Image-grain comparison of core object recognition behavior in humans, monkeys and machines
Rishi Rajalingham*, Elias B. Issa*, Kohitij Kar, Kailyn Schmidt, Pouya Bashivan, James J. DiCarlo
Dept. of Brain and Cognitive Sciences1, McGovern Institute for Brain Research2, MIT

Introduction
Humans can rapidly and accurately recognize objects in spite of variation in viewing parameters (e.g. position, pose, and size) and background conditions. To uncover the algorithms underlying this ability, quantitative benchmarks of human behavior can be used to directly compare animal, neural and computational vision systems.

Previously, we showed that on one such benchmark, the pattern of object-level confusions, monkeys behavior was indistinguishable from human behavior. This pattern was also shared with high-performing object recognition models (convolution neural networks, CNNs) but not models of lower-level visual representations.

Here, we extended on our previous work by systematically comparing the core object recognition behavior of humans against that of monkeys and machines at the image-level using the pattern of image-by-image difficulties.

Methods

Behavioral paradigm: We tested object recognition performance on a two alternative forced choice (2AFC) match-to-sample task using brief (100ms) presentations of naturalistic synthetic images of basic-level objects.

Stimuli: To enforce true object recognition behavior, we generated several thousand naturalistic images, each with one foreground object (by rendering a 3D model of each object with random viewing parameters) on a random natural image background.

Each object can produce myriad images under variations in viewing parameters, with some images being more challenging than others.

High-throughput psychophysics: Over a million behavioral trials across hundreds of humans and five monkeys were aggregated to characterize each species on each image. To collect such large behavioral datasets, we used Amazon Mechanical Turk for humans, and a novel home-cage touchscreen behavioral system (“MonkeyTurk”) for monkeys.

Results

Comparing Humans and Monkeys:

Comparing Humans and Machines (AlexNet):

Behavioral Benchmark:

Our results show that monkeys are highly consistent with the pooled human in their pattern of image difficulties. In contrast, all tested CNN models were significantly less consistent with the pooled human.

Summary

- Our results show that monkeys are highly consistent with humans in their pattern of image difficulties, suggesting that they are a good animal model for studying human object recognition abilities.

- In contrast, all tested CNN models were significantly less consistent with humans at the image grain even though they passed at the object grain. This small gap in image-level consistency to humans could not be easily attributed to the viewing parameters (position, size, pose) of the object, classifier choice, images used in model training, or retinal sampling of images.

- Therefore, high-resolution, image-grain behavioral measures serve as stronger constraints than object-level measures for distinguishing between alternative models of human object recognition.