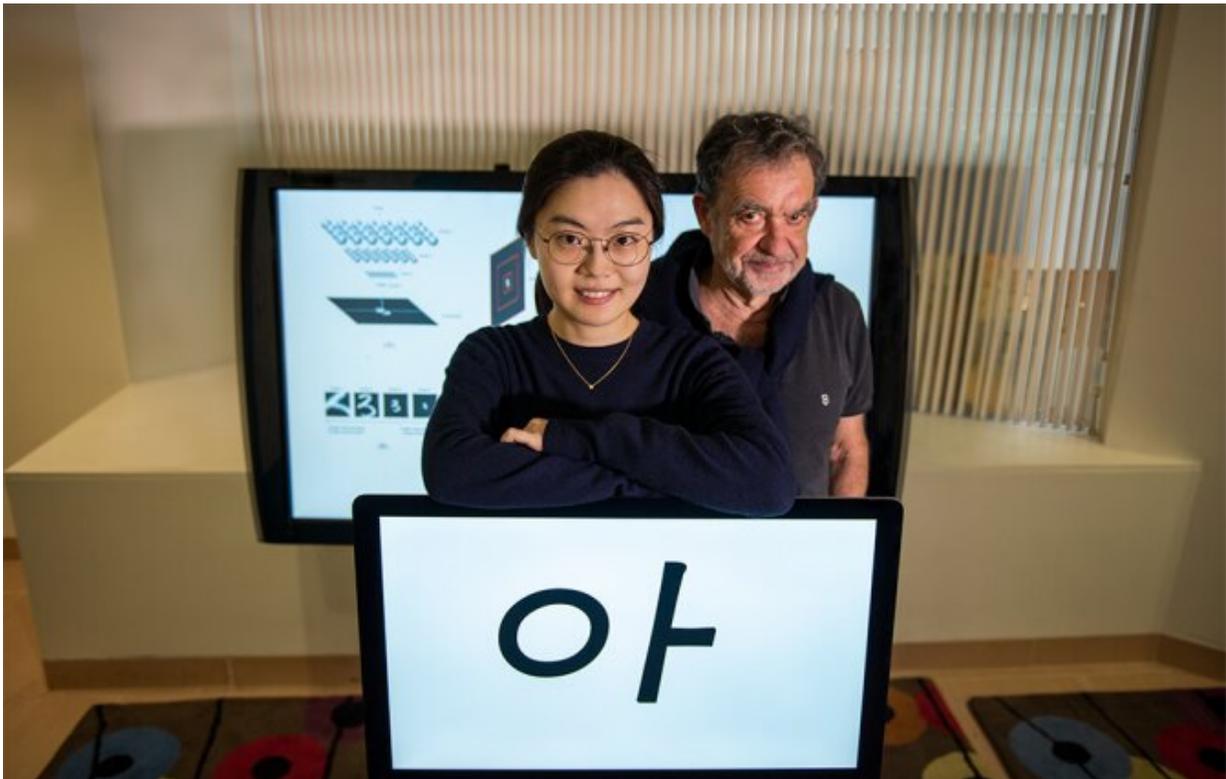


Bridging the gap between human and machine vision

February 12 2020, by Kris Brewer



Yena Han (left) and Tomaso Poggio stand with an example of the visual stimuli used in a new psychophysics study. Credit: Kris Brewer

Suppose you look briefly from a few feet away at a person you have never met before. Step back a few paces and look again. Will you be able to recognize her face? "Yes, of course," you probably are thinking.

If this is true, it would mean that our visual system, having seen a single image of an object such as a specific face, recognizes it robustly despite changes to the object's position and scale, for example. On the other hand, we know that state-of-the-art classifiers, such as vanilla deep networks, will fail this simple test.

In order to recognize a specific face under a range of transformations, [neural networks](#) need to be trained with many examples of the face under the different conditions. In other words, they can achieve invariance through memorization, but cannot do it if only one image is available. Thus, understanding how [human vision](#) can pull off this remarkable feat is relevant for engineers aiming to improve their existing classifiers. It also is important for neuroscientists modeling the primate [visual system](#) with deep networks. In particular, it is possible that the invariance with one-shot learning exhibited by biological [vision](#) requires a rather different computational strategy than that of deep networks.

A new paper by MIT Ph.D. candidate in [electrical engineering](#) and computer science Yena Han and colleagues in *Nature Scientific Reports*, titled "Scale and translation-invariance for novel objects in human vision," discusses how they study this phenomenon more carefully to create novel biologically inspired networks.

"Humans can learn from very few examples, unlike deep networks. This is a huge difference with vast implications for engineering of vision systems and for understanding how human vision really works," states co-author Tomaso Poggio—director of the Center for Brains, Minds and Machines (CBMM) and the Eugene McDermott Professor of Brain and Cognitive Sciences at MIT. "A key reason for this difference is the relative invariance of the primate visual system to scale, shift, and other transformations. Strangely, this has been mostly neglected in the AI community, in part because the psychophysical data were so far less than

clear-cut. Han's work has now established solid measurements of basic invariances of human vision."

To differentiate invariance rising from intrinsic computation with that from experience and memorization, the new study measured the range of invariance in one-shot learning. A one-shot learning task was performed by presenting Korean letter stimuli to human subjects who were unfamiliar with the language. These letters were initially presented a single time under one specific condition and tested at different scales or positions than the original condition. The first experimental result is that—just as you guessed—humans showed significant scale-invariant recognition after only a single exposure to these novel objects. The second result is that the range of position-invariance is limited, depending on the size and placement of objects.

Next, Han and her colleagues performed a comparable experiment in [deep neural networks](#) designed to reproduce this human performance. The results suggest that to explain invariant recognition of objects by humans, neural network models should explicitly incorporate built-in scale-invariance. In addition, limited position-invariance of human vision is better replicated in the network by having the model neurons' receptive fields increase as they are further from the center of the visual field. This architecture is different from commonly used neural [network](#) models, where an image is processed under uniform resolution with the same shared filters.

"Our work provides a new understanding of the brain representation of objects under different viewpoints. It also has implications for AI, as the results provide new insights into what is a good architectural design for deep neural networks," remarks Han, CBMM researcher and lead author of the study.

More information: Yena Han et al. Scale and translation-invariance

for novel objects in human vision, *Scientific Reports* (2020). [DOI: 10.1038/s41598-019-57261-6](https://doi.org/10.1038/s41598-019-57261-6)

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: Bridging the gap between human and machine vision (2020, February 12) retrieved 3 June 2023 from <https://techxplore.com/news/2020-02-bridging-gap-human-machine-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.