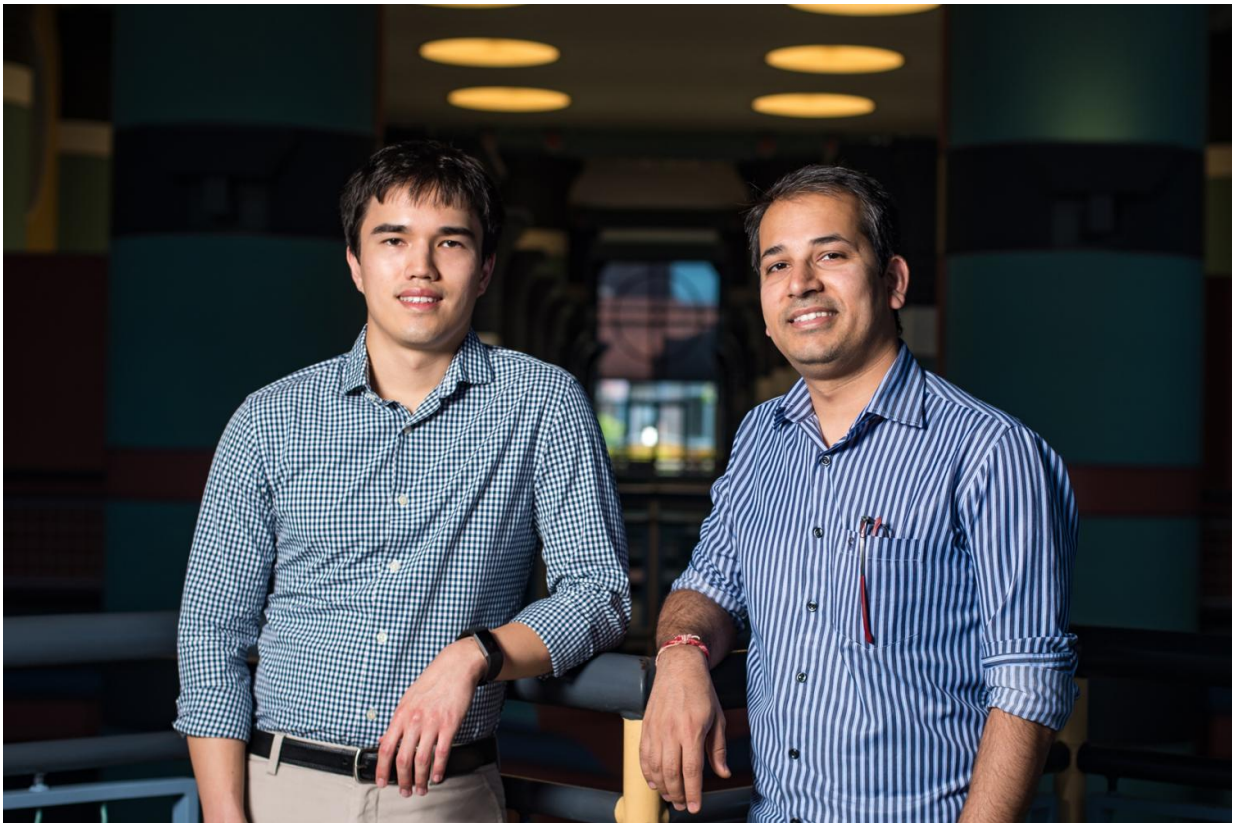


Scientists slash computations for deep learning

June 1 2017



Ryan Spring (left) and Anshumali Shrivastava. Credit: Jeff Fitlow/Rice University

Rice University computer scientists have adapted a widely used technique for rapid data lookup to slash the amount of

computation—and thus energy and time—required for deep learning, a computationally intense form of machine learning.

"This applies to any [deep-learning](#) architecture, and the technique scales sublinearly, which means that the larger the [deep neural network](#) to which this is applied, the more the savings in computations there will be," said lead researcher Anshumali Shrivastava, an assistant professor of computer science at Rice.

The research will be presented in August at the KDD 2017 conference in Halifax, Nova Scotia. It addresses one of the biggest issues facing tech giants like Google, Facebook and Microsoft as they race to build, train and deploy massive deep-learning networks for a growing body of products as diverse as self-driving cars, language translators and intelligent replies to emails.

Shrivastava and Rice graduate student Ryan Spring have shown that techniques from "hashing," a tried-and-true data-indexing method, can be adapted to dramatically reduce the computational overhead for deep learning. Hashing involves the use of smart hash functions that convert data into manageable small numbers called hashes. The hashes are stored in tables that work much like the index in a printed book.

"Our approach blends two techniques—a clever variant of locality-sensitive hashing and sparse backpropagation—to reduce computational requirements without significant loss of accuracy," Spring said. "For example, in small-scale tests we found we could reduce computation by as much as 95 percent and still be within 1 percent of the accuracy obtained with standard approaches."

The basic building block of a deep-learning [network](#) is an artificial neuron. Though originally conceived in the 1950s as models for the biological neurons in living brains, [artificial neurons](#) are just

mathematical functions, equations that act upon an incoming piece of data and transform it into an output.

In [machine learning](#), all neurons start the same, like blank slates, and become specialized as they are trained. During training, the network is "shown" vast volumes of data, and each neuron becomes a specialist at recognizing particular patterns in the data. At the lowest layer, neurons perform the simplest tasks. In a photo recognition application, for example, low-level neurons might recognize light from dark or the edges of objects. Output from these neurons is passed on to the neurons in the next layer of the network, which search for their own specialized patterns. Networks with even a few layers can learn to recognize faces, dogs, stop signs and school buses.

"Adding more neurons to a network layer increases its expressive power, and there's no upper limit to how big we want our networks to be," Shrivastava said. "Google is reportedly trying to train one with 137 billion neurons." By contrast, he said, there are limits to the amount of computational power that can be brought to bear to train and deploy such networks.

"Most machine-learning algorithms in use today were developed 30-50 years ago," he said. "They were not designed with computational complexity in mind. But with 'big data,' there are fundamental limits on resources like compute cycles, energy and memory. Our lab focuses on addressing those limitations."

Spring said computation and energy savings from hashing will be even larger on massive deep networks.

"The savings increase with scale because we are exploiting the inherent sparsity in big data," he said. "For instance, let's say a deep net has a billion neurons. For any given input—like a picture of a dog—only a few

of those will become excited. In data parlance, we refer to that as sparsity, and because of sparsity our method will save more as the network grows in size. So while we've shown a 95 percent savings with 1,000 neurons, the mathematics suggests we can save more than 99 percent with a billion [neurons](#)."

More information: "Scalable and Sustainable Deep Learning via Randomized Hashing" arxiv.org/abs/1602.08194

Provided by Rice University

Citation: Scientists slash computations for deep learning (2017, June 1) retrieved 1 May 2024 from <https://phys.org/news/2017-06-scientists-slash-deep.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.