

Inner Product Layer (500 outputs)

Pool2 Layer (Max, K=2)

Conv2 Layer (50 outputs, K=5)

Pool1 Layer (Max, K=2)

Conv1 Layer (20 outputs, K=5)

MNIST Training Data

- Current networks are "hand engineered" to data sets
- Can we automatically design to a data set?

MENNDL

CIFAR

airplane … truck

Inner Product Layer (10 outputs)

Inner Product Layer (64 outputs)

Pool3 Layer (Avg pooling, K=3)

Conv3 Layer (64 outputs, K=5) (ReLU)

Pool2 Layer (Avg pooling, K=3)

Conv2 Layer (32 outputs, K=5) (ReLU)

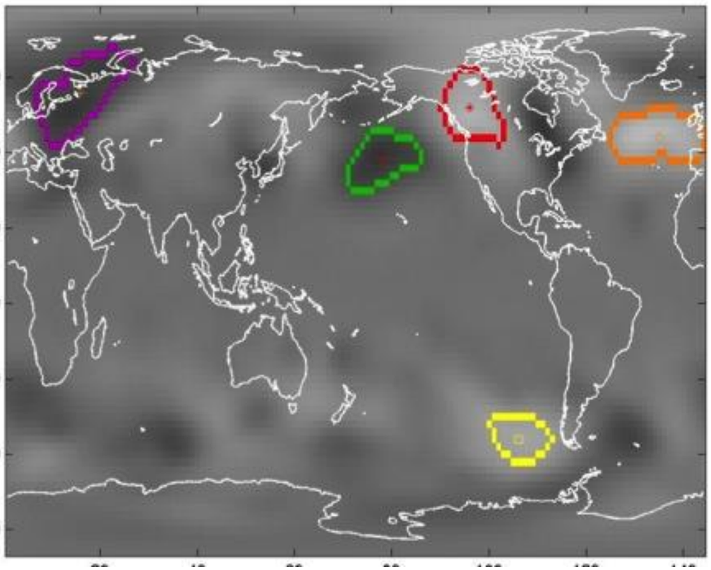
Pool1 Layer (Max pooling, K=3)

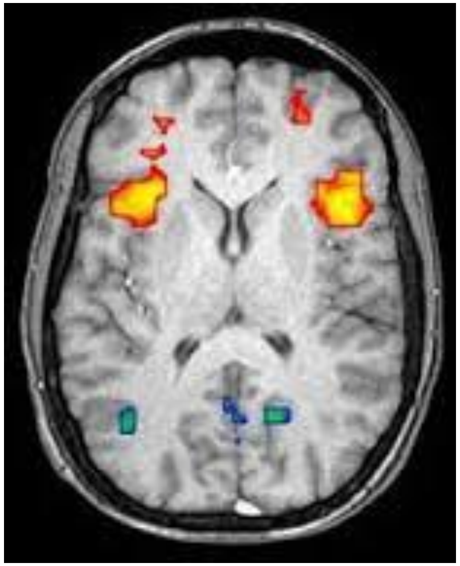
Conv1 Layer (32 outputs, K=5)

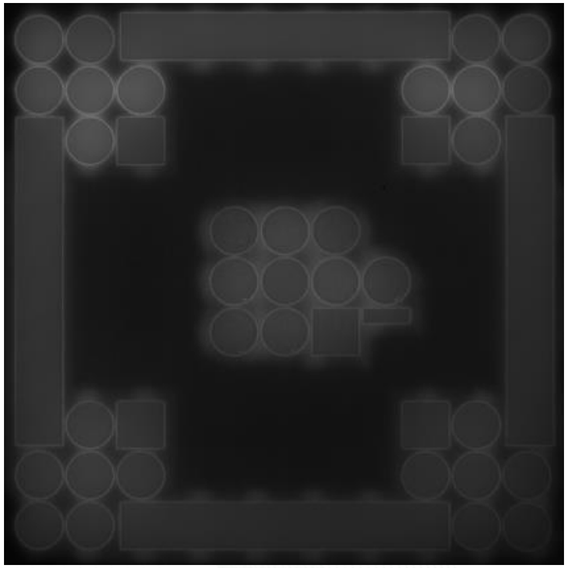
Data

Future Steps: Scientific Discovery

- Apply to extremely large scientific data sets in areas such as climate science, medicine, material science, etc.



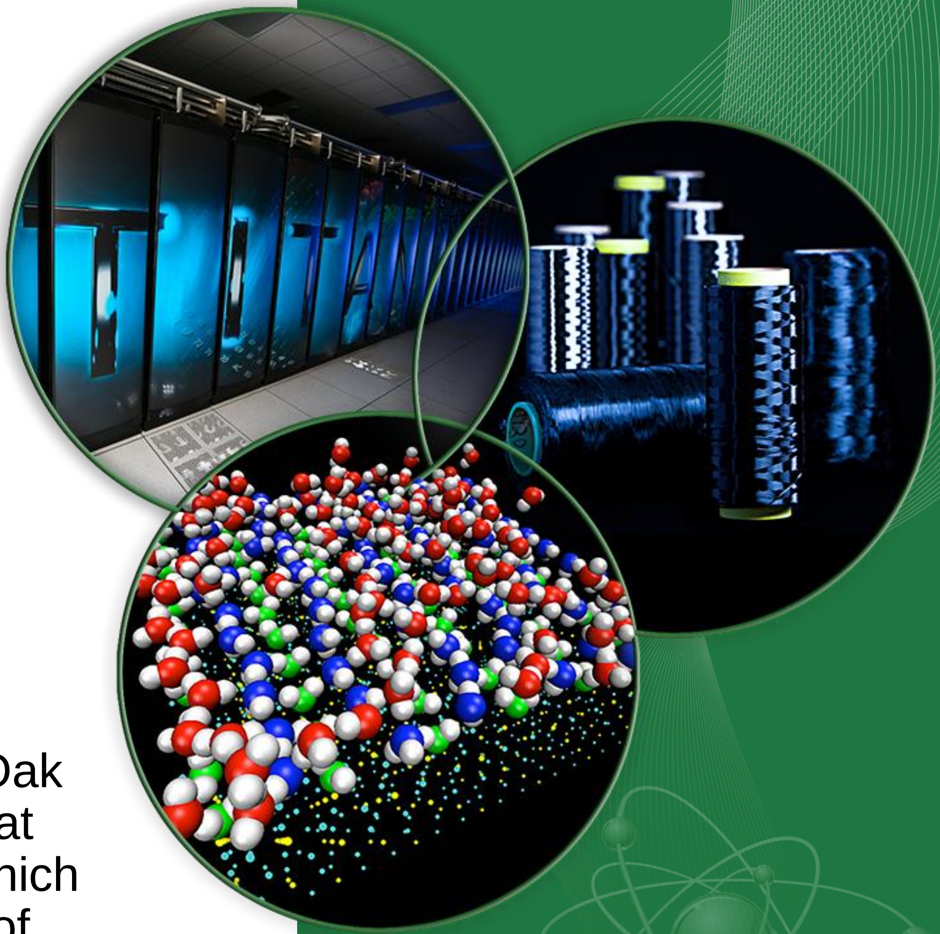




MENNDL team

- Steven Young
- Derek Rose
- Thomas Karnowski
- Seung-Hwan Lim
- Robert Patton

Thank you!



Steven R. Young, PhD
Computational Data Analytics
Oak Ridge National Laboratory
youngsr@ornl.gov
http://stevenryoung.com

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.