

(<https://www.extremetech.com>)

[HOME \(HTTPS://WWW.EXTREMETECH.COM\)](https://www.extremetech.com)

[COMPUTING \(HTTPS://WWW.EXTREMETECH.COM/CATEGORY/COMPUTING\)](https://www.extremetech.com/category/computing)

MIT RESEARCH COULD MAKE NEURAL NETWORKS PORTABLE

MIT Research Could Make Neural Networks Portable

By [Ryan Whitwam \(https://www.extremetech.com/author/rwhitwam\)](https://www.extremetech.com/author/rwhitwam) on July 19, 2017 at 7:30 am

[2 Comments \(https://www.extremetech.com/extreme/252689-mit-research-make-neural-networks-portable#disqus_thread\)](https://www.extremetech.com/extreme/252689-mit-research-make-neural-networks-portable#disqus_thread)

84
shares

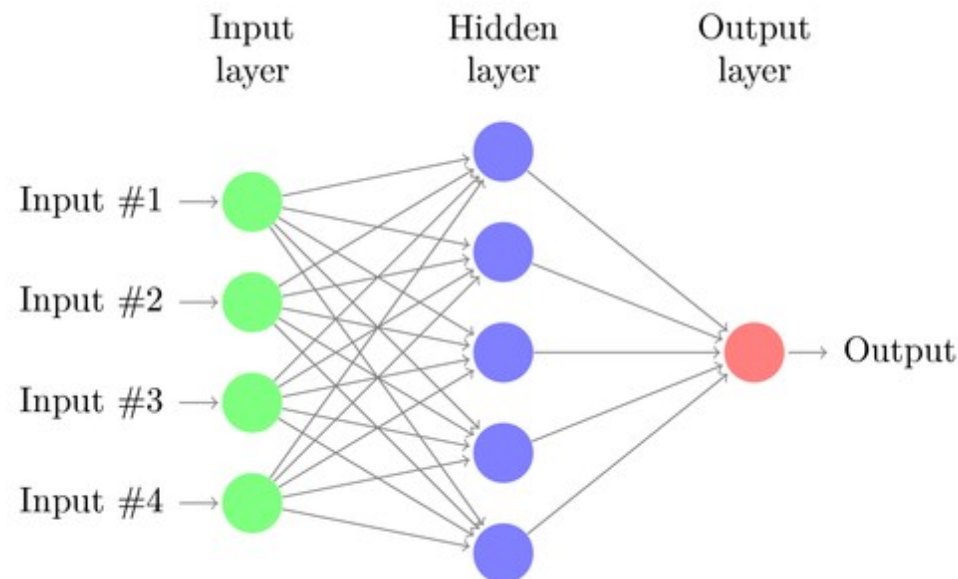


Computers aren't just getting faster these days—they're getting smarter. That's thanks largely to artificial [neural networks](https://www.extremetech.com/tag/neural-networks) (<https://www.extremetech.com/tag/neural-networks>), which consist of multiple nodes of processing designed to interpret and understand

data more like a biological brain. Neural networks are behind modern software features like voice recognition, computer vision, and even beating humans at Go (<https://www.extremetech.com/extreme/249750-googles-alphago-defeats-worlds-top-human-go-player>). However, neural networks need a lot of power, so they usually run in the cloud. Having a neural network locally on your mobile device could be hugely beneficial, and it may be possible thanks to research from a team at MIT.

The researchers, led by MIT associate professor Vivienne Sze, are not new to the idea of running neural networks on the go. In a past project, Sze and her team designed a custom computer chip that could run a neural network more efficiently on a smartphone. However, the adoption of new hardware is a slow and difficult process that affects many other aspects of the design. Sze's new approach is to pare down the neural network (<http://news.mit.edu/2017/bringing-neural-networks-cellphones-0718>) until it can operate efficiently on existing mobile hardware.

The research points to energy savings as high as 73 percent compared with an unaltered neural network, which would be plausible to run on a phone for a subset of “smart” features. Shrinking a neural network to this degree required careful monitoring of energy usage, so the team built a tool that tracks where energy is being used in the network. A neural network is made of many different nodes, some of which are important to the learning and processing and others that are less so. Some of these nodes also consume more or less power, and this is how the team arrived at “energy-aware pruning.”



<https://www.extremetech.com/wp-content/uploads/2016/03/Deep-neural-networks-have-at-least-one-hidden-layer-and-often-hundreds.->

[That-makes-them-expensive-to-emulate-on-traditional-hardware..png\)](#)

Pruning is an established way of shrinking the size of a neural network, wherein less important nodes are removed. That works well to a point, but it doesn't make the best possible impact on power consumption. You could keep trimming low-weight nodes all day, but still end up with an inefficient network. Energy-aware pruning uses the monitoring system devised by Sze's team to remove nodes that do the most to improve efficiency.

The result of this process is a more efficient, functional neural network with fewer nodes than you'd get with traditional pruning. This approach could make neural networks workable on mobile devices where battery life and heat are a concern. Meanwhile, Google has been working on improving its TPU neural network hardware, and it's also exploring the possibility of making its own mobile processors. If those chips include TPU-like capabilities, we could really make use of these more efficient networks.

Now read: [What are artificial neural networks? \(https://www.extremetech.com/extreme/215170-artificial-neural-networks-are-changing-the-world-what-are-they\)](https://www.extremetech.com/extreme/215170-artificial-neural-networks-are-changing-the-world-what-are-they)

Top image credit: Jose-Luis Olivares/MIT

You Might Also Like

Powered By ZergNet

