

Modeling human categorization of natural images using deep feature representations

Ruairidh Battleday

Helen Wills Neuroscience Institute; University of California, Berkeley, Berkeley, California, USA

Joshua Peterson

Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

Tom Griffiths

Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

Abstract: Over the last few decades, cognitive scientists have developed sophisticated formal models of human categorization, and computer vision researchers have achieved increasingly impressive performance in natural image classification. In this paper, we combine the strengths of these approaches, using representations from a convolutional neural network to evaluate cognitive models of categorization against >300,000 human judgments of natural images. We find that a prototype model performs best overall, and that an exemplar model performs best when the network's most abstract features are used. Altogether, our results demonstrate that the optimal categorization strategy over a set of stimuli is deeply linked to how they are represented, suggesting that any satisfying characterization of categorization behavior over naturalistic stimuli must consider it the result of a dual process of feature learning and strategy selection. The paradigm we present herein offers one avenue to begin this undertaking.