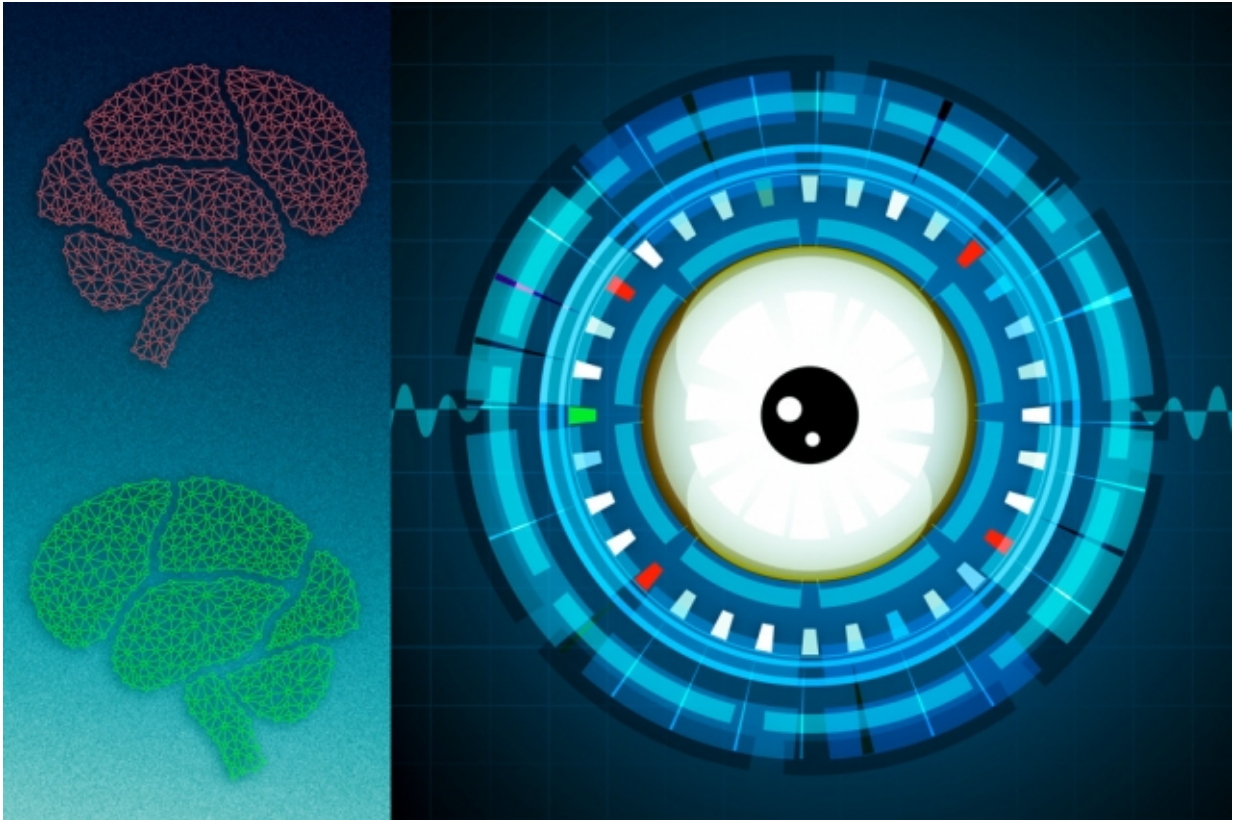


Machines that learn like people

December 23 2015, by Larry Hardesty



Researchers have developed a computational model of visual representation intended to reflect the way the brain works. Credit: MIT News

Object-recognition systems are beginning to get pretty good—and in the case of Facebook's face-recognition algorithms, frighteningly good.

But object-recognition systems are typically trained on millions of visual

examples, which is a far cry from how humans learn. Show a human two or three pictures of an object, and he or she can usually identify new instances of it.

Four years ago, Tomaso Poggio's group at MIT's McGovern Institute for Brain Research began developing a new computational model of visual representation, intended to reflect what the [brain](#) actually does. And in a forthcoming issue of the journal *Theoretical Computer Science*, the researchers prove that a machine-learning system based on their model could indeed make highly reliable object discriminations on the basis of just a few examples.

In both that paper and another that appeared in October in *PLOS Computational Biology*, they also show that aspects of their model accord well with empirical evidence about how the brain works.

"If I am given an image of your face from a certain distance, and then the next time I see you, I see you from a different distance, the image is quite different, and simple ways to match it don't work," says Poggio, the Eugene McDermott Professor in the Brain Sciences in MIT's Department of Brain and Cognitive Sciences. "In order solve this, you either need a lot of examples—I need to see your face not only in one position but in all possible positions—or you need an invariant representation of an object."

An invariant representation of an object is one that's immune to differences such as size, location, and rotation within the lane. Computer vision researchers have proposed several techniques for invariant object representation, but Poggio's group had the further challenge of finding an invariant representation that was consistent with what we know about the brain's machinery.

What nerves compute

Nerve cells, or neurons, are long, thin cells with branching ends. In the cerebral cortex, which is where visual processing happens, each neuron has about 10,000 branches at each end.

Two cortical neurons thus communicate with each other across 10,000 distinct chemical junctions, known as synapses. Each synapse has its own "weight," a factor by which it multiplies the strength of an incoming signal. The signals crossing all 10,000 synapses are then added together in the body of the neuron. Patterns of stimulation and electrical activity change the weights of synapses over time, which is the mechanism by which habits and memories become ingrained.

A key operation in the branch of mathematics known as linear algebra is the dot-product, which takes two sequences of numbers—or vectors—multiplies their elements together in an orderly way, and adds up the results to yield a single number. In the cortex, the output of a single neural circuit could thus be thought of as the dot-product of two 10,000-variable vectors. That's a very large calculation that each neuron in the brain can do at a stroke.

Poggio's group developed an invariant representation of objects that's based on dot-products. Suppose that you make a little digital movie of an object rotating 360 degrees in a plane—say, 24 frames, each depicting the object as rotated a little bit further than it was in the last one. You store the movie as a sequence of 24 stills.

Suppose next that you're presented with a digital image of an unfamiliar object. Because the image can be interpreted as a string of numbers describing the color values of pixels—a vector—you can calculate its dot-product with each of the stills from your movie and store that sequence of 24 numbers.

Invariance

Now, if you're presented with an image of the same object rotated, say, 90 degrees, and you calculate its dot-product with your sequence of stills, you'll get the same 24 numbers. They won't be in the same order: What was the dot-product with the first still will now be the dot-product with the sixth. But they'll be the same numbers.

That list of numbers, then, is a representation of the new object that is invariant to rotation. Similar sequences of stills, which depict an object at various sizes, or at various locations around the frame, will yield sequences of dot-products that are invariant to size and location.

In their new paper, Poggio and his colleagues—first author Fabio Anselmi, a postdoc in Poggio's group; Joel Leibo, a research affiliate at the McGovern Institute and a research scientist at Google DeepMind; Lorenzo Rosasco, a visiting professor in the Department of Brain and Cognitive Science; and Jim Mutch and Andrea Tacchetti, graduate students in Poggio's group—demonstrate that, if the goal is to produce an object representation invariant to rotation, size, and location, then the ideal template is a set of images known as Gabor filters. And Gabor filters, it turns out, are known to offer a good description of the image-processing operations performed by the so-called "simple cells" in the visual cortex.

Three dimensions

While this technique works well for visual transformations within a plane, however, it doesn't work as well for rotation in three dimensions. The dot-product between a new image and that of, say, a car seen straight on would be very different from the dot-product of the same image and that of a car seen from the side.

But Poggio's group has shown that if the template of still images depicts an object of the same type as the new object, dot-products will still yield adequately invariant descriptions. And this observation accords with recent research by MIT's Nancy Kanwisher and others, indicating that the visual cortex has regions specialized for recognizing particular classes of objects, such as faces or bodies.

In the work described in *PLOS Computational Biology*, Poggio and his colleagues—Leibo, Anselmi, and Qianli Liao, a graduate student in electrical engineering and computer science—built a computer system that assembled a set of still images and used the dot-product algorithm to learn to classify thousands of random objects.

For each of the object classes that the system learned, it produced a set of templates that predicted the size and variance of the regions in the human [visual cortex](#) devoted to corresponding classes. That suggests, the researchers argue, that the brain and their system may be doing something similar.

The researchers' invariance hypothesis is "a powerful approach to bridge the large gap between contemporary machine learning, with its emphasis on millions of labeled examples, and the primate visual system that in many instances can learn from a single example," says Christof Koch, a professor of biology and engineering at Caltech and chief scientific officer of the Allen Institute for Brain Science. "This sort of elegant mathematical framework will be necessary if we are to understand existing natural intelligent systems, on the road to building powerful artificial systems."

More information: Fabio Anselmi et al. Unsupervised learning of invariant representations, *Theoretical Computer Science* (2015). [DOI: 10.1016/j.tcs.2015.06.048](https://doi.org/10.1016/j.tcs.2015.06.048)

Joel Z. Leibo et al. The Invariance Hypothesis Implies Domain-Specific Regions in Visual Cortex, *PLOS Computational Biology* (2015). DOI: [10.1371/journal.pcbi.1004390](https://doi.org/10.1371/journal.pcbi.1004390)

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Machines that learn like people (2015, December 23) retrieved 4 May 2024 from <https://phys.org/news/2015-12-machines-people.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.