CSE 40171:
Artificial Intelligence

Artificial Neural Networks with Functional Fidelity:
Internal Behavior of Artificial and Biological Networks

Homework #6 has been released

It is due **tonight** at 11:59PM

Project Updates are Due on 11/25 at 11:59PM

(See Course Website for Instructions)

Is there any correspondence between activity measured in the brain and activity measured in artificial neural networks?

Monkey performing an object recognition task

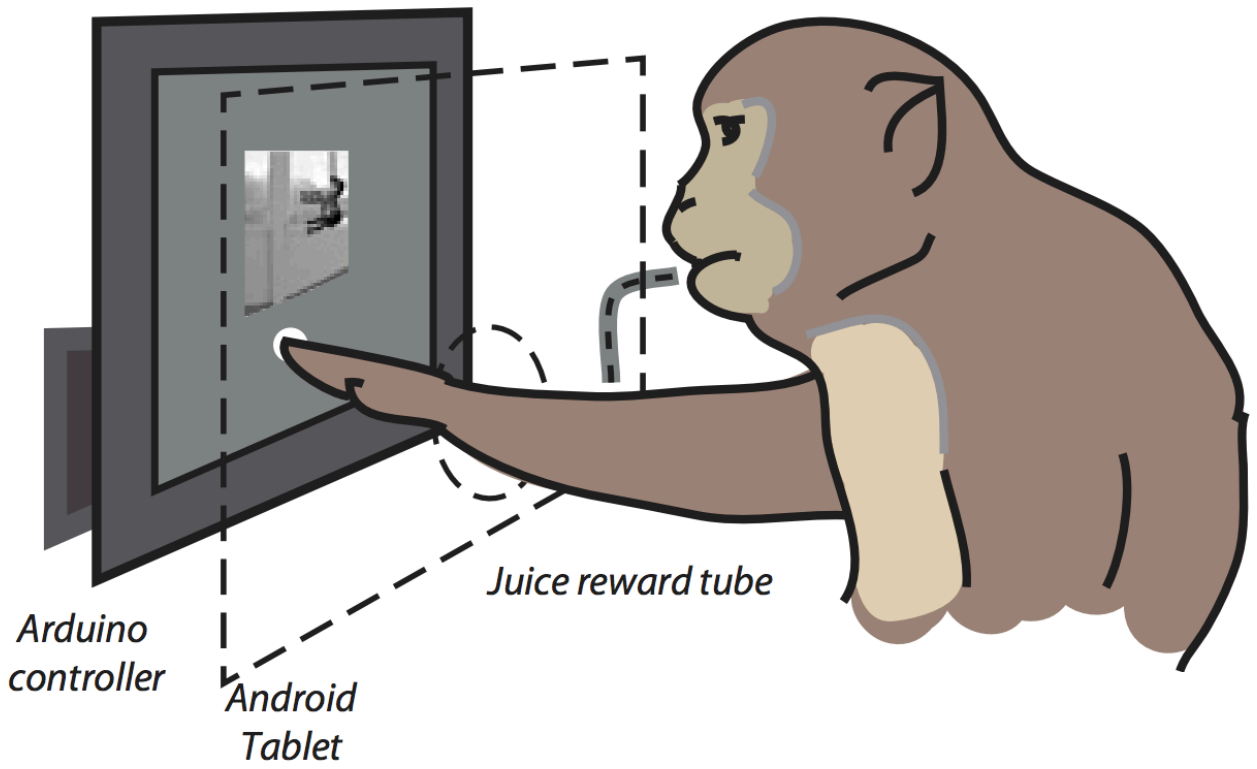
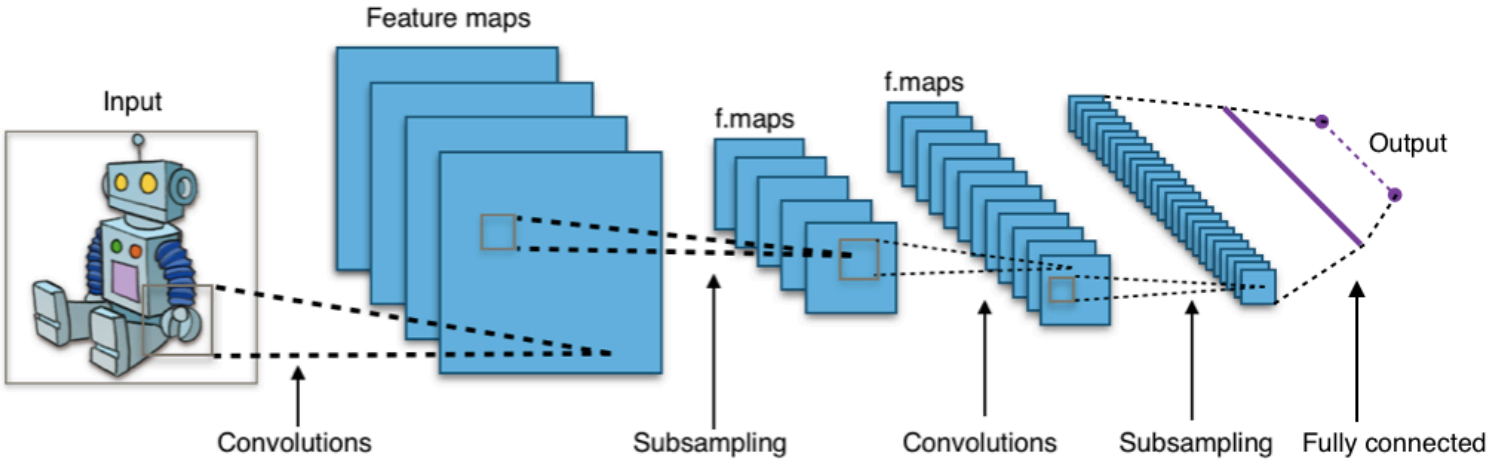


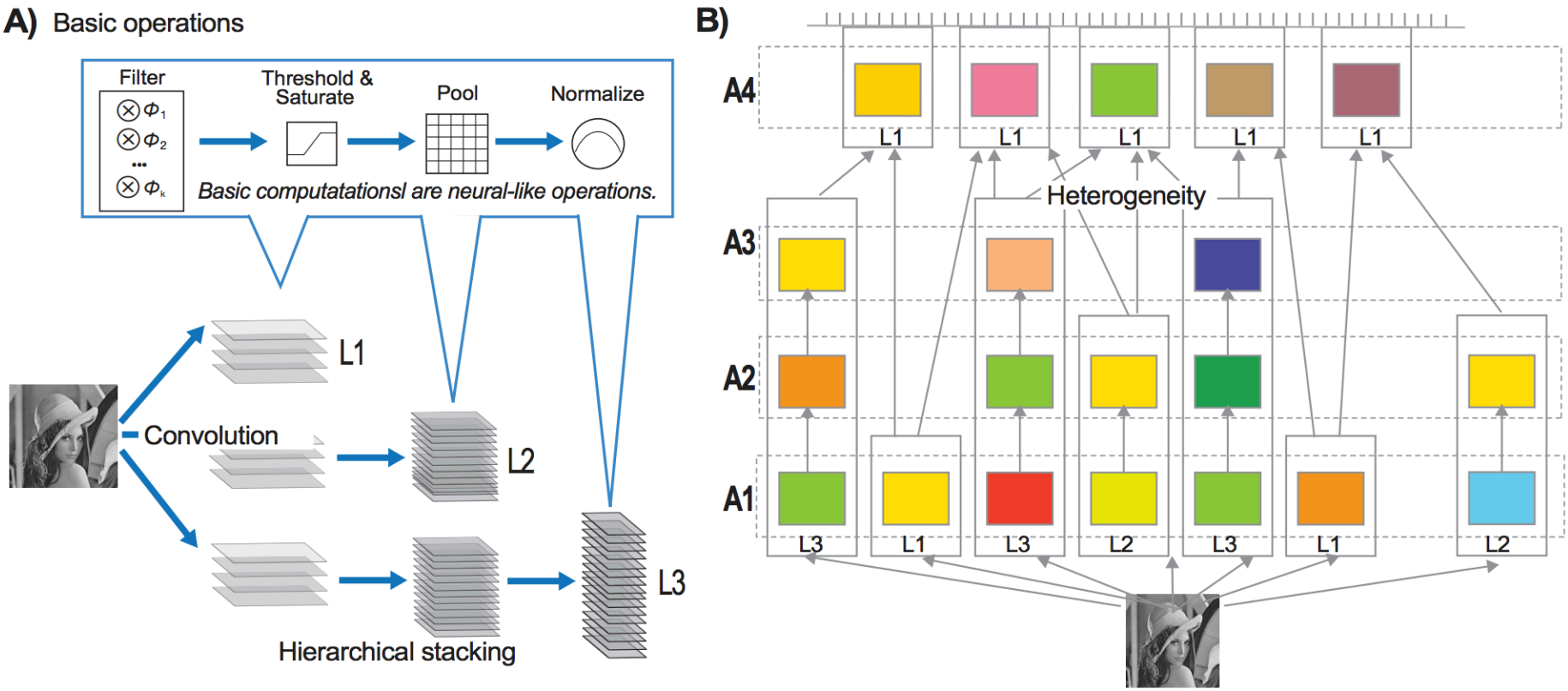
Image adapted from: Rajalingham et al. JNeurosci 2018

CNN for Object Recognition



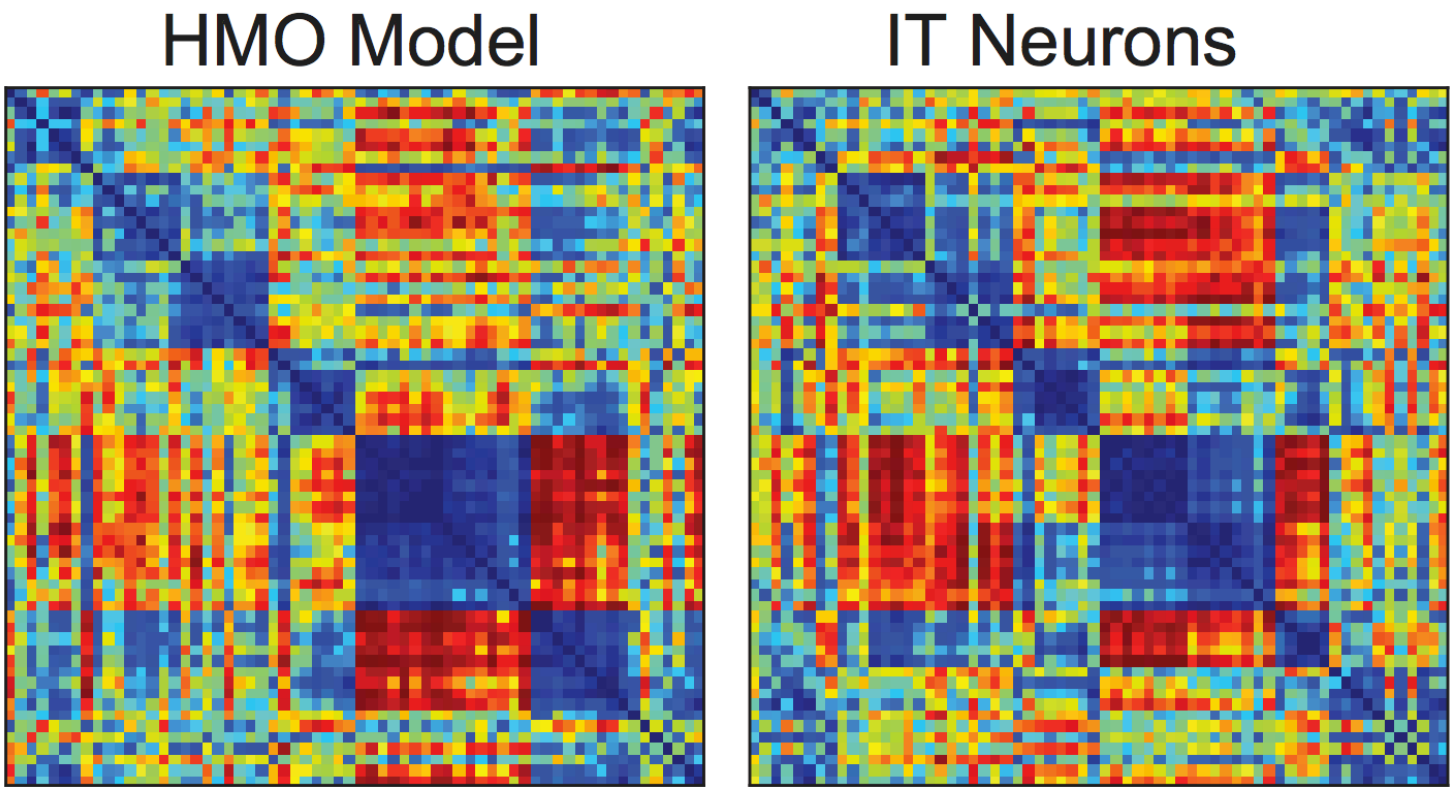
Typical CNN architecture

Heterogeneous Hierarchical CNN



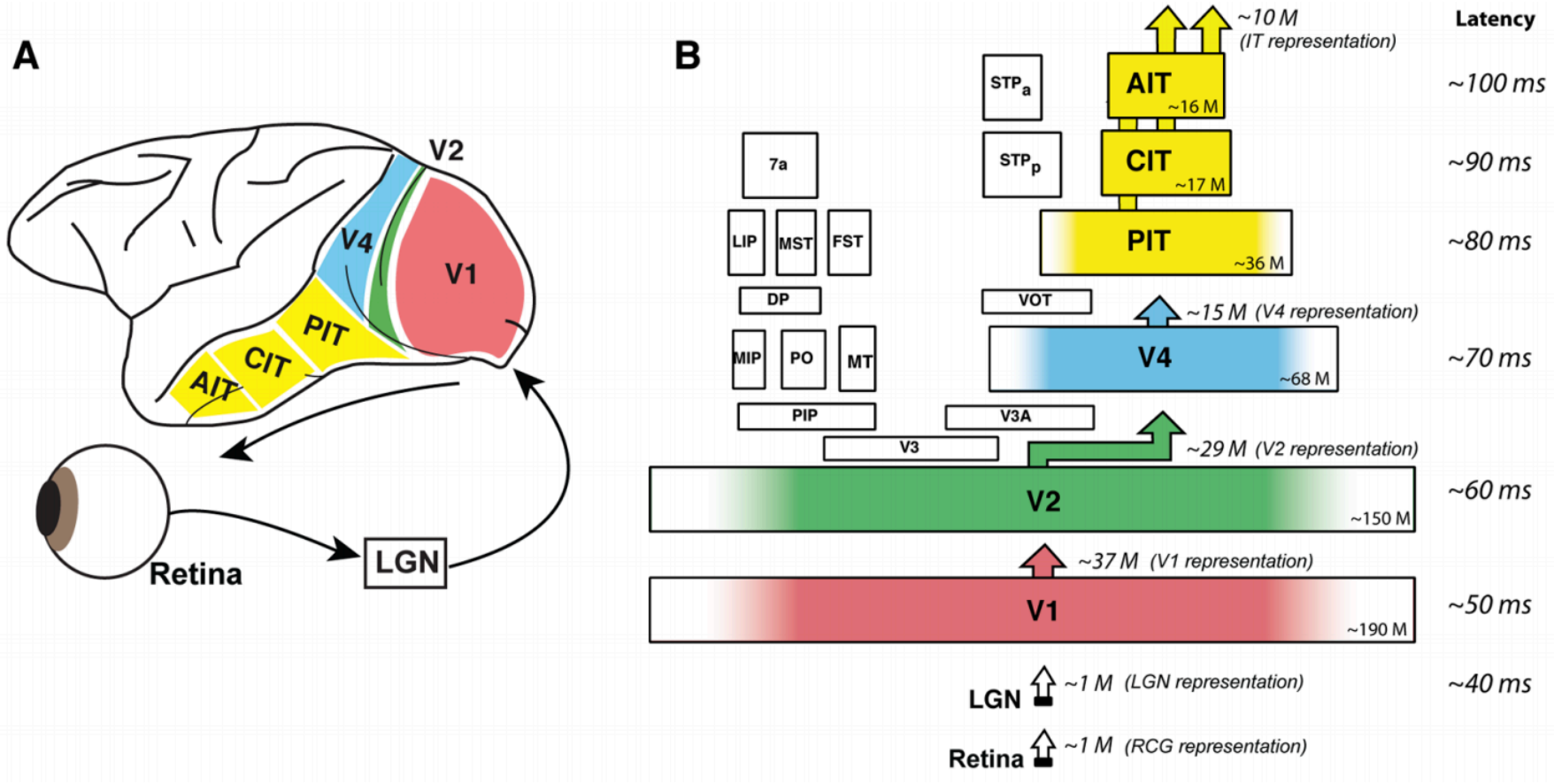
Yamins et al. NeurIPS 2013

Population Responses: Model vs. Brain



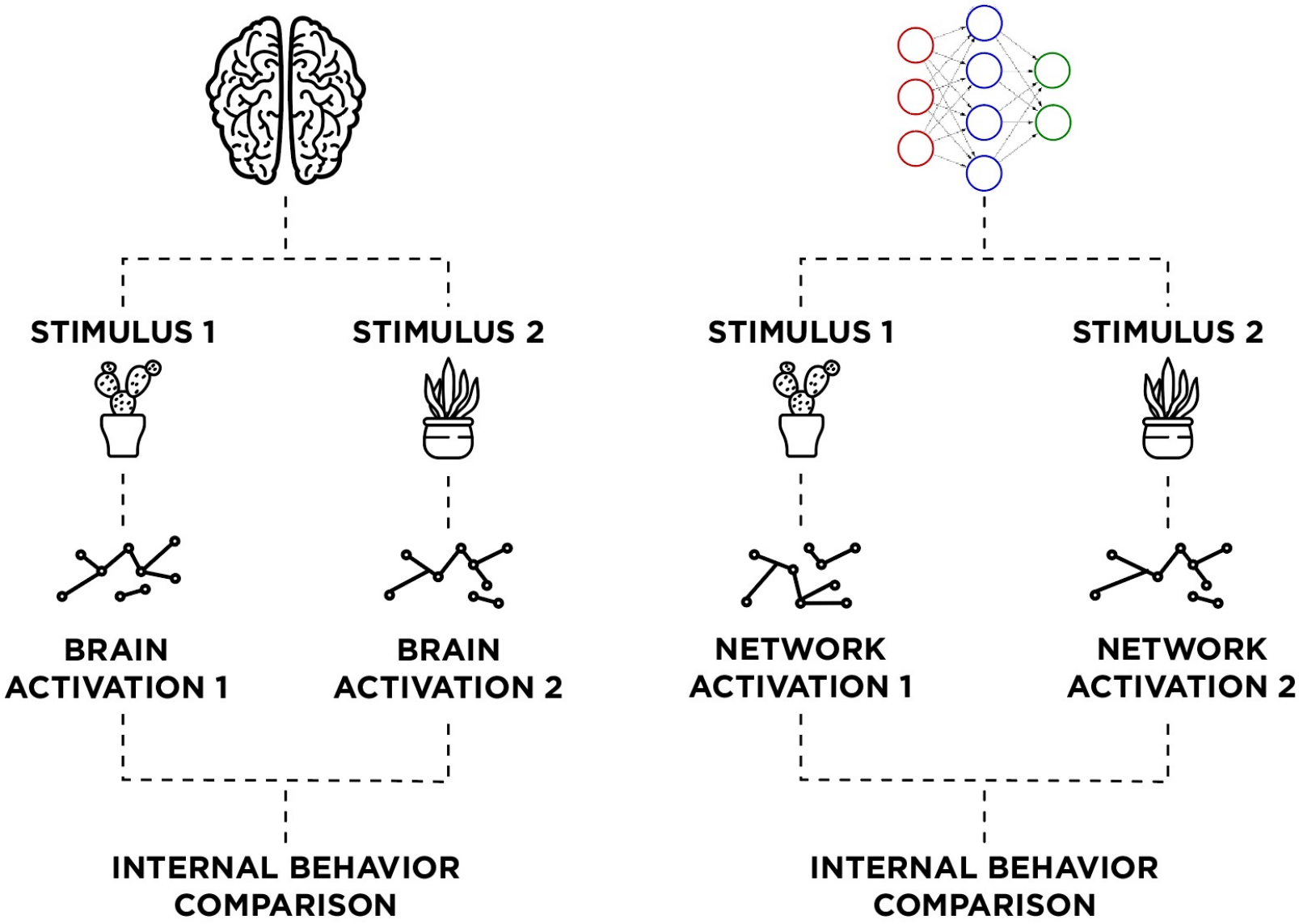
Yamins et al. NeurIPS 2013

Where in the brain is area IT?

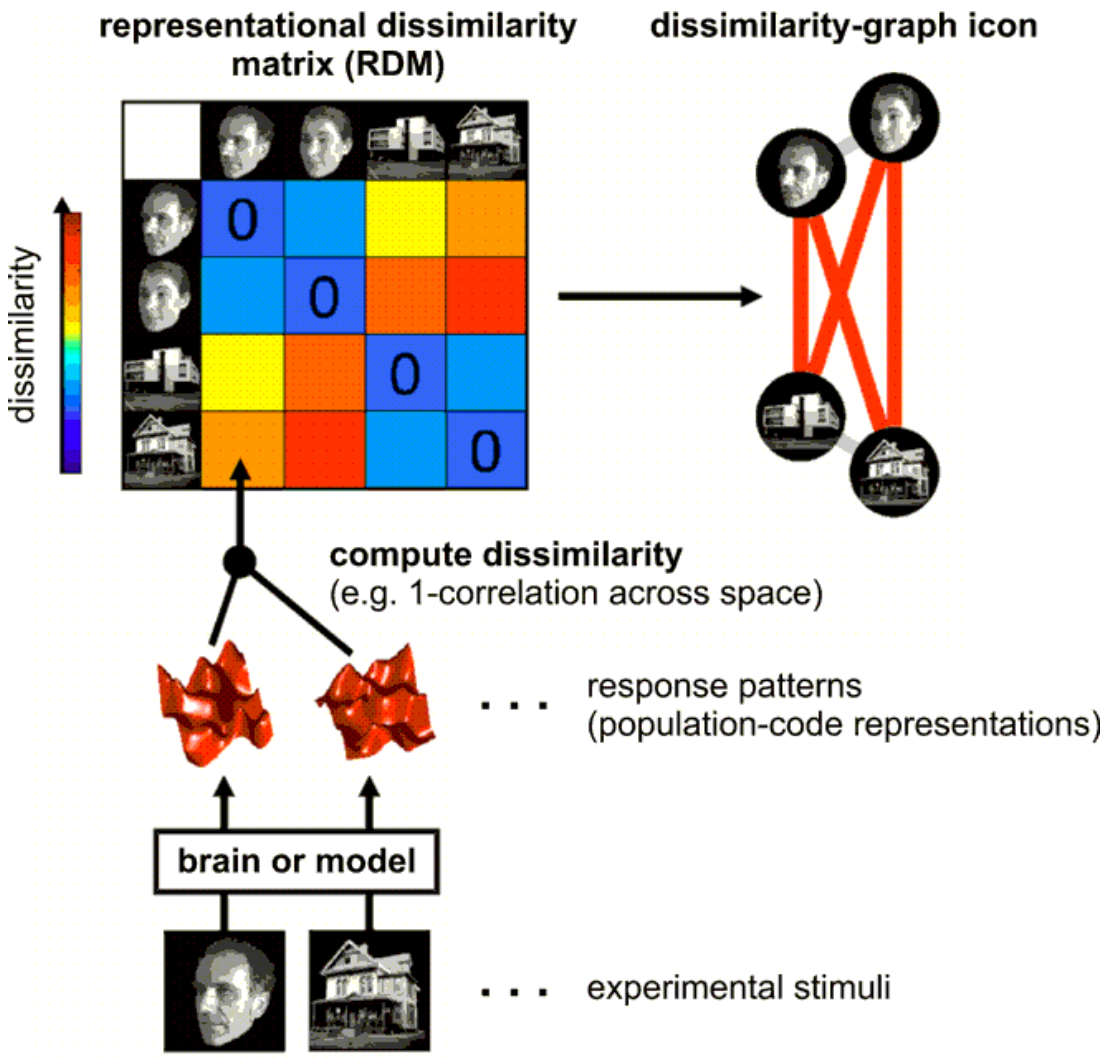


DiCarlo et al. Neuron 2012

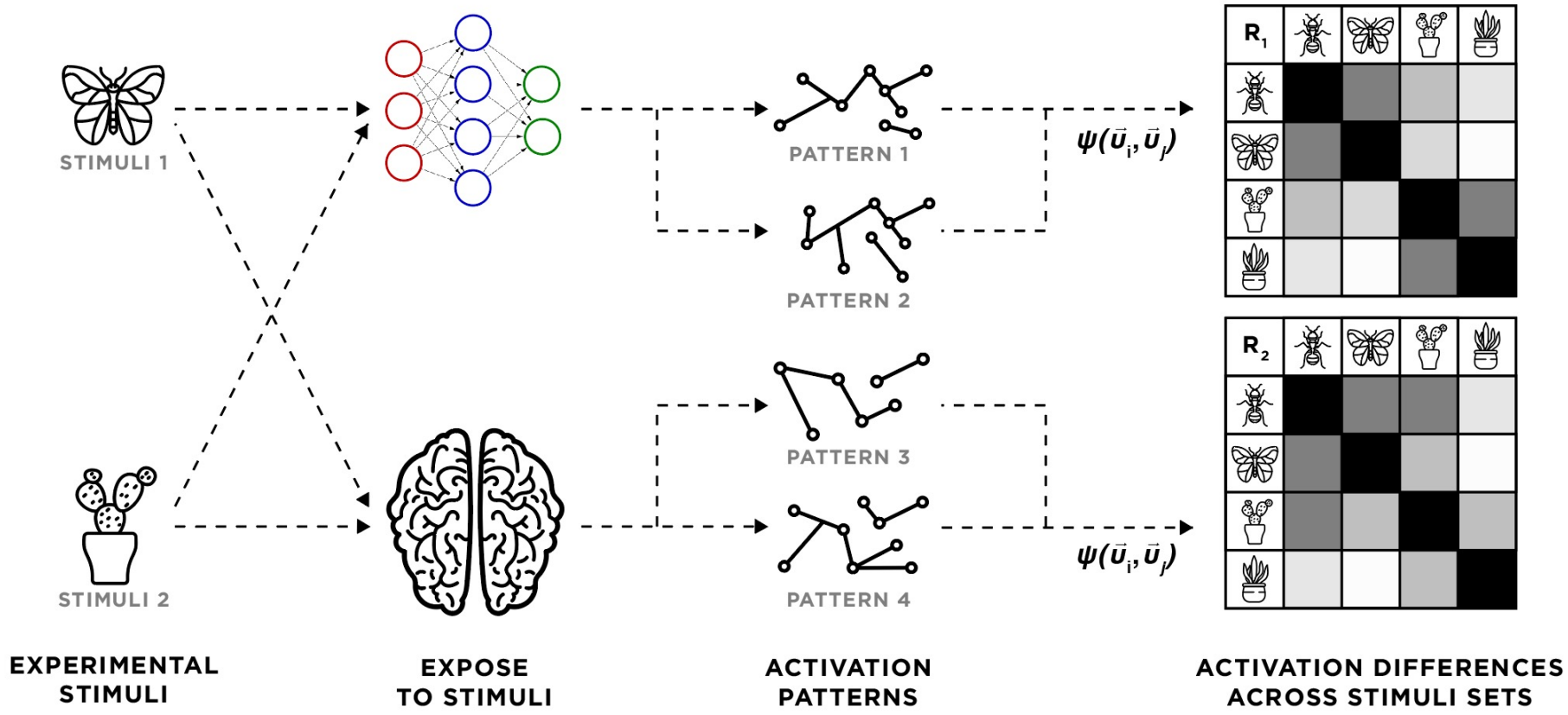
How do we compare the activity in brains with the activity in artificial neural networks?


STIMULUS 1
STIMULUS 2
BRAIN ACTIVATION 1
BRAIN ACTIVATION 2
INTERNAL BEHAVIOR COMPARISON
STIMULUS 1
STIMULUS 2
NETWORK ACTIVATION 1
NETWORK ACTIVATION 2
INTERNAL BEHAVIOR COMPARISON

Representational Similarity Analysis



N. Kriegeskorte, M. Mur and P. A. Bandettini, "Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience," Frontiers in Systems Neuroscience, 2008


STIMULI 1
STIMULI 2
PATTERN 1
PATTERN 2
PATTERN 3
PATTERN 4
$\psi(\vec{v}_i, \vec{v}_j)$
$\psi(\vec{v}_i, \vec{v}_j)$
R_1
R_2
EXPERIMENTAL STIMULI
EXPOSE TO STIMULI
ACTIVATION PATTERNS
ACTIVATION DIFFERENCES ACROSS STIMULI SETS

RDM Step 1: Data Representation

Given a single feature f and a single stimulus s, $v = f(s)$, where v is the value of feature f in response to s.
Likewise, the vector

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}^T = \begin{bmatrix} f_1(s) \\ f_2(s) \\ \vdots \\ f_n(s) \end{bmatrix}^T$$

can represent the feature values of a collection of n features, $f_1, f_2, \ldots, f_n$, in response to s.

RDM Step 1: Data Representation

If one expands the representation of s to a set of m stimuli $S = s_1, s_2, \ldots, s_m$, the natural extension of $\overrightarrow{v}$ is the set of feature value collections $V = \overrightarrow{v_1}, \overrightarrow{v_2}, \ldots, \overrightarrow{v_m}$, in which $s_i \in S$ is paired with $\overrightarrow{v_i} \in V$ for each $i = 1, 2, ..., m$.

RDM Step 2: Dissimilarity

Define the dissimilarity score between any two $\overrightarrow{v_i} \in V$ and $\overrightarrow{v_j} \in V$:

$$\psi(\vec{v}_i, \vec{v}_j) := 1 - \frac{(\vec{v}_i - \bar{v}_i) \cdot (\vec{v}_j - \bar{v}_j)}{\|\vec{v}_i - \bar{v}_i\|_2 \|\vec{v}_j - \bar{v}_j\|_2}$$

RDM Step 3: Construct Matrix

An RDM R may then be constructed from S, V , and ψ as:

$$R = \begin{bmatrix} \psi(\vec{v}_1, \vec{v}_2) & \psi(\vec{v}_1, \vec{v}_3) & \dots & \psi(\vec{v}_1, \vec{v}_m) \\ & \psi(\vec{v}_2, \vec{v}_3) & \dots & \psi(\vec{v}_2, \vec{v}_m) \\ & & \ddots & \vdots \\ & & & \psi(\vec{v}_{m-1}, \vec{v}_m) \\ & & & \end{bmatrix}$$

fMRI Experimental Setup

Data collected by the Kriegeskorte lab at the University of Cambridge*

Eight RDMs were constructed from fMRI recordings of four subjects over two sessions in response to 92 random stimuli

Recordings were from measurements of $1.95 \times 1.95 \times 2mm^3$ within an occipitotemporal measurement slab (5cm thick).

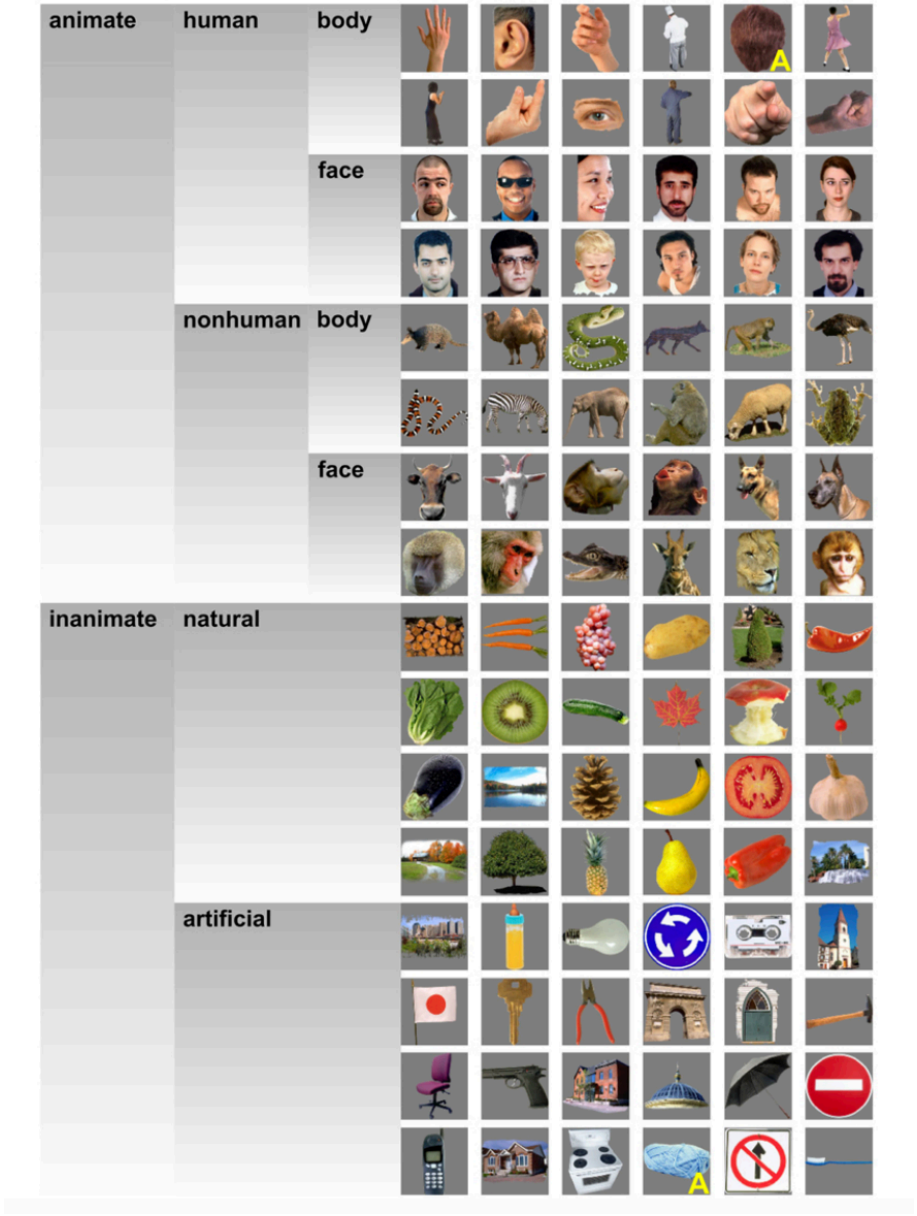
Each stimulus was displayed for 300 milliseconds, every 3700 milliseconds, with four seconds between stimuli.

Subject RDMs were averaged together into a mean human brain RDM, which reduced noise.

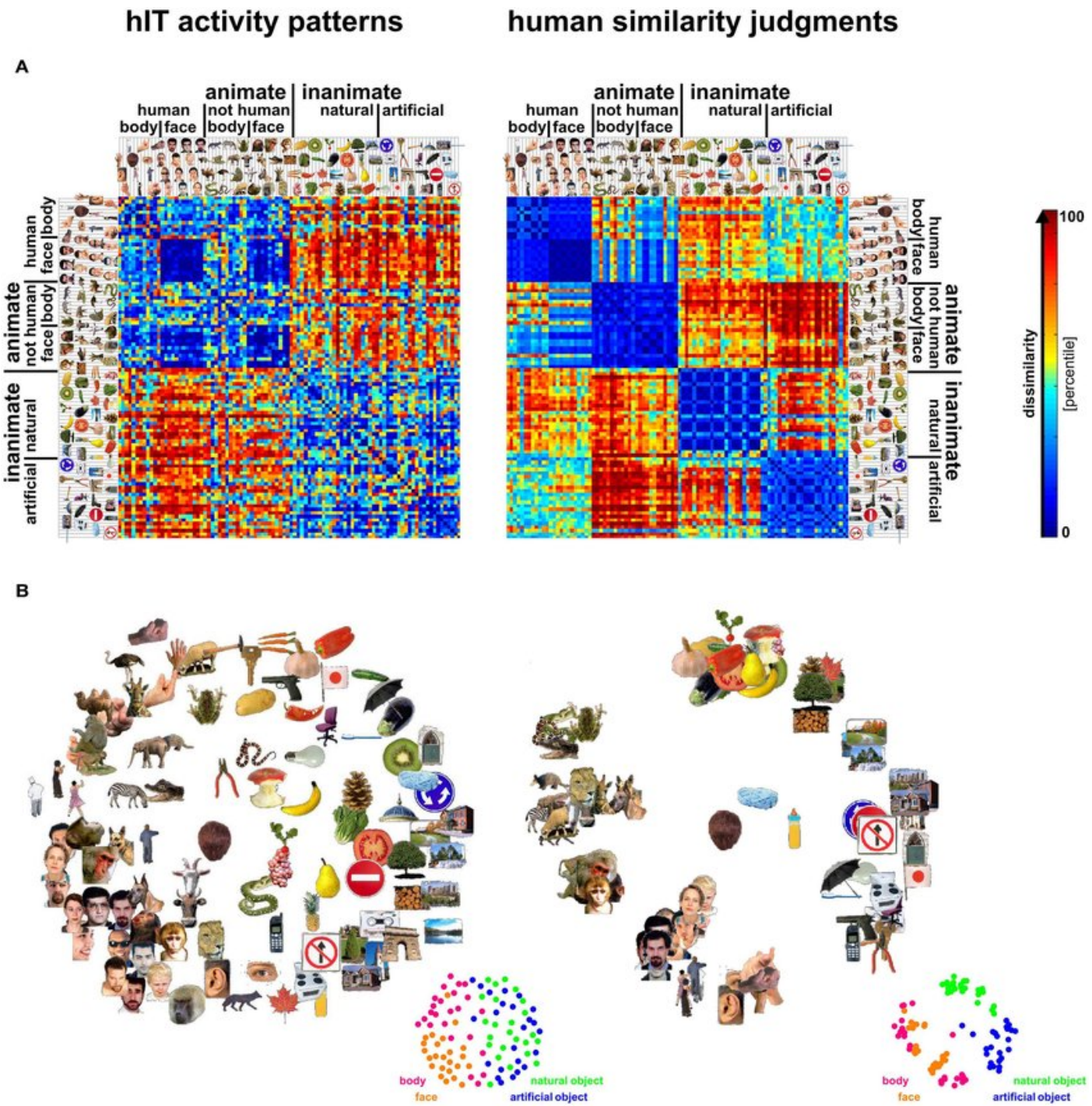
01 Siemens MAGNETOM Trio

H. Nili, C. Wingfield, A. Walther, L. Su, W. Marslen-Wilson, and N. Kriegeskorte. A toolbox for representational similarity analysis. PLoS Computational Biology, 10(4):e1003553, 2014.

fMRI Stimuli Set



Mur et al. Frontiers in Psychology 2013



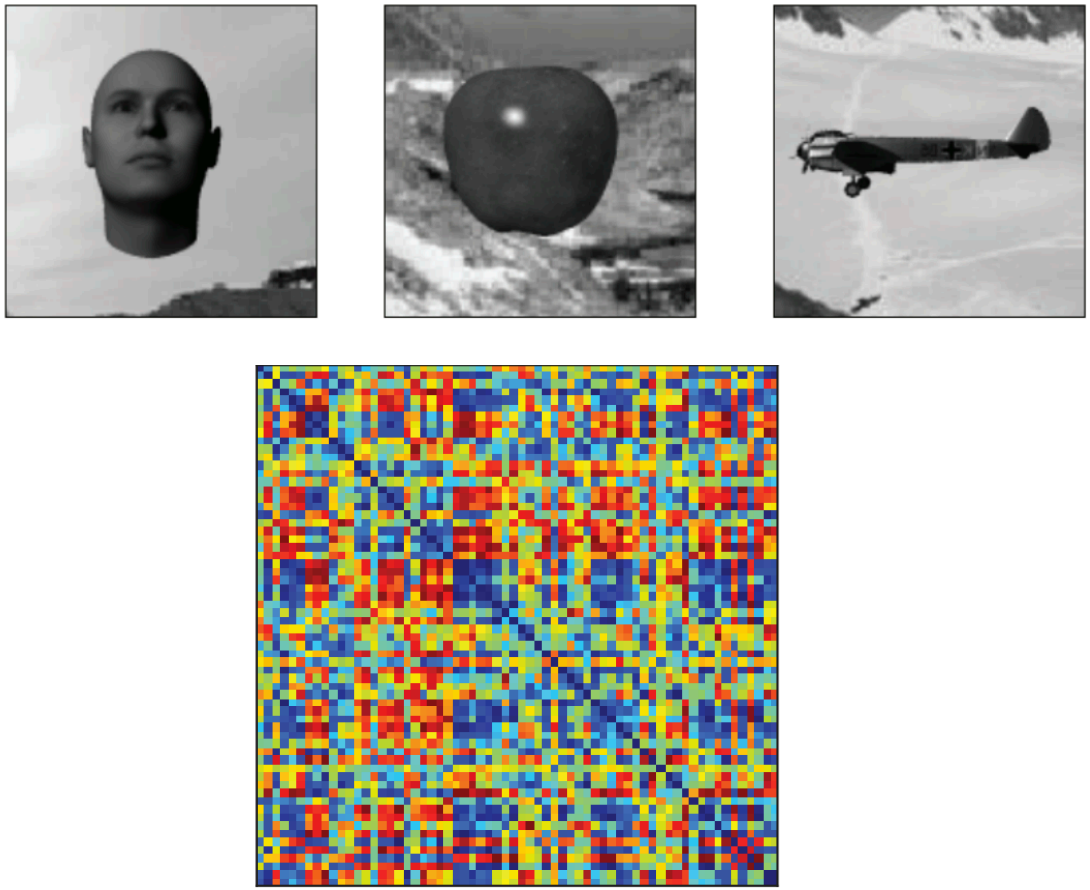
Mur et al. Frontiers in Psychology 2013

Back to the brain vs. model experiment…

Electrophysiology Data: Yamins et al. 2013

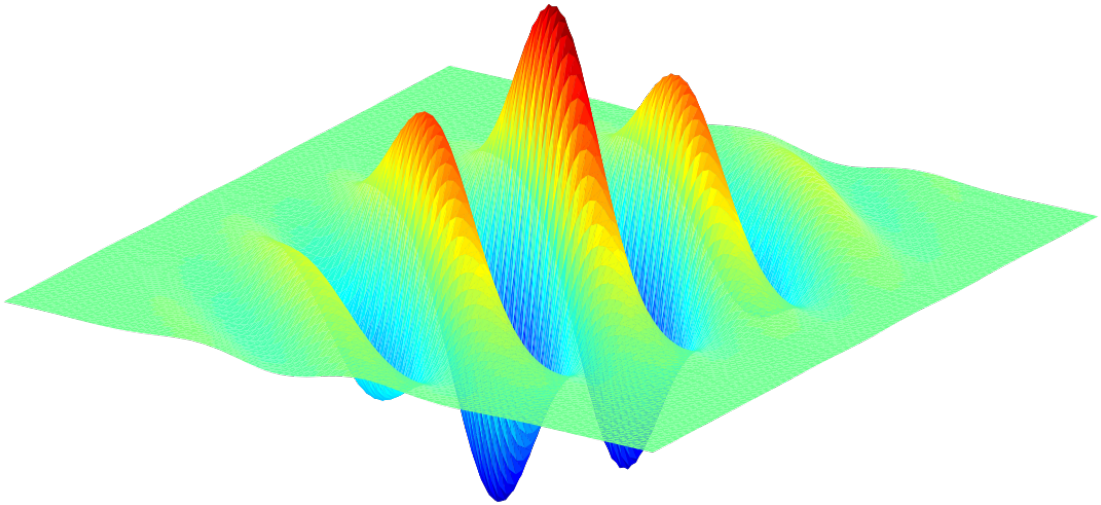
- Large-scale parallel array electrophysiology recordings in the visual cortex of awake behaving macaques
- Passively viewing animals shown random stimulus sequences with durations comparable to those in natural primate fixations (200 ms)
- Electrode arrays were surgically implanted in V4 and IT, and recordings took place daily over a period of months
- A total of 296 multi-unit responses were recorded from two animals
- For each testing stimulus and neuron, final neuron output responses were obtained by averaging data from between 25 and 50 repeated trials

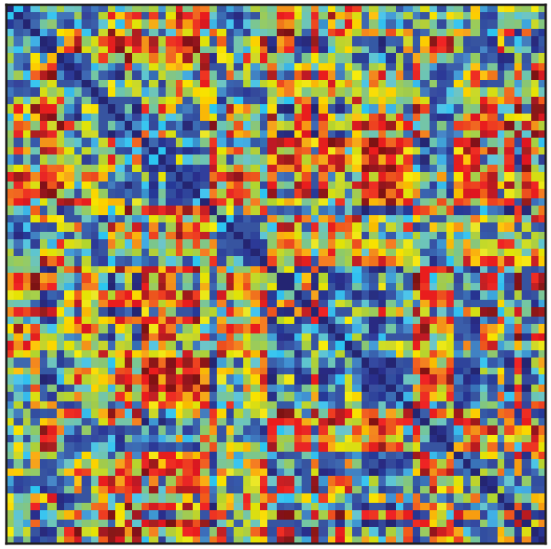
Control Model: Pixels



Yamins et al. NeurIPS 2013

Control Model: V1-Like





Yamins et al. NeurIPS 2013

Control Model: SIFT

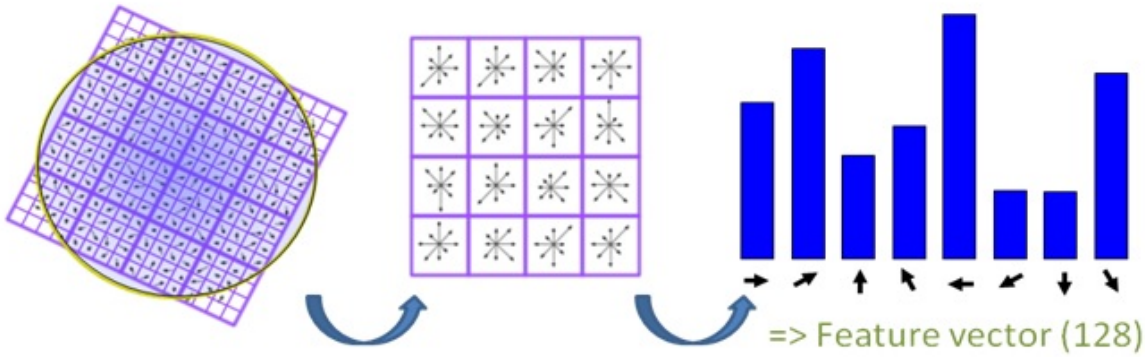
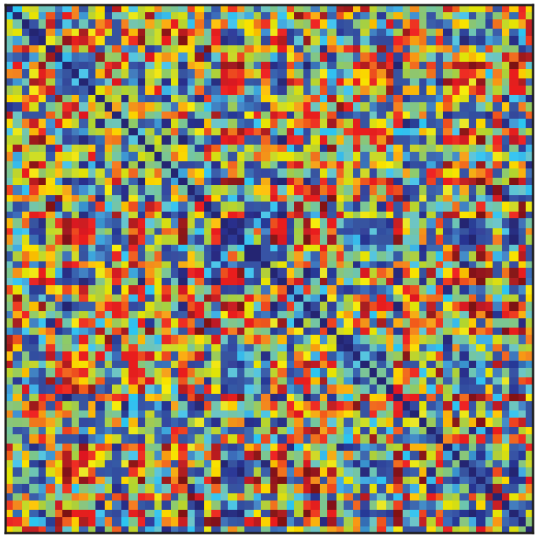
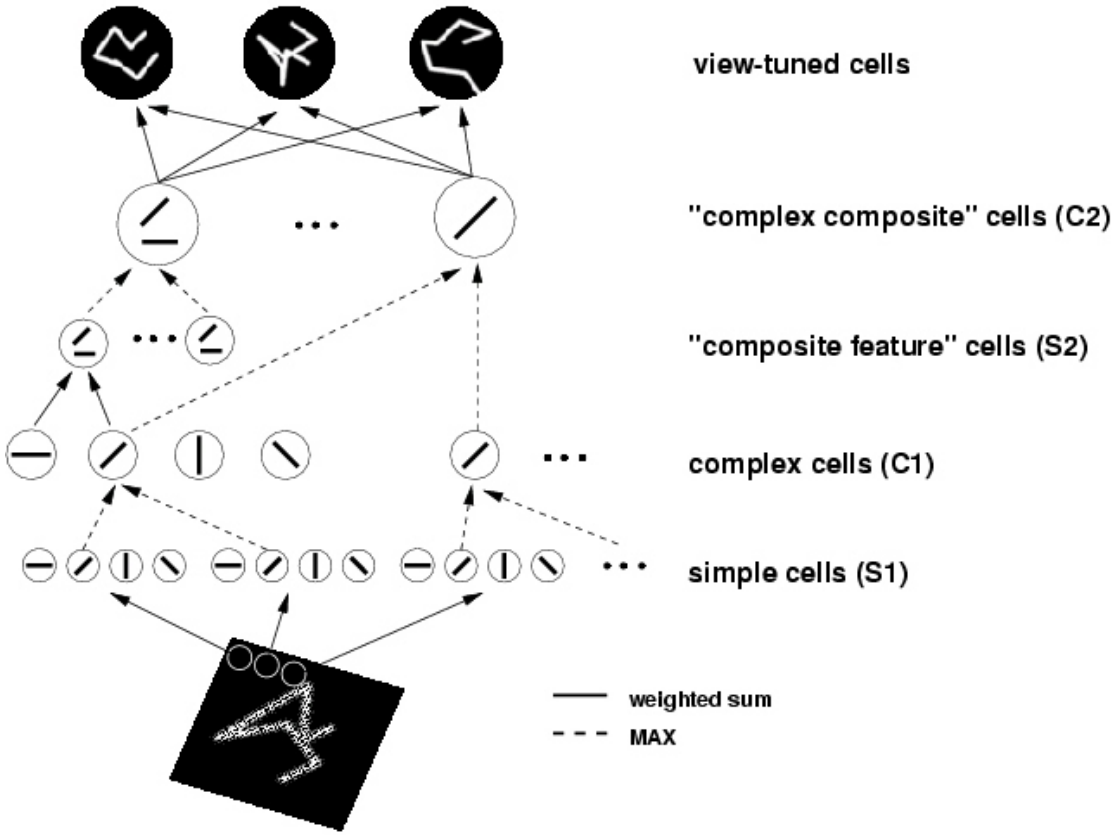


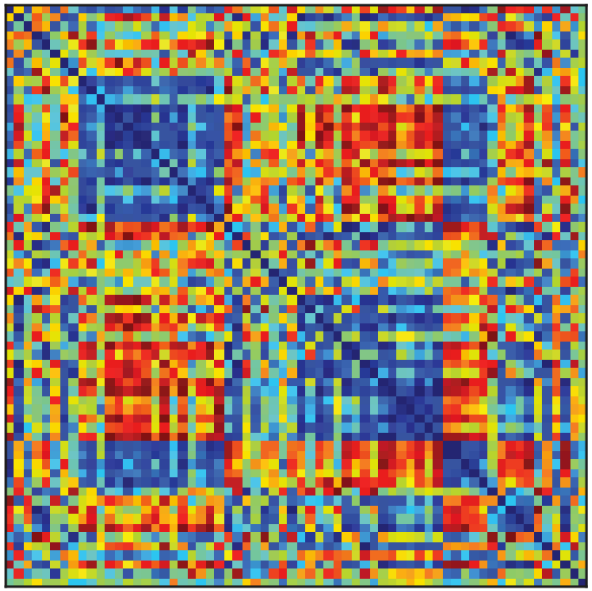
Image Credit: https://gilscvblog.com/2013/08/18/a-short-introduction-to-descriptors/



Yamins et al. NeurIPS 2013

Control Model: HMAX

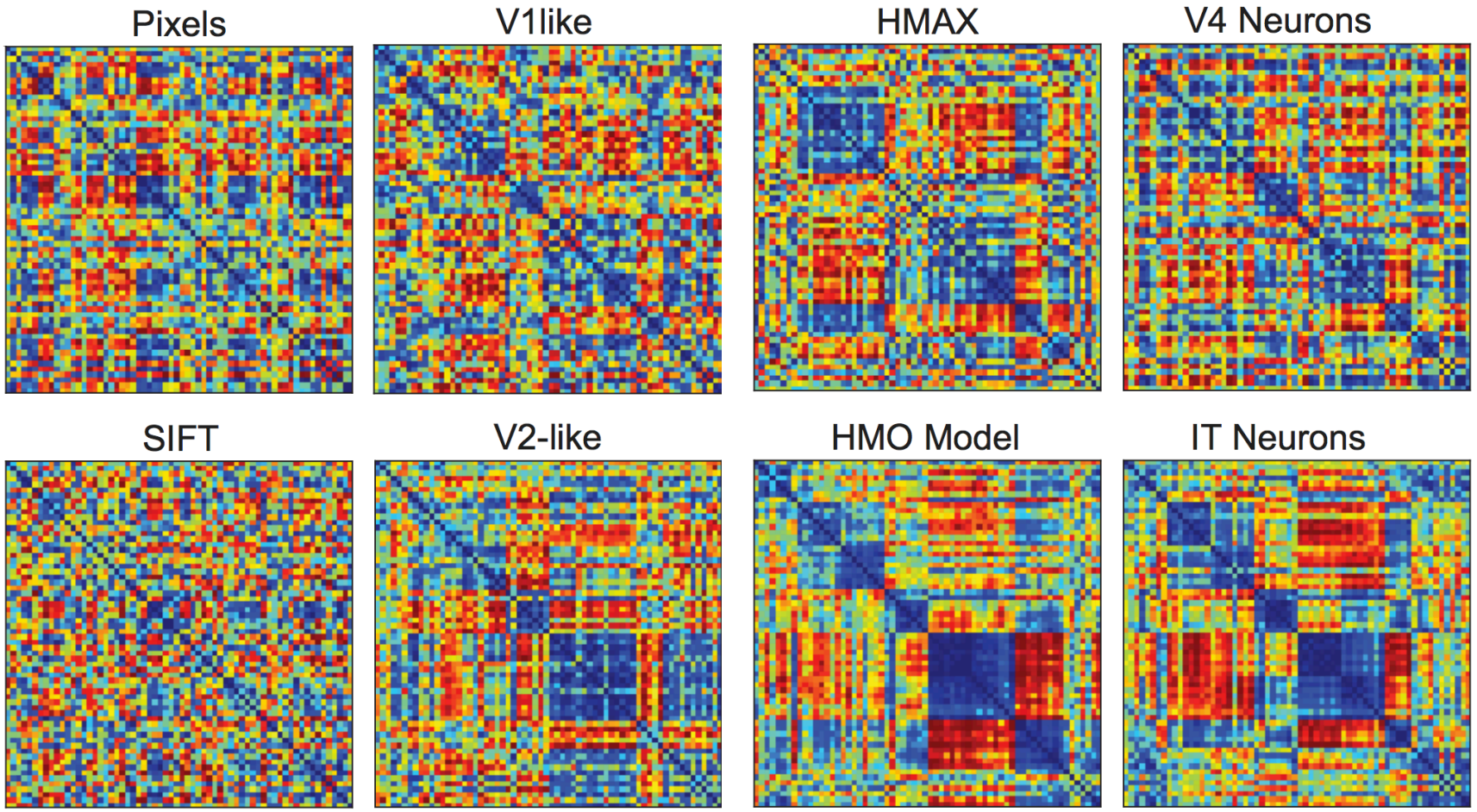




Riesenhuber and Poggio Nat. NeuroSci. 1999

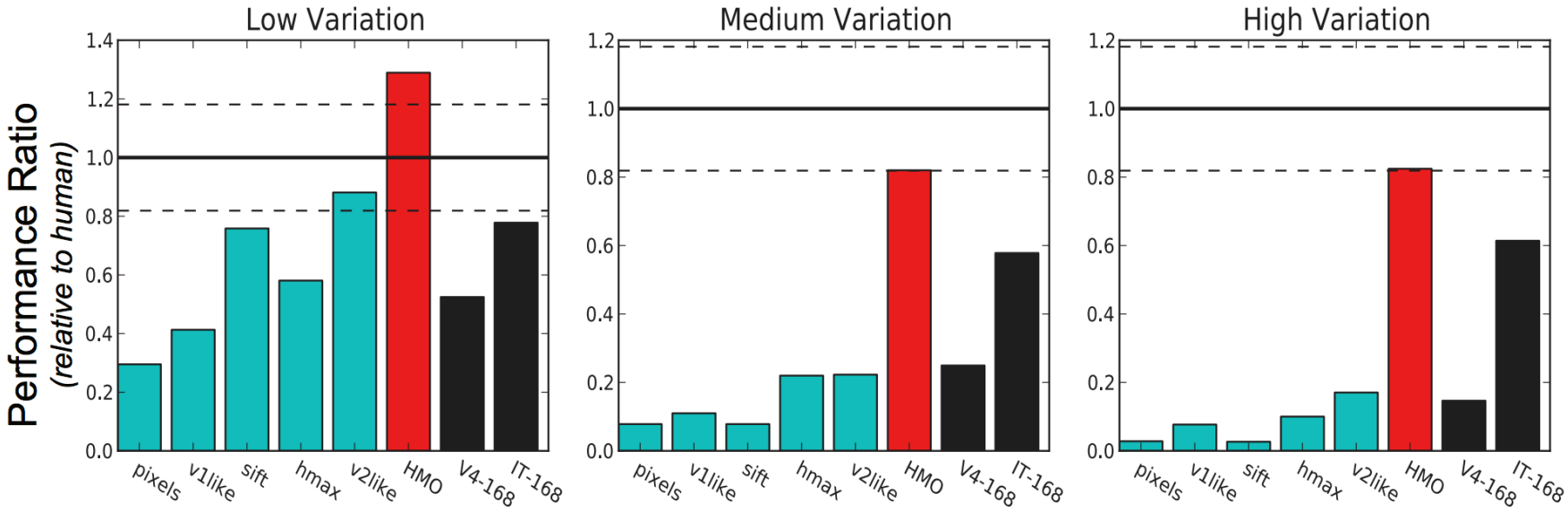
Yamins et al. NeurIPS 2013

Comparison to Monkey Recordings



Yamins et al. NeurIPS 2013

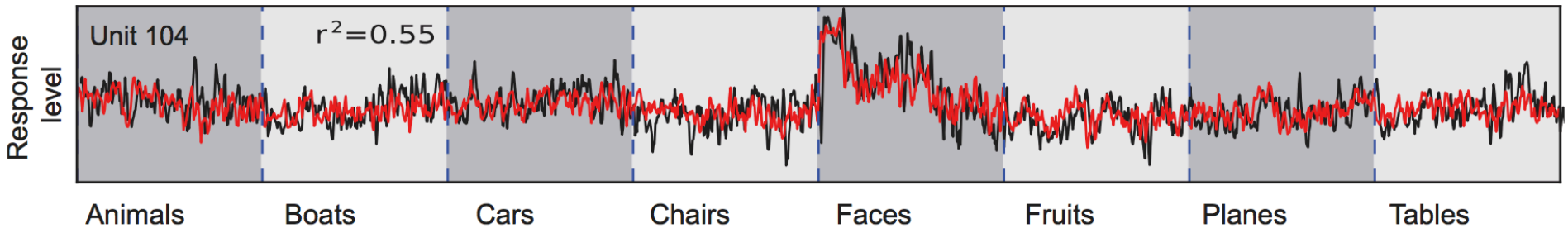
8-way Categorization Performances

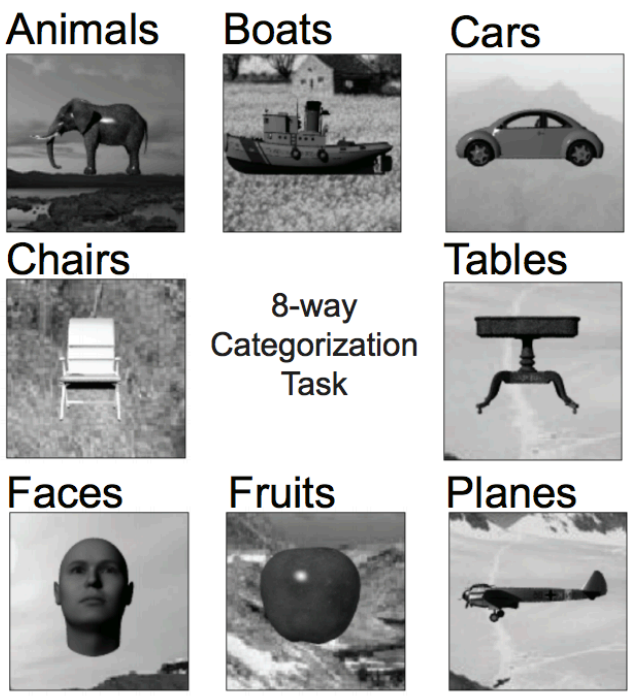


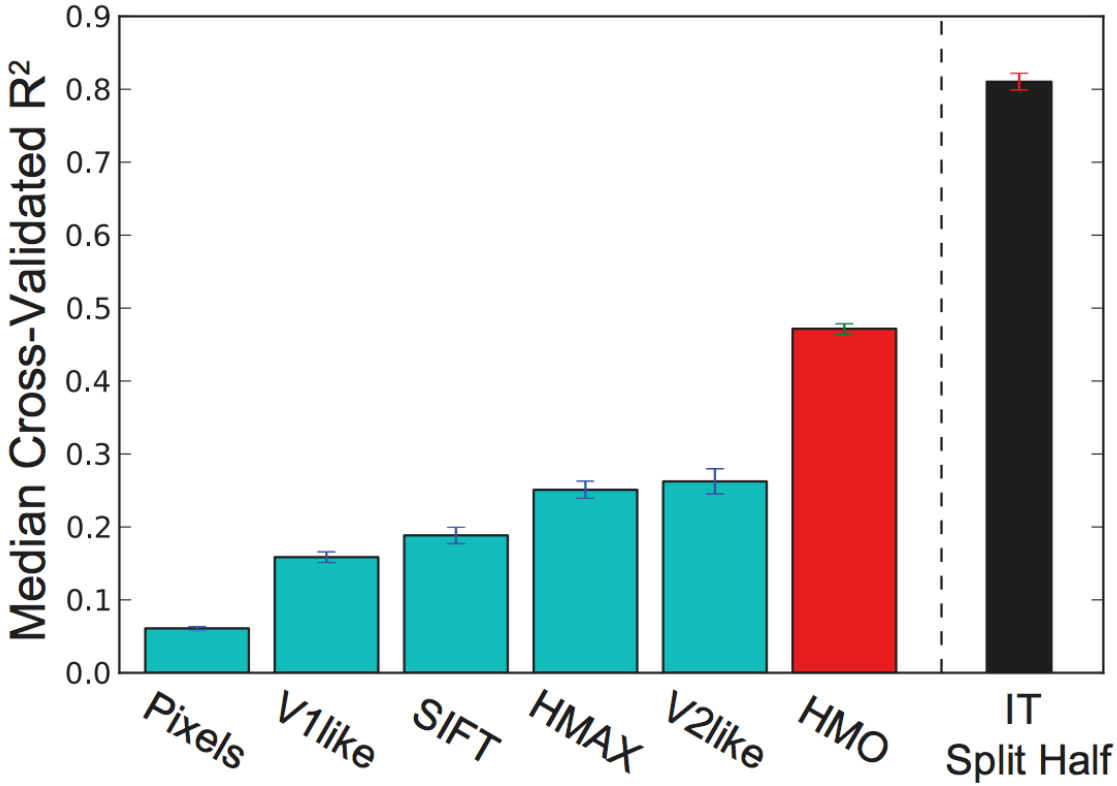
Yamins et al. NeurIPS 2013

Black curve: actual neural response

Red curve: prediction for a single sample IT neuron

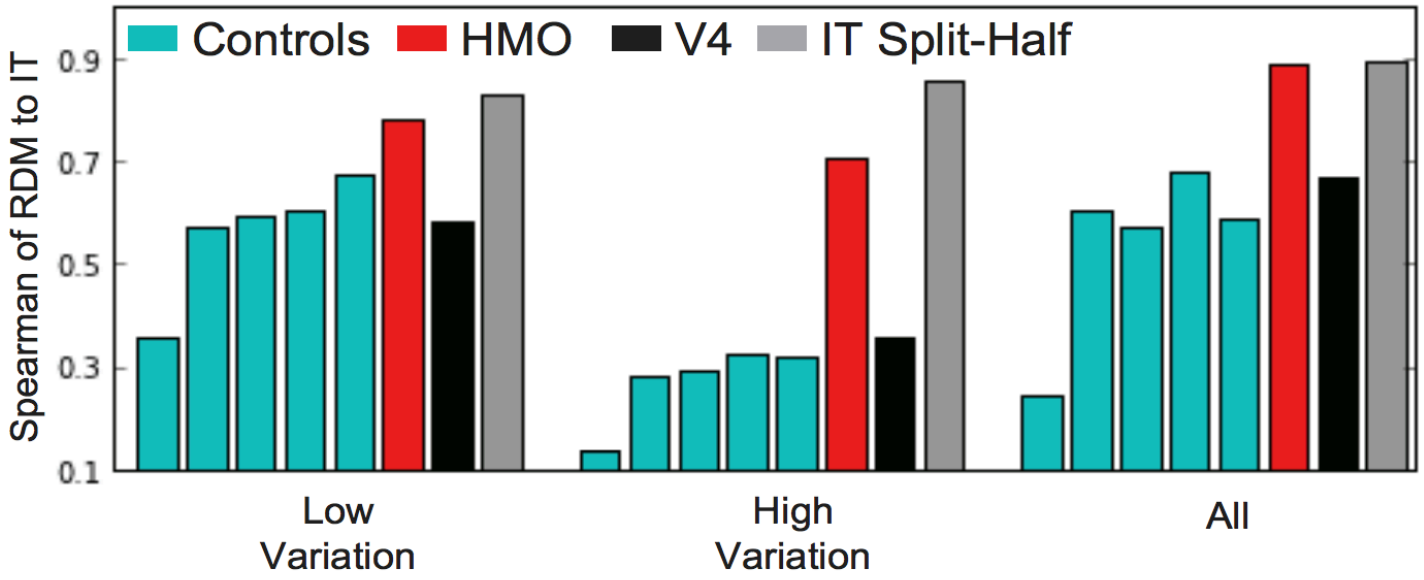






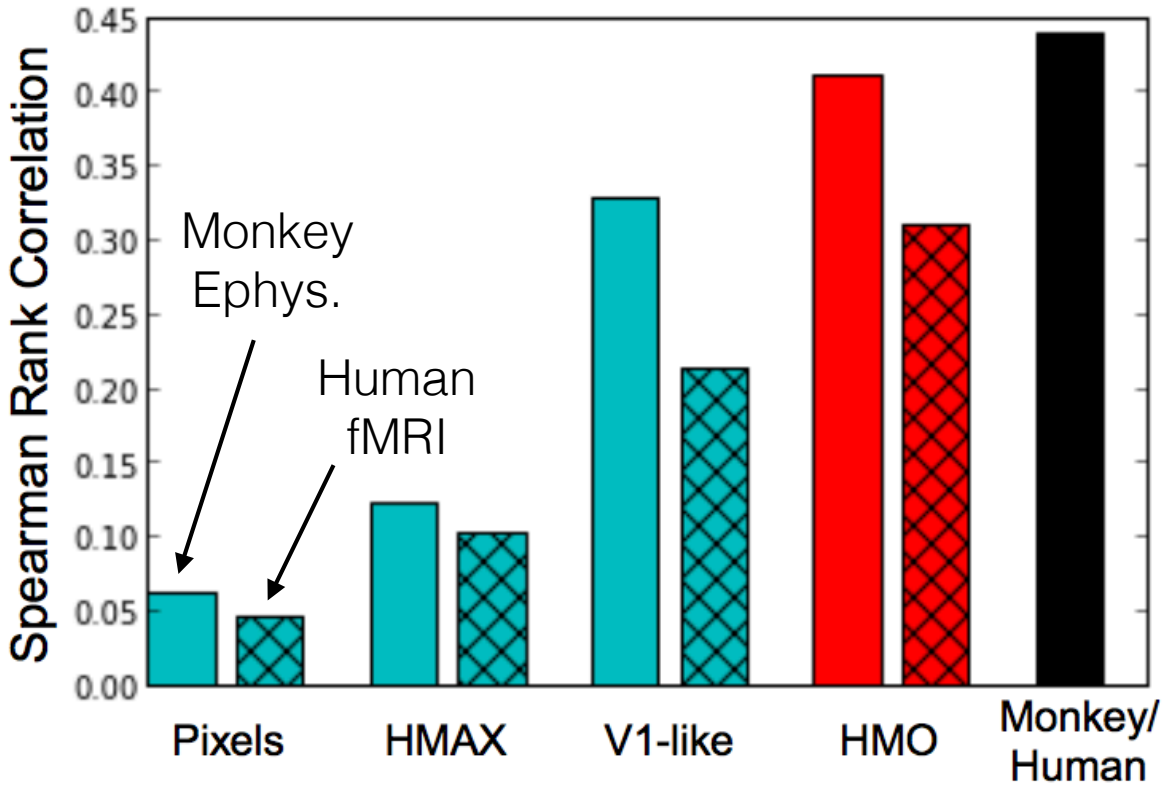
Yamins et al. NeurIPS 2013

Comparison of RDMs



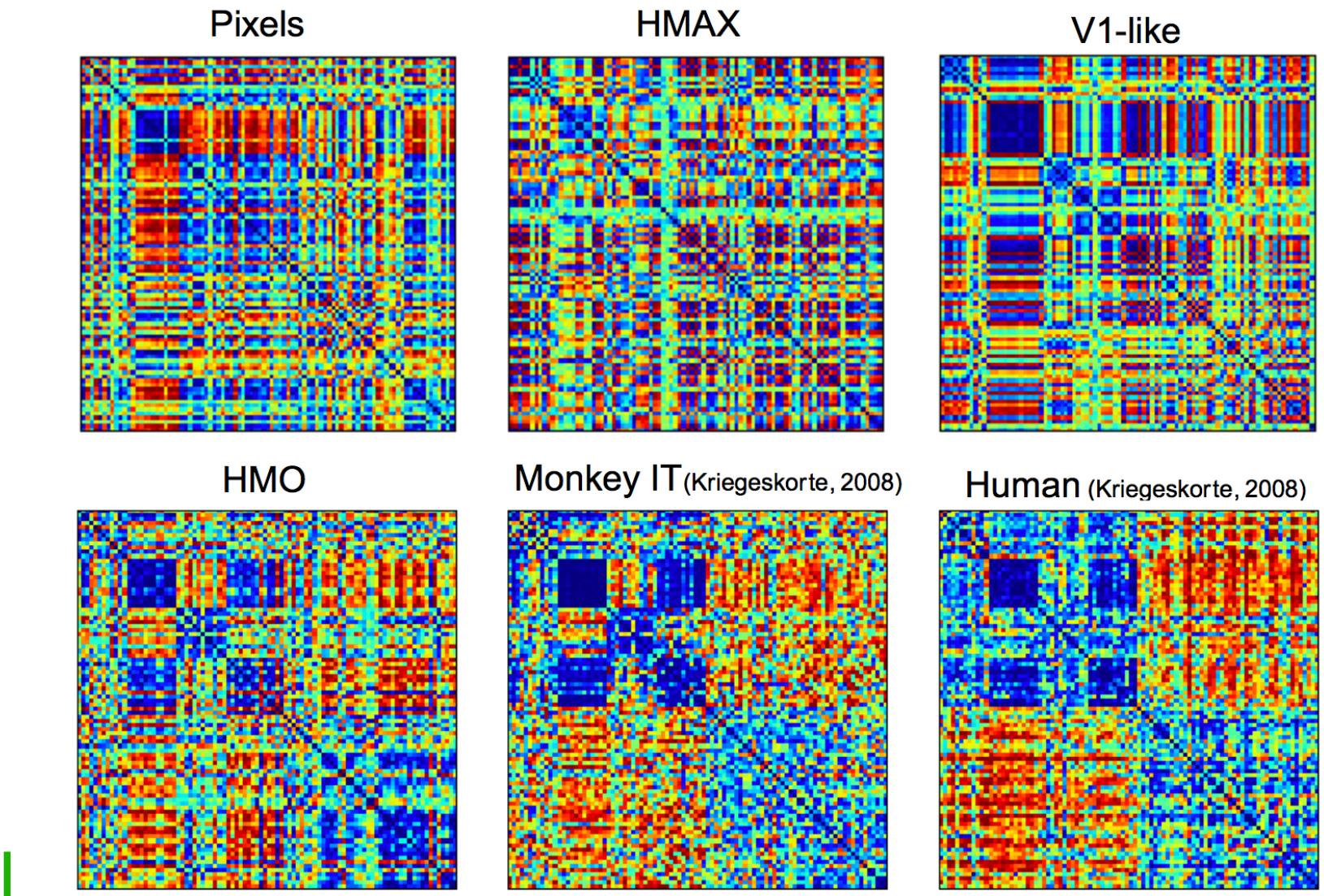
Yamins et al. NeurIPS 2013

Model comparison to monkey IT and human ventral stream



Yamins et al. NeurIPS 2013

RDM Comparison



Yamins et al. NeurIPS 2013