

Summary of “Natural Scene Statistics... in Human Visual Cortex”

Matt Dickenson mcd31

STA571/CS590.01

Due: 9 April, 2014

Stansbury et al. (2013) explore the question, “does the human brain use scene categories to aggregate information about objects?” In doing so, this paper extends theoretical models of human cognition into the realm of prediction under novel conditions. The paper consists of a series of steps that, taken together, suggest that humans use scene categories to represent relationships between visual objects.

The first step in this research process was to create a database of natural scenes with the categories. To do this, Stansbury et al. labeled scenes with objects that appeared in them. These labels were used as input to a Latent Dirichlet Allocation (LDA) model. LDA was performed with $k = 2 \dots 40$ categories, with the vocabulary size ranging from the 25 to the 950 most frequent objects (e.g. car, fish, building). LDA assigned to each scene the probability of belonging to each of the k categories. These category probabilities were then used to construct voxelwise encoding models.

In the next stage of this research, four human subjects each viewed 1,260 natural scene images. While viewing the images, fMRI recorded subjects’ blood oxygenated level-dependent (BOLD) activity. The weights from the voxel encoding models were used to predict the response of the subjects to novel stimulus scenes. The highest correlation with subjects’ response was obtained with encoding models based on 20 categories with a vocabulary of 850 objects (although predictive accuracy is obtained across a wide range of parameter settings). The encoding models based on LDA categories performed significantly better than null models. This suggests that BOLD responses convey information about scene categories.

How does the BOLD activity map to brain regions discovered in previous fMRI studies? The authors discovered that voxels within PPA (previously suggested to respond to the presence of buildings) have large weights on the “Urban/Street” category. OFA is thought to be selective for human faces, and voxels within OFA have high weight on the “Portrait” category. Not all evidence corresponds to prior research, however. For example, voxels within OFA also score highly on the “Plants” category.

In the final stage of this research project, the authors explore whether voxel activity can be used to predict categories in novel images. To do this, they construct a decoder to predict 20 category probabilities for 126 novel scenes. The correlation between subjects’ responses and LDA categorization ranged from about 0.1 (in the worst case, for subject 4) to 0.25 (in the best case, for subject 1). These results suggest that humans use categories to organize visual information about scenes in the natural world.