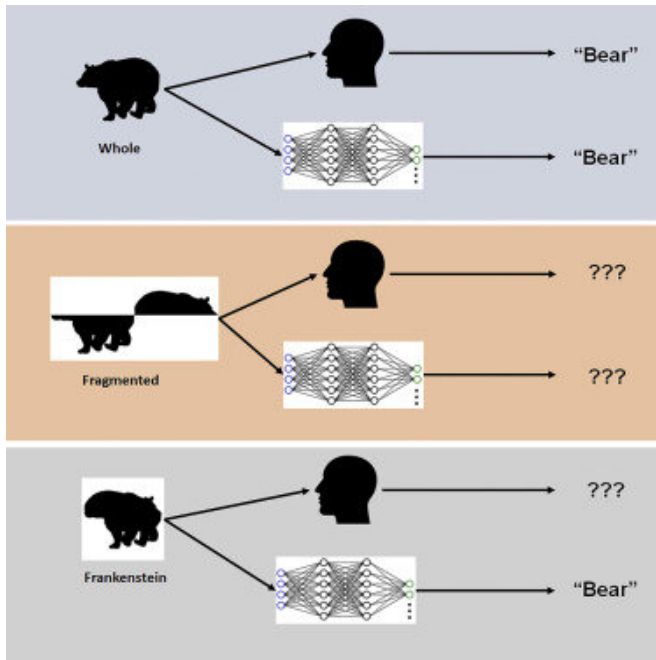


Study highlights how AI models take potentially dangerous 'shortcuts' in solving complex recognition tasks

16 September 2022



Credit: York University

Deep convolutional neural networks (DCNNs) don't see objects the way humans do—using configural shape perception—and that could be dangerous in real-world AI applications, says Professor James Elder, co-author of a York University study published today.

Published in the Cell Press journal *iScience*, Deep learning models fail to capture the configural nature of human shape perception is a collaborative study by Elder, who holds the York Research Chair in Human and Computer Vision and is Co-Director of York's Centre for AI & Society, and Assistant Psychology Professor Nicholas Baker at Loyola College in Chicago, a former VISTA postdoctoral fellow at York.

The study employed novel visual stimuli called "Frankensteins" to explore how the [human brain](#) and DCNNs process holistic, configural object properties.

"Frankensteins are simply objects that have been taken apart and put back together the wrong way around," says Elder. "As a result, they have all the right local features, but in the wrong places."

The investigators found that while the human visual system is confused by Frankensteins, DCNNs are not—revealing an insensitivity to configural object properties.

"Our results explain why deep AI models fail under certain conditions and point to the need to consider tasks beyond object recognition in order to understand [visual processing](#) in the brain," Elder says. "These deep models tend to take 'shortcuts' when solving complex recognition tasks. While these shortcuts may work in many cases, they can be dangerous in some of the real-world AI applications we are currently working on with our industry and government partners," Elder points out.

One such application is traffic video safety systems: "The objects in a busy traffic scene—the vehicles, bicycles and pedestrians—obstruct each other and arrive at the eye of a driver as a jumble of disconnected fragments," explains Elder. "The brain needs to correctly group those fragments to identify the correct categories and locations of the objects. An AI system for traffic safety monitoring that is only able to perceive the fragments individually will fail at this task, potentially misunderstanding risks to vulnerable road users."

According to the researchers, modifications to training and architecture aimed at making networks

more brain-like did not lead to configural processing, and none of the networks were able to accurately predict trial-by-trial human object judgements. "We speculate that to match human configural sensitivity, networks must be trained to solve broader range of object tasks beyond category recognition," notes Elder.

More information: Nicholas Baker et al, Deep learning models fail to capture the configural nature of human shape perception, *iScience* (2022). [DOI: 10.1016/j.isci.2022.104913](https://doi.org/10.1016/j.isci.2022.104913)

Provided by York University

APA citation: Study highlights how AI models take potentially dangerous 'shortcuts' in solving complex recognition tasks (2022, September 16) retrieved 2 December 2022 from <https://techxplore.com/news/2022-09-highlights-ai-potentially-dangerous-shortcuts.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.