Data-driven clustering of neural responses to a large set of natural images

James Campbell[1], Zijin Gu[1], Keith Jamison[2], Mert Sabuncu[1], Amy Kuceyeski[2]

[1]Cornell University, Ithaca, NY
[2]Weill Cornell Medicine, New York, NY

Introduction

One way to understand the brain is as a function mapping, wherein external stimuli are transformed into activation patterns. Accordingly, much experimental work has been done in presenting subjects with stimuli, such as images or auditory input, and observing subsequent neural responses through functional MRI (fMRI). Images act as a particularly powerful choice of stimulus as they represent high-dimensional data consisting of both semantic and morphological content, and therefore enable exploration and modelling of the whole brain more generally (Gu *et al.,* 2021) and of the visual regions in particular (Schrimpf *et al.*, 2020). Now, if we regard the neural response to an image as a vector, where each entry represents the activation of a particular region, then a natural question might be: how might we characterize the space of all such vectors? Our approach in this work is to employ data-driven k-means clustering to identify and describe the qualitative properties of images that give rise to characteristic clusters of brain activation patterns.

Methods

This investigation utilized the Natural Scenes Dataset (NSD) (Allen *et al*., 2021), which consists of 7T whole-brain, high-resolution fMRI of eight healthy subjects, collected over a year's time (30-40 MRIs, 60 mins each). Each subject was shown 9,000-10,000 images from the Microsoft Common Objects in Context (COCO) dataset (Lin *et al.,* 2014), resulting in 22,500-30,000 trials (with repetition of images). The fMRI data underwent post-processing, including motion correction, artifact distortion, and eddy motion correction. Moreover, a generalized linear model was used to estimate the voxel-level activation induced by each image; regional activation values were found by averaging the voxel-level beta maps in each region's mask, where regions were determined with the help of a functional localizer. To clarify, our data consisted of 22,500-30,000 neural response vectors of dimension 23-28, depending on the subject (some visual regions were missing in some individuals). We then employed k-means clustering with correlation distance (implemented with Lloyd's algorithm) to identify groups of similar brain responses across all individuals and images. Setting the elbow criterion as an R^2 gain of less than .01, the optimal value of k was determined to be six. Upon identification of each cluster's centroid, we found the images corresponding to the 10 and 100 closest points to them (still using correlation distance). Finally, using the COCO API (Lin *et al.,* 2014), we pulled the classification labels for each image and aggregated them for each cluster.

Results

By qualitatively examining the top 10 and 100 images closest to the centroid for each cluster, we can ascertain markedly distinct themes amongst the clusters. Specifically, we consistently find that the clusters divide into images of 1) food/high frequency patterns, 2) bodies/outdoor scenes, 3) faces, 4) sky/landscapes, 5) animals, and 6) downward-facing/indoor scenes. We confirm these observations quantitatively by counting the labels of the images (e.g. there are around 50 times more instances of broccoli in the first cluster, etc.) and moreover provide visualizations of these counts in the form of wordclouds. In addition to establishing the semantic content of these clusters, we verify their consistency by computing correlations between centroids for various values of k and across subjects, finding in nearly all cases strong correlations (greater than .8).

Conclusions

In this work, we reveal, in a purely data-driven manner, six characteristic clusters of regional activation patterns whose images have semantically distinct content. Such results illustrate an essential premise in vision neuroscience: that semantically similar stimuli give rise to similar activation patterns. It also sheds light on which visual regions may activate together most frequently, and if this joint activation is related to the semantic content of natural images, e.g. which items appear together generally.

References

Gu, Z., Jamison, K.W., Khosla, M., Allen, E.J., Wu, Y., Naselaris, T., Kay, K., Sabuncu, M.R. and Kuceyeski, A. (2021), ‘NeuroGen: activation optimized image synthesis for discovery neuroscience’, *ArXiv*: 2105.07140

Schrimpf, M., Kubilius, J., Lee, M.J., Murty, N., Ajemian, R. and DiCarlo, J.J. (2020), ‘Integrative Benchmarking to Advance Neurally Mechanistic Models of Human Intelligence’, Neuron Perspective, Vol. 108, No. 3, pp. 413-423. doi: https://doi.org/10.1016/j.neuron.2020.07.040

Allen, E.J., St-Yves, G., Wu, Y., Breedlove, J.L., Prince, J.S., Dowdle, L.T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J.B., Naselaris, T. and Kay, K. (2021), ‘A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence’, Nature Neuroscience, doi: https://doi.org/10.1038/s41593-021-00962-x

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L. (2014), ‘Microsoft COCO: Common Objects in Context’, European Conference on Computer Vision, pp. 740-755, doi: https://doi.org/10.1007/978-3-319-10602-1_48

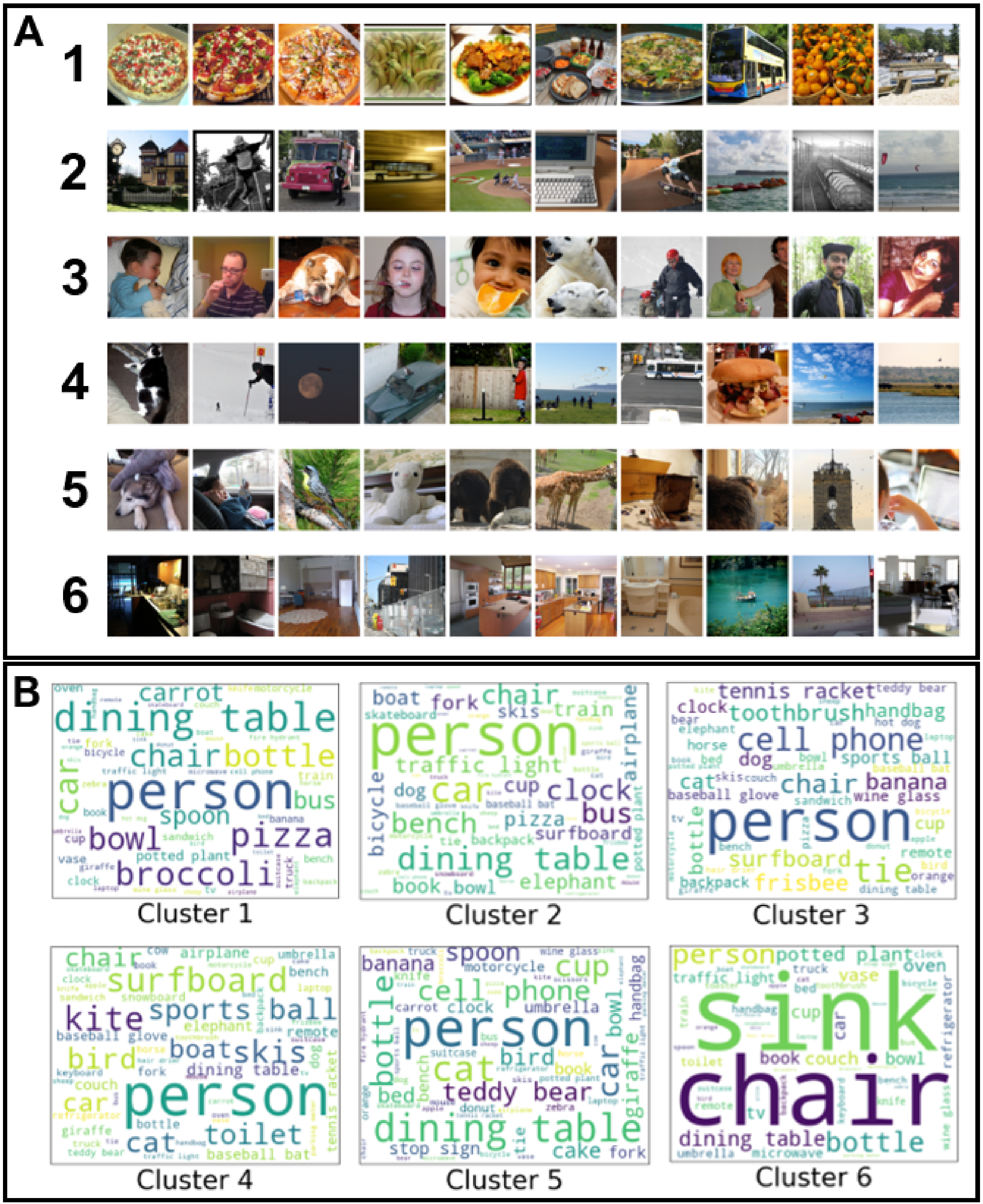

Figure 1: (A) Images corresponding to the 10 closest points (ordered from left to right) to each centroid. **(B)** Wordclouds depicting frequencies of image labels for the 100 closest points to each centroid.

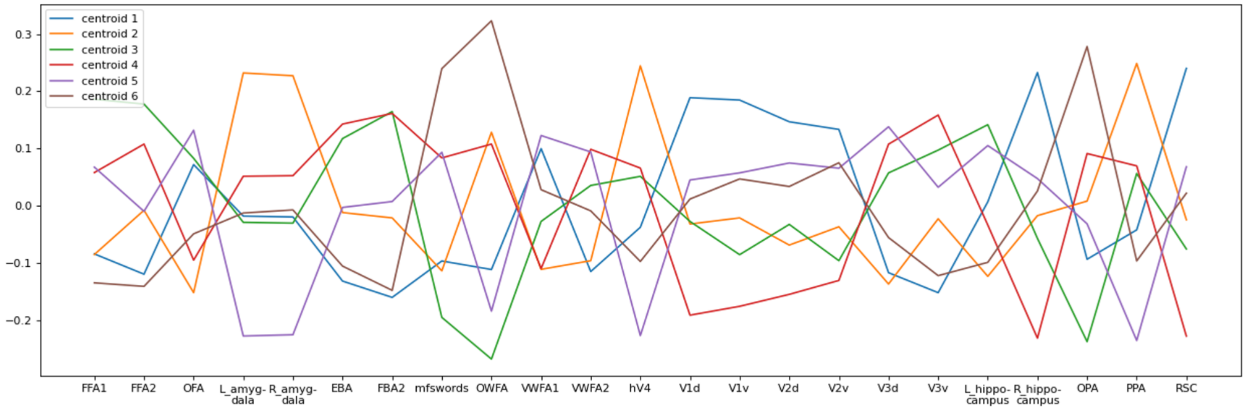

Figure 2: Plot of the centroids for each of the six clusters. Displayed are the activation values at 23 regions.