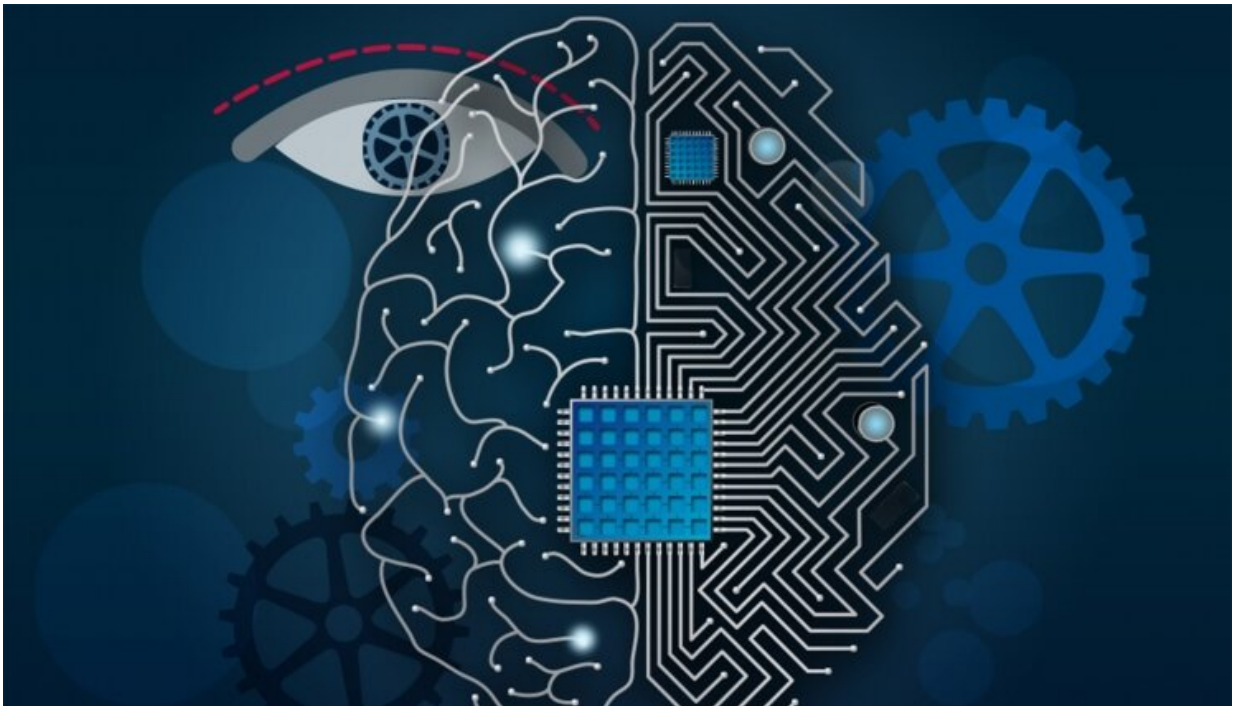# For better deep neural network vision, just add feedback (loops)

April 30 2019, by Sabbi Lall

Artistic rendering of artificial intelligence and deep learning visual recognition systems. Credit: Christine Daniloff

Your ability to recognize objects is remarkable. If you see a cup under unusual lighting or from unexpected directions, there's a good chance that your brain will still compute that it is a cup. Such precise object recognition is one holy grail for artificial intelligence developers, such as those improving self-driving car navigation.

While modeling primate object recognition in the [visual cortex](#) has revolutionized artificial visual recognition systems, current deep learning systems are simplified, and fail to recognize some objects that are child's play for primates such as humans.

In findings published in *Nature Neuroscience*, McGovern Institute investigator James DiCarlo and colleagues have found evidence that feedback improves recognition of hard-to-recognize objects in the [primate brain](#), and that adding feedback circuitry also improves the performance of [artificial neural network](#) systems used for vision applications.

Deep convolutional neural networks (DCNN) are currently the most successful models for accurately recognizing objects on a fast timescale (less than 100 milliseconds) and have a general architecture inspired by the primate ventral visual stream, cortical regions that progressively build an accessible and refined representation of viewed objects. Most DCNNs are simple in comparison to the primate ventral stream, however.

"For a long period of time, we were far from an model-based understanding. Thus our field got started on this quest by modeling visual recognition as a feedforward process," explains senior author DiCarlo, who is also the head of MIT's Department of Brain and Cognitive Sciences and research co-leader in the Center for Brains, Minds, and Machines (CBMM). "However, we know there are recurrent anatomical connections in brain regions linked to object recognition."

Think of feedforward DCNNs, and the portion of the visual system that first attempts to capture objects, as a subway line that runs forward through a series of stations. The extra, recurrent brain networks are instead like the streets above, interconnected and not unidirectional. Because it only takes about 200 ms for the brain to recognize an object

quite accurately, it was unclear if these recurrent interconnections in the brain had any role at all in core object recognition. Perhaps those recurrent connections are only in place to keep the visual system in tune over long periods of time. For example, the return gutters of the streets help slowly clear it of water and trash, but are not strictly needed to quickly move people from one end of town to the other. DiCarlo, along with lead author and CBMM postdoc Kohitij Kar, set out to test whether a subtle role of recurrent operations in rapid visual object recognition was being overlooked.

Challenging recognition

The authors first needed to identify objects that are trivially decoded by the primate brain, but are challenging for artificial systems. Rather than trying to guess why deep learning was having problems recognizing an object (is it due to clutter in the image? a misleading shadow?), the authors took an unbiased approach that turned out to be critical.

Kar explains further that "we realized that AI models actually don't have problems with every image where an object is occluded or in clutter. Humans trying to guess why AI models were challenged turned out to be holding us back."

Instead, the authors presented the deep learning system, as well as monkeys and humans, with images, homing in on "challenge images" where the primates could easily recognize the objects in those images, but a feedforward DCNN ran into problems. When they, and others, added appropriate recurrent processing to these DCNNs, object recognition in challenge images suddenly became a breeze.

Processing times

Kar used neural recording methods with very high spatial and temporal precision to determine whether these images were really so trivial for primates. Remarkably, they found that although challenge images had initially appeared to be child's play to the human brain, they actually involve extra neural processing time (about an additional 30 ms), suggesting that recurrent loops operate in our brain, too.

"What the computer vision community has recently achieved by stacking more and more layers onto artificial neural networks, evolution has achieved through a brain architecture with recurrent connections," says Kar.

Diane Beck, professor of psychology and co-chair of the Intelligent Systems Theme at the Beckman Institute and not an author on the study, explains further. "Since entirely feedforward deep convolutional nets are now remarkably good at predicting primate brain activity, it raised questions about the role of feedback connections in the primate brain. This study shows that, yes, feedback connections are very likely playing a role in object recognition after all."

What does this mean for a self-driving car? It shows that deep learning architectures involved in object recognition need recurrent components if they are to match the primate brain, and also indicates how to operationalize this procedure for the next generation of intelligent machines.

"Recurrent models offer predictions of neural activity and behavior over time," says Kar. "We may now be able to model more involved tasks. Perhaps one day, the systems will not only recognize an object, such as a person, but also perform cognitive tasks that the human brain so easily manages, such as understanding the emotions of other people."

More information: Kohitij Kar et al. Evidence that recurrent circuits

are critical to the ventral stream's execution of core object recognition behavior, *Nature Neuroscience* (2019). [DOI: 10.1038/s41593-019-0392-5](#)

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