Mapping visual categorical selectivity across the whole brain using transformer-based encoders and large-scale generative models

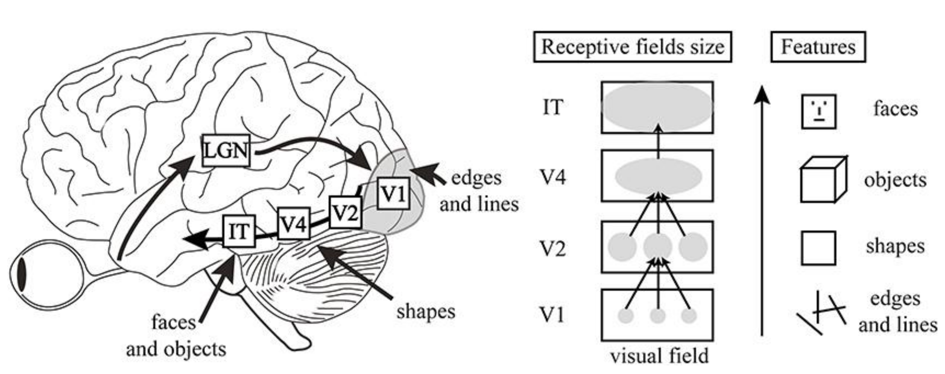
Columbia University in the City of New York

Columbia | Zuckerman Institute

Visual Inference Lab Kriegeskorte

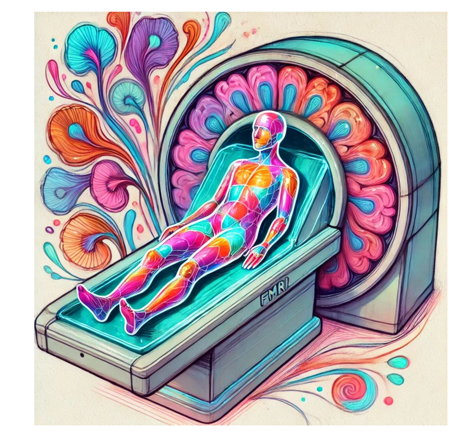
Ethan Hwang, Hossein Adeli, Nikolaus Kriegeskorte

Zuckerman Mind Brain Behavior Institute, Columbia University, New York, USA



Extensive neuroimaging experiments, especially functional magnetic resonance imaging (fMRI), have mapped a few prominent categories—faces, places, words, bodies, and food—to dedicated brain regions. But visual perception goes beyond simple categories and it remains an open question what higher-level visual concepts enable humans to make sense of the complex world.

Here, we leverage recent advances in artificial intelligence and a large-scale fMRI dataset to explore the visual selectivity of brain regions beyond the well-studied visual cortex. We found many parcels with more complex selectivity transcending simple categorical concepts.



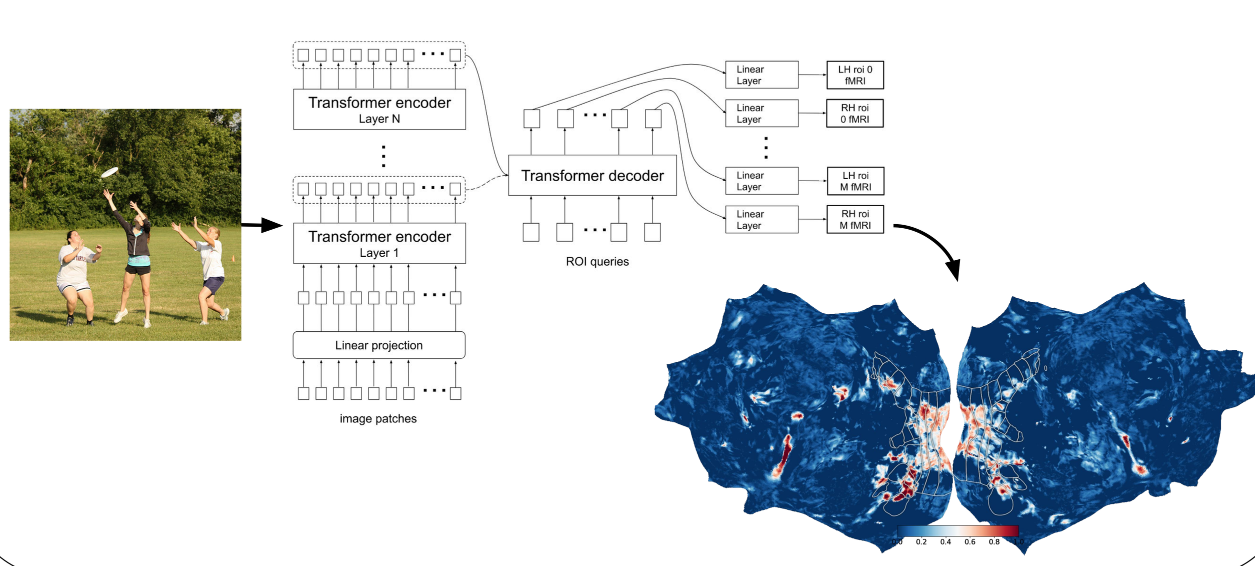
Data-driven functional parcellation

Each hemisphere is divided up into ~500 parcels using k-means clustering. Each parcel groups 200–400 voxels based on correlated responses.

Since the parcellation is determined with an unsupervised method and no spatial constraints, some parcels are not contiguous. We will explore the selectivity of these parcels in the rest of the experiment.

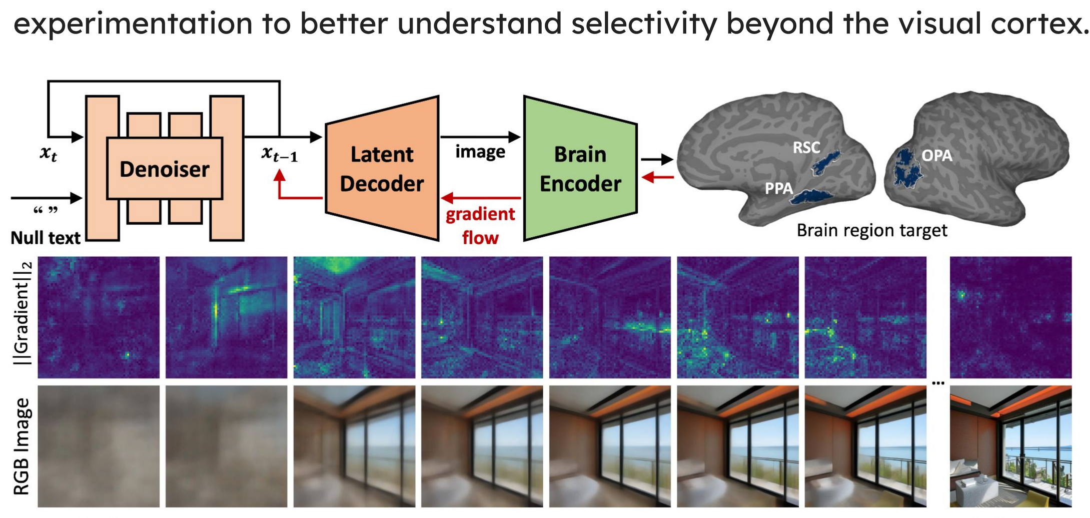
Brain encoder

We built a transformer-based brain encoder to predict brain activity from visual features [3] [5]. We trained the model on the NSD dataset [1], which includes up to 30,000 image-fMRI pairs from 8 subjects.



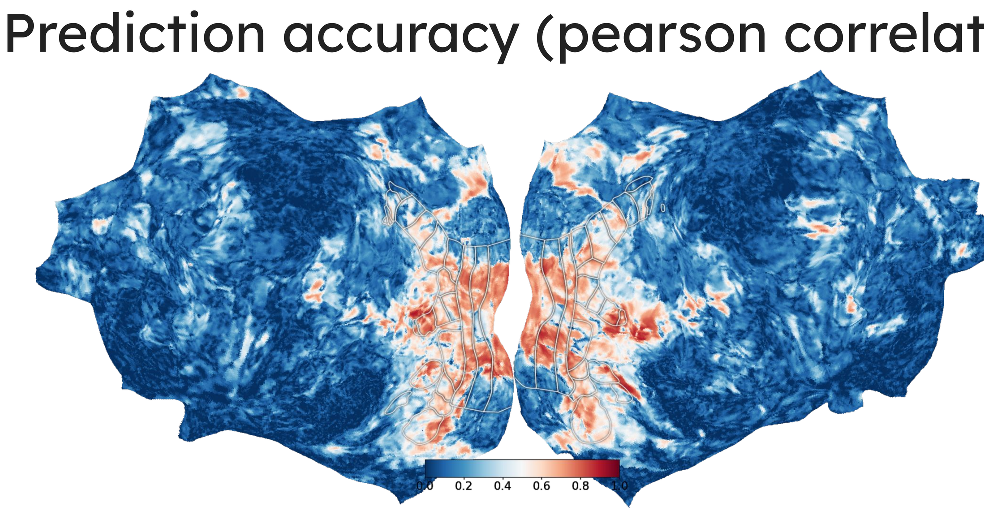
Generating images with diffusion

We used a diffusion model, BrainDIVE [2], to generate images that would maximally activate a given parcel. The encoder model serves as a "digital twin" that is fully observable, upon which we perform extensive experimentation to better understand selectivity beyond the visual cortex.



Model results

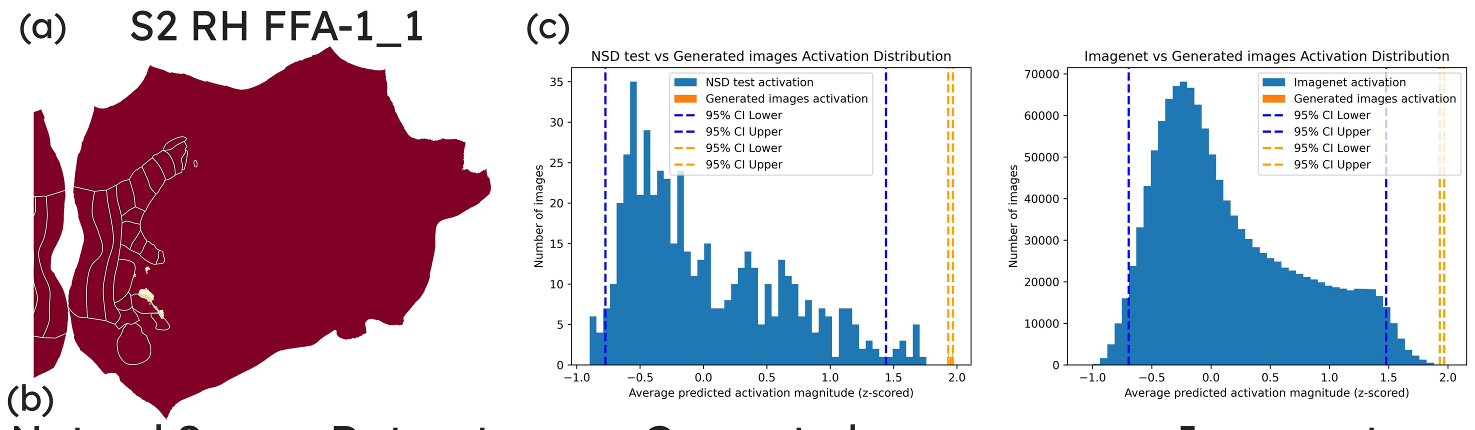
Prediction accuracy (pearson correlation) on held-out set:



As expected, our model performs well on predicting activity in the visual cortex. Several areas beyond the visual cortex are also well-predicted, which may offer insight into downstream visual processing.

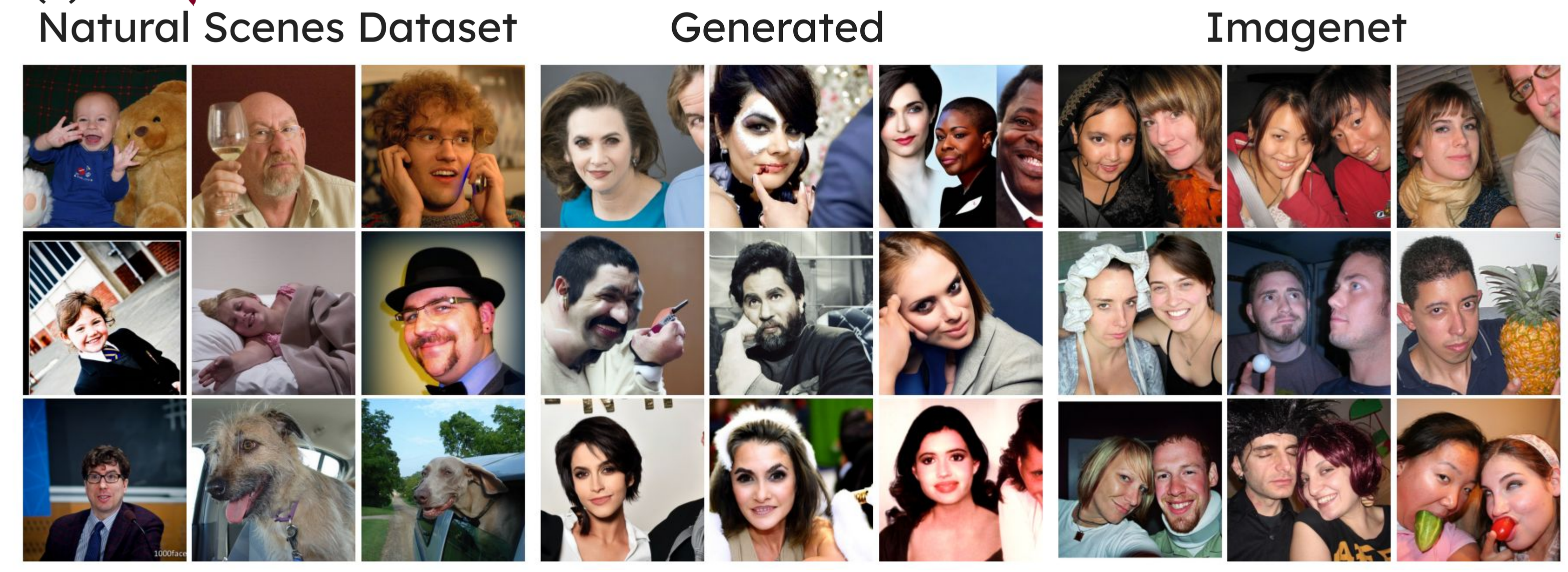
Sanity check: selectivity of known regions (FFA)

(a) S2 RH FFA-1_1 (c)



(b)

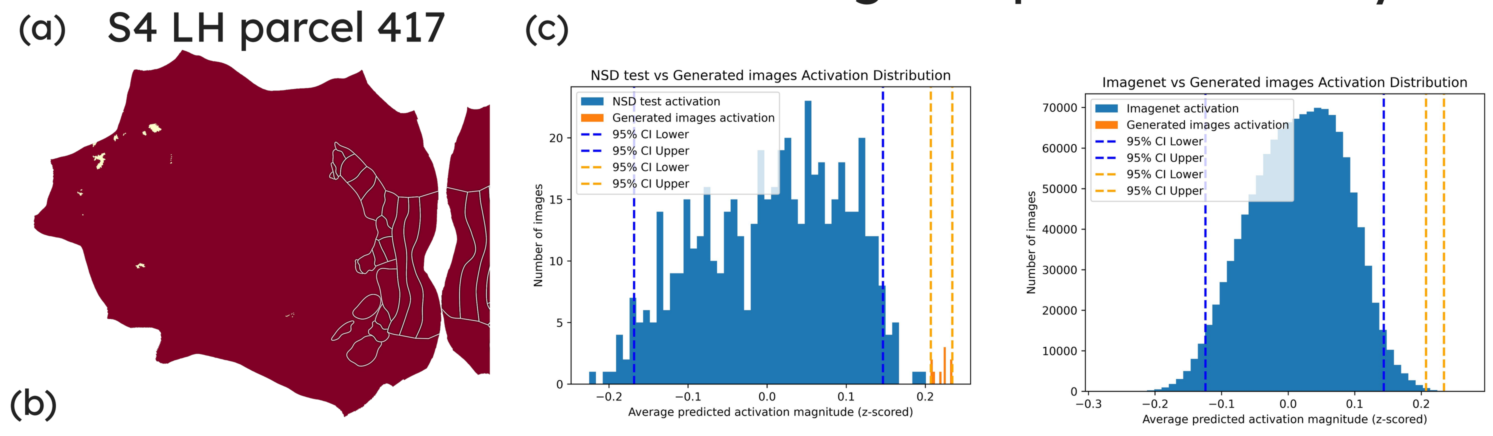
Natural Scenes Dataset | Generated | Imagenet



We start by demonstrating the selectivity of a parcel that overlaps significantly with a known region (a), the fusiform face area, which is known to respond to faces [7]. We chose images from the NSD, our generated set, and Imagenet [4] that maximally activate the parcel (c). The images consistently include faces, as expected (b). The distribution of generated images is at the very tail end of both datasets (c).

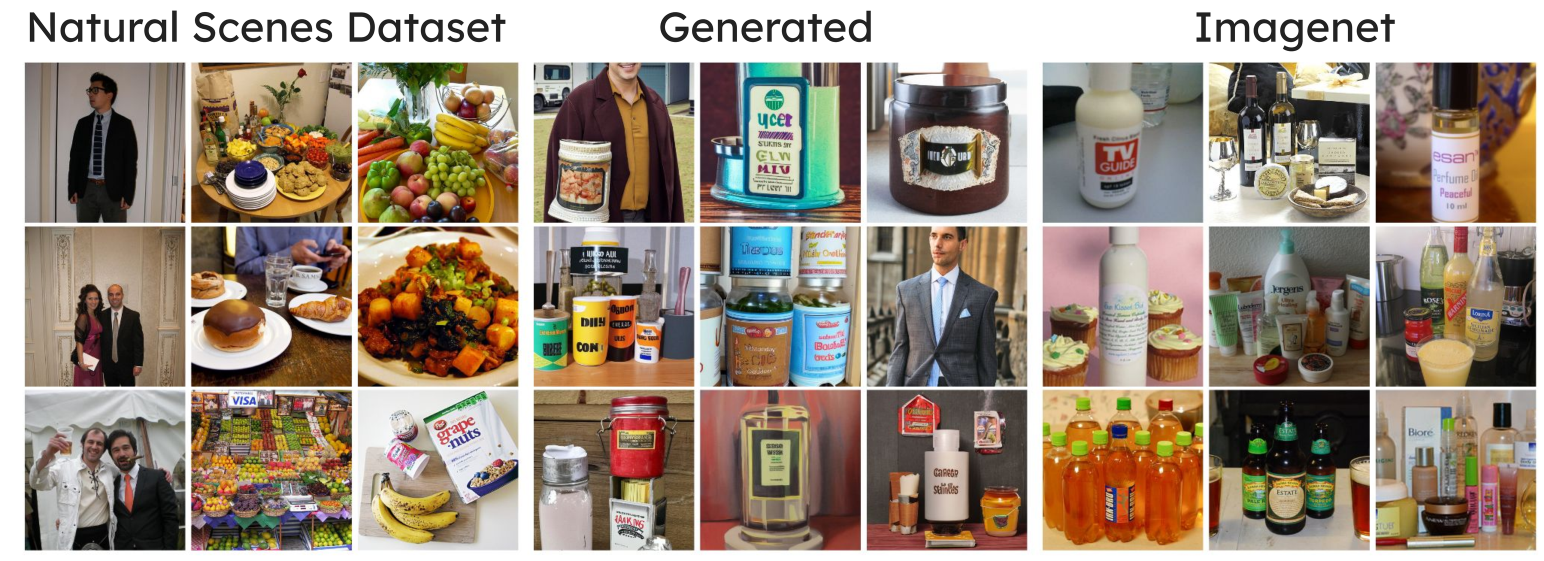
Main result: demonstrating complex selectivity

(a) S4 LH parcel 417 (c)



(b)

Natural Scenes Dataset | Generated | Imagenet



We next explore the selectivity of a parcel in the frontal region of brain (a), chosen because it is visually responsive and well-predicted by our model. The images that maximally activate the area seem to include bottles with words (c). The model trained on NSD (a relatively small dataset) reveals complex selectivity that can only be discovered using generative models or large-scale image retrieval.

References

[1] E. J. Allen et al., "A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence," Nat Neurosci, vol. 25, no. 1, pp. 116–126, Jan. 2022, doi: 10.1038/s41593-021-00962-x.
[2] A. F. Luo, M. M. Henderson, L. Wehbe, and M. J. Tarr, "Brain Diffusion for Visual Exploration: Cortical Discovery using Large Scale Generative Models," Nov. 28, 2023, arXiv: arXiv:2306.03089. doi: 10.48550/arXiv.2306.03089.
[3] M. Oquab et al., "DINOv2: Learning Robust Visual Features without Supervision," Feb. 02, 2024, arXiv: arXiv:2304.07193. doi: 10.48550/arXiv.2304.07193.
[4] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL: IEEE, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
[5] H. Adeli, S. Minni, and N. Kriegeskorte, "Predicting brain activity using Transformers," Aug. 05, 2023, Neuroscience. doi: 10.1101/2023.08.02.551743.
[6] J. S. Gao, A. G. Huth, M. D. Lescroart, and J. L. Gallant, "Pycortex: an interactive surface visualizer for fMRI," Front. Neuroinform., vol. 9, Sep. 2015, doi: 10.3389/fninf.2015.00023.
[7] N. Kanwisher, J. McDermott, and M. M. Chun, "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception".
[8] M. H. Herzog and A. M. Clarke, "Why vision is not both hierarchical and feedforward," Front. Comput. Neurosci., vol. 8, Oct. 2014, doi: 10.3389/fncom.2014.00135.