Distributed population activity in the macaque inferior temporal cortex but not current deep neural networks predict the Ponzo illusion

Vivian C. Paulun1,2, Kristine Zheng1, and Kohitij Kar1,2

McGovern Institute for Brain Research, Department of Brain and Cognitive Science, and 2Center for Brains, Minds and Machines, MIT, Cambridge, MA. Contact: vmpaulun@mit.edu

**Introduction**

An object’s size can be approximated from the population activity across the macaque inferior temporal (IT) cortex (Hong et al., 2016). These neural predictions were highly consistent with human estimates of object size.

**Aims of this study**

- Does population activity in IT reflect the retinal or perceived size?
- Do deep neural networks (DNNs) as state-of-the-art models of the ventral stream reflect the same?
- Does population activity in IT reflect the retinal or perceived size?
- Do deep neural networks (DNNs) as state-of-the-art models of the ventral stream reflect the same?

**Stimulus set**

- Objects were placed near, far or randomly with respect to a linear perspective background. Control stimuli: same objects on a background with little perspective.
- 400 images, across 10 objects

**Population activity in IT reflects perceived size**

- Exhibits Ponzo illusion
- Higher consistency with perceived than with retinal size

**Feature activation in DNNs reflects retinal size**

- No Ponzo illusion
- Higher consistency with retinal than with perceived size

**Conclusions**

- Object-size readout from IT population activity exhibits the Ponzo illusion.
- Population activity in macaque IT reflects the perceived rather than retinal object size.
- Our results support a linear IT readout model of object size perception.
- DNNs trained on object classification, and depth estimation do not capture the Ponzo illusion.
- Our results demonstrate yet another explanatory gap in current DNNs as models of primate vision.