Hierarchical sparse coding of objects in deep convolutional neural networks

Presented by Hal Rockwell 8/8/20

Overview

- Distributed vs. sparse vs. local coding strategies

- The paper evaluates sparsity of Imagenet-trained CNNs, and compares them to the brain

- Also looks at how sparsity varies with depth and classification accuracy

Population sparseness index

$$\mathrm{PSI} = \frac{1-a}{1-\frac{1}{N_u}}, \text{ where } a = \frac{\left(\left(\sum r_u\right) / N_u\right)^2}{\sum\left(r_u{}^2 / N_u\right)},$$

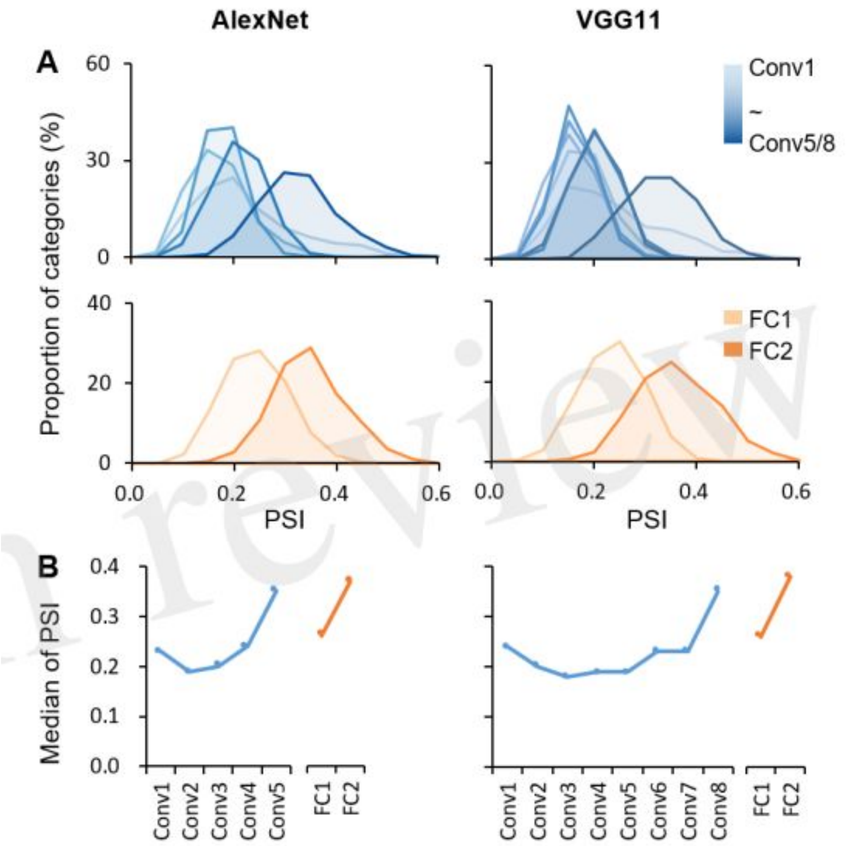
- Equivalent to fraction of units participating in coding in the case of binary responses; from Vinge and Gallant 2000

- 1 is maximally sparse, 0 is minimally sparse

- Requires nonzero activations, so they’re z-scored then normalized to 0-1 range (a little weird)

Sparsity and depth

- PSI is evaluated separately for each object category
- Increasing pattern is significant by Kendall's tau
- Comparable to PSI of 0.36 in macaque extrastriate cortex, according to a paper they cite

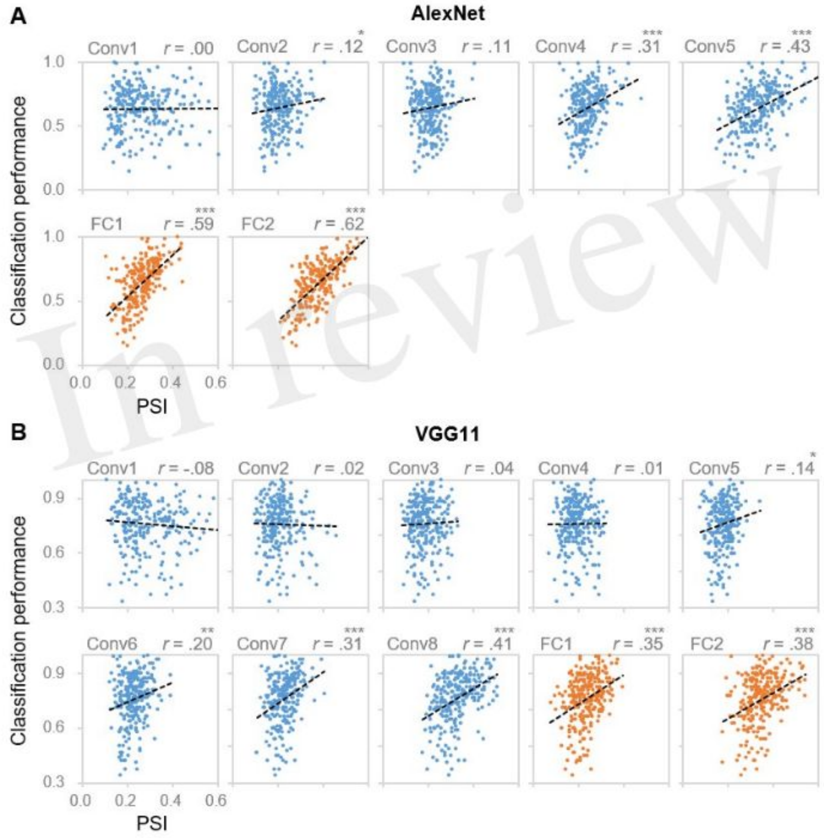


Sparsity and performance

- Significant correlations (across object categories) from middle on, increasing by Kendall’s tau

- Weaker in VGG11, though they don’t highlight this

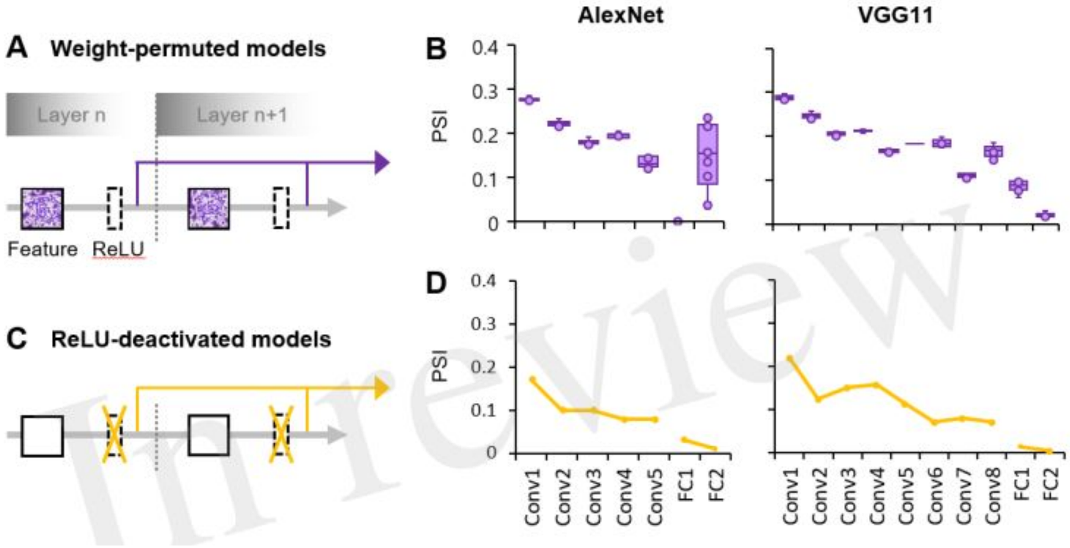
- Multiple regression of both FC layers in Alexnet gives R^2=0.52



Checks ruling out simple causes

- Shuffling weights or removing nonlinearities destroys the findings

- Mostly; the actual sparsity levels are still moderately high for the weight-permuted ones



Conclusions/Complaints

- Overall pretty cool, interesting results and very straightforward
- Would this hold for more modern CNN architectures?
- What explains the dip before the rise in sparsity, and does it correspond to something biological?