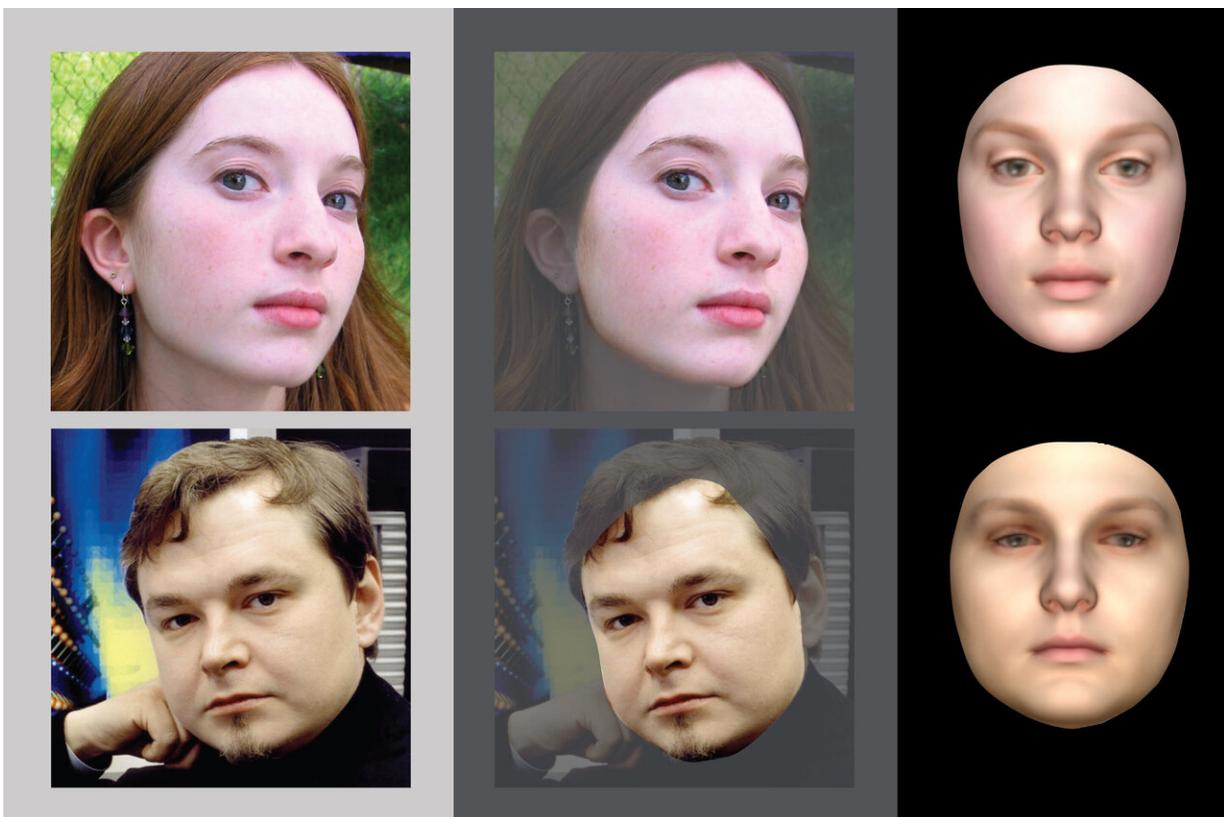


Computer model of face processing could reveal how the brain produces richly detailed visual representations so quickly

March 5 2020, by Anne Trafton



MIT cognitive scientists have developed a computer model of face recognition that performs a series of computations that reverse the steps that a computer graphics program would use to generate a 2D representation of a face. MIT cognitive scientists have developed a computer model of face recognition that performs a series of computations that reverse the steps that a computer graphics program would use to generate a 2D representation of a face. Credit: MIT

When we open our eyes, we immediately see our surroundings in great detail. How the brain is able to form these richly detailed representations of the world so quickly is one of the biggest unsolved puzzles in the study of vision.

Scientists who study the brain have tried to replicate this phenomenon using [computer](#) models of vision, but so far, leading models only perform much simpler tasks such as picking out an object or a face against a cluttered background. Now, a team led by MIT cognitive scientists has produced a computer model that captures the human visual system's ability to quickly generate a detailed scene description from an image, and offers some insight into how the brain achieves this.

"What we were trying to do in this work is to explain how perception can be so much richer than just attaching semantic labels on parts of an image, and to explore the question of how do we see all of the physical world," says Josh Tenenbaum, a professor of computational cognitive science and a member of MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and the Center for Brains, Minds, and Machines (CBMM).

The new model posits that when the brain receives visual input, it quickly performs a series of computations that reverse the steps that a computer graphics program would use to generate a 2-D representation of a face or other object. This type of model, known as efficient inverse graphics (EIG), also correlates well with electrical recordings from face-selective regions in the brains of nonhuman primates, suggesting that the primate visual system may be organized in much the same way as the [computer model](#), the researchers say.

Ilker Yildirim, a former MIT postdoc who is now an assistant professor

of psychology at Yale University, is the lead author of the paper, which appears today in *Science Advances*. Tenenbaum and Winrich Freiwald, a professor of neurosciences and behavior at Rockefeller University, are the senior authors of the study. Mario Belledonne, a graduate student at Yale, is also an author.

Inverse graphics

Decades of research on the brain's visual system has studied, in great detail, how light input onto the retina is transformed into cohesive scenes. This understanding has helped artificial intelligence researchers develop computer models that can replicate aspects of this system, such as recognizing faces or other objects.

"Vision is the functional aspect of the brain that we understand the best, in humans and other animals," Tenenbaum says. "And computer vision is one of the most successful areas of AI at this point. We take for granted that machines can now look at pictures and recognize faces very well, and detect other kinds of objects."

However, even these sophisticated artificial intelligence systems don't come close to what the human visual system can do, Yildirim says.

"Our brains don't just detect that there's an object over there, or recognize and put a label on something," he says. "We see all of the shapes, the geometry, the surfaces, the textures. We see a very rich world."

More than a century ago, the physician, physicist, and philosopher Hermann von Helmholtz theorized that the brain creates these rich representations by reversing the process of image formation. He hypothesized that the visual system includes an image generator that would be used, for example, to produce the faces that we see during

dreams. Running this generator in reverse would allow the brain to work backward from the image and infer what kind of face or other object would produce that image, the researchers say.

However, the question remained: How could the brain perform this process, known as inverse graphics, so quickly? Computer scientists have tried to create algorithms that could perform this feat, but the best previous systems require many cycles of iterative processing, taking much longer than the 100 to 200 milliseconds the brain requires to create a detailed visual representation of what you're seeing. Neuroscientists believe perception in the brain can proceed so quickly because it is implemented in a mostly feedforward pass through several hierarchically organized layers of neural processing.

The MIT-led team set out to build a special kind of deep neural network model to show how a neural hierarchy can quickly infer the underlying features of a scene—in this case, a specific face. In contrast to the standard deep neural networks used in computer vision, which are trained from labeled data indicating the class of an object in the image, the researchers' network is trained from a model that reflects the brain's internal representations of what scenes with faces can look like.

Their model thus learns to reverse the steps performed by a computer graphics program for generating faces. These graphics programs begin with a three-dimensional representation of an individual face and then convert it into a two-dimensional image, as seen from a particular viewpoint. These images can be placed on an arbitrary background image. The researchers theorize that the brain's visual system may do something similar when you dream or conjure a mental image of someone's face.

The researchers trained their deep neural network to perform these steps in reverse—that is, it begins with the 2-D image and then adds features

such as texture, curvature, and lighting, to create what the researchers call a "2.5D" representation. These 2.5D images specify the shape and color of the face from a particular viewpoint. Those are then converted into 3-D representations, which don't depend on the viewpoint.

"The model gives a systems-level account of the processing of faces in the brain, allowing it to see an image and ultimately arrive at a 3-D object, which includes representations of shape and texture, through this important intermediate stage of a 2.5D image," Yildirim says.

Model performance

The researchers found that their model is consistent with data obtained by studying certain regions in the brains of macaque monkeys. In a study published in 2010, Freiwald and Doris Tsao of Caltech recorded the activity of neurons in those regions and analyzed how they responded to 25 different faces, seen from seven different viewpoints. That study revealed three stages of higher-level face processing, which the MIT team now hypothesizes correspond to three stages of their inverse graphics model: roughly, a 2.5D viewpoint-dependent stage; a stage that bridges from 2.5 to 3-D; and a 3-D, viewpoint-invariant stage of face representation.

"What we show is that both the quantitative and qualitative response properties of those three levels of the brain seem to fit remarkably well with the top three levels of the network that we've built," Tenenbaum says.

The researchers also compared the model's performance to that of humans in a task that involves recognizing faces from different viewpoints. This task becomes harder when researchers alter the faces by removing the face's texture while preserving its shape, or distorting the shape while preserving relative texture. The new model's performance

was much more similar to that of humans than computer models used in state-of-the-art face-recognition software, additional evidence that this [model](#) may be closer to mimicking what happens in the human [visual system](#).

The researchers now plan to continue testing the modeling approach on additional images, including objects that aren't faces, to investigate whether inverse graphics might also explain how the brain perceives other kinds of scenes. In addition, they believe that adapting this approach to computer vision could lead to better-performing AI systems.

"If we can show evidence that these models might correspond to how the brain works, this work could lead computer vision researchers to take more seriously and invest more engineering resources in this inverse graphics approach to perception," Tenenbaum says. "The [brain](#) is still the gold standard for any kind of machine that sees the world richly and quickly."

More information: "Efficient inverse graphics in biological face processing" *Science Advances* (2020).

advances.sciencemag.org/content/6/10/eaax5979

Provided by Massachusetts Institute of Technology

Citation: Computer model of face processing could reveal how the brain produces richly detailed visual representations so quickly (2020, March 5) retrieved 27 April 2024 from

<https://techxplore.com/news/2020-03-reveal-brain-richly-visual-representations.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.