A conversation with Professor Won Mok Shim, August 7, 2019

Participants

- Professor Won Mok Shim - Associate Professor at the Centre for Neuroscience Imaging Research, Sungkyunkwan University
- Joseph Carlsmith - Research Analyst, Open Philanthropy

Note: These notes were compiled by Open Philanthropy and give an overview of the major points made by Prof. Shim.

Summary

Open Philanthropy spoke with Prof. Won Mok Shim of Sungkyunkwan University as part of its investigation of what we can learn from the brain about the computational power ("compute") sufficient to match human-level task performance. The conversation focused on information-processing in V1.

Function of V1

Traditional view

There is a traditional view of V1, on which it is the front end of a hierarchical information-processing pipeline, and is responsible for processing simple, low-level features of bottom-up visual input from the retina/LGN. However, many feedback processes and connections have been discovered in V1 over the last decade, and most vision scientists would agree that V1's information-processing cannot be entirely explained using bottom-up inputs.

For example, Prof. Shim and her collaborators have used fMRI data to show that V1 fills in gaps in an apparent motion stimulus -- an effect that cannot be explained using purely bottom-up inputs. There are also visual illusions in which, due to the perspective of the scene, objects that are the same size on the retina appear to the brain to be of different sizes. The representation in V1, however, reflects the perceived size. There are many possible explanations of this, but the traditional, feedforward story does not explain it.

The anatomy of the visual system also suggests an important role for feedback. For example, there are more feedback connections from V1 to the LGN, than there are feedforward connections from the LGN to V1. V1 receives a large number of connections from other brain areas, like V2, and there are also many lateral connections between cells within V1. The direction of these connections can be identified using neuroanatomical trace studies, mostly from monkeys or cats.

Alternative view

On an alternative to the traditional view, V1 is receiving top-down, high-level predictions, which it then compares with the bottom-up input. The difference between the two is an error signal, which is then conveyed from the low-level areas to the high-level areas. The origins of this idea are in computational theory (predictive coding). There is some empirical support as well, but the evidence is not completely clear.

People originally thought that different neural structures had to be responsible for the visual representation and the error signal, but people are also now considering the possibility that the same pyramidal neurons can be involved in both. These cells can collect input from all the layers of the cortex, including layers that receive long-range connections from other brain areas.

You may be able to put some numbers on what percentage of V1's anatomical organization is captured by the traditional story of feedforward processing. Putting numbers on the percentage of V1's *function* that this story captures, however, would be quite difficult.

Comparisons between deep neural networks and the human visual system

Convolutional neural networks are largely feedforward. As noted above, though, the human visual system involves many feedback connections, and we still don't really know what these feedback connections are doing. One possibility is that they convey the kind of top-down predictions discussed above.

Over a lifetime, you have accumulated a lot of domain knowledge, which would be encoded in these top down predictions. It's not clear to Prof. Shim how this domain knowledge could be incorporated into current deep neural networks.

Also, even with the exact same input, your visual system will be extracting different types of information depending on the goal/task. And the tasks you need to do in a naturalistic environment (e.g., searching for a certain object, trying to understand the body language or gestures of a stranger, etc), are very diverse.

Images per second

There is a fair amount of consensus in the field that the human visual system can recognize about ten images per second (e.g., one image per 100 ms). However, this doesn't mean that it takes 100 ms to recognize an image. For example, you might be able to recognize an image shown very briefly (e.g., for less than 100 ms), but without sequences of other images before and afterwards.

Bits per second receives from the retina

Estimates of the number of bits sent to the brain from the retina would need to incorporate the fact that the retina's inputs are not random -- rather, they reflect the statistics of visual scenes. What's more, the spiking patterns of retinal ganglion cells are not independent.

All Open Philanthropy conversations are available at
http://www.openphilanthropy.org/research/conversations