Proceedings of the 8th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'22)
Prague, Czech Republic - July 28- 30, 2022
DOI: 10.11159/mvml22.001

The Model 2.0: An Anatomically-Inspired Model of the Primate Visual System

Garrison W. Cottrell
University of California, USA

Abstract

Over the last thirty or so years, my lab has used variants of a relatively simple biologically-inspired neurocomputational model of face and object recognition (The Model™) to explain a number of behavioral, developmental, and neurophysiological phenomena. These results include, for example, fits to data supporting both the categorical and continuous theories of facial expression perception ("one model to rule them all"), a novel explanation of hemispheric asymmetries (local and global perception of hierarchical stimuli), and my favorite result, why the fusiform face area is recruited for other domains of visual expertise. Here, I report on some results of The Model 2.0, a deep version of The Model that includes a foveated retina, the log-polar mapping from the visual field to V1, sampling from the image via a salience map, and dual pathways from V1, central and peripheral. First, I describe some previously reported results on how The Model 2.0 can explain behavioral data in human scene perception under scotoma and tunnel vision conditions (Wang & Cottrell, 2017). Second, I provide a novel explanation of the face inversion effect. Contrary to the generally accepted wisdom that this occurs deep in the visual stream, our hypothesis is that the face inversion effect can be accounted for by the representation in V1 combined with the reliance on the configuration of features in face recognition.

The log-polar mapping, when used as input to a convolutional neural network (CNN), provides two kinds of invariances. Scale is just a left-right shift in this representation (see images of Geoff Hinton (top row) and their log-polar representation (bottom row)). Similarly, rotation in the image plane is an up-down shift. Because CNNs are (somewhat) translation invariant, the network as a whole becomes scale and rotation invariant. However, translation invariance is lost. We make up for this by sampling from the image at multiple points, just as humans use multiple fixations to recognize a face (Hsiao & Cottrell, 2008). I end by explaining the puzzle of why a network that is rotation invariant shows a face inversion effect.