

The benefits of peripheral vision for machines

2 March 2022, by Adam Zewe



New research from MIT suggests that a certain type of computer vision model that is trained to be robust to imperceptible noise added to image data encodes visual representations similarly to the way humans do using peripheral vision. Credit: Jose-Luis Olivares, MIT

Perhaps computer vision and human vision have more in common than meets the eye?

Research from MIT suggests that a certain type of robust [computer-vision model](#) perceives [visual representations](#) similarly to the way humans do using peripheral vision. These models, known as adversarially robust models, are designed to overcome subtle bits of noise that have been added to [image data](#).

The way these models learn to transform images is similar to some elements involved in human peripheral processing, the researchers found. But because machines do not have a [visual periphery](#), little work on computer vision models has focused on peripheral processing, says senior author Arturo Deza, a postdoc in the Center for Brains, Minds, and Machines.

"It seems like peripheral vision, and the textural

representations that are going on there, have been shown to be pretty useful for human vision. So, our thought was, OK, maybe there might be some uses in machines, too," says lead author Anne Harrington, a graduate student in the Department of Electrical Engineering and Computer Science.

The results suggest that designing a machine-learning model to include some form of peripheral processing could enable the model to automatically learn visual representations that are robust to some subtle manipulations in image data. This work could also help shed some light on the goals of peripheral processing in humans, which are still not well-understood, Deza adds.

The research will be presented at the International Conference on Learning Representations.

Double vision

Humans and computer vision systems both have what is known as foveal vision, which is used for scrutinizing highly detailed objects. Humans also possess peripheral vision, which is used to organize a broad, spatial scene. Typical computer vision approaches attempt to model foveal vision—which is how a machine recognizes objects—and tend to ignore peripheral vision, Deza says.

But foveal computer vision systems are vulnerable to adversarial noise, which is added to image data by an attacker. In an adversarial attack, a malicious agent subtly modifies images so each pixel has been changed very slightly—a human wouldn't notice the difference, but the noise is enough to fool a machine. For example, an image might look like a car to a human, but if it has been affected by adversarial noise, a computer vision model may confidently misclassify it as, say, a cake, which could have serious implications in an autonomous vehicle.

To overcome this vulnerability, researchers conduct what is known as adversarial training, where they create images that have been manipulated with adversarial noise, feed them to the neural network, and then correct its mistakes by relabeling the data and then retraining the model.

"Just doing that additional relabeling and training process seems to give a lot of perceptual alignment with human processing," Deza says.

He and Harrington wondered if these adversarially trained networks are robust because they encode object representations that are similar to human [peripheral vision](#). So, they designed a series of psychophysical human experiments to test their hypothesis.

Screen time

They started with a set of images and used three different computer vision models to synthesize representations of those images from noise: a "normal" machine-learning model, one that had been trained to be adversarially robust, and one that had been specifically designed to account for some aspects of human peripheral processing, called Texforms.

The team used these generated images in a series of experiments where participants were asked to distinguish between the original images and the representations synthesized by each model. Some experiments also had humans differentiate between different pairs of randomly synthesized images from the same models.

Participants kept their eyes focused on the center of a screen while images were flashed on the far sides of the screen, at different locations in their periphery. In one experiment, participants had to identify the oddball image in a series of images that were flashed for only milliseconds at a time, while in the other they had to match an image presented at their fovea, with two candidate template images placed in their periphery.

In the experiments, participants kept their eyes focused on the center of a screen while images were flashed on the far sides of the screen, at different locations in their periphery, like these animated gifs. In one experiment, participants had to identify the oddball image in a series that of images that were flashed for only milliseconds at a time. Credit: Massachusetts Institute of Technology

When the synthesized images were shown in the far periphery, the participants were largely unable to tell the difference between the original for the adversarially robust model or the Texform model. This was not the case for the standard machine-learning model.

However, what is perhaps the most striking result is that the pattern of mistakes that humans make (as a function of where the stimuli land in the periphery) is heavily aligned across all experimental conditions that use the stimuli derived from the Texform model and the adversarially robust model. These results suggest that adversarially robust models do capture some aspects of human peripheral processing, Deza explains.

In this experiment, researchers had humans match the center template with one of the two peripheral ones, without moving their eyes from the center of the screen. Credit: Massachusetts Institute of Technology

The researchers also computed specific machine-learning experiments and image-quality assessment metrics to study the similarity between images synthesized by each model. They found that those generated by the adversarially robust model and the Texforms model were the most similar, which suggests that these models compute similar image transformations.

"We are shedding light into this alignment of how humans and machines make the same kinds of mistakes, and why," Deza says. Why does adversarial robustness happen? Is there a biological equivalent for adversarial robustness in machines that we haven't uncovered yet in the brain?"

Deza is hoping these results inspire additional work in this area and encourage computer vision researchers to consider building more biologically inspired models.

These results could be used to design a computer [vision](#) system with some sort of emulated visual periphery that could make it automatically robust to adversarial noise. The work could also inform the development of machines that are able to create more accurate visual representations by using some aspects of human peripheral processing.

"We could even learn about [human vision](#) by trying to get certain properties out of artificial neural networks," Harrington adds.

Previous work had shown how to isolate "robust" parts of images, where training models on these images caused them to be less susceptible to adversarial failures. These robust images look like scrambled versions of the real images, explains Thomas Wallis, a professor for perception at the Institute of Psychology and Centre for Cognitive

Science at the Technical University of Darmstadt.

"Why do these robust images look the way that they do? Harrington and Deza use careful human behavioral experiments to show that peoples' ability to see the difference between these images and original photographs in the periphery is qualitatively similar to that of images generated from biologically inspired models of peripheral information processing in humans," says Wallis, who was not involved with this research. "Harrington and Deza propose that the same mechanism of learning to ignore some visual input changes in the periphery may be why robust images look the way they do, and why training on robust images reduces adversarial susceptibility. This intriguing hypothesis is worth further investigation, and could represent another example of a synergy between research in biological and machine intelligence."

More information: Anne Harrington, Arturo Deza, Finding biological plausibility for adversarially robust features via metameric tasks, openreview.net/forum?id=yeP_zx9vqNm

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

APA citation: The benefits of peripheral vision for machines (2022, March 2) retrieved 13 August 2022 from <https://techxplore.com/news/2022-03-benefits-peripheral-vision-machines.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.