

Researchers design novel method for energy-efficient deep neural networks

March 14 2018, by Scott Jones



ORNL's Hong-Jun Yoon, Mohammed Alawad and Gina Tourassi have developed a novel method for more efficiently training large numbers of networks capable of solving complex science problems. Credit: Jason Richards/Oak Ridge National Laboratory, US Department of Energy

An Oak Ridge National Laboratory method to improve the energy efficiency of scientific artificial intelligence is showing early promise in efforts to parse insights from volumes of cancer data.

Researchers are realizing the potential of deep learning to rapidly advance science, but "training" the underlying [neural networks](#) with large volumes of data to tackle the task at hand can require large amounts of energy. These networks also require complex connectivity and enormous amounts of storage, both of which further reduce their [energy efficiency](#) and potential in real-world applications.

To address this issue, ORNL's Mohammed Alawad, Hong-Jun Yoon, and Georgia Tourassi developed a novel method for the development of energy-efficient [deep neural networks](#) capable of solving complex science problems. They [presented](#) their research at the 2017 IEEE Conference on Big Data in Boston.

The researchers demonstrated that by converting deep learning neural networks (DNNs) to "deep spiking" neural networks (DSNNs), they can improve the energy efficiency of network design and realization.

DSNNs imitate neurons in the human brain via pulses or "spikes" in the place of actual signals, with the individual spikes indicating where to perform the computations. This process minimizes the necessary calculations and maximizes the network's energy efficiency. However, energy efficiency comes at the cost of task performance, and the authors' novel stochastic method for implementing DSNNs overcomes this tradeoff.

The results were impressive: the team's approach achieved nearly the same accuracy as the original DNN and performed better than a state-of-the-art spiking neural network. The team's stochastic-based DSNN, which distributes spikes uniformly over time, consumed 38 times less

energy than the original DNN and almost 2 times less energy than a conventional DSNN while delivering markedly better task performance.

The researchers trained their network on clinical text data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program, which provides cancer statistics such as incidence, prevalence, and mortality across the population associated by age, sex, race, year of diagnosis, and geographic areas.

The ORNL team applied the newly trained networks to clinical pathology reports, the main source of information for the national cancer surveillance program. These reports contain vast amounts of unstructured text, said Yoon, and researchers are developing intelligent language understanding systems to extract the most relevant clinical concepts in the sea of text.

The clinical reports represent a "sparse" dataset, which typically pose unique challenges to spiking networks. Most DSNN techniques have focused on computer vision tasks such as the MNIST dataset, which consist of a series of handwritten digits to train image processing networks. These datasets are typically "dense," meaning all variables in the dataset are populated with values, a characteristic that often simplifies analyses.

Traditional techniques for improving the performance and energy efficiency of spiking networks often preserve the structures of conventional neural networks, a practice that sacrifices accuracy and performance. These shortcomings drove the team to develop a new methodology, which relies on simple circuit hardware to perform complex calculations.

"Spiking the network lowers energy consumption because we disregard the unnecessary computations and we look only for the relevant nodes of

the [network](#)," said Yoon, "and this is one way we get energy efficiency improvements while identifying important clinical information with high accuracy."

The team's technique will help ORNL researchers in the CANcer Distributed Learning Environment (CANDLE) project, which aims to use the lab's world-class big data expertise and computing facilities to scan millions of clinical reports in search for insights on causes of cancer, best courses of treatment, and improved outcomes. They will soon attempt to parallelize the algorithm for more computational efficiency.

The spiking networks were optimized on graphics processing units (GPUs), the processors of choice for artificial intelligent applications, particularly those that utilize machine learning and deep learning. However, the methodology can be extended for training spiking networks, further increasing the [energy](#) efficiency of these novel networks while expediting scientific discovery via [deep learning](#).

Provided by Oak Ridge National Laboratory

Citation: Researchers design novel method for energy-efficient deep neural networks (2018, March 14) retrieved 19 April 2024 from <https://techxplore.com/news/2018-03-method-energy-efficient-deep-neural-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.