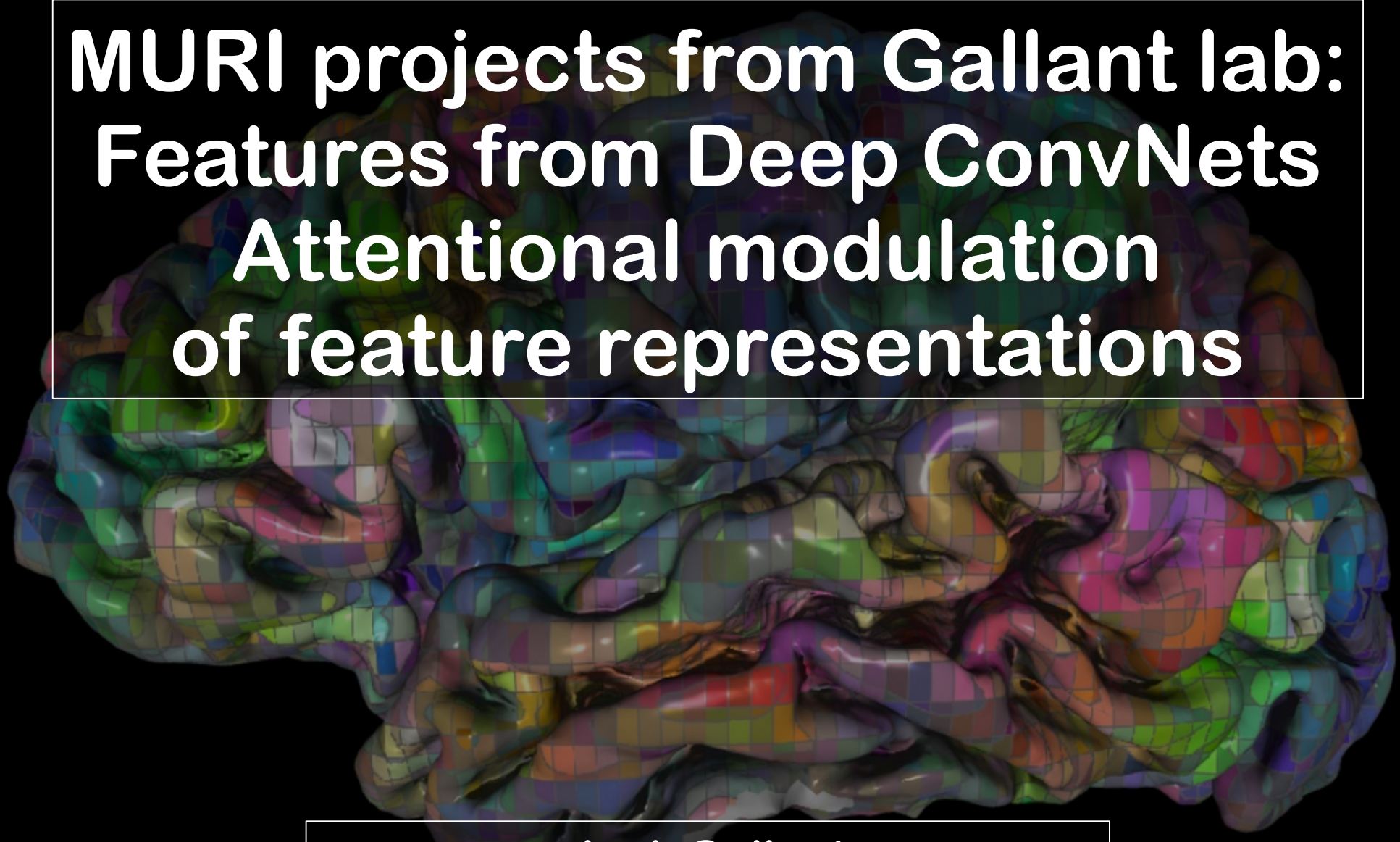
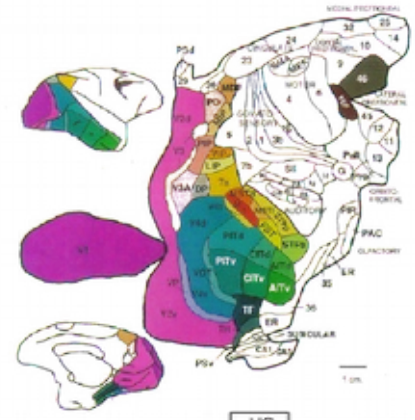
MURI projects from Gallant lab: Features from Deep ConvNets Attentional modulation of feature representations

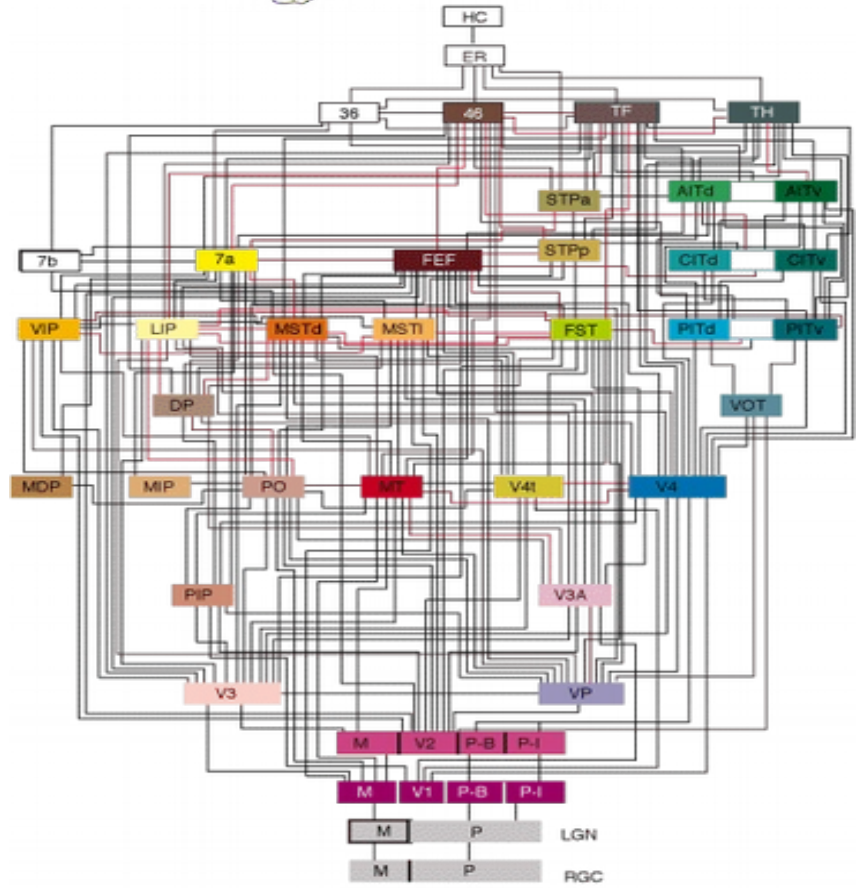


Jack Gallant
University of California at Berkeley

A hierarchical parallel network for vision

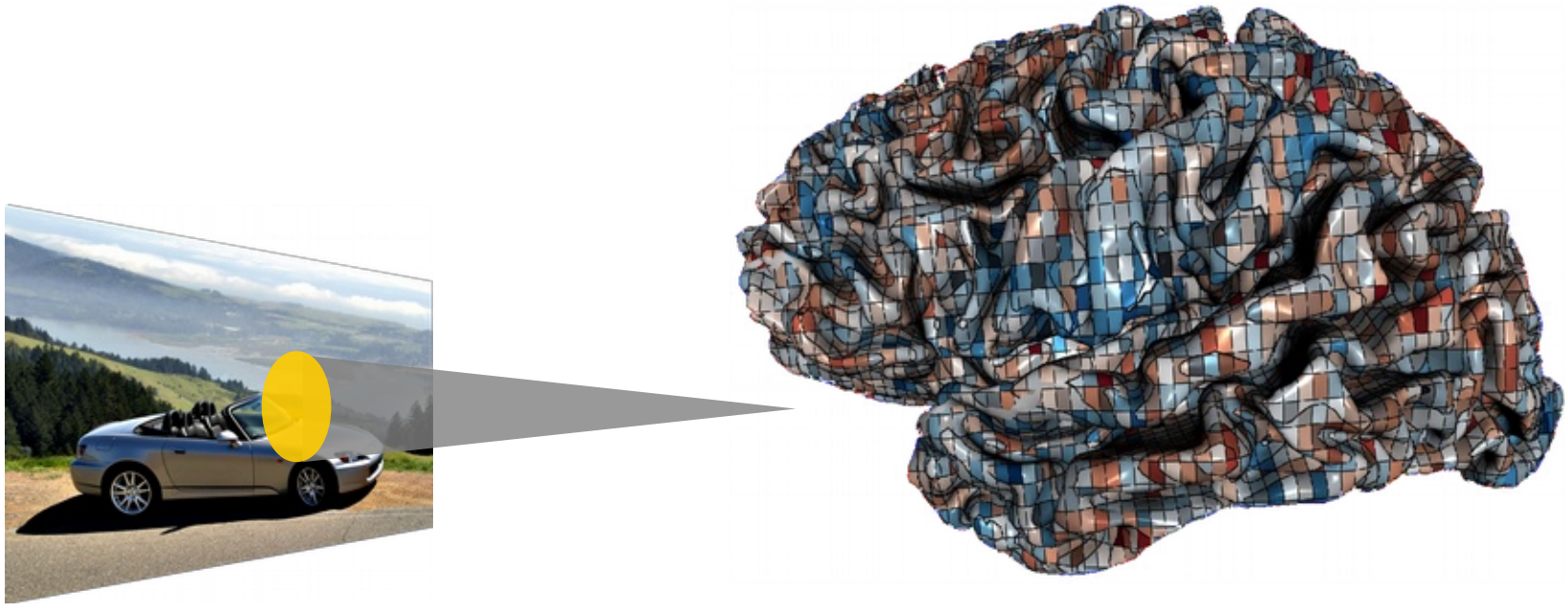
- Dozens of anatomical and functional areas arranged in a hierarchical, parallel network with recurrent feedback.
- Transformations between areas are nonlinear.
- Each area represents a different high-dimensional feature space.
- Each feature space is projected/distributed across the cortical surface.
- Each area is differentially affected by bottom-up and top-down information.





Felleman and Van Essen, *Cerebral Cortex,* 1992

Visual neuroscience as system identification



$$s(x,y,t) \rightarrow H_v(x,y,t,a) \rightarrow r_v(t)$$

The amount of data required to fit the model
Is proportional to the degree of nonlinearity.

State variables (e.g. Attention) multiply
The dimensionality of the problem.

A unique feature space for each visual area



Mid-level visual areas

Retina

Thalamus (LGN)

V1

V2

V3

V4

Lateral Occipital Cortex

FFA (Fusiform Face Area)

EBA (Extrastriate Body Area)

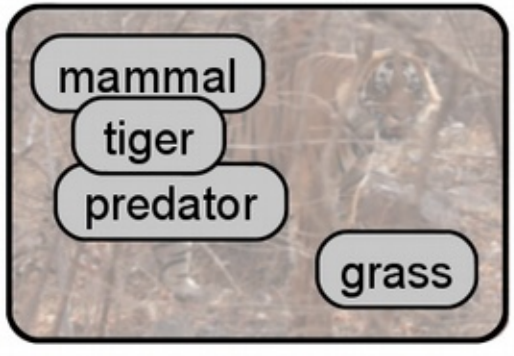
PPA (Parahippocampal Place Area)

Luminance edges





Semantic categories

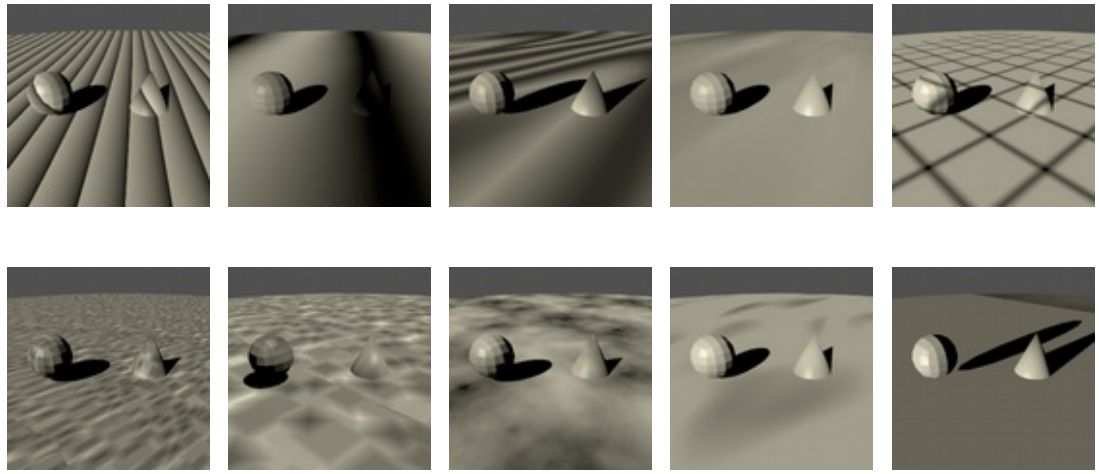


A unique feature space for each visual area

448 Objects in 8 Categories



Variation in background, lighting & Shadow



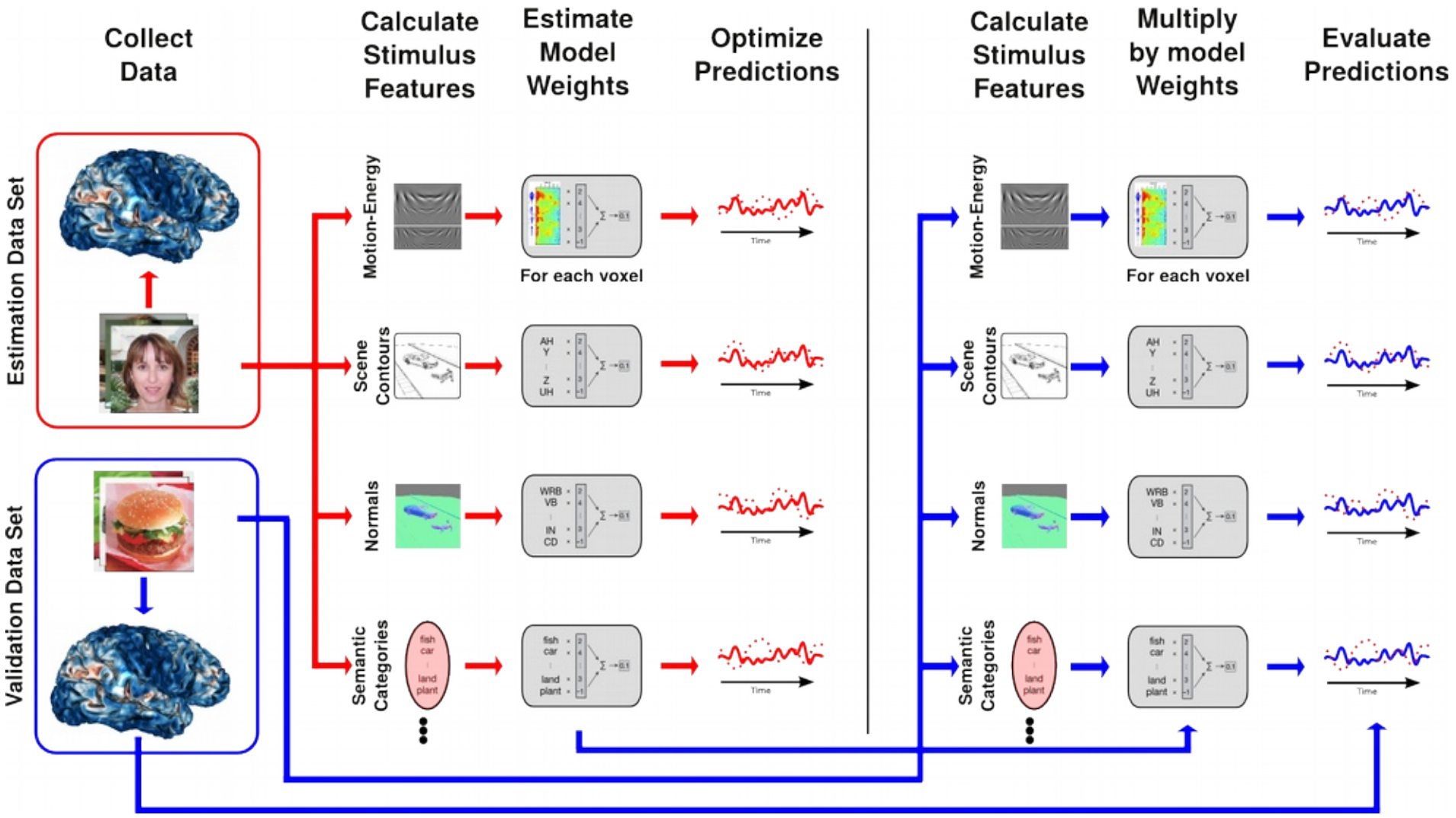
Rendered movies with
3D camera motion





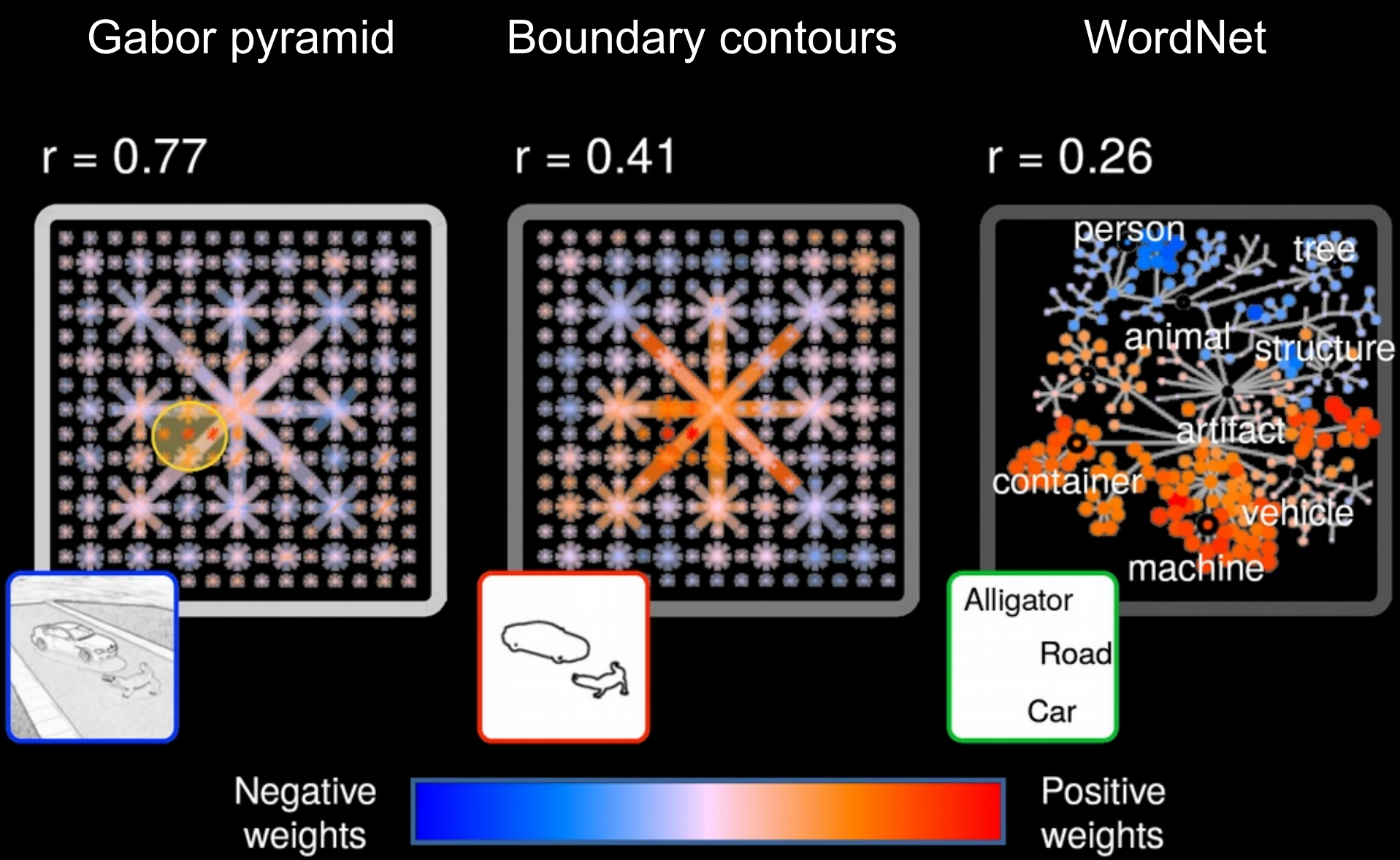
Mark
Lescroart

Feature spaces from rendered images



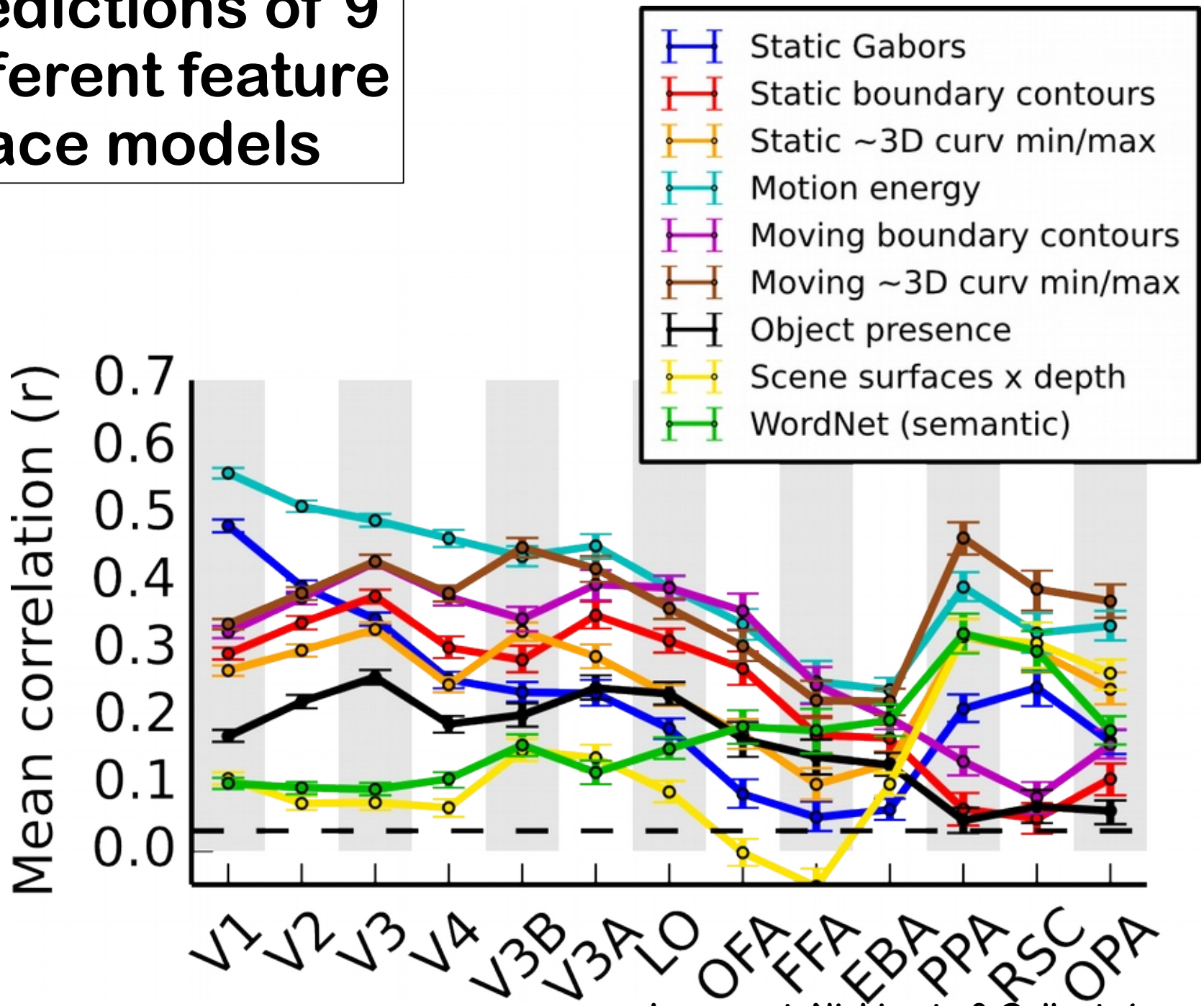
Lescroart, Nishimoto & Gallant, *in preparation*

The “receptive field” of a single V1 voxel



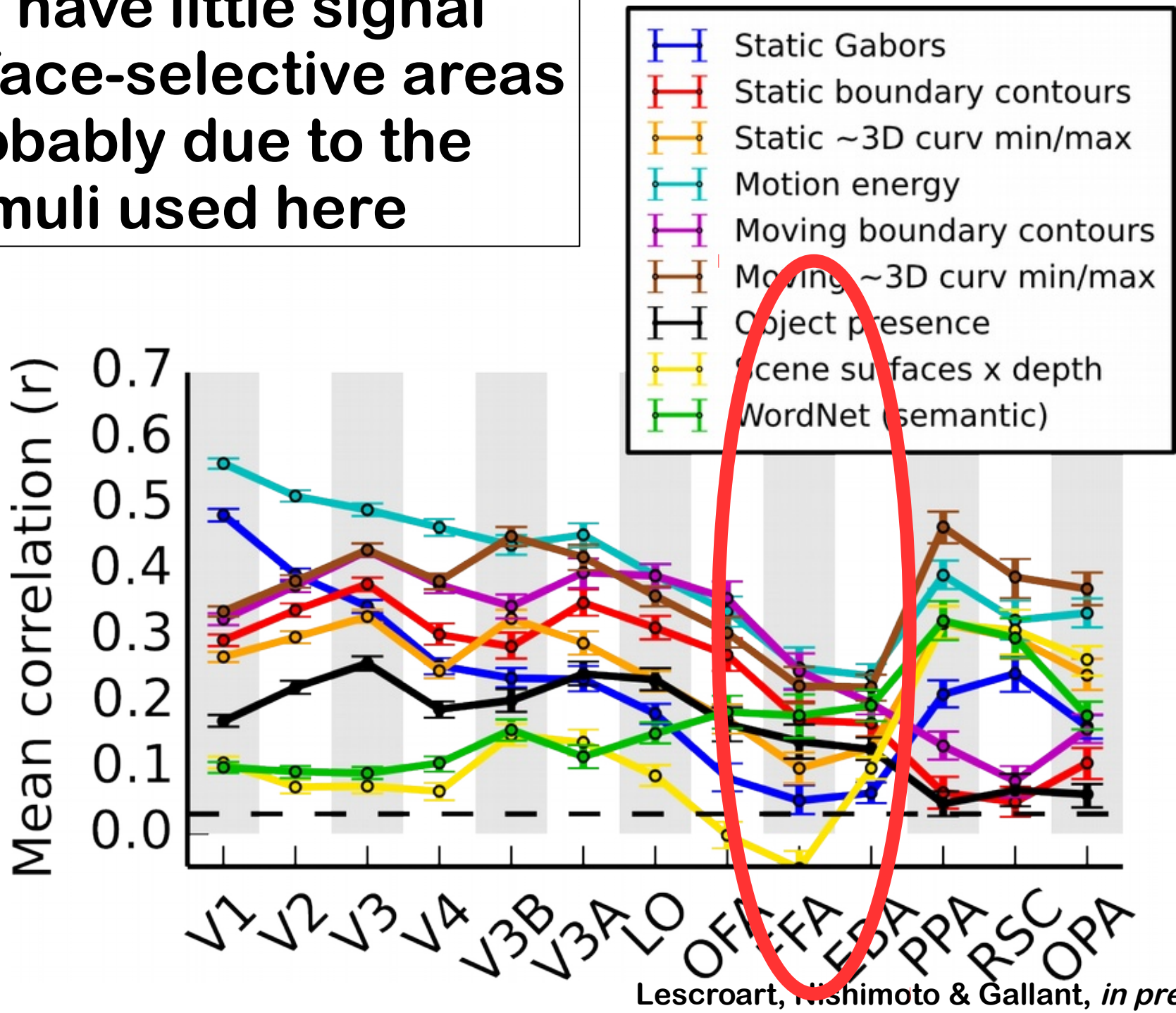
Lescroart, Nishimoto & Gallant, *in preparation*

Predictions of 9 different feature space models



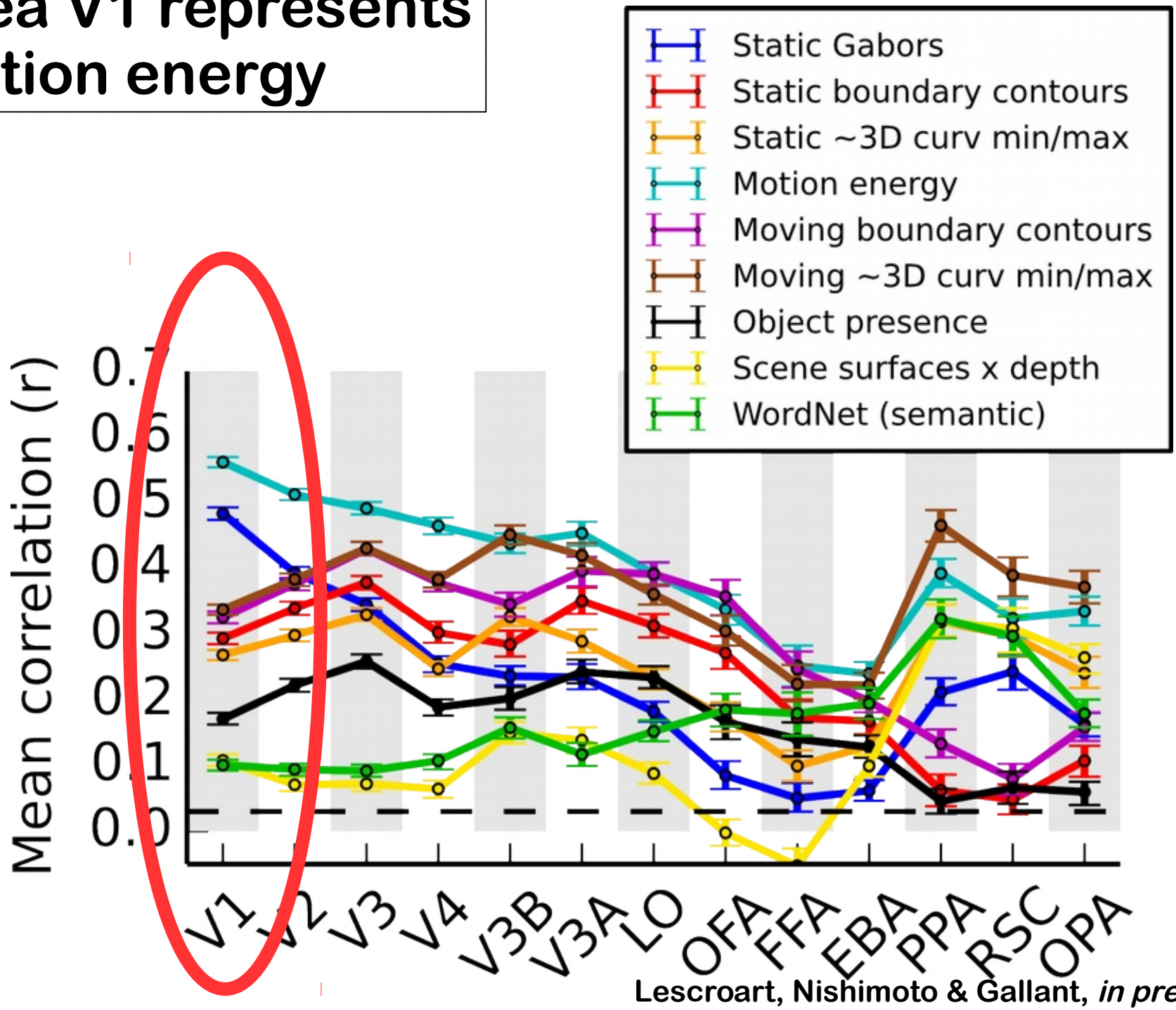
Lescroart, Nishimoto & Gallant, *in preparation*

We have little signal in face-selective areas probably due to the stimuli used here



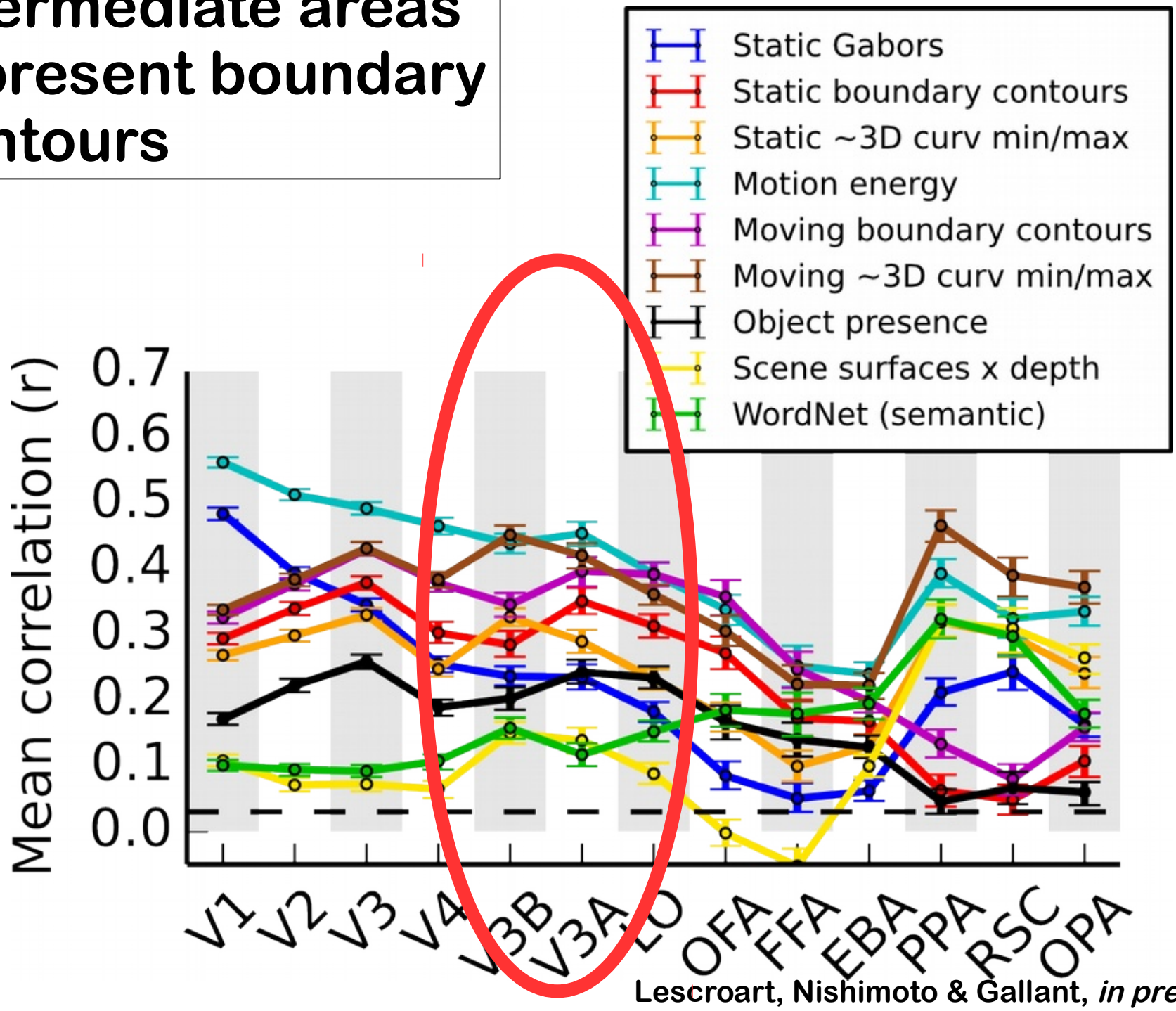
Lescroart, Nishimoto & Gallant, *in preparation*

Area V1 represents motion energy



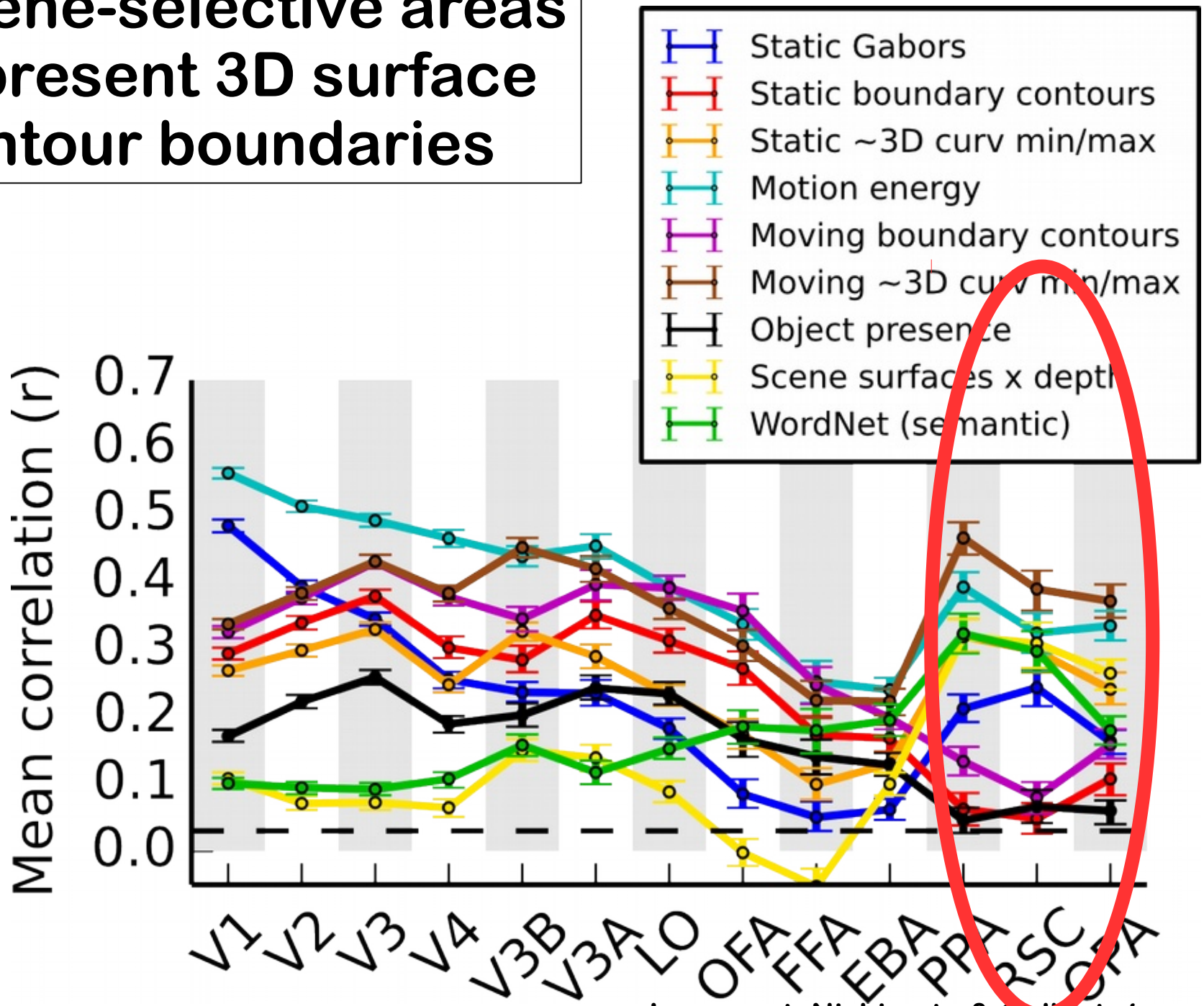
Lescroart, Nishimoto & Gallant, *in preparation*

Intermediate areas represent boundary contours



Lescroart, Nishimoto & Gallant, *in preparation*

Scene-selective areas represent 3D surface contour boundaries

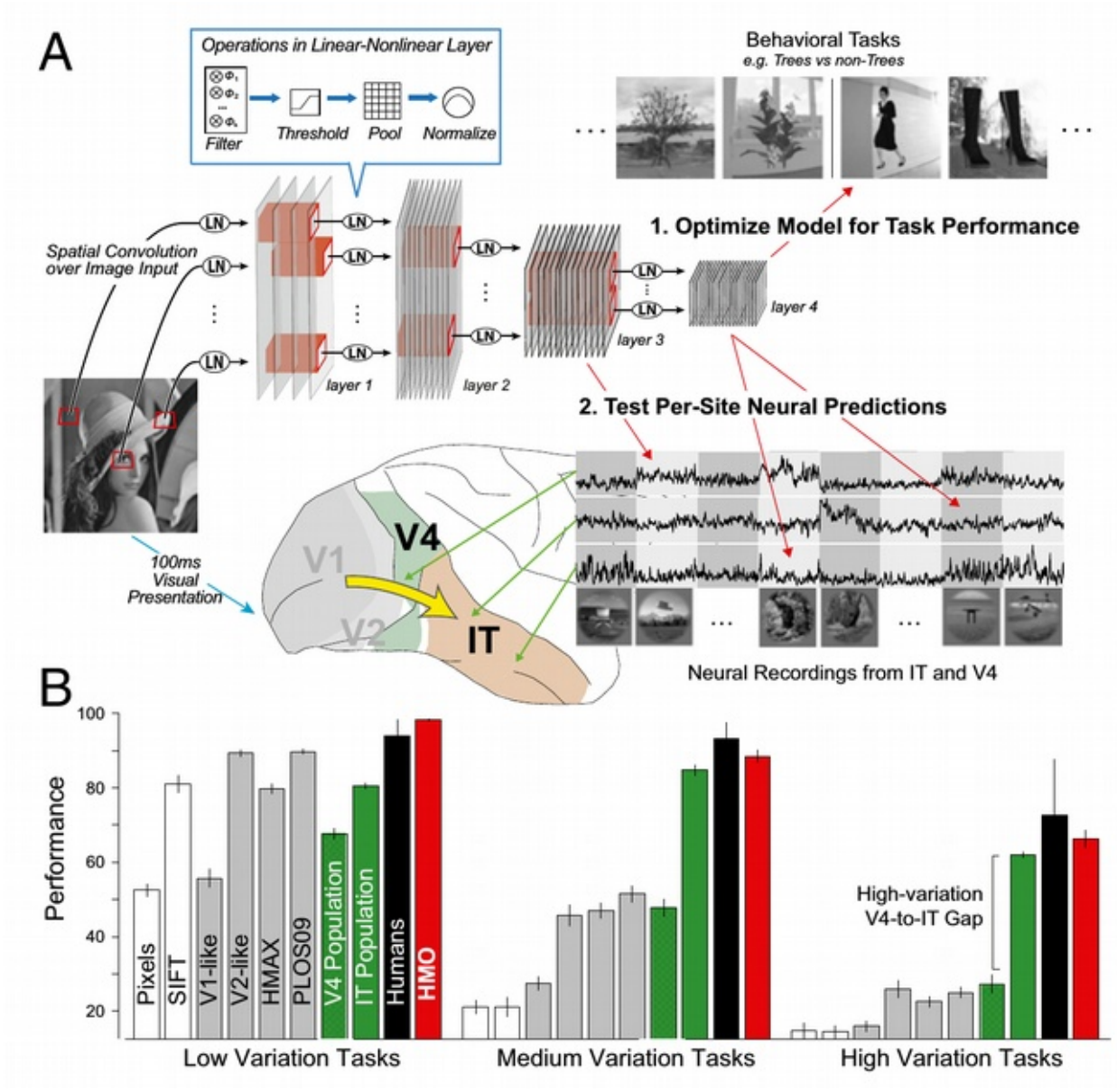


Lescroart, Nishimoto & Gallant, *in preparation*

Plans and future directions

- Explore many additional feature spaces.
- Disentangle correlations between various feature spaces.
- Examine effects of attention to foreground, background, features, surfaces and objects.
- Create more complicated rendered stimuli that include object motion, flexible objects and so on.

Convolutional Networks can provide useful features for modeling neurons

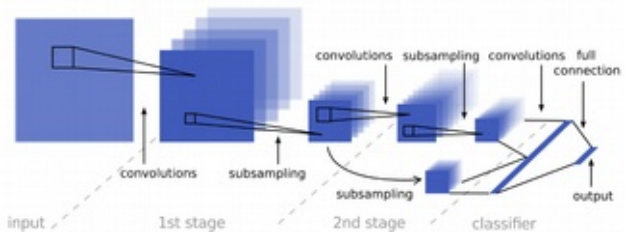


Yamins, Hong, Cadieu, Solomon, Seibert & DiCarlo, *PNAS*, 2014

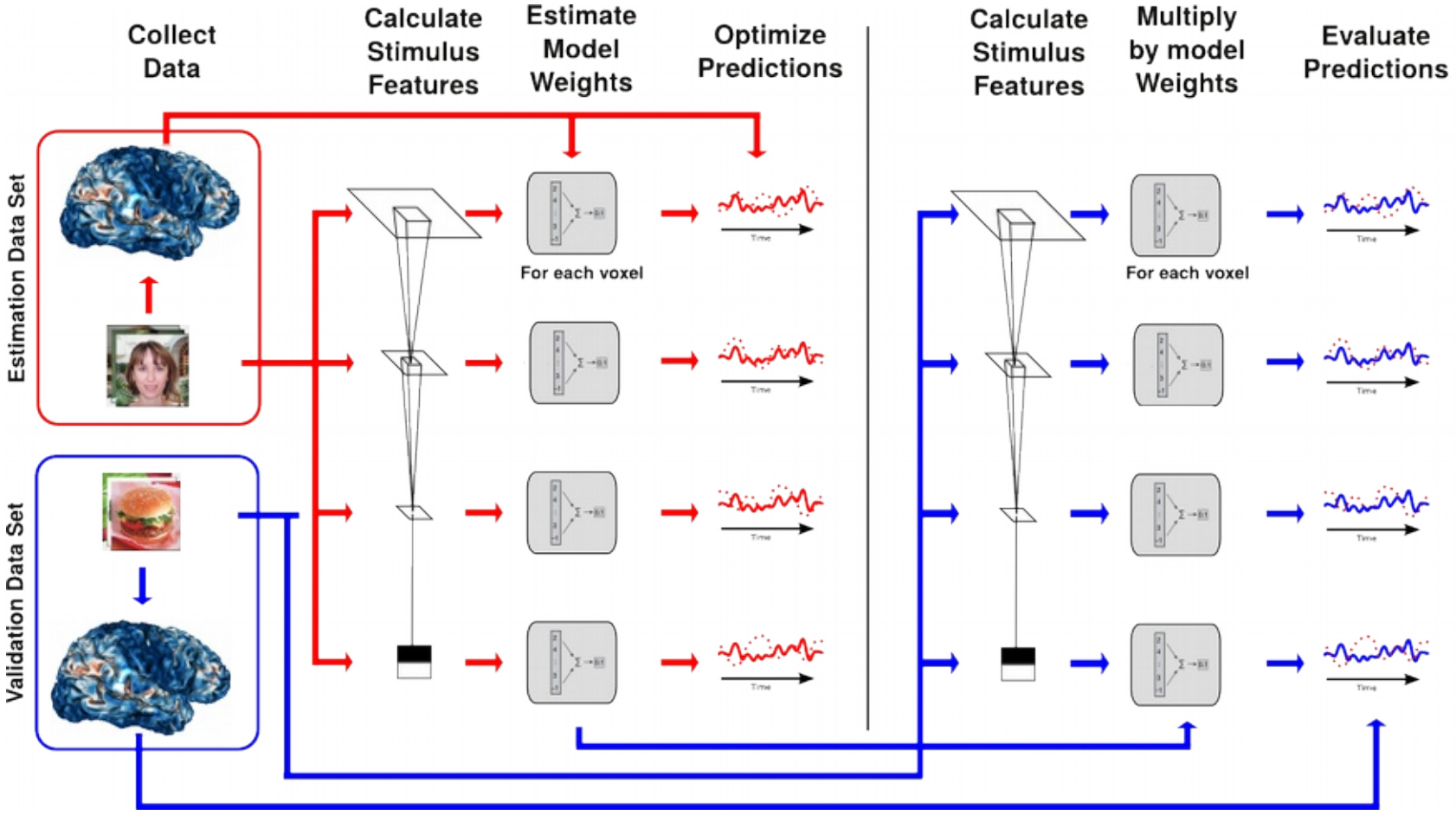
Using ConvNet layers as feature spaces



Pulkit Agrawal

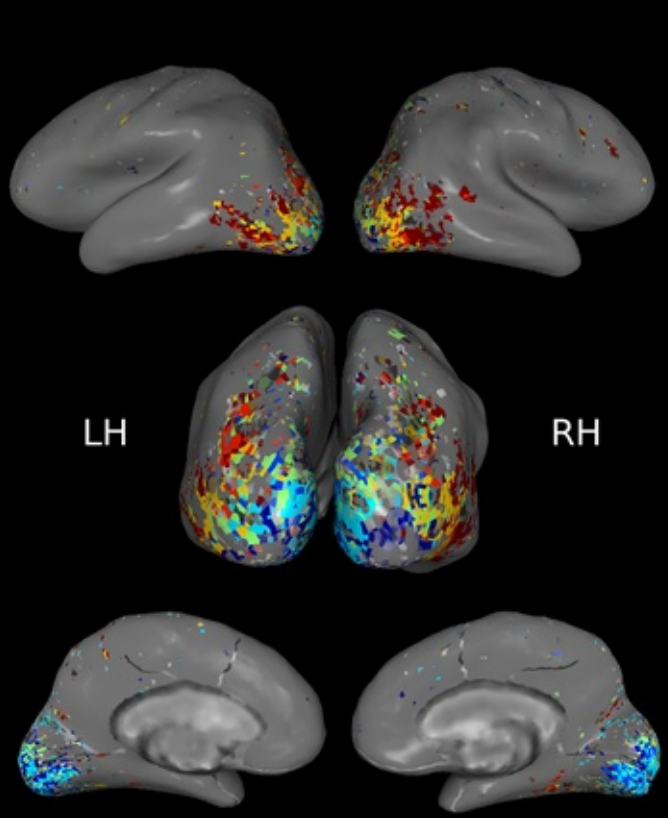


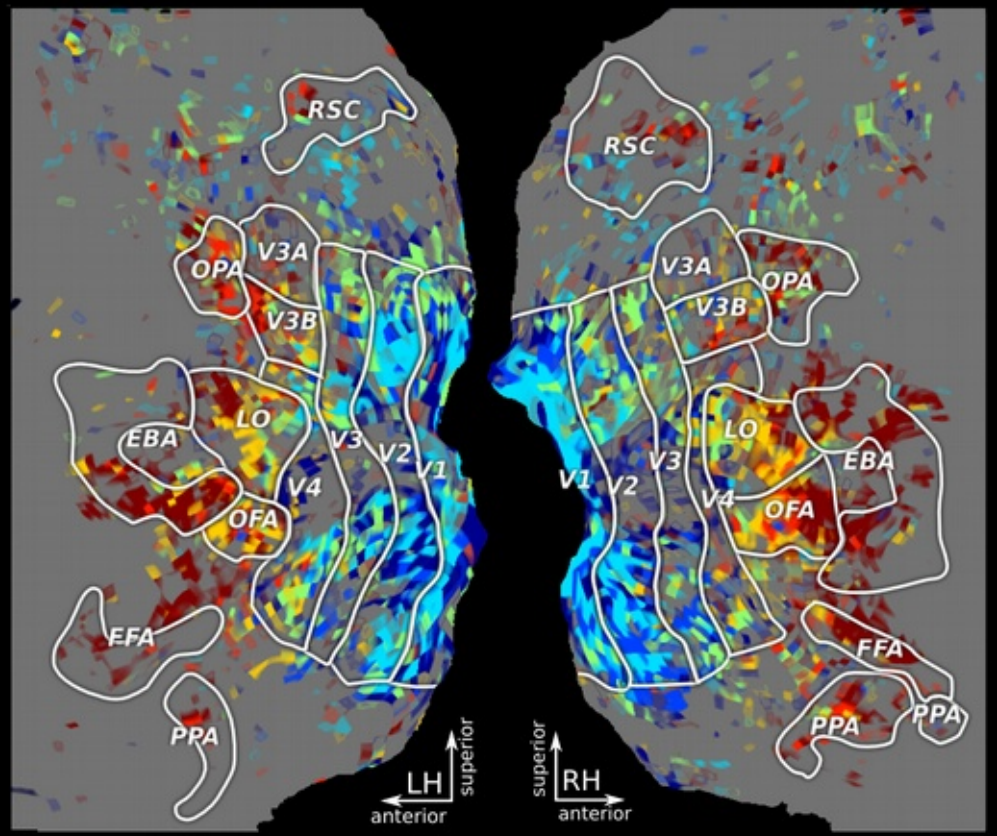
Note: The CNN used here contained 5 convolutional layers and 2 classifier layers.

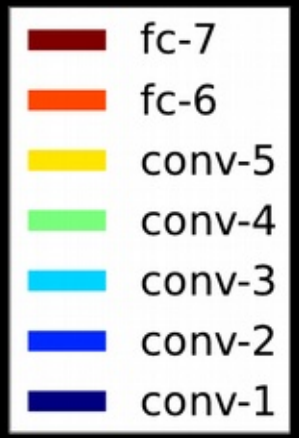


Agrawal, Stansbury, Malik & Gallant, *in preparation*

Predictions of Deep ConvNet layers match the visual hierarchy

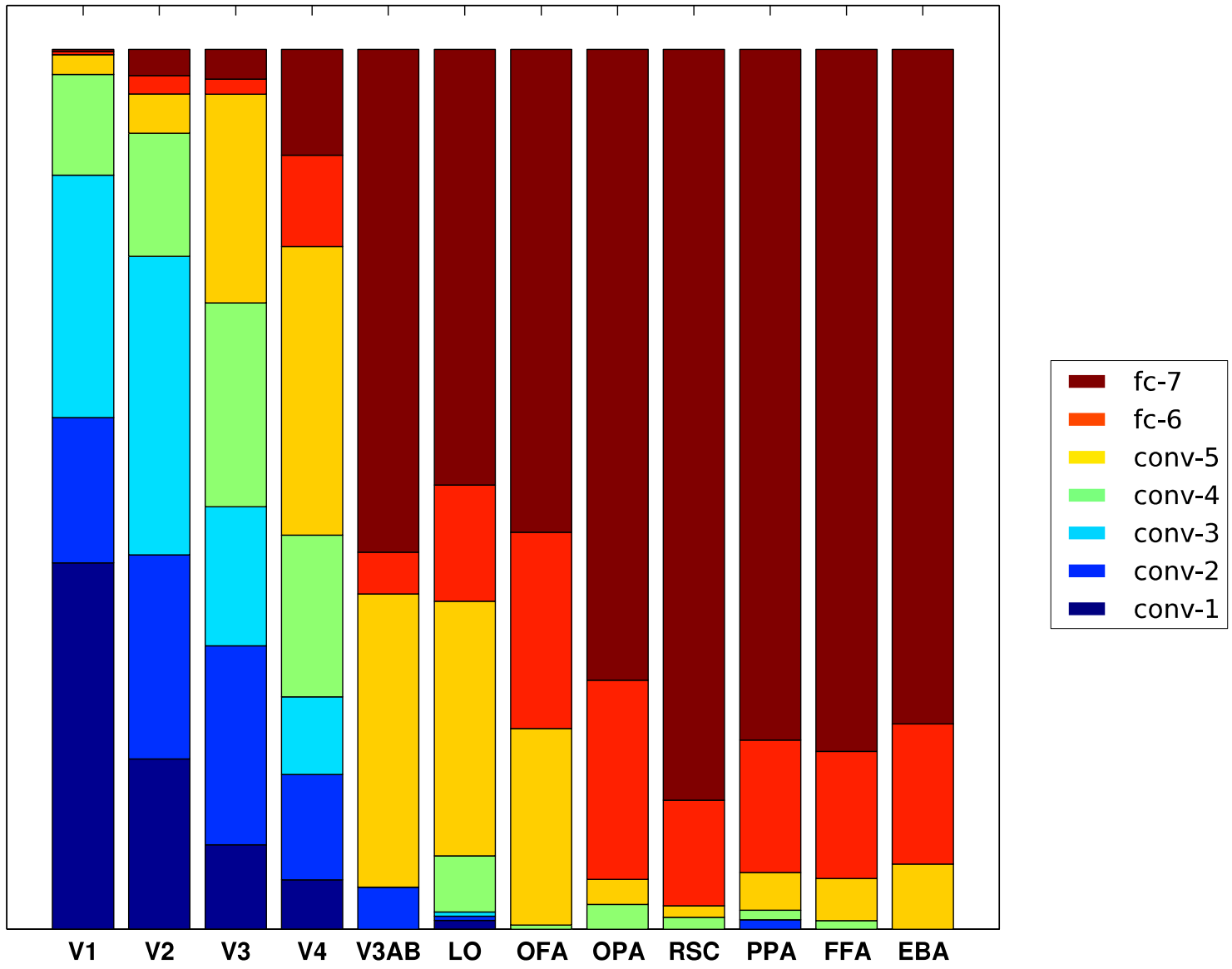






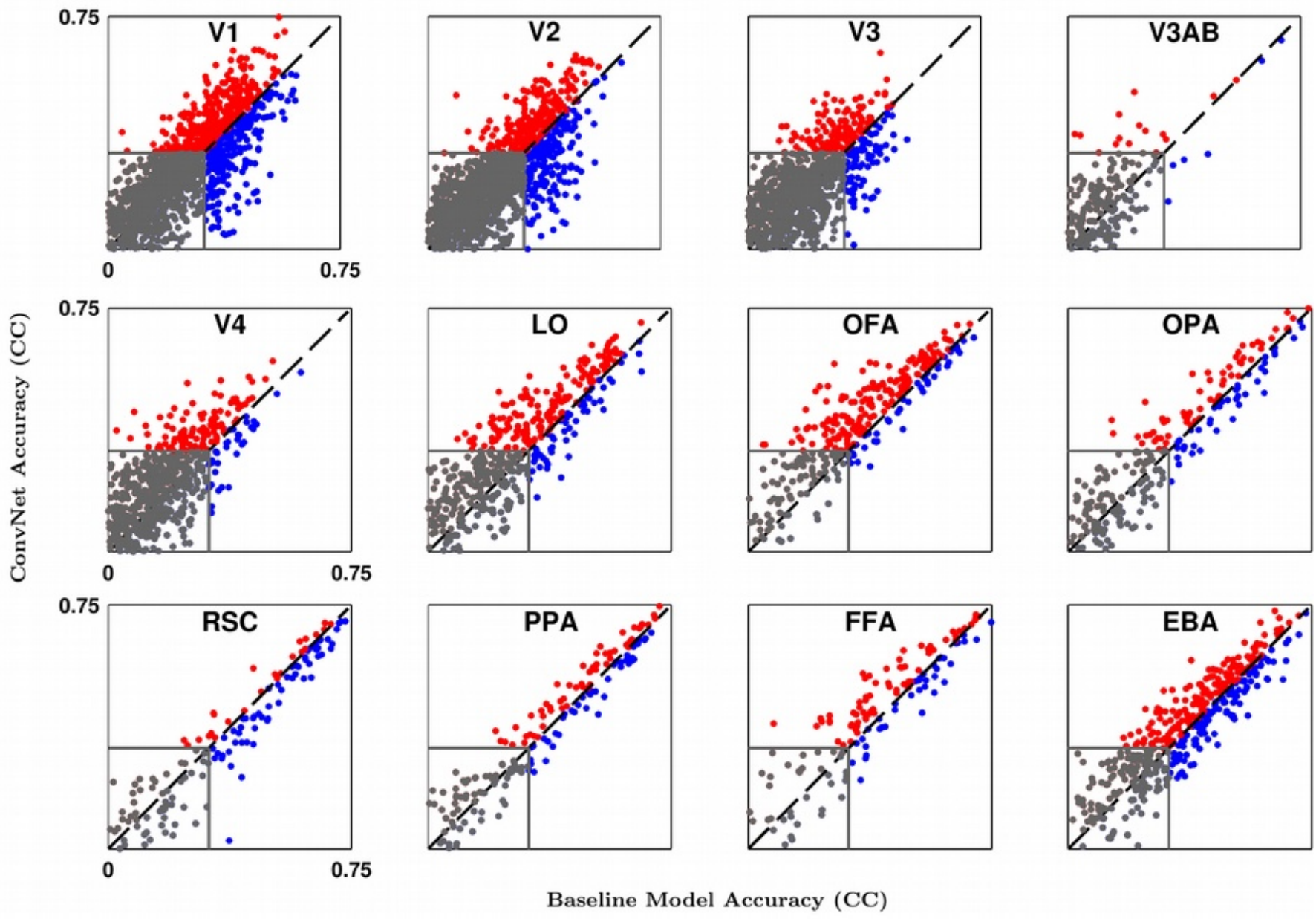
Agrawal, Stansbury, Malik & Gallant, *in preparation*

Complexity across the cortical hierarchy



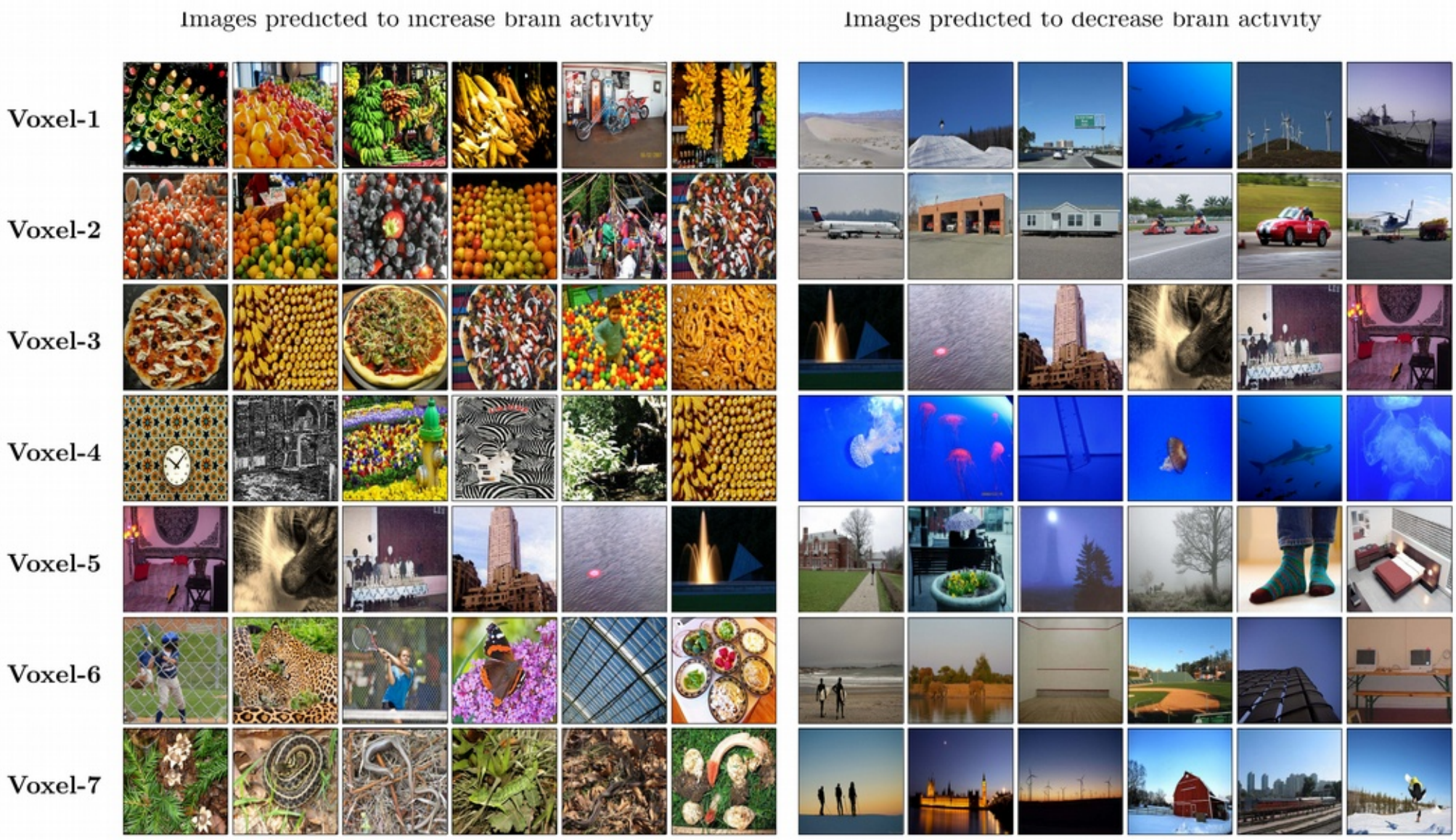
Agrawal, Stansbury, Malik & Gallant, *in preparation*

Deep ConvNet predict all areas well



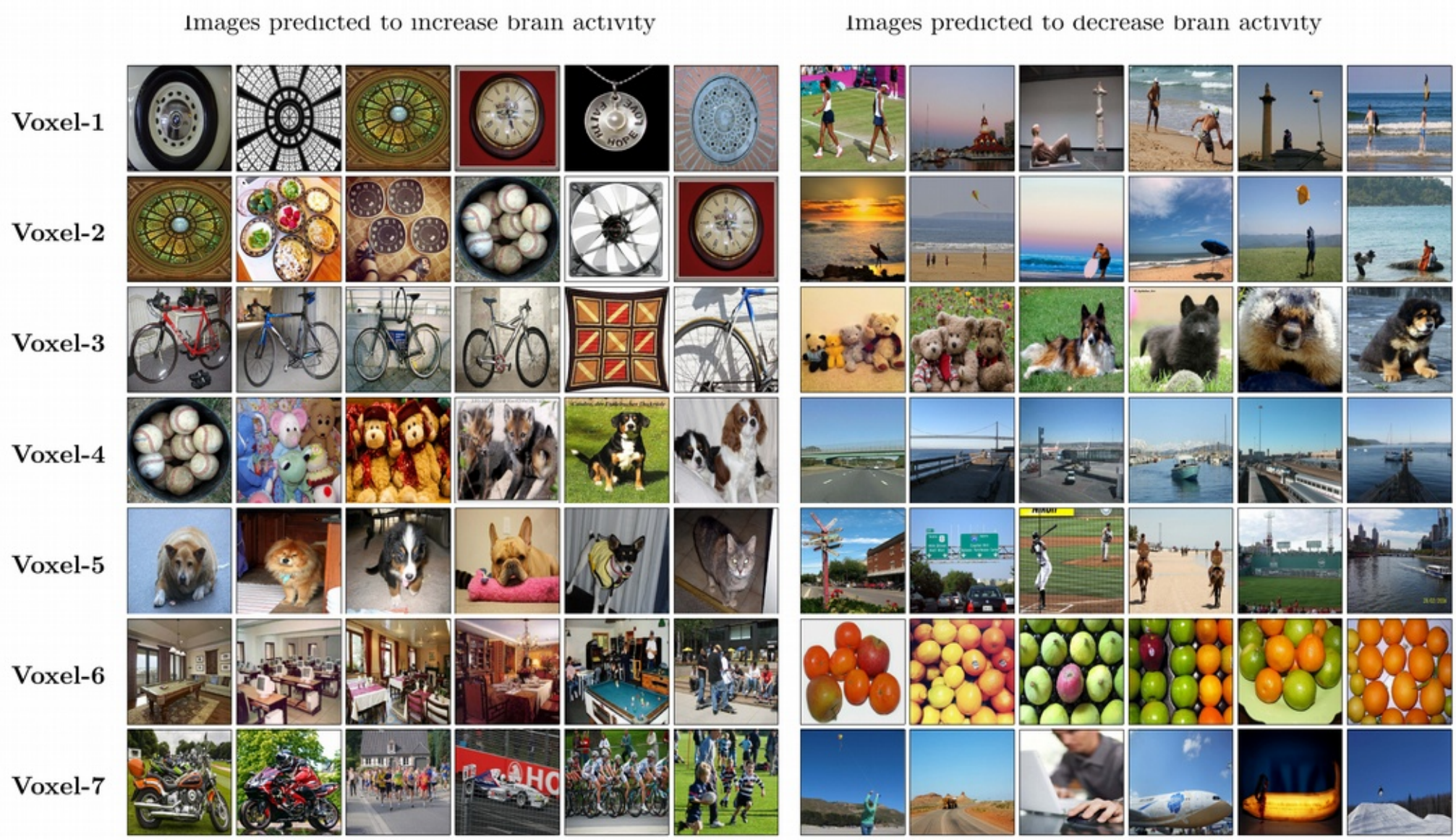
Agrawal, Stansbury, Malik & Gallant, *in preparation*

V1 voxels are selective for fine texture



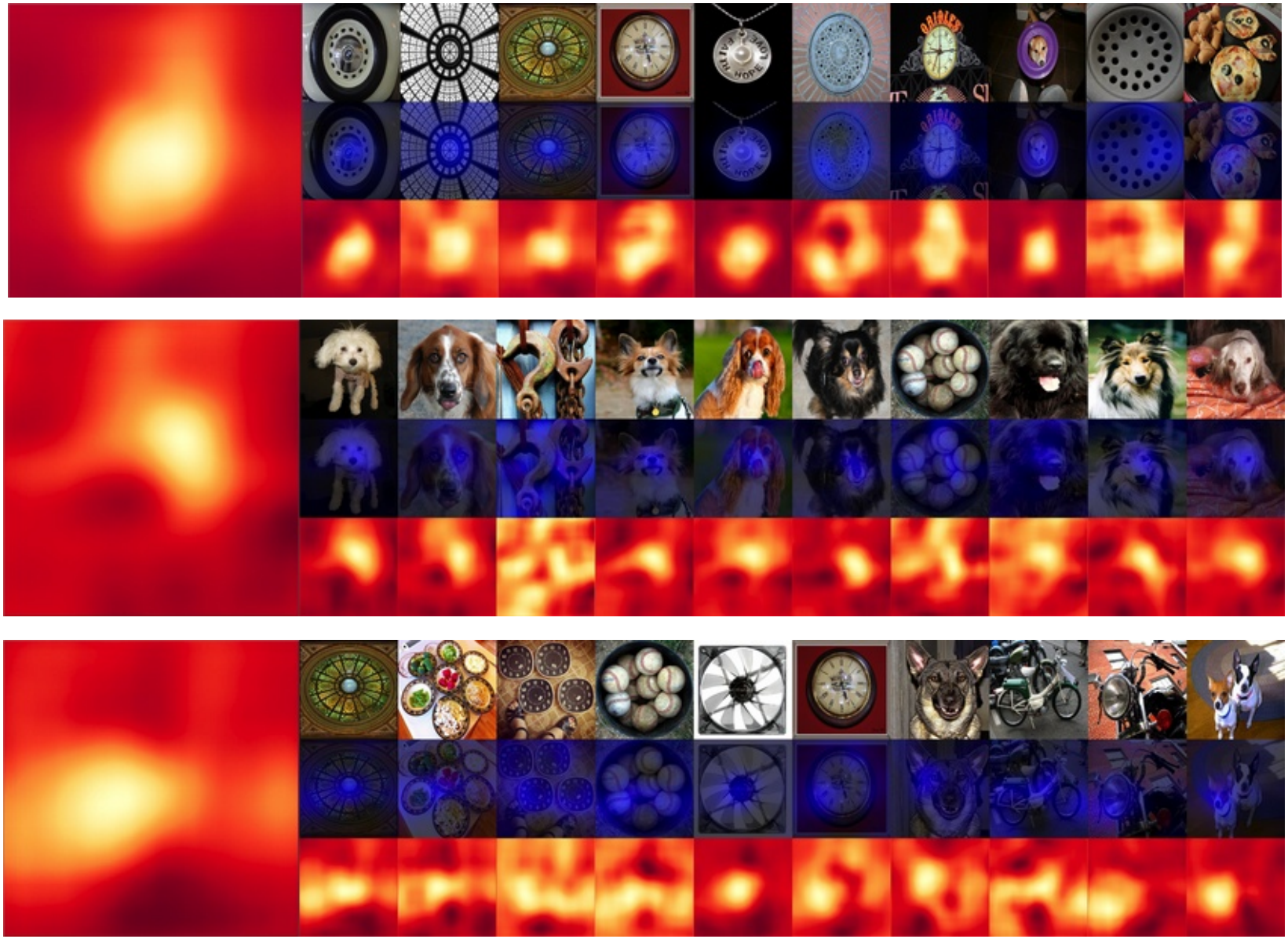
Agrawal, Stansbury, Malik & Gallant, *in preparation*

V4 voxels are selective for curvature & texture



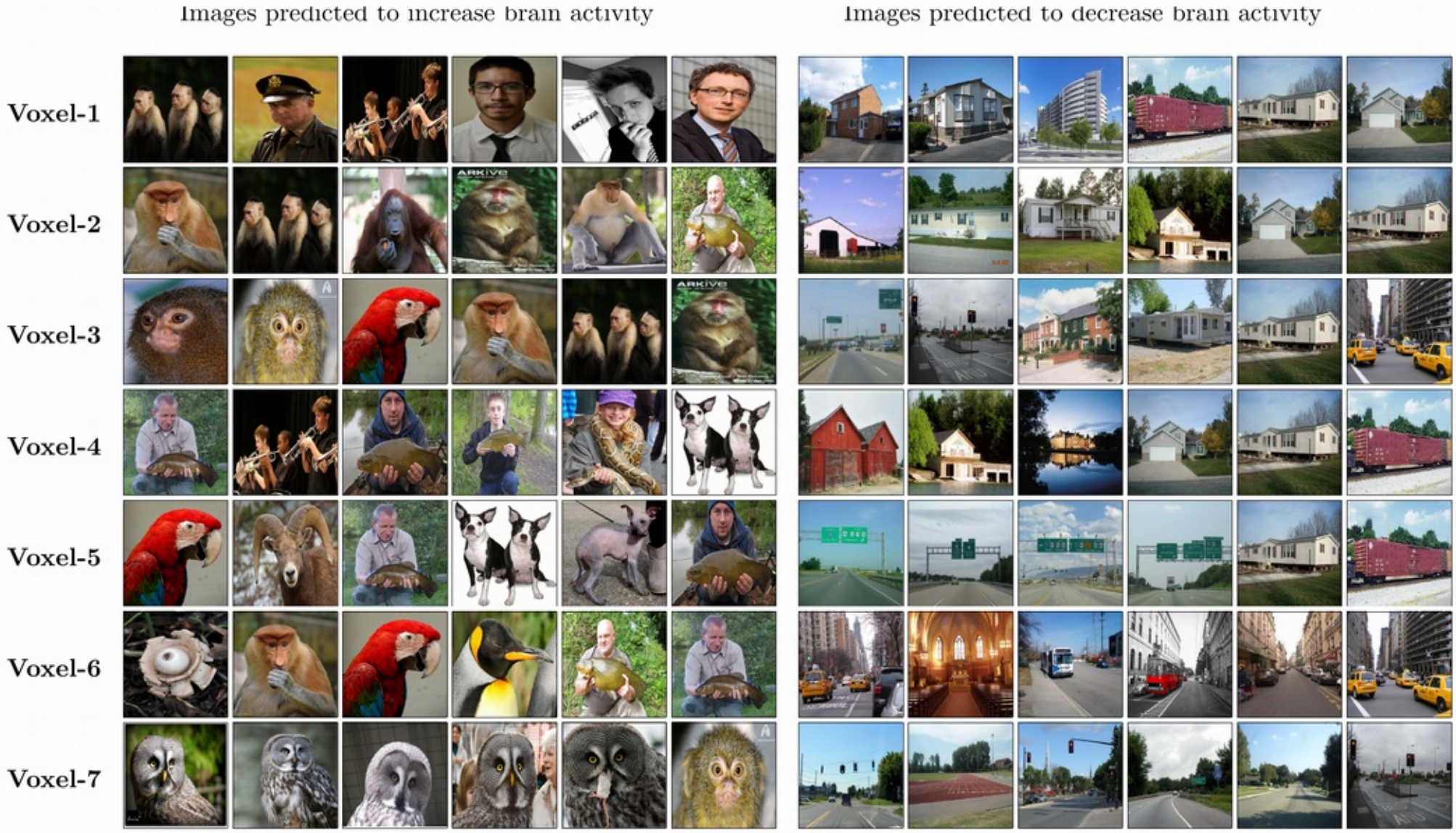
Agrawal, Stansbury, Malik & Gallant, *in preparation*

V4 voxels are selective for curvature & texture



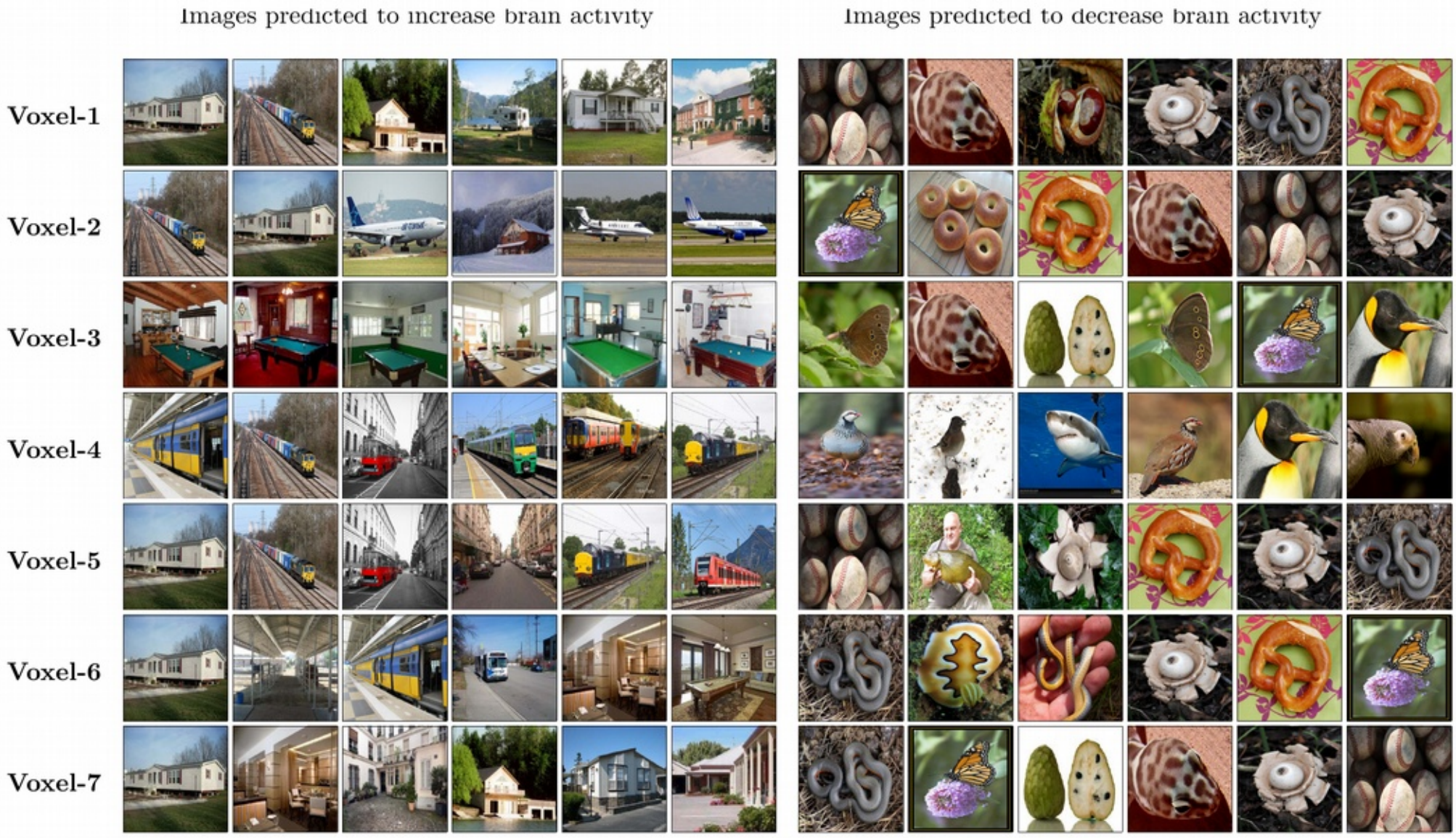
Agrawal, Stansbury, Malik & Gallant, *in preparation*

FFA voxels are selective for faces



Agrawal, Stansbury, Malik & Gallant, *in preparation*

PPA voxels are selective for places

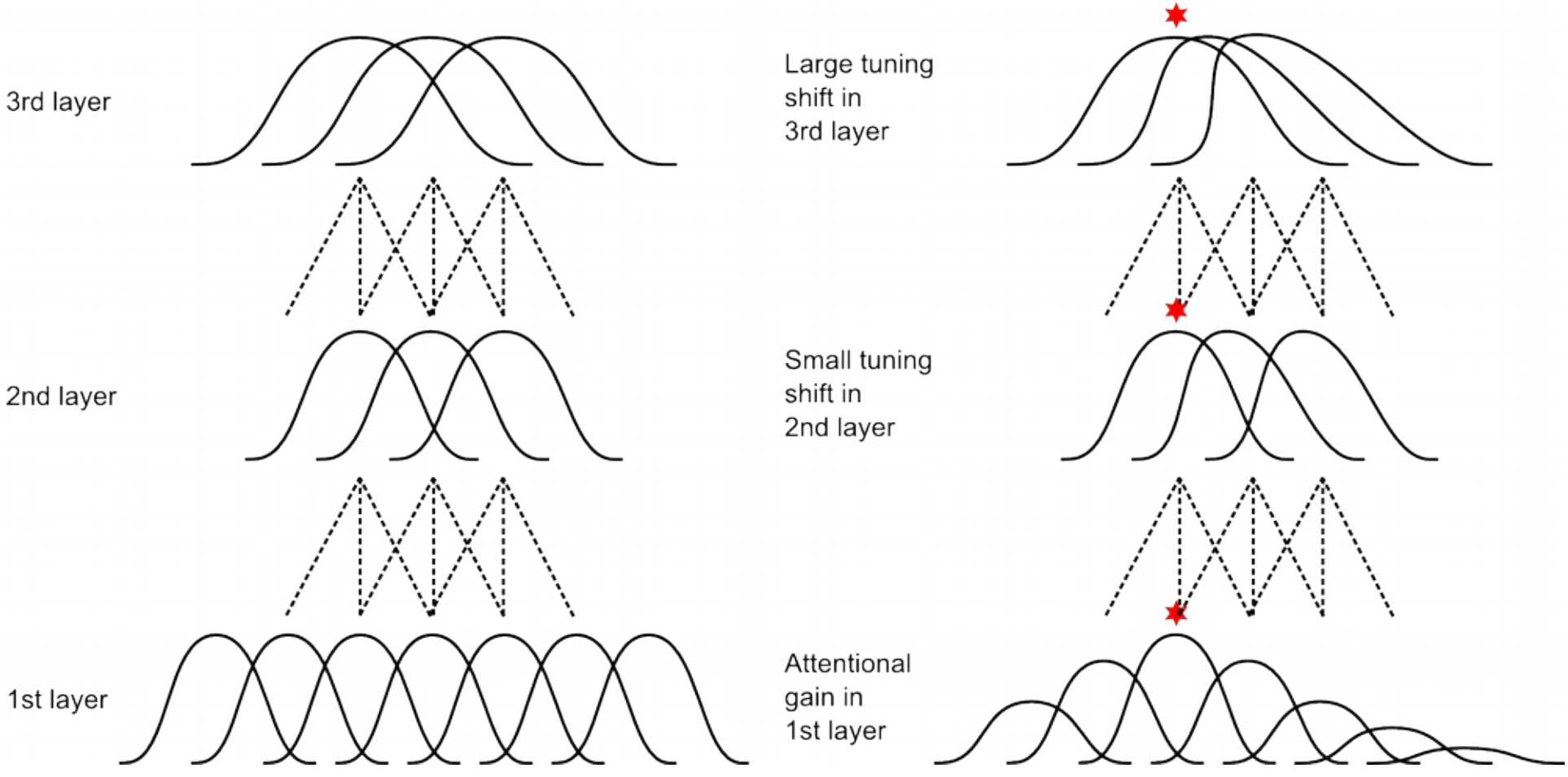


Agrawal, Stansbury, Malik & Gallant, *in preparation*

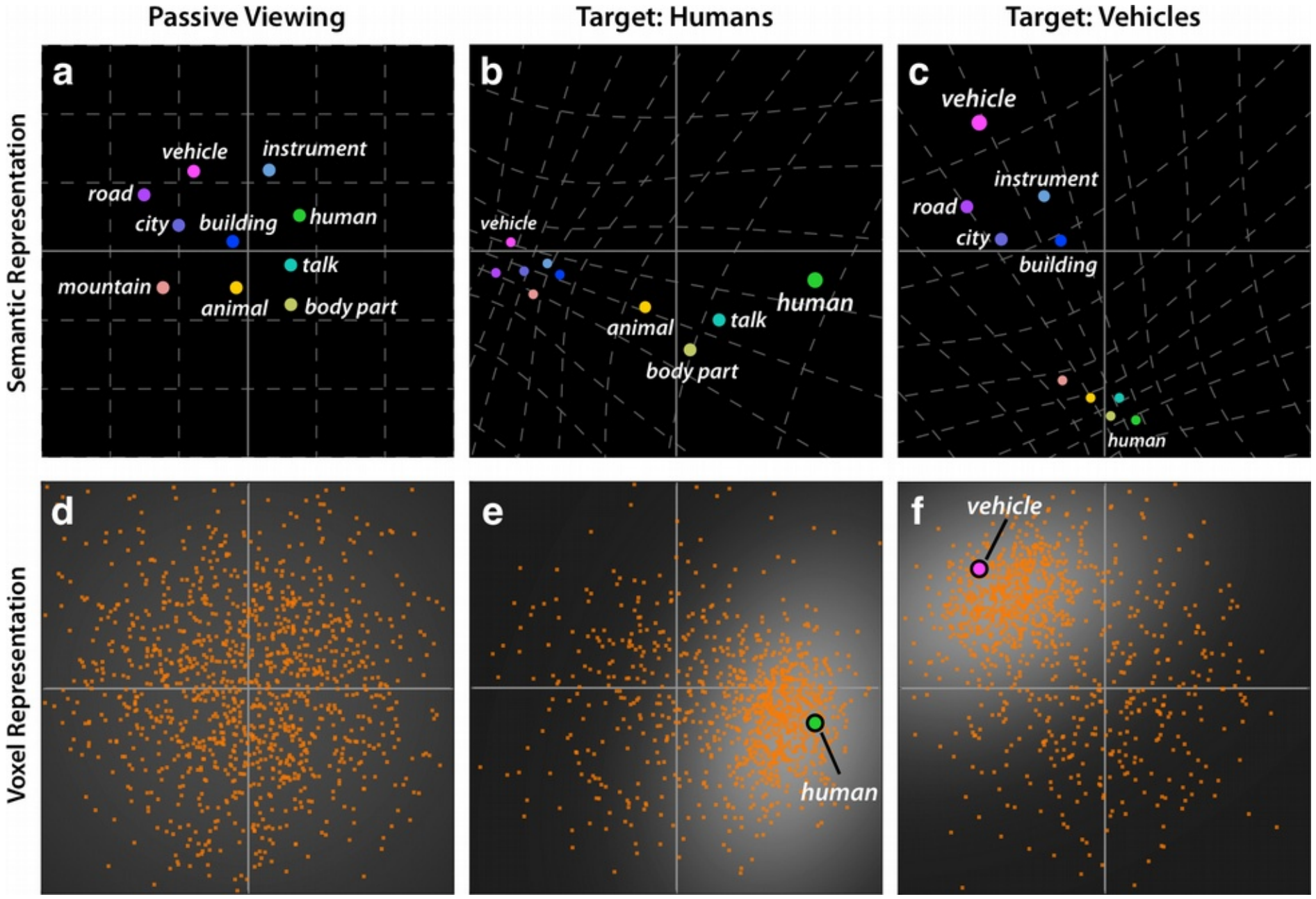
Plans and future directions

- Examine other artificial neural network architectures that might produce better voxel-wise models than we get from Caffe/AlexNet.
- Perform spatial clustering on the fit models to verify existing functional parcellations and to discover new ones.
- Use the fit models to better understand what information is represented in each visual area.
- Use this modeling network to analyze data obtained in attention experiments.

Changes in attentional gain will change tuning



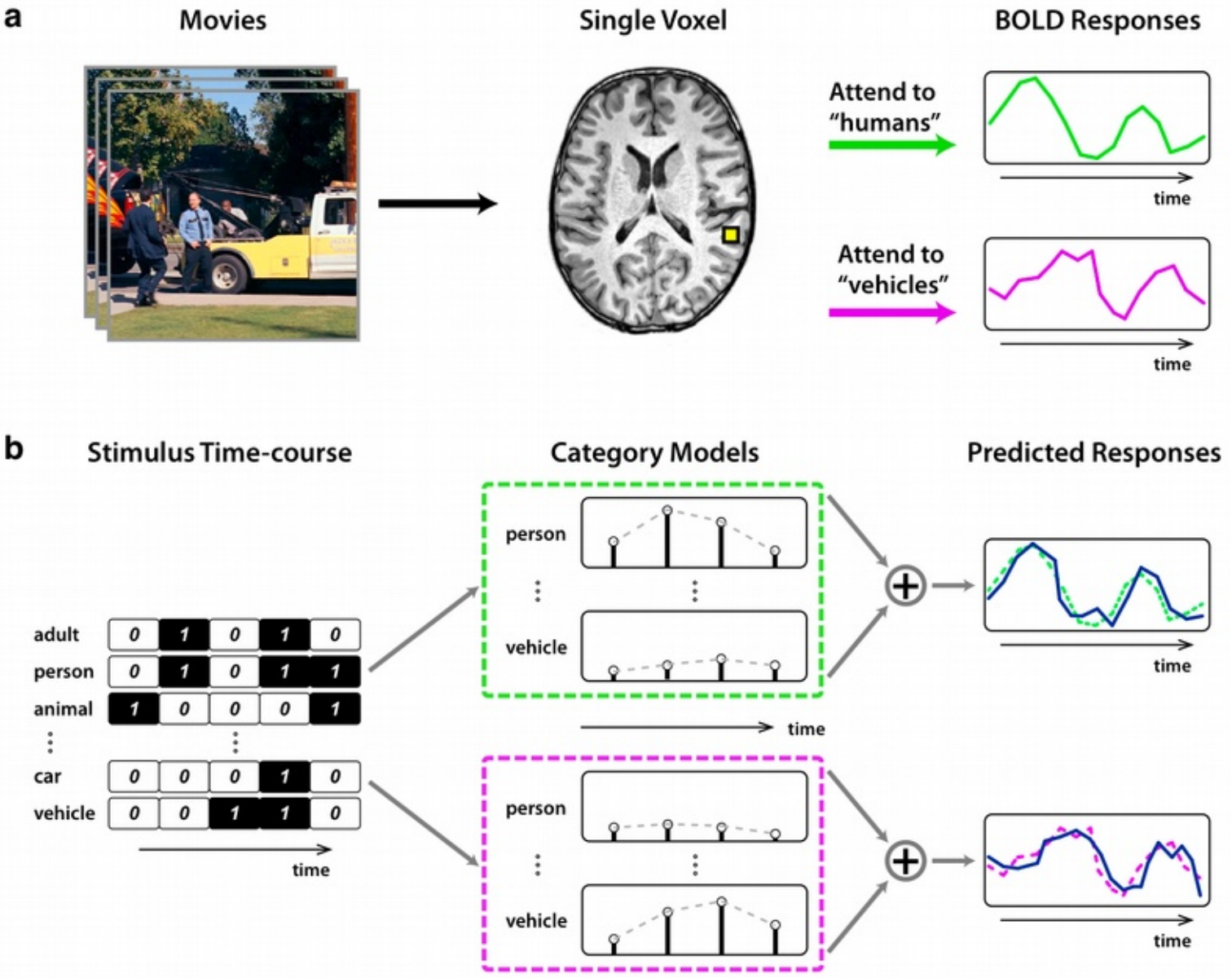
Top-down influence and the matched filter



Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

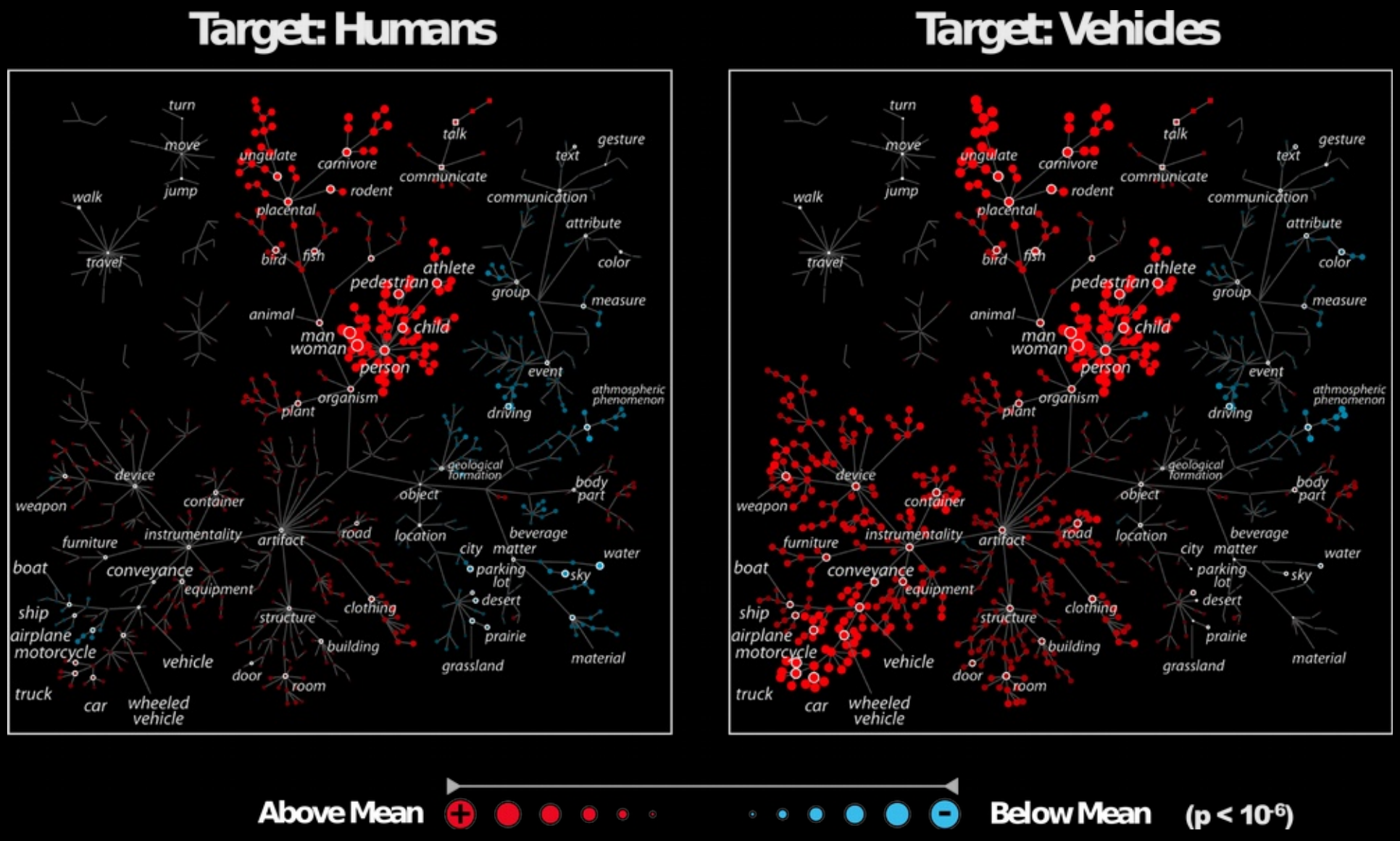
Fitting the category attention model





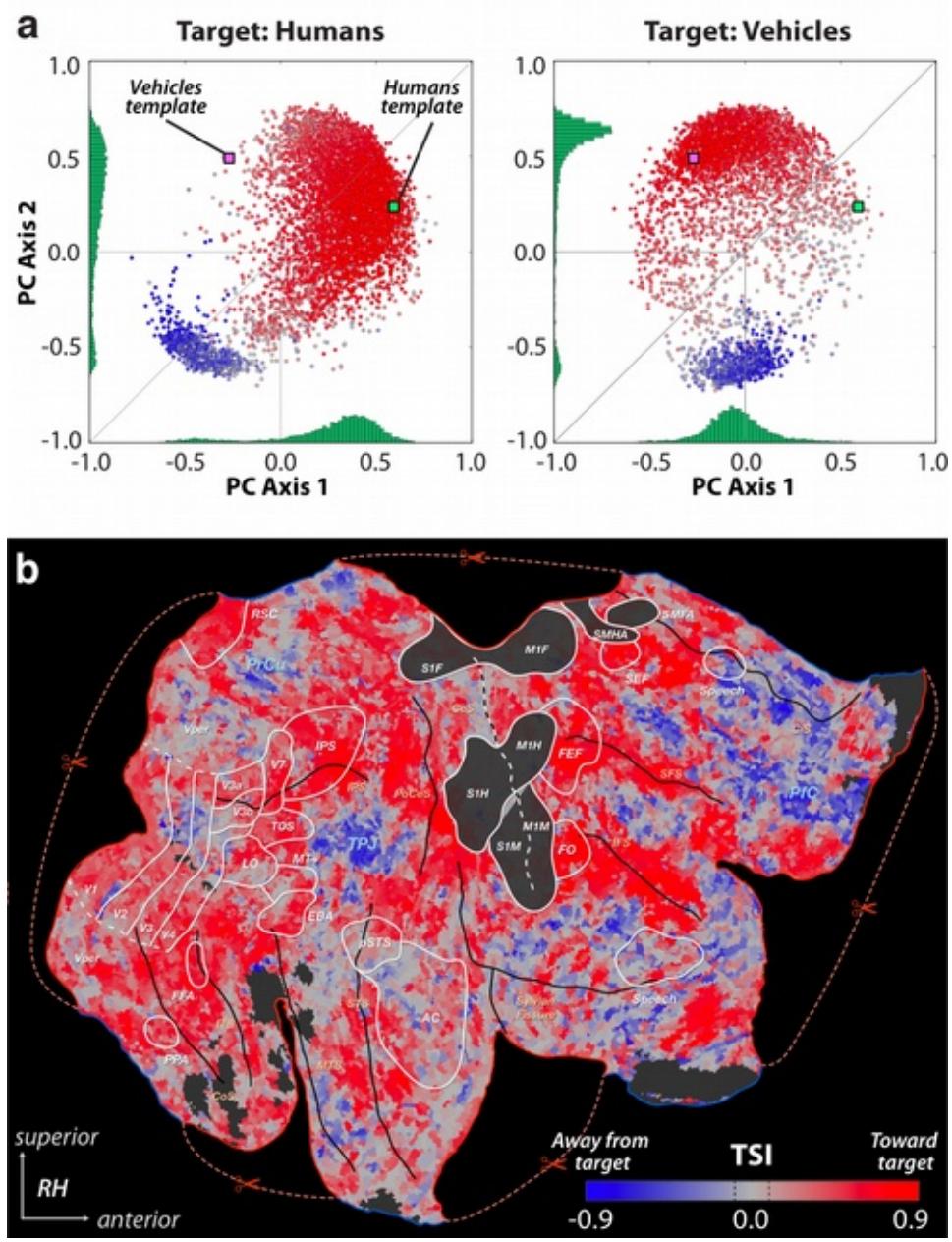
Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

Attentional tuning shifts in a single FFA voxel



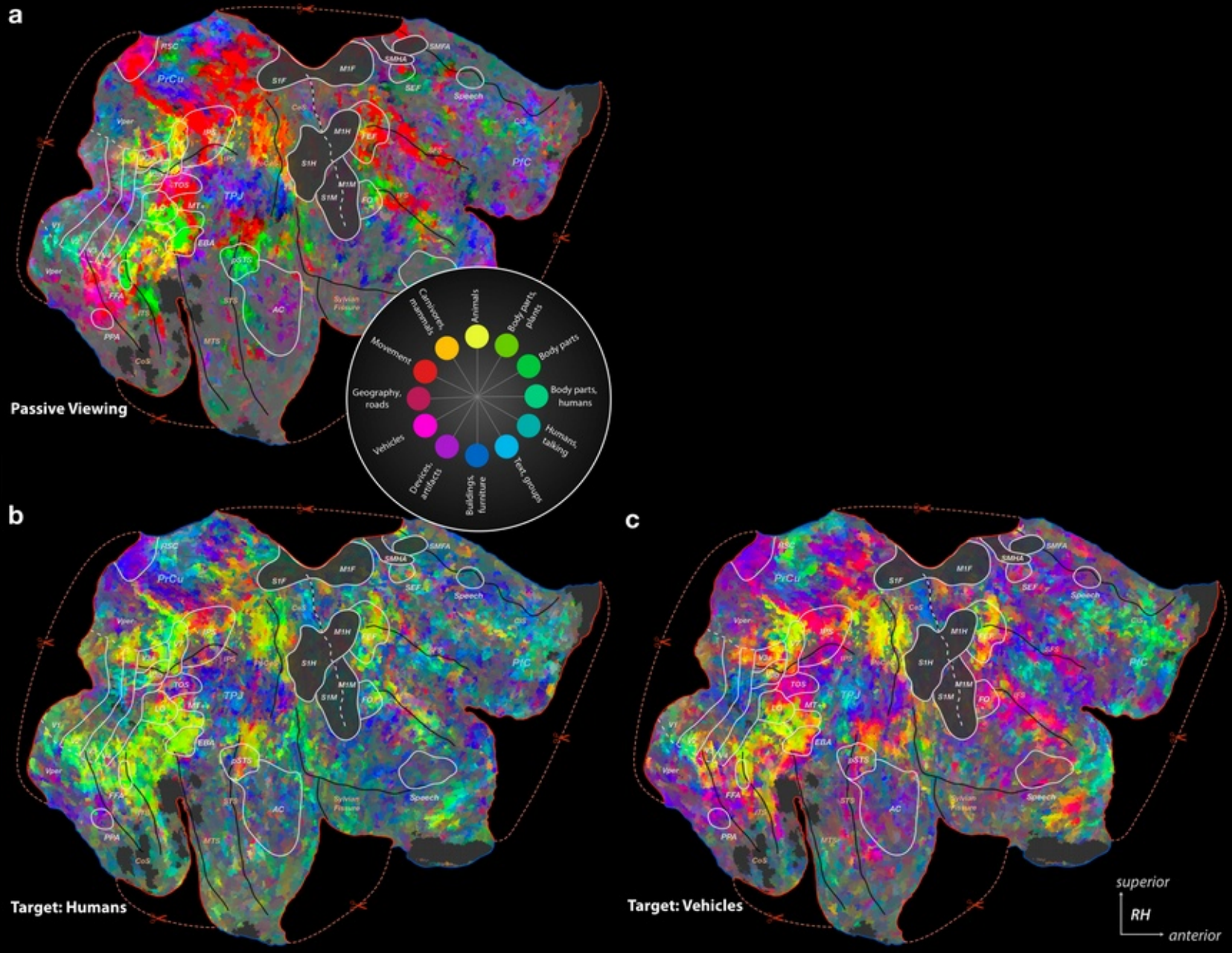
Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

Attention shifts category tuning in single voxels



Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

Attention to object categories changes maps



Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

Plans and future directions

- Use good and bad images from the VGG set.
- Evaluate several different semantic hierarchies from different regions of wordnet.

Vertebrates
    Mammals
        Primates
        Carnivores
    Birds
    Reptiles
        Snakes
        Lizards
Bugs
    Butterflies
    Worms
    Spiders

- Explore fine-grained discriminations ("Is there a person wearing glasses?" "Is there a person who looks angry?")
- Use image sets where the distractors are relatively homogeneous and highly similar to the target (e.g., horses in other large four legged mammals).
- Examine attention to pose.

We can use a scene model to decode objects

Image	Likely Scene	Likely Objects	Image	Likely Scene	Likely Objects
	land animals few people food fields water animals	man woman person head animal		urban areas large crowds lecture hall	sky car building road people
	sporting event large crowds few people lecture hall fields	athlete people man woman person		fields fenced areas large crowds birds living area	grass trees fence field ground
	signs/text living area food manmade objects	text sign washbowl beverage background		food few people human indoors	food fruit vegetables vegetable container

Stansbury, Naselaris & Gallant, *Neuron*, 2013

Current Lab Members

Natalia Bilenko
James Gao
Alex Huth
Fatma Imamoglu
Mark Lescroart
Lydia Majur
Anwar Nunez
Michael Oliver

Collaborating labs

Frederic Theunissen, Bin Yu, Jitendra Malik, Tom Griffiths, Ralph Adolphs, Bertrand Thirion, Essa Yacoub, Kamil Ugurbil



Past Lab Members

Tolga Cukur (Bilkent U)
Stephen David (OHSU))
Kate Gustavsen (Vet)
Kathleen Hansen (OHSU)
Ben Hayden (URochester)
Kendrick Kay (WashU)
James Mazer (Yale)
Thomas Naselaris (MUSC)
Shinji Nishimoto (NIST Japan)
Ryan Prenger (Baidu)
Dustin Stansbury (8Trax)
Ben Willmore (Cambridge)
Bill Vinje (Salk)
An Vu (UMinn)
Michael Wu (Lithium)

This work was supported by NEI, NIMH, NSF & ONR