

# Researchers use Titan to accelerate design, training of deep learning networks

January 10 2018, by Scott Jones

---



ORNL's Steven Young (left) and Travis Johnston used Titan to prove the design and training of deep learning networks could be greatly accelerated with a capable computing system. Credit: Oak Ridge National Laboratory, US Department of Energy. Photographed by Jason Richards

A team of researchers from the Department of Energy's Oak Ridge National Laboratory has married artificial intelligence and high-performance computing to achieve a peak speed of 20 petaflops in the generation and training of deep learning networks on the laboratory's Titan supercomputer.

Deep learning is a burgeoning field of [artificial intelligence](#) that uses networks modeled after the human brain to "learn" how to distinguish features and patterns in vast datasets. Such networks hold great promise in the realization of numerous technologies, from self-driving cars to intelligent robots.

Due to its ability to make sense of massive amounts of data, researchers across the scientific spectrum are eager to refine deep learning and apply it to some of today's most challenging science problems. One such effort is ORNL's Advances in Machine Learning to Improve Scientific Discovery at Exascale and Beyond (ASCEND) project, which aims to use deep learning to make sense of the massive datasets produced by the world's most sophisticated scientific experiments, such as those located at ORNL.

Analysis of such datasets generally requires existing [neural networks](#) to be modified, or novel networks designed and then "trained" so that they know precisely what to look for and can produce valid results.

This is a time-consuming and difficult task, but one that an ORNL team led by Robert Patton and including Steven Young and Travis Johnston recently demonstrated can be dramatically expedited with a capable computing system such as ORNL's Titan, the nation's fastest supercomputer for science.

To efficiently design neural networks capable of tackling scientific datasets and expediting breakthroughs, Patton's team developed two

codes for evolving (MENNDL) and fine-tuning (RAvENNA) [deep neural network](#) architectures.

Both codes can generate and train as many as 18,600 neural networks simultaneously. Peak performance can be estimated by randomly sampling, and then carefully profiling, several hundred of these independently trained networks.

Both codes achieved a [peak performance](#) of 20 petaflops, or 20 thousand trillion calculations per second, on Titan (or just under half of Titan's single precision total peak performance). In practical terms, that translates to training 40-50,000 networks per hour.

"The real measure of success in the deep learning community is time-to-solution," said Johnston. "And with a machine like Titan we are able to train an unparalleled number of highly accurate networks."

Titan is a Cray hybrid system, meaning that it uses both traditional CPUs and graphics processing units (GPUs) to tackle complex calculations for big science problems efficiently; the GPUs also happen to be the processor of choice for training deep learning networks.

The team's work demonstrates that with the right [high-performance computing](#) system researchers can efficiently train large numbers of networks, which can then be used to help them tackle today's increasingly data-heavy experiments and simulations.

This efficient design of deep neural networks will enable researchers to deploy highly accurate, custom-designed models, saving both time and money by freeing the scientist from the task of designing a network from the ground up.

And because the OLCF's next leadership computing system, Summit,

features a deep-learning friendly architecture with enhanced GPUs and complementary Tensor cores, the team is confident both codes will only get faster.

"Out of the box, without tuning to Summit's unique architecture, we are expecting an increase in performance up to 50 times," said Johnston.

With that sort of network training capability, Summit could be indispensable to researchers across the scientific spectrum looking to deep learning to help them tackle some of science's most immense challenges.

Patton's team is not waiting for the improved hardware to start tackling current scientific data challenges; they have already deployed their codes to assist domain scientists at the Department of Energy's Fermilab in Batavia, Illinois.

Researchers at Fermilab used MENNDL to better understand how neutrinos interact with ordinary matter by producing a classification network to support their Main Injector Experiment for  $\nu$ -A (MINER $\nu$ A), a neutrino scattering experiment. The task, known as vertex reconstruction, required a network to analyze images and precisely identify the location where neutrinos interact with one of many targets—a task akin to finding the aerial source of a starburst of fireworks.

In only 24 hours, MENNDL produced optimized networks that outperformed any previously handcrafted network—an achievement that could easily have taken scientists months to accomplish. To identify the high-performing [network](#), MENNDL evaluated approximately 500,000 neural networks, training them on a data set consisting of 800,000 images of neutrino events, steadily using 18,000 of Titan's nodes.

"You need something like MENNDL to explore this effectively infinite space of possible networks, but you want to do it efficiently," Young said. "What Titan does is bring the time to solution down to something practical."

And with Summit to come online this year, the future of [deep learning](#) in big science looks bright indeed.

Provided by Oak Ridge National Laboratory

Citation: Researchers use Titan to accelerate design, training of deep learning networks (2018, January 10) retrieved 24 April 2024 from <https://phys.org/news/2018-01-titan-deep-networks.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.