UC SANTA BARBARA

THE Current

December 15, 2025
Sonia Fernandez

AI helps explain how covert attention works and uncovers new neuron types

Shifting focus on a visual scene without moving our eyes — think driving, or reading a room for the reaction to your joke — is a behavior known as covert attention. We do it all the time, but little is known about its neurophysiological foundation. Now, using convolutional neural networks (CNNs), UC Santa Barbara researchers Sudhanshu Srivastava, Miguel Eckstein and William Wang have uncovered the underpinnings of covert attention and, in the process, have found new, emergent neuron types, which they confirmed in real life using data from mouse brain studies.

“This is a clear case of AI advancing neuroscience, cognitive sciences and psychology,” said Srivastava, a former graduate student in the lab of Eckstein, now a postdoctoral researcher at UC San Diego.

Their findings are published in the Proceedings of the National Academy of Sciences.

Emergent properties and new neuron types

We think of attention as a spotlight or zoom in our brain that focuses on something in our visual field and devotes resources to that area, improving how we see it. Even more contemporary computational and neurobiological models have a built-in attention mechanism that turns on and changes visual processing (ups the volume,

or reduces the noise) at the attended location.

In the lab, scientists study covert attention by having a flash or arrow appear before or simultaneously with a briefly presented target and measuring how target detection gets faster and more accurate when it appears with the cue. The idea is that the cue orients the attention brain mechanism to the cued location, altering visual processing.

“Because it’s so natural to us humans, these covert attention behaviors seem deceptively and necessarily associated with some ability to move our awareness across the visual world,” said Eckstein, a professor of psychological and brain sciences at UCSB. For decades, this often split-second behavior was thought to be the sole domain of primates, made possible by the parietal lobes of our brains, and even associated with consciousness.

However, more recently, this behavior has been documented in other species as well, including archer fish, mice and bees — animals with simpler brain architectures. All of which has led the researchers to wonder if some types of covert attention could be an emergent phenomenon, the result of neurons across the brain working together, as opposed to the work of specialized attention brain modules.

But mapping how we process information in the brain, particularly how we optimize our attention for accuracy, is a difficult task. A human brain contains billions of neurons; it is a highly dynamic and variable system. Meanwhile, imaging techniques currently cannot reach the resolution necessary to measure individual neuronal activity.

However, we have the next best thing: artificial intelligence models. By building a relatively simple model of the brain and turning it on tasks the human brain typically performs, it is possible to peek under the AI’s hood and gain insight into how the brain may be organized to perform such tasks.

Such has been the case for the Srivastava, Eckstein and Wang, who in 2024 demonstrated how 200,000 to 1 million neuron CNNs — a very rudimentary version of a brain — exhibited the hallmarks of human covert attention when presented in various target detection tasks, though it had no built-in mechanism for orienting attention. In doing so, they proved that covert attention could be an emergent property of an artificial or biological organism learning to detect the target as best

as it can.

But what inside the CNN makes the emergent covert attention possible? The new paper dives deeper into that line of inquiry, investigating the inner workings of the CNN, how “the emergent neuronal mechanisms in CNNs give rise to the behavioral signatures of covert attention.”

“For this paper we thought we could analyze these convolutional neural networks instead of treating them like a black box,” senior author Eckstein said. “If you’re doing single-cell physiology, trying to isolate individual neurons, you can record activity from maybe thousands of neurons, but a million is not presently possible. But in the CNN, we could characterize every single of its units, and that might guide our understanding of the real neurons in brains.”

Taking a population of 1.8 million artificial neurons (180k neuronal units over 10 trained CNNs), the researchers put their AI’s “brain cells” to a Posner cueing task, a visual test that measures the accuracy or speed at which the participants detect the target when it appears with or without the cue (box or arrow).

Among their findings were CNN units that, despite lacking any built-in attention mechanism, parallel those reported by neurophysiologists in primate and mouse brains. Importantly, they found several CNN “neuron” types that had response properties never before highlighted. For example, most studies focus on how neurons are excited by attention, but they found units in the CNN whose response is diminished by the presence of the cue (“cue inhibitory”).

“The most surprising one is a ‘location opponent,”’ Eckstein said. According to the researchers, this neuron type is excitatory, boosting activity by the presence of the target and cue at one location, while suppressing activity at the other locations, in effect turning up the signal where the target was expected and dampening it in places it wasn’t expected to show up. While unheard of in the scientific understanding of covert attention, opponency cells are common in other areas of vision. For example, there are cells that get excited with red light but inhibited with green (color opponent) and other neurons that respond to things moving but inhibited by downward motion.

“It’s kind of a push-pull,” Eckstein explained. Studies of the effects of attention on neural activity, he added, tend to focus on excitatory responses, in which neurons

increase activity, so mechanisms that dampen activity can often go unnoticed.

Still, the researchers weren't sure whether the CNNs they were studying would have any correspondence to real biological neurons, so they dove into neural data from studies of mice during a cueing task. They found that in fact these location-opposing neurons did exist in the mouse superior colliculus (a structure in the mouse's midbrain), along with the other previously unreported neuron types implicated in attention, such as cue-inhibitory and location-summation neurons.

"These neurons might be one of a variety mediating emergent attentional behavior," Eckstein said.

Interestingly, one neuron type present in CNN, which combines opponency for the cue but excitatory summation for the target at both locations, was not found in the mouse, suggesting that there might be biological constraints that the AI might not have.

Just how far these findings can be stretched to apply to humans remains to be seen; the scientists are still at the early stages of this research arc. However, this work proves that there is far more to covert attention than previously thought. Not only have the researchers shown that there are emergent attentional behaviors, they also show that there are emergent neural mechanisms, and that CNNs can predict neural types with unique properties that have not been reported before.

"It fundamentally changed how we think about attention," Srivastava said. "So we'll see how these new concepts evolve through time."

Eckstein and Wang are deeply interested in the interface of human and machine intelligence, heading UCSB's Mind & Machine Intelligence Initiative (made possible by a gift from Duncan and Suzanne Mellichamp) to bring together people working at the intersection of AI and the study of the mind.

Tags
Artificial Intelligence

Media Contact
Sonia Fernandez
Senior Science Writer
(805) 893-4765
sonia.fernandez@ucsb.edu

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.