

## **UC Merced**

# **Proceedings of the Annual Meeting of the Cognitive Science Society**

### **Title**

Can deep convolutional networks explain the semantic structure that humans see in photographs?

### **Permalink**

<https://escholarship.org/uc/item/9s283743>

### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

### **Authors**

Suresh, Siddharth

Huang, Wei-Chun

Mukherjee, Kushin

et al.

### **Publication Date**

2024

Peer reviewed

# Can deep convolutional networks explain the semantic structure that humans see in photographs?

**Siddharth Suresh**

University of Wisconsin-Madison, Madison, Wisconsin, United States

**Wei-Chun Huang**

University of Wisconsin-Madison, Madison, Wisconsin, United States

**Kushin Mukherjee**

University of Wisconsin-Madison, Madison, Wisconsin, United States

**Timothy Rogers**

University of Wisconsin-Madison, Madison, Wisconsin, United States

## Abstract

In visual cognitive neuroscience, there are two main theories about the function of the ventral visual system. One suggests that it serves to classify objects (H1); the other suggests that it generates intermediate representations from which people can generate verbal descriptions, actions, and other kinds of information (H2). To adjudicate these, we trained two deep convolutional AlexNet models on 330,000 images belonging to 86 classes, representing the intersection of Ecoset images and the semantic norms collected by the Leuven group. One model was trained to produce category labels (H1), the other to generate all of an item's semantic features (H2). The two models learned very different representational geometries throughout the network. The representations acquired by the feature-generating model aligned better with human-perceived similarities amongst images, and better predicted human judgments in a triadic comparison task. The results thus support H2.