Evidence for Visual Representation of Numerosity in Natural Scenes

Maggie Mae Mell (mell@musc.edu)
Ghislain St-Yves (stayves@musc.edu)
Department of Neuroscience, 173 Ashley Avenue
Charleston, SC 29425 USA

Emily J. Allen (prac0010@umn.edu)
Department of Psychology, 75 E River Rd
Minneapolis, MN 55455 USA

Yihan Wu (wuxx0468@umn.edu)
Kendrick Kay (kay@umn.edu)
Department of Radiology, 2021 6th Street SE
Minneapolis, MN 55455 USA

Thomas Naselaris (tnaselar@musc.edu)
Department of Neuroscience, 173 Ashley Avenue
Charleston, SC 29425 USA

Abstract
In visual cortex of human and non-human primates, high-level visual areas near intraparietal sulcus have been shown to explicitly encode the number of objects in visual displays. To date, evidence for this numerosity code has come from experiments that use simple dot-like visual stimuli, raising the question of whether the numerosity code persists during perception of natural scenes. Here, we assessed evidence for a numerosity code in high-resolution fMRI measurements of responses to thousands of natural scenes in 3 human subjects. We constructed an encoding model that predicted voxelwise responses as a function of local object counts in each natural scene. Our model was able to accurately predict voxelwise activity in visual cortex. To test if local object counts were acting as a proxy for simple low-level image features, we constructed voxelwise encoding models based on Gabor wavelet filtering of the natural scenes. For voxels in anterior visual cortex, the numerosity encoding model generated more accurate predictions than the Gabor model. These results offer preliminary evidence for a numerosity code in anterior visual cortex during natural scene stimulation.

Keywords: fMRI; encoding models; natural scenes; numerosity

Introduction
Humans and non-human primates are able to quickly estimate the number of objects in visual space. Previous studies have suggested that this ability is linked to maps in parietal cortex that are tuned to the number of objects in simple dot-like displays (Harvey, Klein, Petridou, & Dumoulin, 2013; Nieder & Miller, 2004; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Tudusciuc & Nieder, 2007). However, in the natural world we are rarely presented with cleanly segmented and separated objects. In contrast, natural scenes contain many different objects with varied shapes, sizes, occlusion, etc. It is not known if the numerosity representation found in previous studies persists when viewing complex natural stimuli. In this paper we developed a numerosity-based voxelwise encoding model to explore the representation of numerosity in the human brain in response to natural scene stimulation. We present preliminary evidence for a numerosity representation that appears to be distinct from low-level, wavelet-like features, but is subsumed by more complex feature representation in a performance-optimized deep neural network.

Figure 1: Experiment. fMRI BOLD responses were recorded in response to thousands of natural scene images. Model. The Numerosity Encoding Model was constructed by calculating local object counts for every natural scene image. Object counts were transformed into one-hot maps for numerals 1-9 and 10+. This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0
高分辨率全脑fMRI BOLD测量是在人类受试者对自然场景图像的响应下获得的（图1顶部）。每个受试者被呈现了9000张唯一的图像和1000张图像，这些图像在所有受试者中共享，总数为10000张图像。每张图像被显示3次，非连续显示，总共30000次刺激试验在40次单独的运行中呈现。图像持续时间为3秒，后接1秒的刺激间隔。受试者被指示专注于一个中心的固定点，并按下按钮指示该图像是新的还是重复的。本文展示了N=3名受试者的数据。进一步的数据收集正在进行中。

编码模型

三个独立的编码模型被用于预测像素响应。所有模型都使用了基于特征权重的感知机的岭回归方法（St-Yves & Naselaris, 2018）进行训练。简而言之，fwRF是基于每个体素的编码模型，估计刺激生成的特征映射的感知机的中心位置和大小。fwRF根据特征权重和特征池化字段生成对视觉刺激的脑活动预测。特征池化字段表示每个体素的活动最依赖于视觉空间的区域。相同的特征映射被用于所有体素，但是分配给每个特征的权重将不同，表示每个体素中编码的特征。fwRF中心和大小，以及特征权重的值使用岭回归估计。三个不同的特征映射集被用于生成三种脑活动模型。

数量编码模型

为了测试数量编码，我们构建了一个编码模型，该模型作为本地物体计数的函数预测像素响应（图1底部）。我们基于全景场景分割的公开图像（从COCO数据集（Lin et al., 2014）中生成的图像）。分割图像是被裁剪并下采样到128x128像素。每个像素的物体计数是基于该像素周围128x128像素的唯一分割映射计算的。一个热映射图用于每个数目的1-9和10+，生成80个特征映射。一个热映射图是通过判断每个像素周围的空间频率带的对比度和功率变化来确定本地物体计数的。对于包含3个和0的皮质区域，以及所有其他为该窗口大小的热映射图。将像素编号为1的热映射图与数目的3和0对应。

交替编码模型

为了对比，我们构建了两种基于Gabor小波的交替编码模型和来自深度神经网络的特征映射（Krizhevsky, Sutskever, & Hinton, 2012），分别。

预测准确率和交叉验证

所有编码模型在大约80%的响应中进行训练。大约20%的训练集被留作优化岭超参数以及fwRF位置和大小的选择。剩下的约20%用于交叉验证。预测准确率是模型预测和测量脑活动之间的皮尔逊相关系数。
local number of objects in natural scenes, and that this representation cannot be entirely accounted for by simple low-level image attributes.

A recent study (DeWind, 2019) suggests that numerosity representations can arise “spontaneously” in feature maps of deep convolutional neural networks (DCNN) trained to categorize objects (Krizhevsky et al., 2012). This result suggests that the Numerosity encoding model should exhibit no advantage in prediction accuracy over an encoding model based on the feature maps of an optimized DCNN. Indeed, the DCNN-based encoding model outperforms the Numerosity encoding model for all voxels (Figure 3). This result clearly demonstrates that a representation of the local number of objects in natural scenes is only a subset of the representations maintained in all visual areas.

**Figure 3: Prediction Accuracy Model Comparison.** The joint distribution of prediction accuracy (Pearson correlation between predicted and measured brain activity) for Deep Convolutional Neural Network (DCNN)-based encoding model (x-axis) and Numerosity based encoding model (y-axis). As in Figure 2, voxels are binned into hexagons and colored by average coronal slice location. The numerosity model is completely subsumed by the DCNN model, regardless of slice location, suggesting numerosity may be a feature generated by DCNNs. Data from Subject 1 only.

**Conclusion**

Our results offer evidence for a representation of numerosity in the human visual cortex during natural scene viewing. This representation appears to be distinct from low-level visual features and maintained primarily in anterior parts of the visual system, in line with previous studies in humans and non-human primates (Harvey et al., 2013; Nieder & Miller, 2004; Piazza et al., 2004; Tudusciuc & Nieder, 2007). Like many other potentially useful and behaviorally relevant representations, the particular representation of numerosity built into our encoding model appears to be subsumed by the features encoded in a performance optimized DCNN. How or if the numerosity representation studied here is utilized to guide behavior or cognition will be an interesting topic for future exploration.

**Acknowledgments**

NIH R01 EY023384  
NSF IIS-1822683

**References**