Examining the Role of Parts in Object Recognition for Artificial Neural Networks

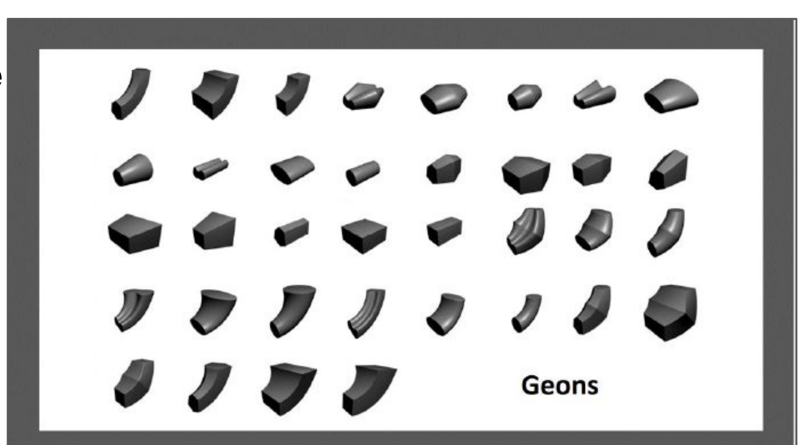
Hart Fogel,
Neuroscience
Vision Sciences Laboratory

Abstract

One ongoing area of inquiry within the realm of machine learning research is the extent to which biological brains and artificial neural networks use analogous methods to process visual information and recognize objects. Specifically, the question whether artificial neural networks rely on an object's constituent parts to classify it in a manner akin to the geon-mediated recognition-by-components model in humans remains largely unexamined. This thesis represents a first foray into this subject.

Background/Research Question

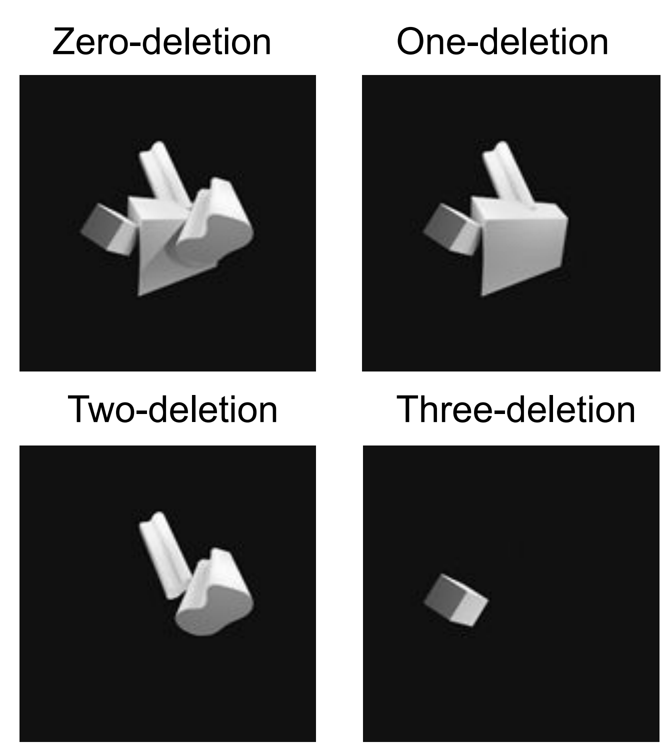
According to recognition-by-components (RBC) theory, humans rely a set of 36 simple geometric forms called geons in order to perform bottom-up object recognition. Artificial neural networks are inspired by the structure of the brain and display strikingly humanlike (or even superhuman) performance on certain visual tasks but respond unintuitively and ineffectively to others. How will these models fare on a challenge that assesses their ability to use geons to recognize objects?



Digital renderings of the 36 geons proposed by RBC theory. I used these to construct my experimental stimuli. (O'Brien, 2018).

Methods

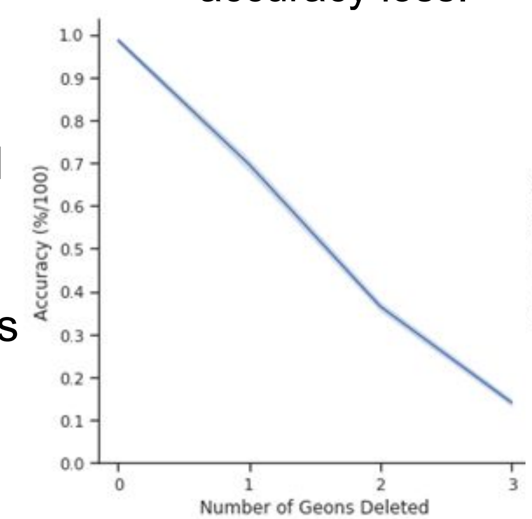
Using animation software, I generated 50 simple four-geon objects and rendered hundreds of thousands of images of them from many perspectives. Then, I trained seven artificial neural networks (five variants of AlexNet, a SE-ResNet 50 fixed feature extractor pre-trained on an object recognition dataset and a SE-ResNet 50 fixed feature extractor pre-trained on a facial recognition dataset) to recognize these stimuli and tested each algorithm on its ability to identify the objects as their individual parts were progressively removed. 55 human participants were given a similar task.

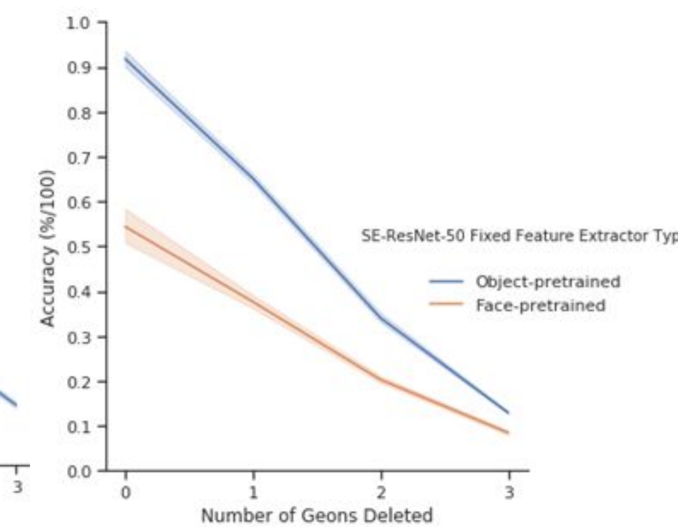


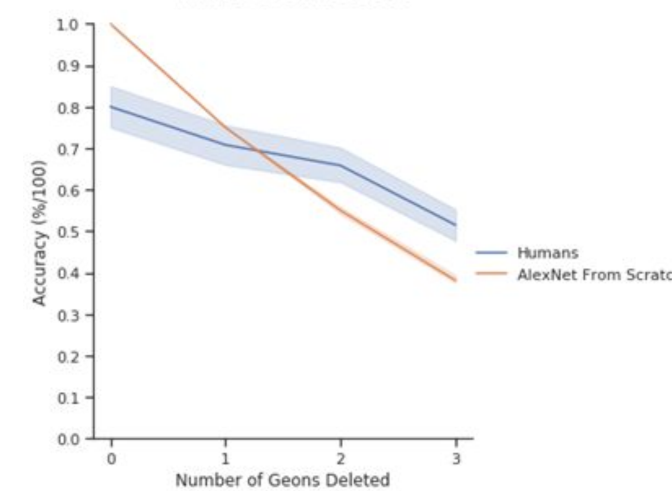
Example test images corresponding to one of the 50 objects that the neural networks were trained to recognize.

Key Results

The AlexNet variants suffered surprisingly gradual and linear accuracy loss in accordance with geon removal rather than immediately failing. The object-pretrained SE-ResNet-50 outperformed its face-pretrained counterpart on all test sets. Humans beat AlexNet on the harder test sets but displayed a similar overall pattern of accuracy loss.







(Clockwise) Performance of AlexNet trained from scratch, SE-ResNet-50 fixed feature extractor comparison and AlexNet versus humans on a subset of the stimuli.

Conclusion

There is reason to believe that with the right training, neural networks may be capable of humanlike component-based recognition. With improvement and augmentation, neural networks will likely be able to match or surpass human performance on the geon deletion task. The experimental paradigm established here helps pave the way for significant further exploration.