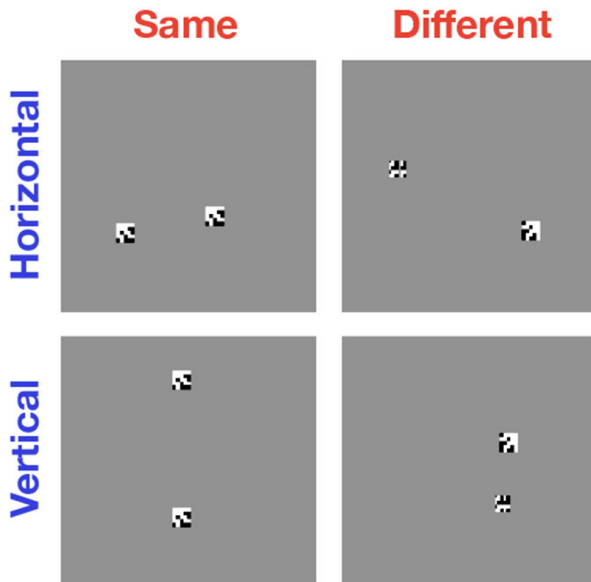


Research identifies key weakness in modern computer vision systems

30 July 2018, by Kevin Stacey



Computers are great at categorizing images by the objects found with them, but they're surprisingly bad at figuring out when two objects in a single image are the same or different from each other. New research helps to show why that task is so difficult for modern computer vision algorithms. Credit: Serre lab / Brown University

Computer vision algorithms have come a long way in the past decade. They've been shown to be as good or better than people at tasks like categorizing dog or cat breeds, and they have the remarkable ability to identify specific faces out of a sea of millions.

But research by Brown University scientists shows that computers fail miserably at a class of tasks that even young children have no problem with: determining whether two objects in an image are the same or different. In a paper presented last week at the annual meeting of the Cognitive Science Society, the Brown team sheds light on why computers are so bad at these types of tasks

and suggests avenues toward smarter computer [vision systems](#).

"There's a lot of excitement about what computer vision has been able to achieve, and I share a lot of that," said Thomas Serre, associate professor of cognitive, linguistic and psychological sciences at Brown and the paper's senior author. "But we think that by working to understand the limitations of current computer vision systems as we've done here, we can really move toward new, much more advanced systems rather than simply tweaking the systems we already have."

For the study, Serre and his colleagues used state-of-the-art computer [vision algorithms](#) to analyze simple black-and-white images containing two or more randomly generated shapes. In some cases the objects were identical; sometimes they were the same but with one [object](#) rotated in relation to the other; sometimes the objects were completely different. The computer was asked to identify the same-or-different relationship.

The study showed that, even after hundreds of thousands of training examples, the algorithms were no better than chance at recognizing the appropriate relationship. The question, then, was why these systems are so bad at this task.

Serre and his colleagues had a suspicion that it has something to do with the inability of these computer vision algorithms to individuate objects. When computers look at an image, they can't actually tell where one object in the image stops and the background, or another object, begins. They just see a collection of pixels that have similar patterns to collections of pixels they've learned to associate with certain labels. That works fine for identification or categorization problems, but falls apart when trying to compare two objects.

To show that this was indeed why the algorithms were breaking down, Serre and his team performed

experiments that relieved the computer from having human visual processing. to individuate objects on its own. Instead of showing the computer two objects in the same image, the researchers showed the computer the objects one at a time in separate images. The experiments showed that the algorithms had no problem learning same-or-different relationship as long as they didn't have to view the two objects in the same image.

Provided by Brown University

The source of the problem in individuating objects, Serre says, is the architecture of the machine learning systems that power the algorithms. The algorithms use convolutional neural networks—layers of connected processing units that loosely mimic networks of neurons in the brain. A key difference from the brain is that the artificial networks are exclusively "feed-forward"—meaning information has a one-way flow through the layers of the network. That's not how the visual system in humans works, according to Serre.

"If you look at the anatomy of our own visual system, you find that there are a lot of recurring connections, where the information goes from a higher visual area to a lower visual area and back through," Serre said.

While it's not clear exactly what those feedbacks do, Serre says, it's likely that they have something to do with our ability to pay attention to certain parts of our visual field and make mental representations of objects in our minds.

"Presumably people attend to one object, building a feature representation that is bound to that object in their working memory," Serre said. "Then they shift their attention to another object. When both objects are represented in working memory, your visual system is able to make comparisons like same-or-different."

Serre and his colleagues hypothesize that the reason computers can't do anything like that is because feed-forward neural networks don't allow for the kind of recurrent processing required for this individuation and mental representation of objects. It could be, Serre says, that making [computer](#) vision smarter will require neural networks that more closely approximate the recurrent nature of

APA citation: Research identifies key weakness in modern computer vision systems (2018, July 30)
retrieved 19 October 2021 from <https://techxplore.com/news/2018-07-key-weakness-modern-vision.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.