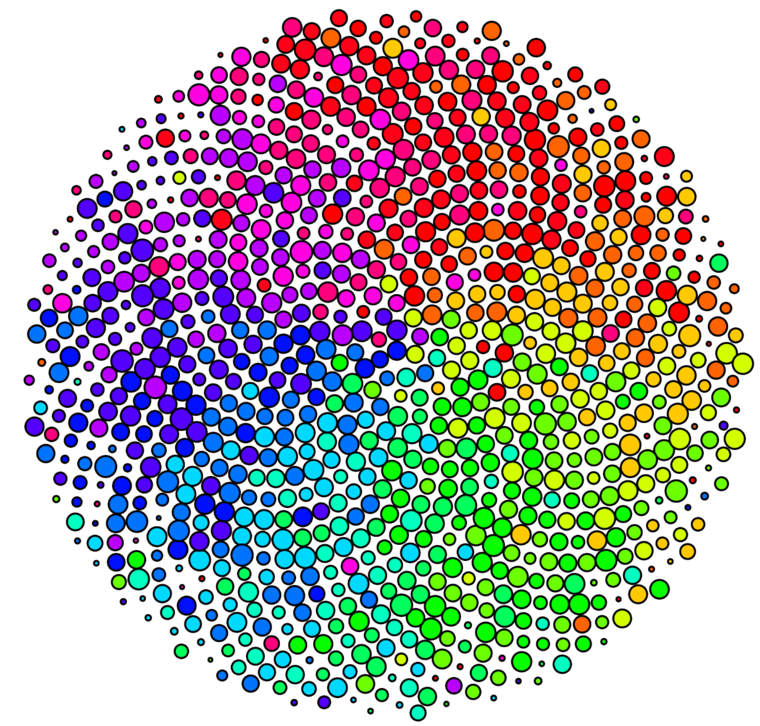
Topographic deep neural networks predict the functional organization of the primate ventral visual pathway

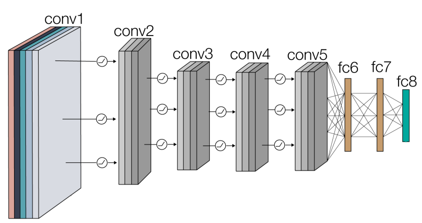


Eshed Margalit, Hyodong Lee, James J. DiCarlo, Kalanit Grill-Spector, and Daniel L.K. Yamins

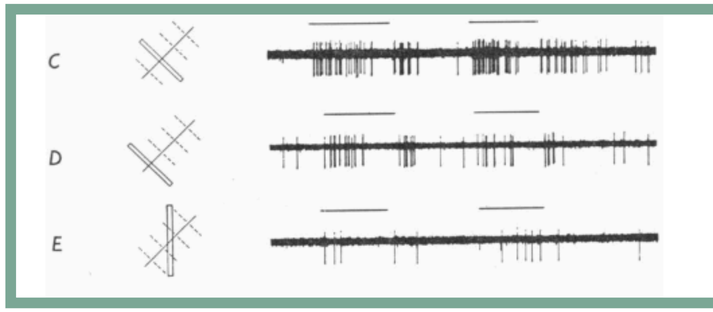
The Ventral Visual Pathway: Features in Space

Response Properties

Well-predicted by task-optimized deep convolutional neural networks (DCNNs)[1,2,3]

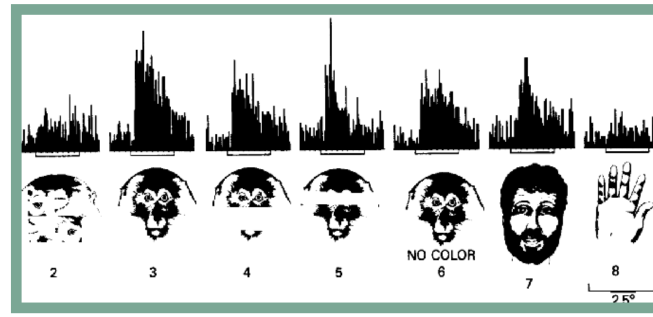


Orientation Detectors



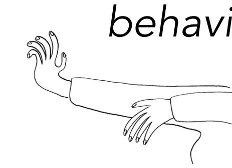
Hubel and Wiesel, 1962

Face Detectors



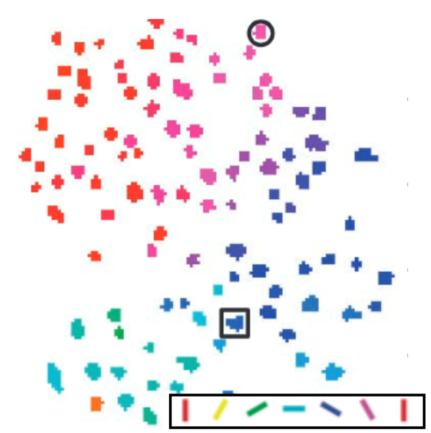
Desimone et al., 1984

macaque — Retina — LGN — V1 — V2 — V4 — IT → *behavior*



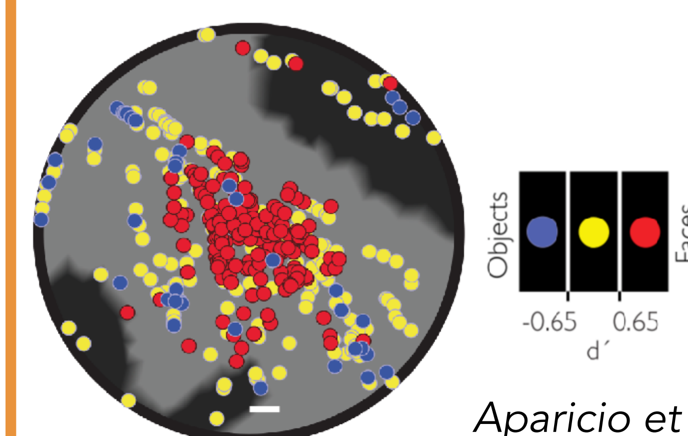
Topographic Properties

Not predicted by *any* single model, including DCNNs



Nauhaus et al., 2012

Orientation Clustering



Aparicio et al., 2016

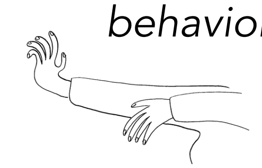
Face Clustering

[1] Yamins et al., 2014
[2] Cadena et al., 2019
[3] Schrimpf et al., 2020

Hypothesis

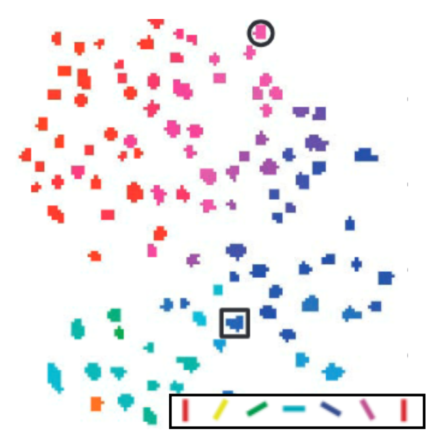
Topographic properties emerge from **a bias for nearby neurons to be correlated** in their responses to natural images during representation learning

macaque — Retina — LGN — V1 — V2 — V4 — IT → *behavior*



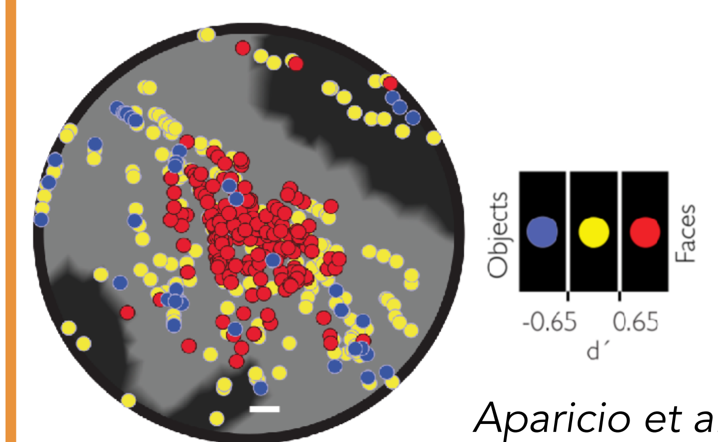
Topographic Properties

Not predicted by *any* single model, including DCNNs



Nauhaus et al., 2012

Orientation Clustering



Aparicio et al., 2016

Face Clustering

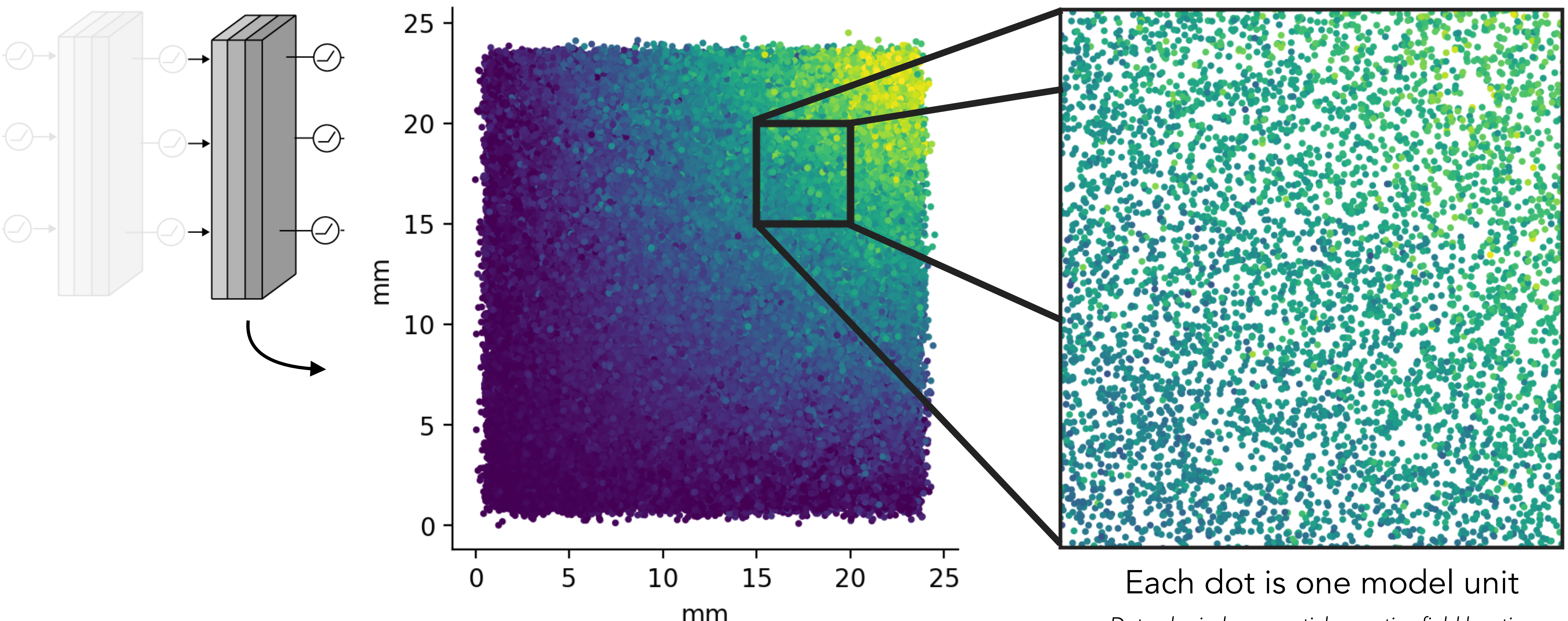
Approach

1. Augment DCNNs by assigning a spatial position to each model neuron
2. Train the model to learn useful representations from natural images, while keeping nearby model neurons correlated
3. Test the model for topographic properties using the same stimuli and metrics used in the lab

1 Each model unit is assigned a position

Placement of units in convolutional layers respects retinotopy

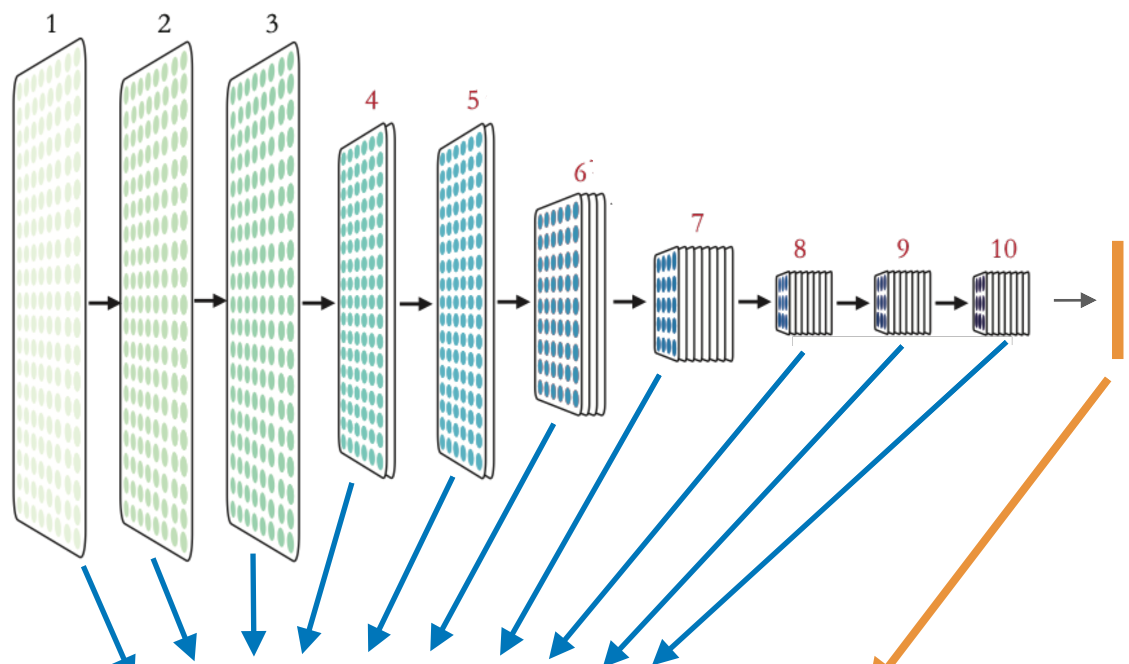
Simulated 2D Cortical Sheet



Each dot is one model unit

Dot color indexes spatial receptive field location

2 Train model to minimize the sum of task + spatial losses



$$\textbf{Loss} = \sum_{l \in \textbf{layers}} \underbrace{\alpha_l L_{\textbf{spatial}}(y_l, \phi_l)} + \underbrace{L_{\textbf{task}}}$$

L_{task} encourages learning of useful representations, while L_{spatial} encourages nearby units to have high response correlations

L_{spatial}

L_{spatial} is minimized when nearby units are correlated

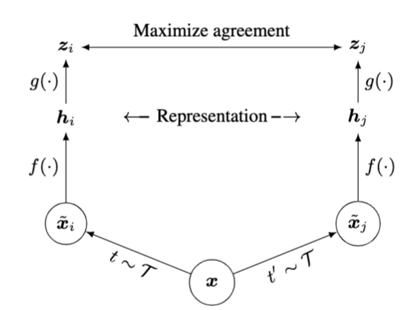
α_l → Magnitude of spatial loss at layer l

y_l → Population response at layer l

ϕ_l → Unit positions in layer l

L_{task}

Unsupervised Representation Learning

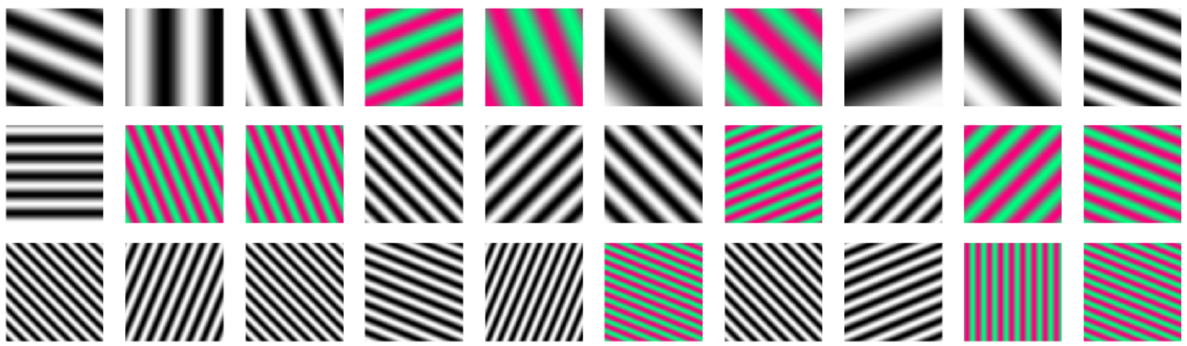


Chen et al., 2020

3

Evaluate model with test stimuli

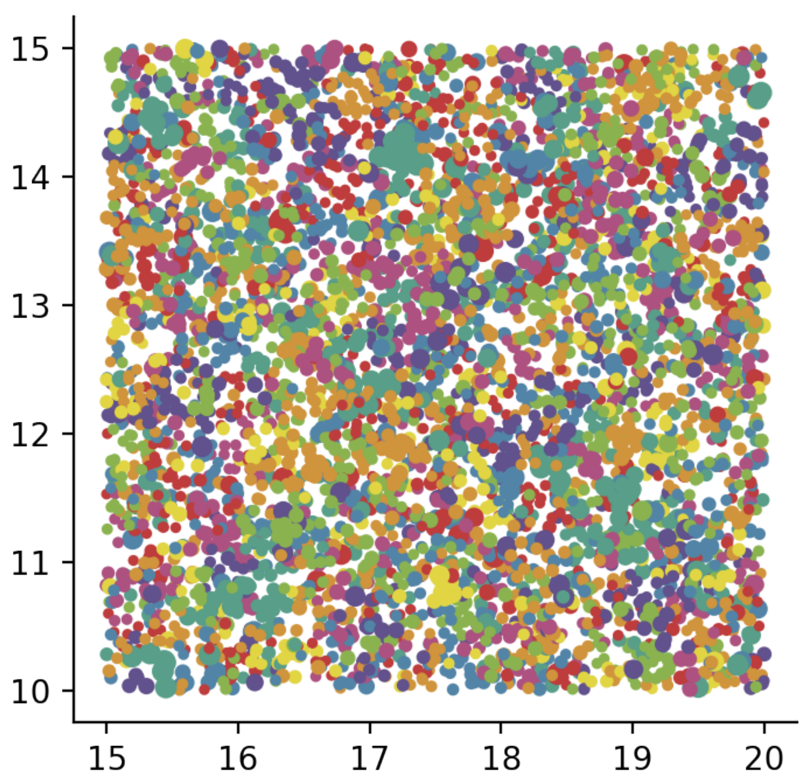
[1/2] V1-like topography | 40% through model depth



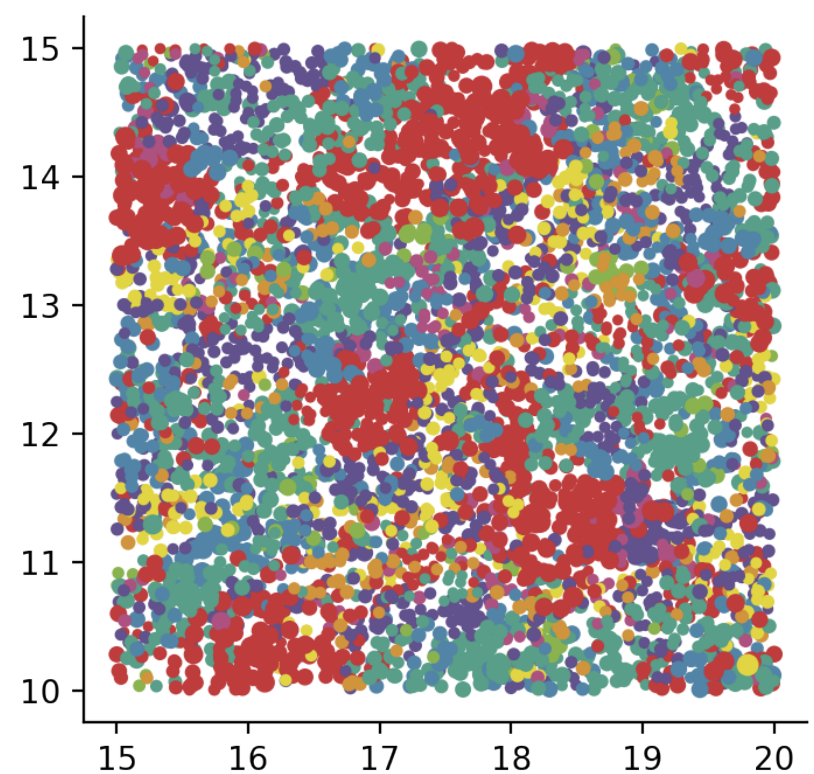
Model reproduces:

1. Smooth orientation maps
2. Clustering by spatial frequency
3. Color-tuned “blobs”
4. Cardinal orientation bias

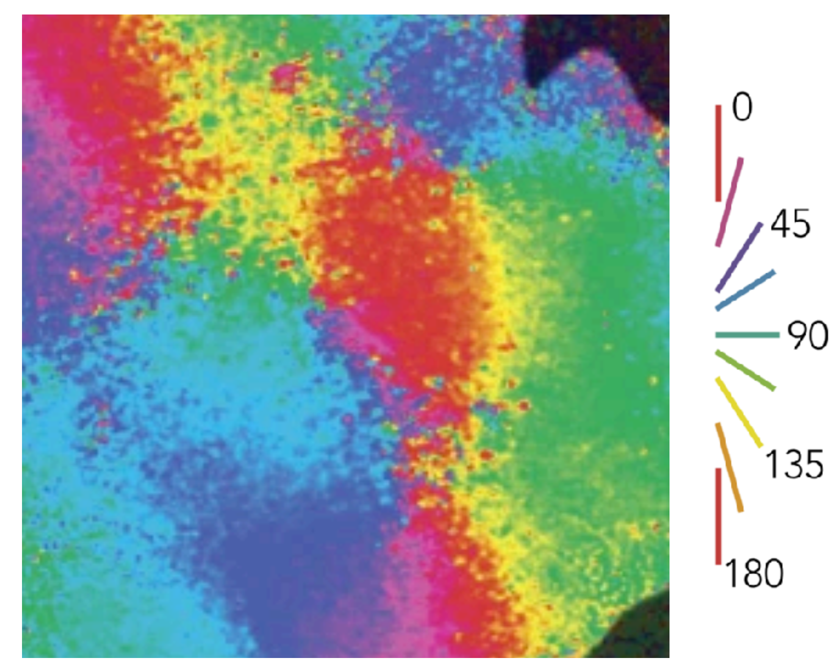
Untrained Model



Trained Model



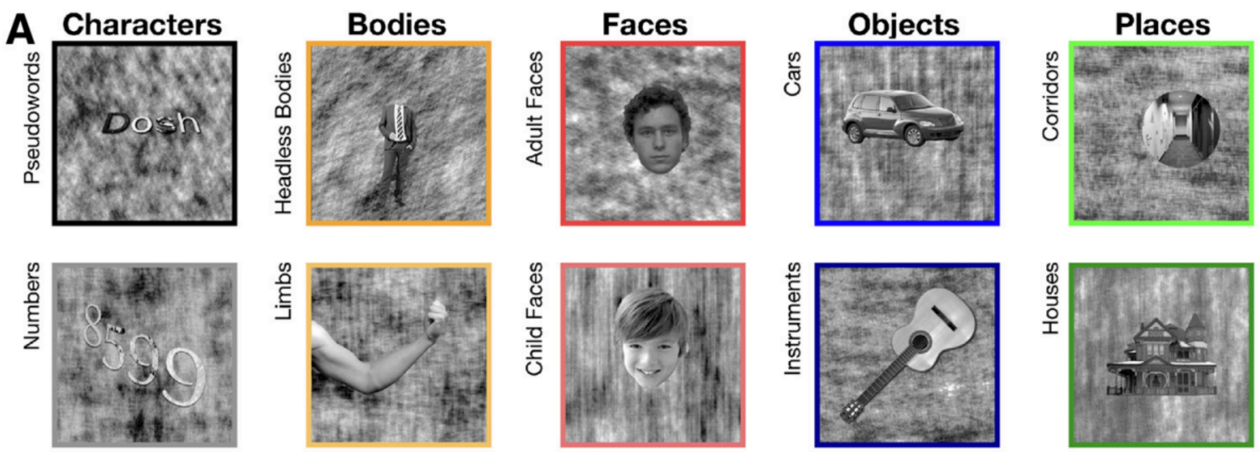
Macaque V1



Nauhaus et al., 2012

3 Evaluate model with test stimuli

[2/2] IT-like topography | 90% through model depth

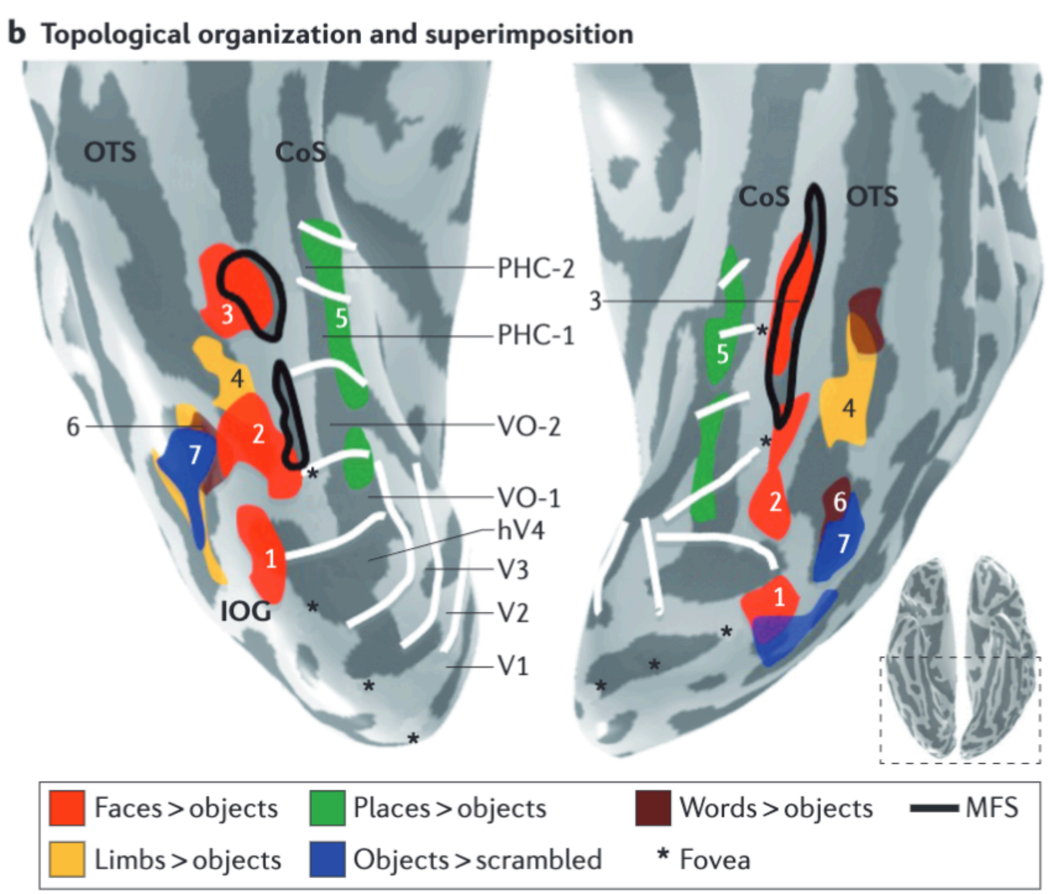


Stigliani et al., 2015; Margalit et al., 2020

Model reproduces:

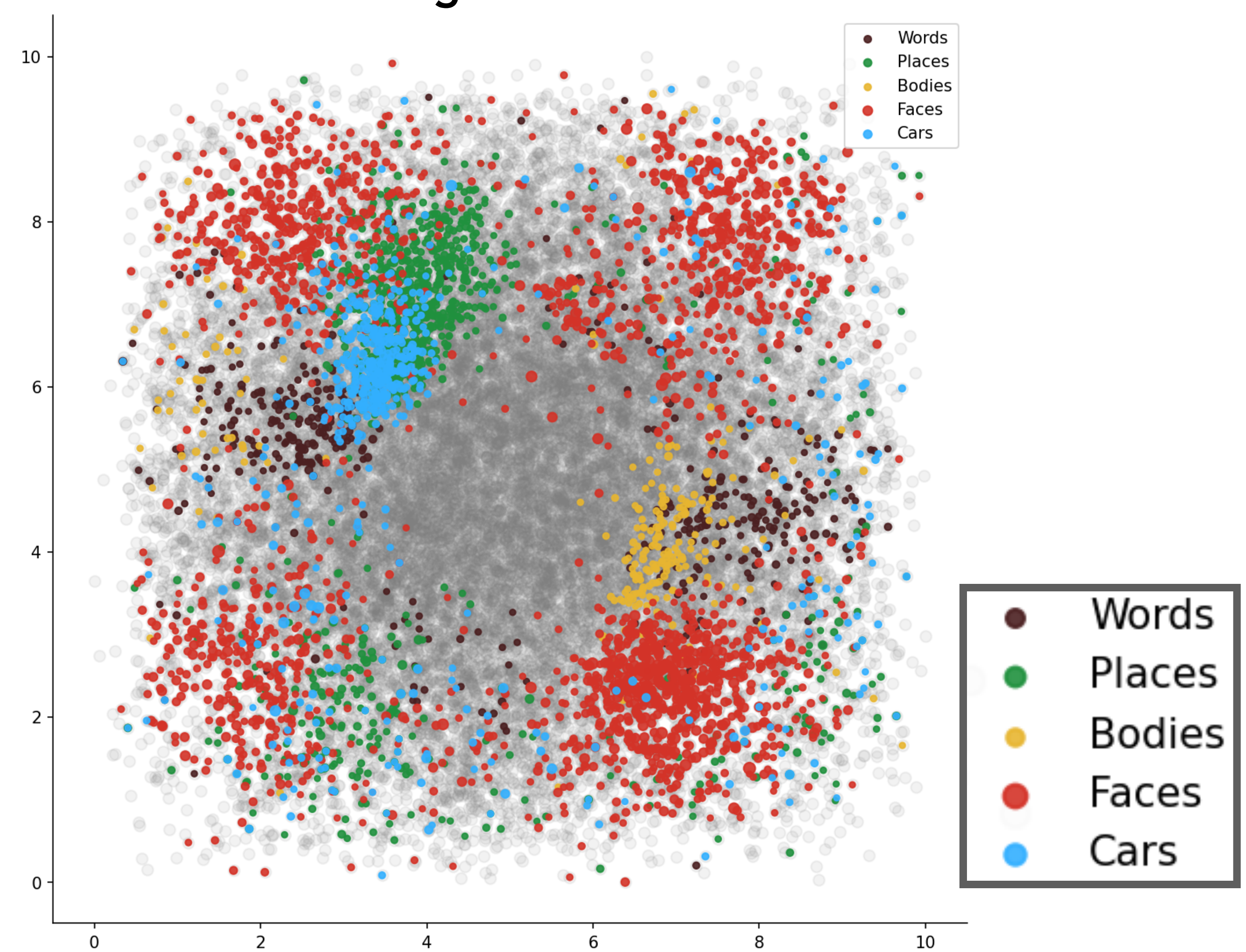
1. Multiple face patches
2. Body patches between face patches
3. Word patches near faces and bodies
4. Place-selectivity far from strong face selectivity

Human Higher Visual Cortex

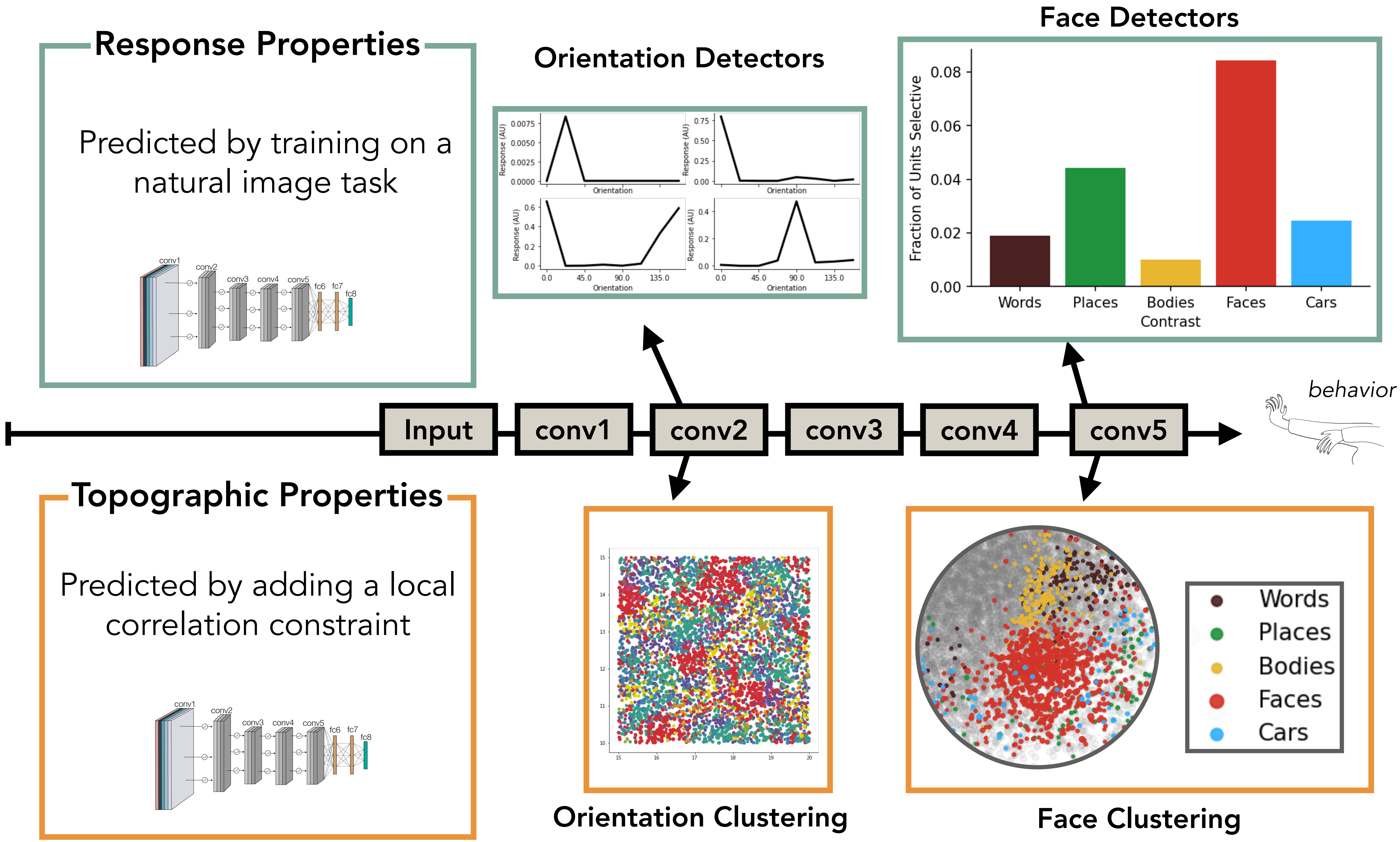


Grill-Spector and Weiner, 2014

Model Higher Visual Cortex



Topographic DCNNs are a unified model of the ventral visual pathway



Thank you!

Eshed Margalit
www.eshedmargalit.com
eshedmargalit

Ask me about…

Whether supervised and unsupervised models yield similar results

(They do not!)

Why might there be differences between supervised and unsupervised models?

Quantitative brain-model comparison

How can you compare orientation preference maps in brains and models?

Do topographic models predict neuronal responses to unseen images?

Performance-constraint tradeoffs

Does topographic structure come at a cost to model performance?

How might wiring length change with a spatial cost?