Neural Coding and Scene Statistics

By Yue Xu[1,2] and Tai Sing Lee[1,2]
[1]Center for the Neural Basis of Cognition and [2]School of Computer Science, Carnegie Mellon University

Introduction:

Problem: From the perspective of image statistics, what are the logics behind sparseness in V1 population response?

Observation: A large percentage of **V1** neurons of a macaque monkey are **highly selective** to local complex patterns. Only **0.5%** of the 1225 neurons in a specific layer within a cortical hyper-column responded strongly to any **visual stimulus** (Fig 1).[1]

Hypothesis: The more frequently occurring visual stimuli are encoded by more sparse and in general weaker responses in neuronal population (red curve in Fig 2), and the less frequent are encoded by more distributed responses (gray curve in Fig 2).

Data description: In **V1** of an awake macaque monkey, **1225 neurons**' responses were recorded using two-photon calcium imaging. The neurons' receptive fields covered the same center area of the tested **2250 natural scene images** (Fig 3).

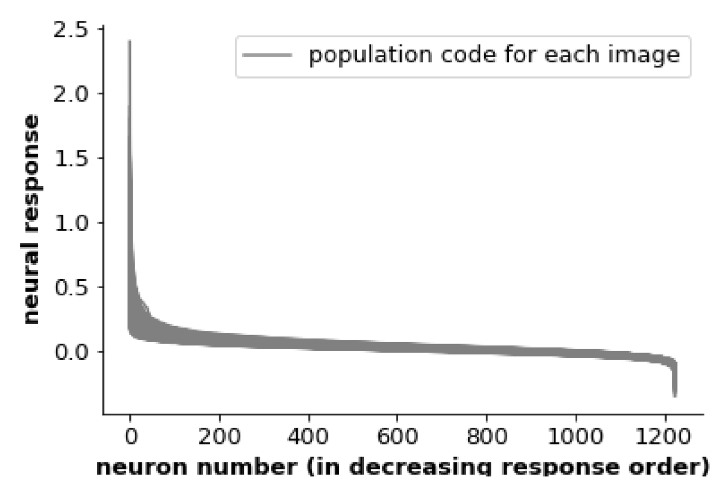

Fig 1. Observation: The population code for each image is very sparse.

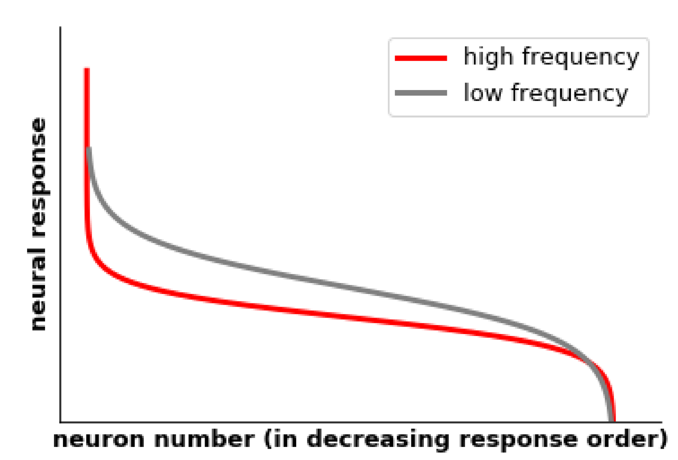

Fig 2. Hypothesis demonstration.

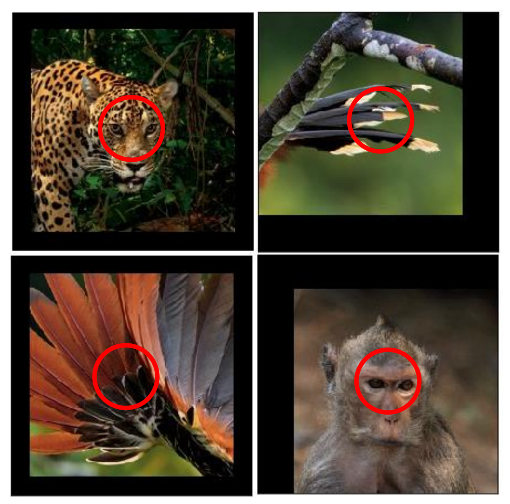

Fig 3. Data: example tested natural scene images. Red circles indicate receptive field areas of 1 visual degree.

Methods:

Goal: To study the correlation between the **sparsity** of neuronal population response and **occurring frequency** of tested visual stimuli with same **image complexity**.

To measure **occurring frequency:**

The tested images are encoded by **feature vectors** extracted from three layers of the **VGG19** deep neural networks, so as all 1000-class, 1.2 million **Imagenet** images (Fig 4.2). The feature vector columns used are directly encoding each image's area of the neurons' receptive fields (Fig 3) . Similarity between a visual stimulus and an Imagenet image is measured as the **Euclidean distance** of their feature vectors.

Frequency of a visual stimulus is the number of Imagenet images with distances smaller than a defined threshold value.

Assumption and justification:

- It's impossible to obtain the exact occurring frequency of each tested image in nature. The **Imagenet** images are assumed to be a representative sample of all images in nature.
- VGG19 **"Conv3_1"** layer provides a good model to predict V1 neural responses using transfer learning.[2]
- The receptive field size of a **"Conv3_1"** layer's unit is similar to that of a V1 neuron (Fig 4.1).
- The "**Conv5_4**" and "**Pool5**" layers are chosen as control layers because both layers provide good measurements of image similarity for classification purpose.

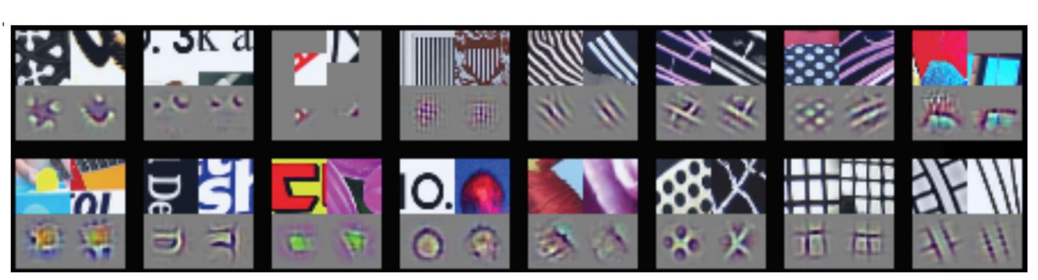

Fig 4.1 The visualization of some Conv3_1 units with the images that activate the unit most in the first row and the image features that caused the activations in the second row. [2]

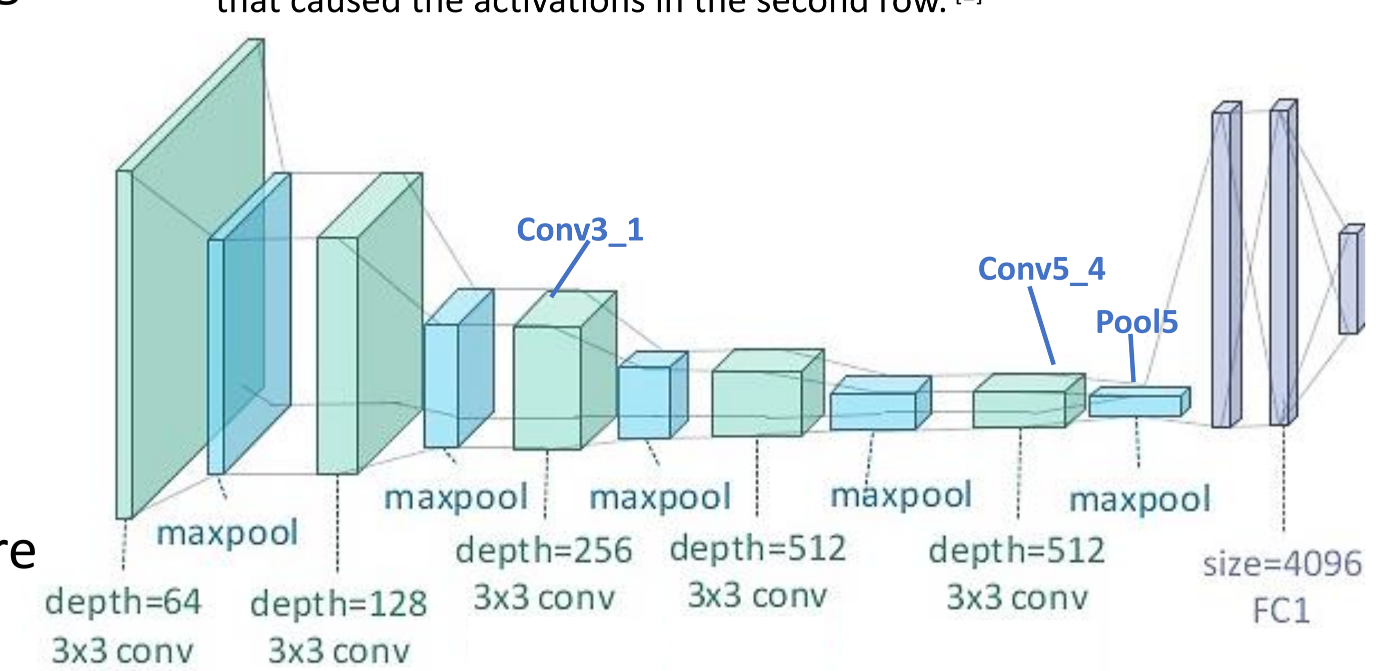

Fig 4.2. The three layers used in VGG19 neural network. (photo credit to Mark Chang from SlideShare)

Results:

1) The distribution of tested images' sparsity and layer "**Conv3_1**" frequency shows a correlation of **0.30,** as the baseline (Fig 5).
2) For each complexity range, the correlation between "**Conv3_1**" frequency and population response sparsity is above baseline, with p-value << 0.05 (Fig 6).
3) The neuronal population responses are generally weaker for more frequently occurring visual stimuli (Fig 7).

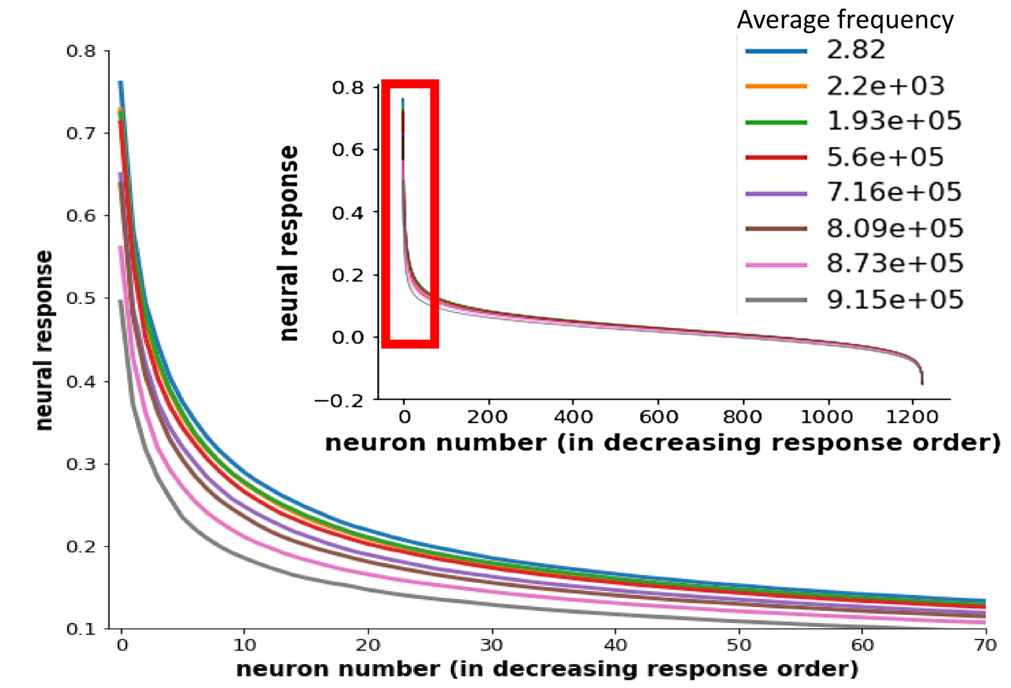

Fig 7. Results (3). Each colored line represents 300 neurons within the same frequency range.

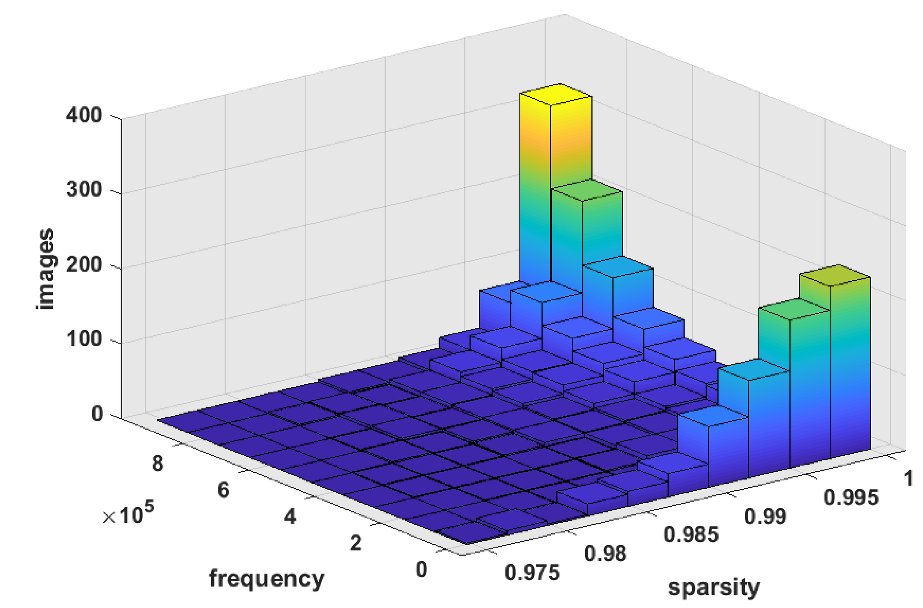

Fig 5. Results (1): 3D distribution plot of "Conv3_1" frequency vs sparsity with baseline correlation = 0.30.

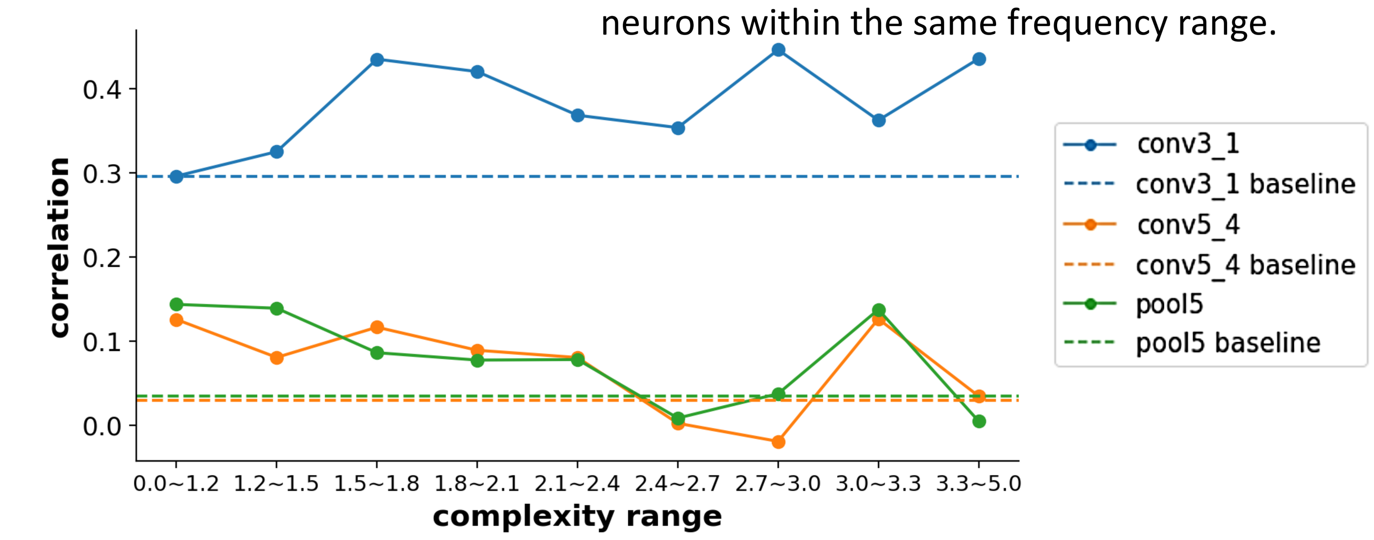

Fig 6. Results (2): Correlations between sparsity and frequency are above baseline and significant only for frequencies calculated using Conv3_1.

Discussion:

- The more frequently occurring visual stimuli are more common, so they are less informative for object recognition. Therefore, shown in the results, neuronal population responses are **weaker and more sparse** to visual patterns with higher frequency (Results (3)), resulting in a **positive correlation** between frequency and sparsity (Results (1,2)).
- However, Results (3) contradicts with part of the **hypothesis** . It doesn't show that more frequent visual stimulus has a few very high neural responses, i.e. not having neurons dedicate to encode this frequently occurring pattern.

Future work:

- The measurements of sparsity and complexity can be refined.
- The logics behind population response sparsity can be studied from the perspective of a visual stimulus' significance in object recognition, by comparing it to **visual concepts** (visual features that are useful for image classification).

Methods (continued):

To measure **sparsity:**

Sparsity of the population response of a visual stimulus is one minus percentage of the neurons that each responded above **half of its overall maximum** response.

To measure **complexity**:

Each visual stimulus is convolved with a set of **Gabor filters** to get a response vector. **Complexity** is measured as the L2-norm of the response vector.

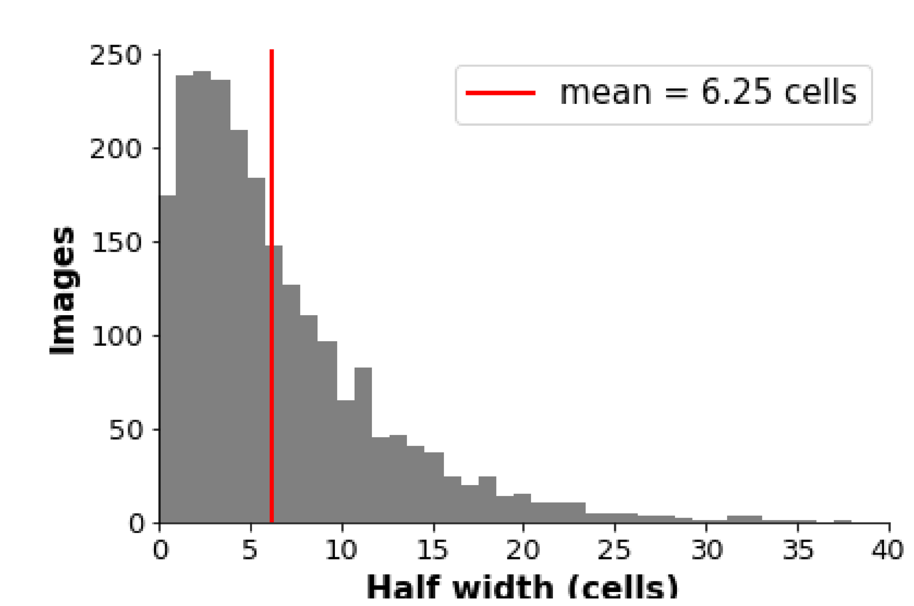

Fig 8. Sparsity measure: In average, 6.25 neurons responded above their half-maxima for each image.

References:

[1] Tang, S., Zhang, Y., Li, Z., Li, M., Liu, F., Jiang, H., Lee, T. S. (2018). Large-scale two-photon imaging revealed super-sparse population codes in the V1 superficial layer of awake monkeys. *eLife 2018;7*:e33370.

[2] Zhang, Y., Lee, T. S., Li, M., Liu, F., Tang, S. (2018). Convolutional neural network models of V1 responses to complex patterns. *J Comput Neurosci:* 1-12.

Acknowledgement:

The author would like to thank Dr. Lee for advising and supporting the entire project, and Dr. Tang for providing the data. The author would like to thank various members of the Lee lab for commenting and providing technical help. And the author would like to acknowledge the funding and guidance provided by the UPNC fellowship of the CNBC.