Comprehensive and Compact Perceptual Space of Visual Objects

Ha Hong[1,2*], Ethan Solomon[1,3*], Dan Yamins[1*], and James J. DiCarlo[1] (*equal contribution)

[1]Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, MIT, Cambridge, MA. [2]Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA. [3]Present address: School of Medicine, University of Pennsylvania, Philadelphia, PA.

We describe a novel approach to construct a perceptual space of visual objects and show that it robustly explains human behaviors across a wide range of visual perception tasks. We also demonstrate that the perceptual space bears a significant similarity to the neural representation in the macaque inferior temporal cortex (IT), a key visual area for object recognition, and has a relatively low dimensionality.

We measured human observers' confusion patterns in an object identification task with two-alternative forced-choice method (2AFC), where each observer saw a subset of an image set with 64000 images of 64 photorealistic objects placed in complex natural scene backgrounds with substantial variation in object position, size, and pose. We constructed a perceptual space that is optimized to match the observers' responses by modeling the image distributions of objects in the space with multi-dimensional Gaussians and the observers' object identification process with simple linear classifiers. The resulting perceptual space accurately predicted the confusion patterns in held-out dataset and these in a separate 8AFC object identification task. Moreover, with simple mathematical readout rules, it robustly predicted: the subjective ratings of the objects with adjectives; the pairwise similarity judgments for the objects in ImageNet image set; and the pairwise distances of objects in the IT neural representation recorded with our image set. Remarkably, the perceptual space outperformed control models in all these tasks. In addition, an extrapolation analysis suggested that ~47 dimensions would be needed to reliably embed 10000 objects in the perceptual space.

Our results show that the perceptual space is comprehensive and compact. Taken together, we speculate that the perceptual space and simple readout rules might reflect a biologically plausible mechanism of how the human brain represents visual objects and solves visual tasks.