Single-kernel models of single-voxel visual selectivities in convolutional neural networks

Daniel D Leeds, PhD; Ivan Iotzov
Fordham University - Computer and Information Science Department

FORDHAM UNIVERSITY
THE JESUIT UNIVERSITY OF NEW YORK

COMPUTER AND INFORMATION SCIENCES DEPARTMENT

Background

Visual object perception in the human brain is understood to employ a network of brain regions selective for increasingly complex visual properties. Beyond simple visual properties in primary visual cortex (V1), **the nature of more complex visual properties encoded in the brain is unclear.**

Recent studies have illustrated the power of computer vision models, and particularly **Convolutional Neural Networks (CNNs)**, to **predict cortical region responses to visual stimuli**. (Yamins 2014; Leeds 2013). However, selectivities of individual neural populations within each region require further study.

Building on Guclu (2015), **we explore the ability of single artificial CNN "neurons" to model single fMRI voxel selectivities for mid-level visual properties.**

Methods: Data collection and model definition

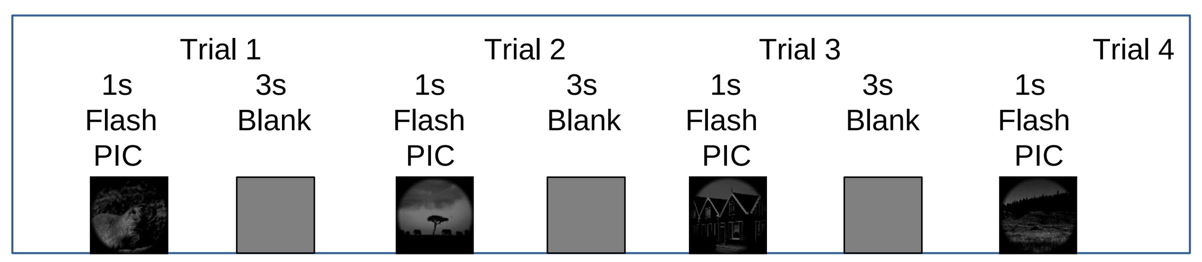
Stimuli and fMRI data were obtained from Kay (2008) and Naselaris (2009)
Their methods are summarized in the first three section below:

Stimuli

1750 grayscale images
Photographs of scenes & objects

Presentation

1s display: stimulus flashed 3 times
Every eighth trial was null



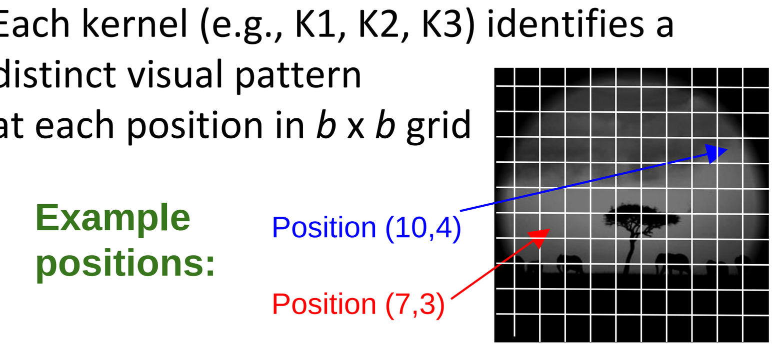
Cortical fMRI data

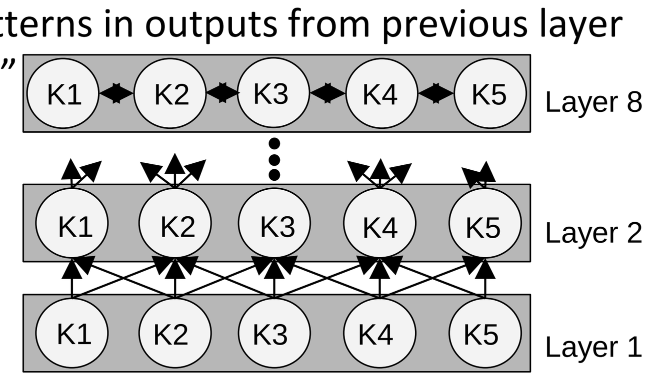
2 subjects. fMRI coverage of posterior cortex: ventral and dorsal visual pathways.
Voxels 2x2x2.5mm

Visual processing model

We used Caffe implementation of Convolutional Neural Network (CNN; Jia et al. 2014)
Network composed of 8 layers, each layer finds patterns in outputs from previous layer
Each layer consists of artificial neurons, or "kernels"

Each kernel (e.g., K1, K2, K3) identifies a distinct visual pattern at each position in *b* x *b* grid





Layer 2 kernel responses are computed for all 1750 stimuli

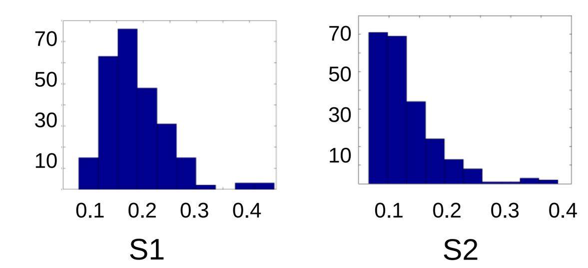
Methods: Voxel-model comparisons

Each of the 256 CNN **Layer 2 kernels serve as candidate models for mid-level visual properties.** All stimuli are divided into a 13x13 grid and a kernel response is computed at each position for each kernel. To test the correspondence of a candidate kernel to visual encoding in one cortical location (one voxel), **we find the correlation between the kernel's responses to the 1750 stimuli with the voxel's responses to the same 1750 stimuli.**
Correlations were computed separately for each image position across the 13x13 grid. Maximum (positive or negative) correlation across grid positions is typically reported below. We focus on $|r|>0.3$ as significant ($p^* \ll 1e\text{-}5$). We also study effect of image location on correlation is studied.

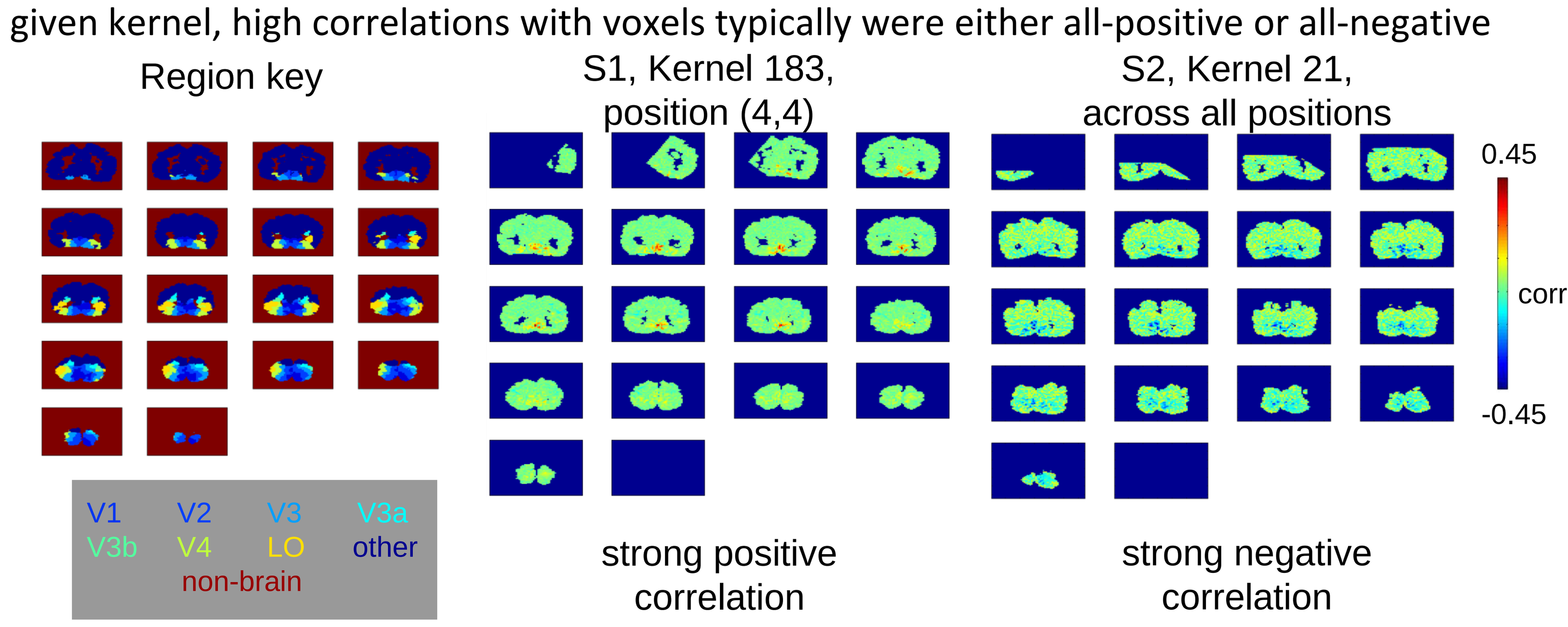
Results: Distribution of kernels-voxel correlations

47% of kernels in S1 had significant voxel correlations ($|r|>0.3$), and 28% in S2.
The number of high correlations ranged from 1 voxel per kernel to 662 voxels per kernel.
"Near top correlations": For each voxel, we sort absolute correlation values $|r|$ for all voxels and find the 100th largest $|r|$ value as the "near top" correlation. **The median "near top" correlations across 256 kernels are 0.18 for S1 and S2.**

Distribution of "near top" correlations

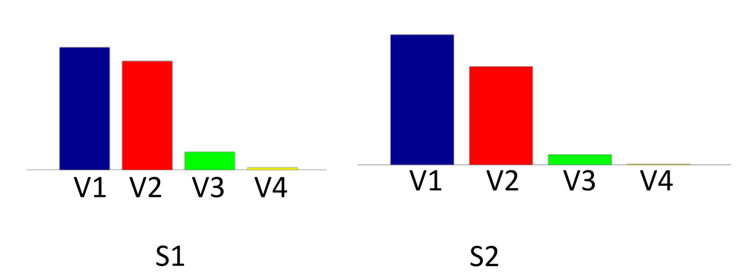


Significant correlations were found to be both positive and negative. For a given kernel, high correlations with voxels typically were either all-positive or all-negative



The vast majority of high correlations are found in V1 and V2, 49%/43% for S1 and 54%/41% for S2, respectively. Smaller amount of high correlations located in V3, and very small proportion in V4.
No high correlations found outside visual areas.

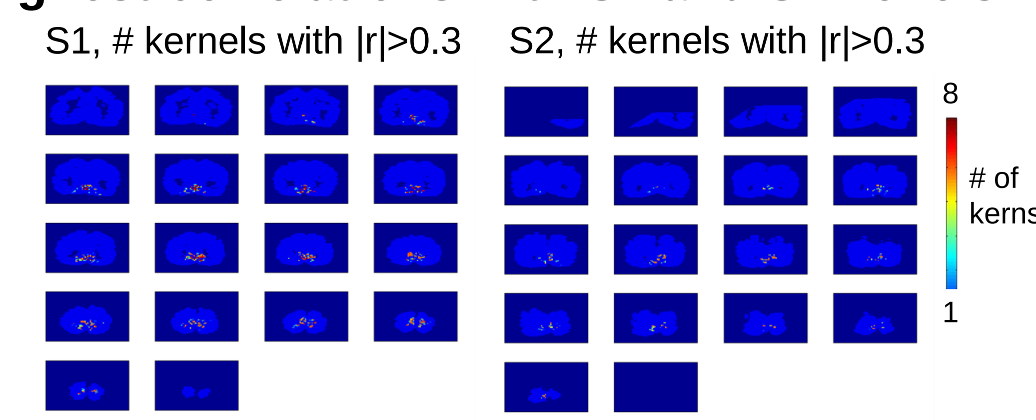
Distribution of $|r|>0.3$ across all voxels



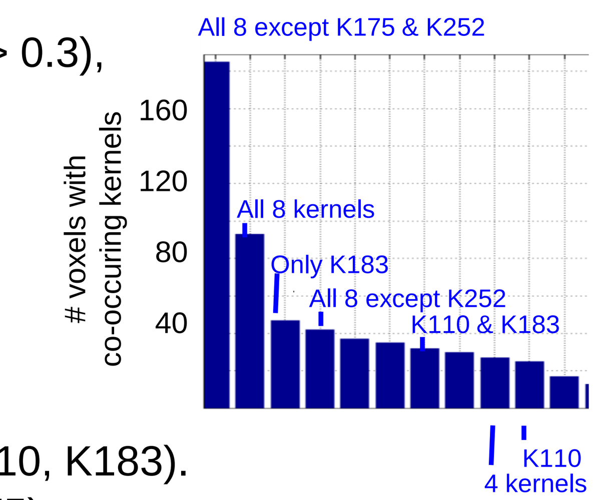
Results: Cross-subject comparisons, kernel overlaps

- **Same kernels have highest correlations with S1 and S2 voxels**

S1 kernel	S1 Corr	S2 kernel	S2 Corr
K59	0.45	K163	0.38
K163	0.43	K59	0.36
K82	0.43	K82	0.34
K183	0.40	K252	0.34
K252	0.39	K183	0.32

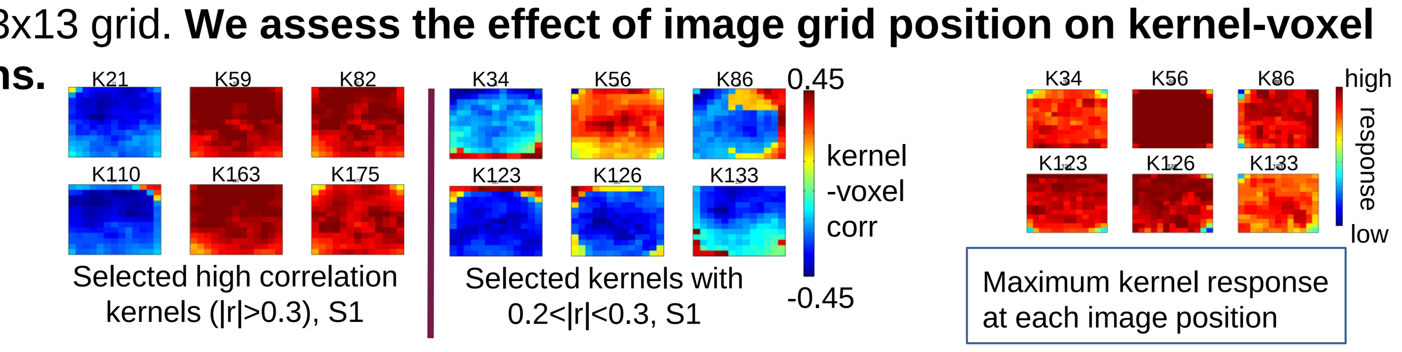


- Looking at kernels with strongest correlations (8 kernels for S1, 6 for S2; near-top correlation > 0.3), **most voxels with $|r|>0.3$ for at least 1 kernel correlate with most of the other strong-correlation kernels too.**
- For S1, **only two kernels have high correlations with 20+ voxels where <u>no other</u> kernels correlate equally** (K110, K183). Only one kernel has a unique match for S2 (K175).



Results: Effects of kernel image-position

Both CNNs and biological neural populations are sensitive to the position of visual patterns in an image. Each CNN Layer 2 kernel is applied to 169 image positions across a 13x13 grid. **We assess the effect of image grid position on kernel-voxel correlations.**



- **Top-correlation kernels** relatively **invariant to image position** for voxel correlation
- **Lower-correlation kernels show dependence on image position** for voxel correlation strength, with strongest correlations near borders of image
- Position-based voxel correlation effects not reflect position-based changes in kernel responses across stimuli

Discussion

- 25%-50% of individual Layer 2 CNN kernels can serve as models for individual voxels - **more than 50% of kernels are relatively weak biological models, despite strength of whole Layer 2 as model**
- Strong inverse correlations for several kernels indicate **voxel activities may be inhibited for complex kernel-relevant patterns**
- Typically many single kernels each model the same voxel, indicating **potential functional redundancy among Layer 2 kernels**
- Position of image analysis affects correlations for weaker Layer 2 kernels - **possible inflated effects of circle stimulus boundary for some supposed voxel-kernel links**

Contact

Daniel D Leeds
Fordham University
dleeds@fordham.edu
storm.cis.fordham.edu/leeds

Ivan V Iotzov
Now at: City University of New York
ivaniotzov@gmail.com

References

1. Güçlü, Umut, and Marcel AJ van Gerven. "Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream." The Journal of Neuroscience 35.27 (2015): 10005-10014.
2. Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.
3. Kay, Kendrick N., et al. "Identifying natural images from human brain activity." Nature 452.7185 (2008): 352-355.
4. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
5. Leeds, Daniel D., et al. "Comparing visual representations across human fMRI and computational vision." Journal of vision 13.13 (2013): 25-25.
6. Naselaris, Thomas, et al. "Bayesian reconstruction of natural images from human brain activity." Neuron 63.6 (2009): 902-915.
7. Yamins, Daniel LK, et al. "Performance-optimized hierarchical models predict neural responses in higher visual cortex." Proceedings of the National Academy of Sciences 111.23 (2014): 8619-8624.

Acknowledgements

This work was supported in part by funds to Daniel Leeds from the Fordham University Faculty Research Grant and funds to Ivan Iotzov through the Fordham University Undergraduate Research Grant.