Computational Models of Vision

Pouya Bashivan and Kohitij Kar

Encoding Edition

CBMM Summer School 2018

Overview - Encoding Models

- Why studying vision?
- What do we know about brain’s representation of the visual world?
- What models are are there for these processes?

Why Models?

- Why do we need models?
- How can we use models in science?

(hypotheses) Models → Falsifiable Predictions → Errors → Fix/Reinvent Models

"Truth will sooner come out from error than from confusion." Francis Bacon

Why Vision?

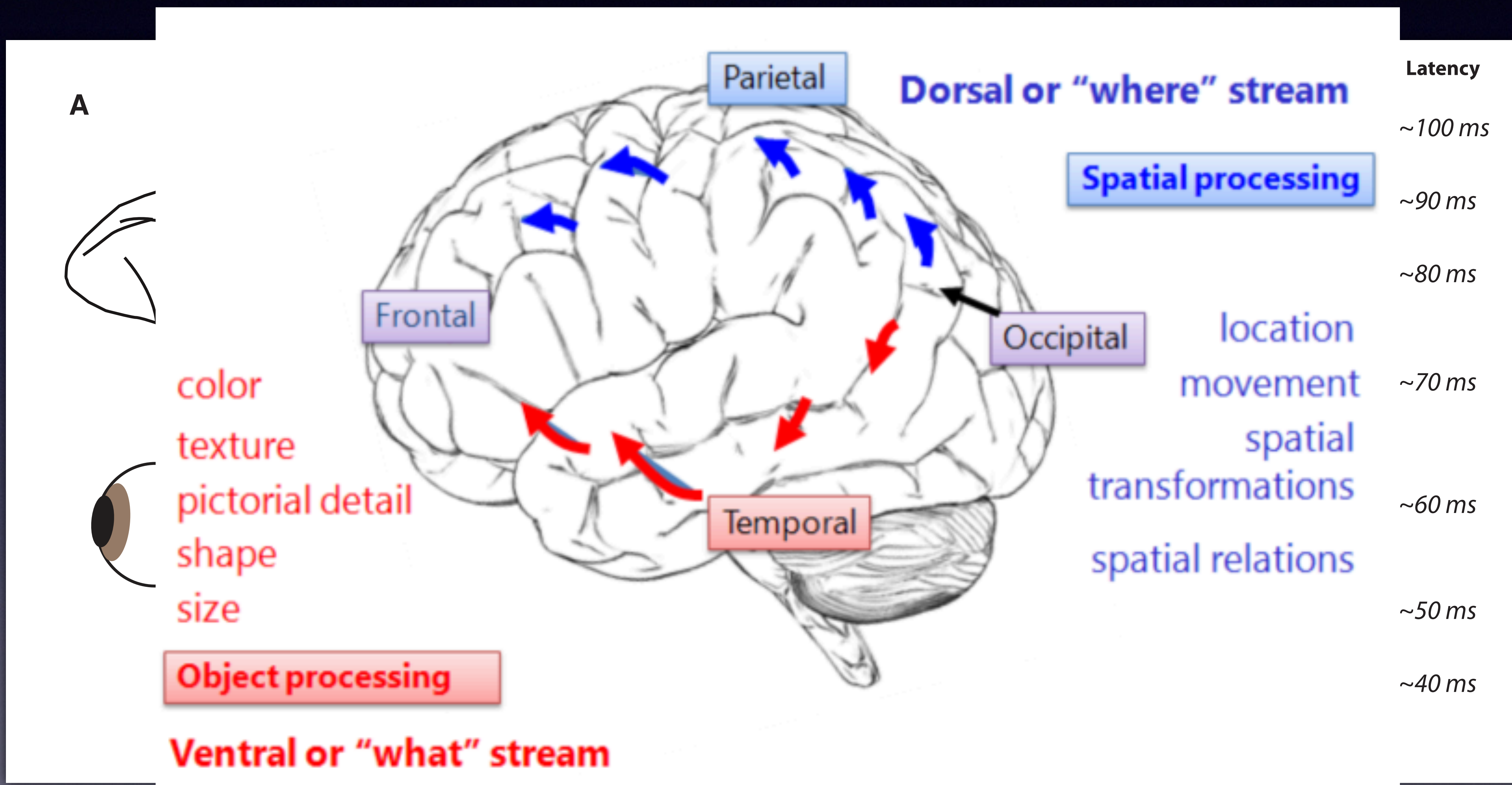
1. A window into how neural networks build a compact representation of the world.

2. How the encoded image is represented by the neural response within the peripheral and early cortical visual pathways.

3. What's the role of these representations in efficient image coding.

What we know about vision in primates

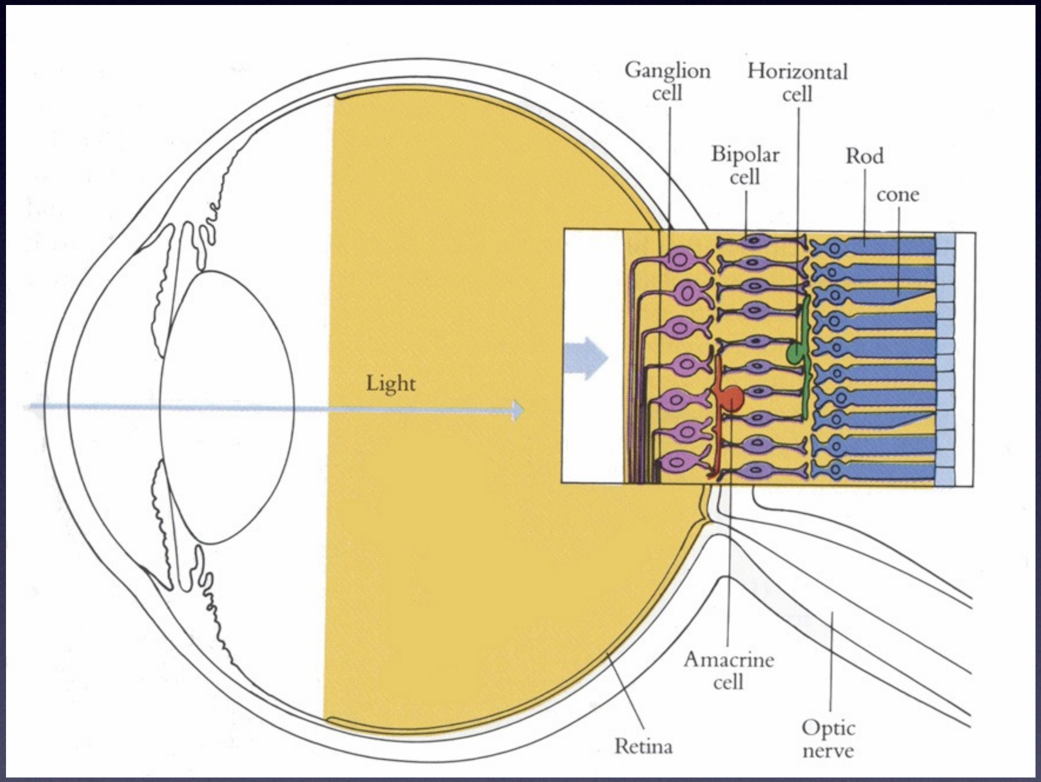
Visual Processing Streams



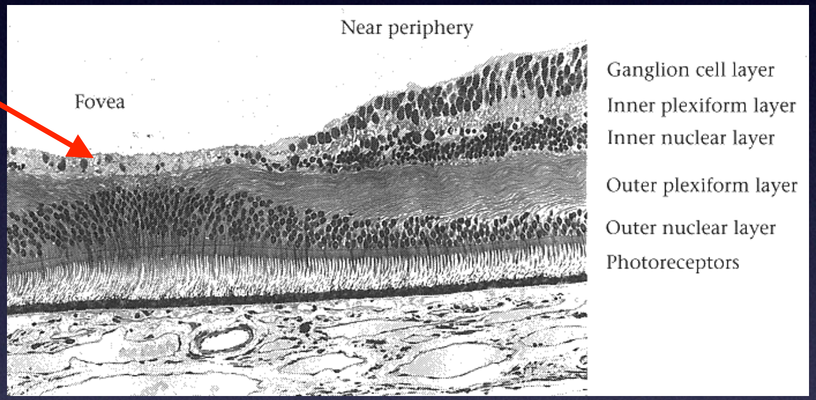
Ventral Stream

Adapted from DiCarlo et al. 2012

Image Formation - Retina



Only one layer
Cone photoreceptors



Rods and Cones: encode the image in different intensity ranges.

1. There is only a narrow region of high visual acuity in the fovea.
2. Dynamic range of sensors is very small.
3. Representation of wavelength is coarse.

Adapted from Hubel

adapted from A. Hendrickson

Things We Know
Lateral Geniculate Nucleus (LGN)

- Receives input from Retina through the optic nerve

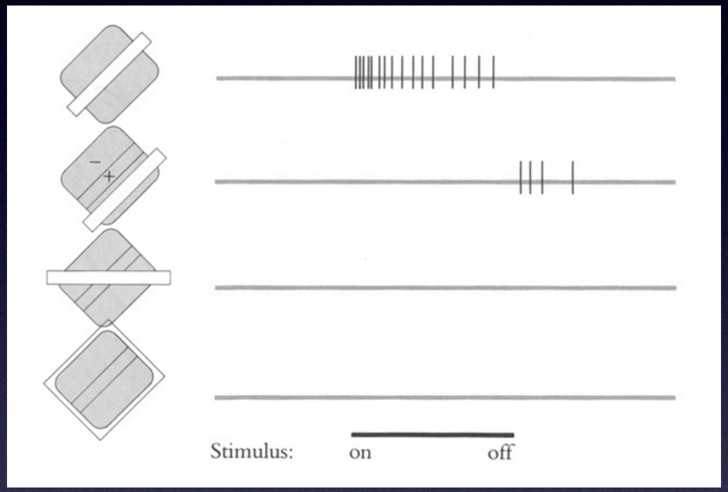
- Two major streams:

  1. **Parvocellular**: high-spatial freq + low-temporal-freq

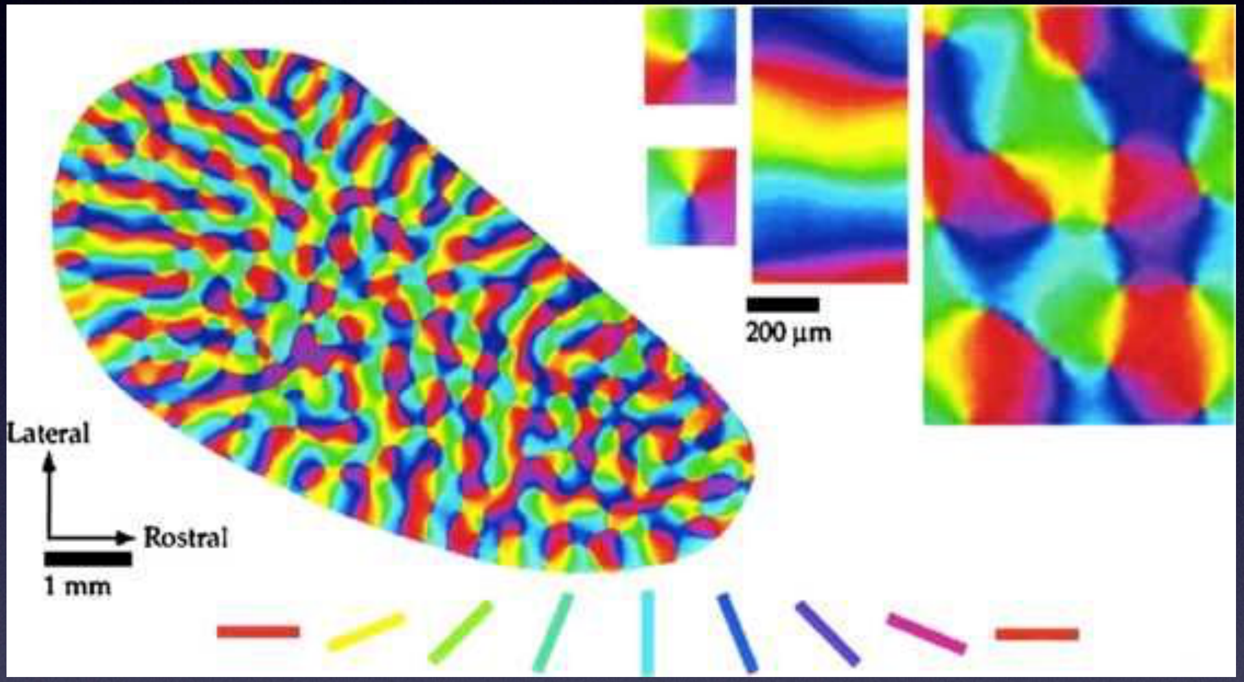
  2. **Magnocellular**: low-spatial freq + high-temporal-freq

Primary Visual Cortex - V1

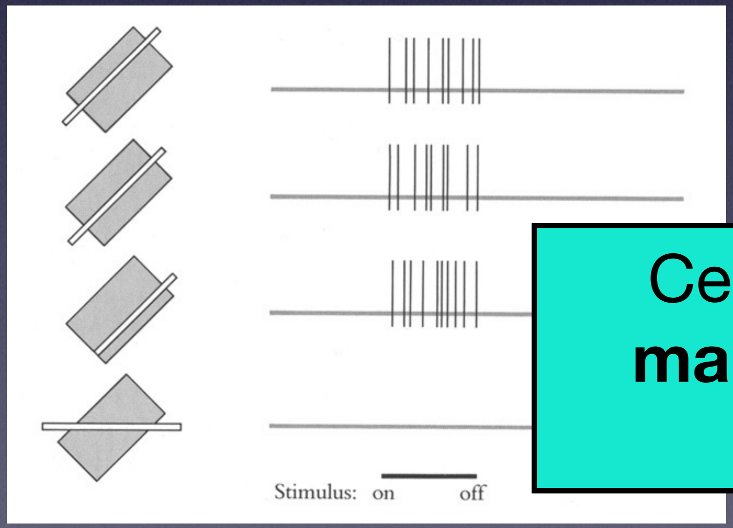
Simple cell



Orientation selectivity



Complex cell



Orientation

Cells are **topographically mapped** according to their preferred pattern.

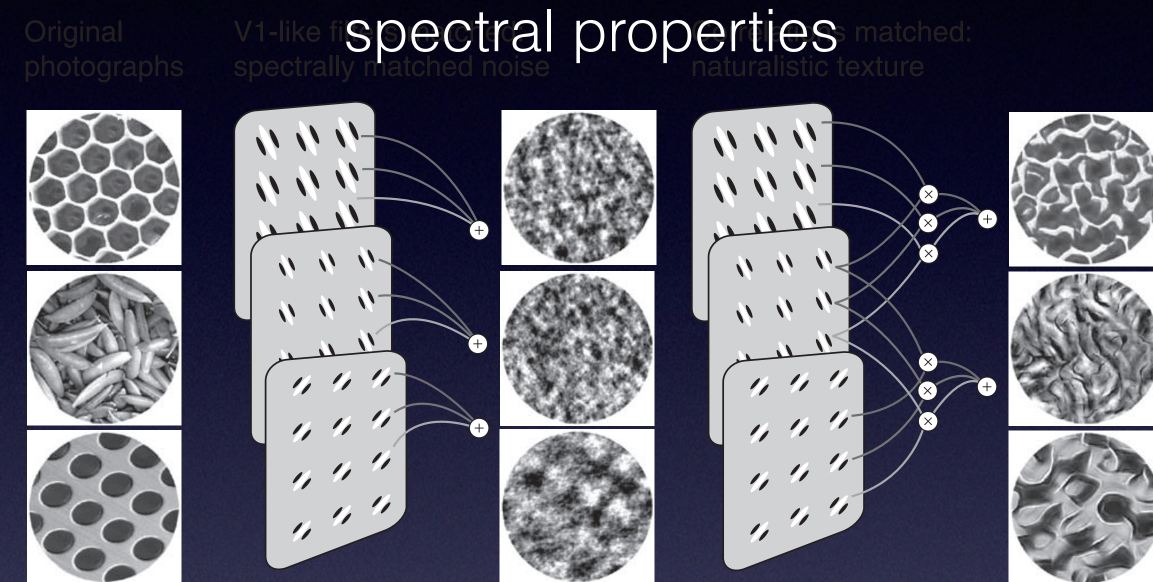
Adapted from Kandel , Schwartz and Jessell

Adapted from Bosking 1997

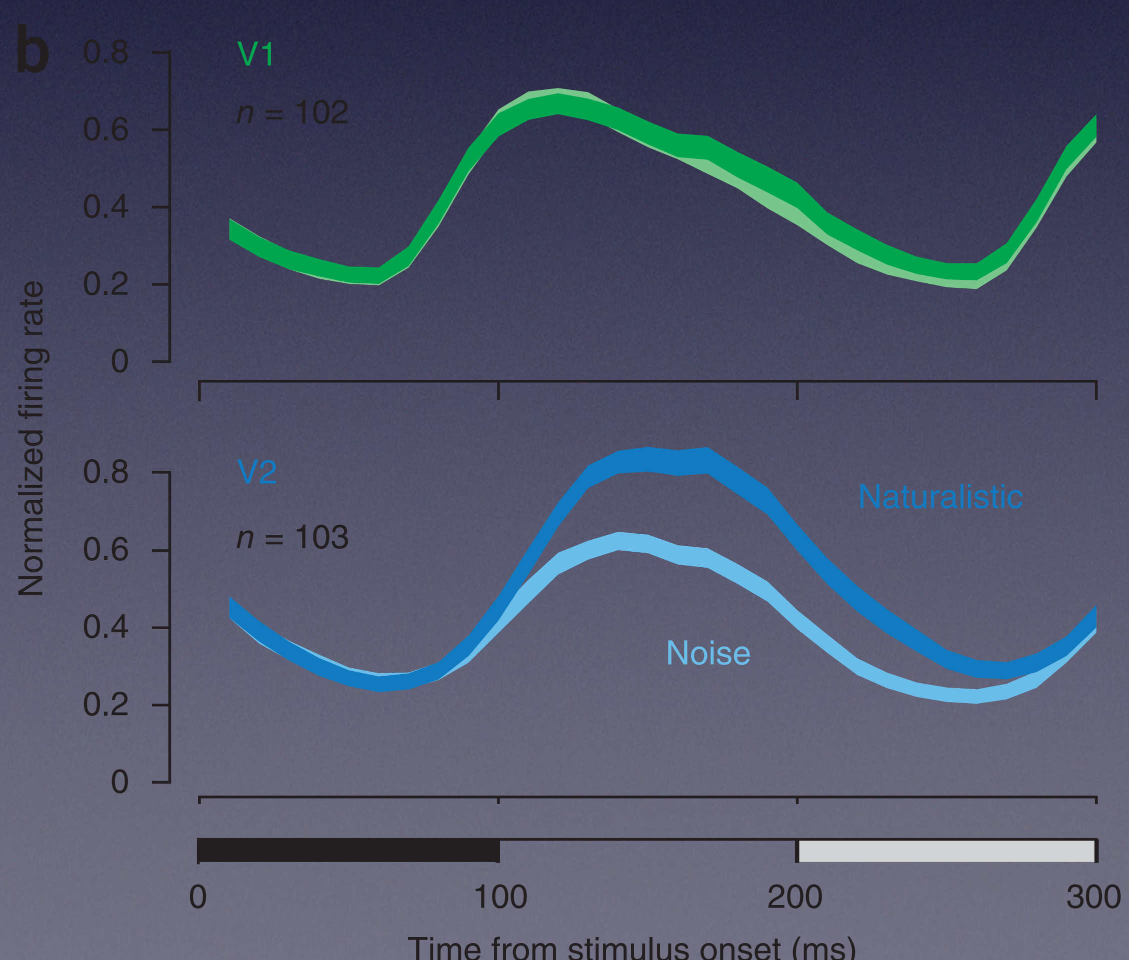
Primary Visual Cortex - V2

Area V2 (first cortical area after V1):

Noise images with matching spectral properties



Synthetic images with matched correlation in V1 responses

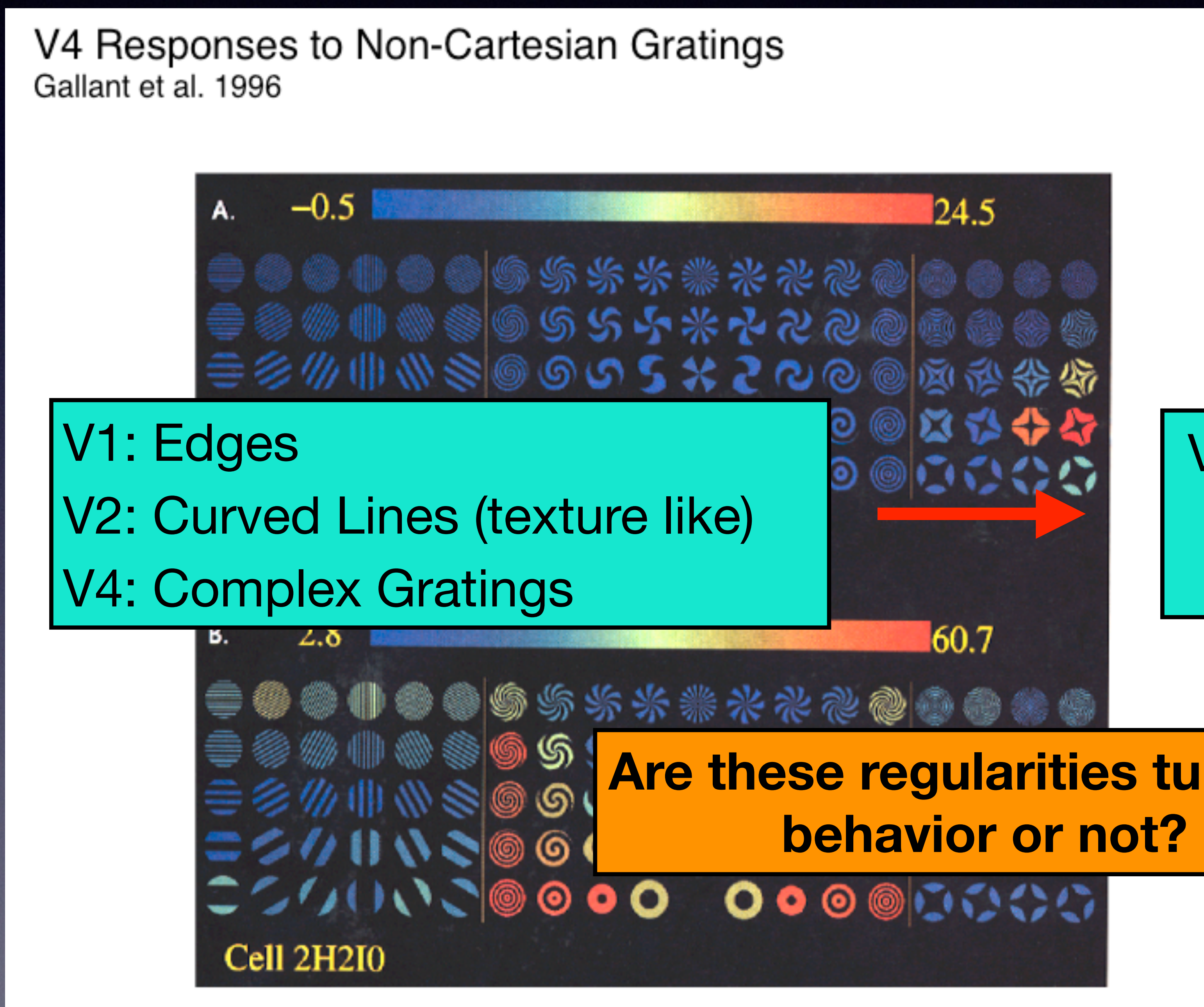


Interpretation:

- **V2 neurons apply "and-like" operators on V1 outputs**
- **those "ands" are tuned toward natural co-occurring V1 statistics**

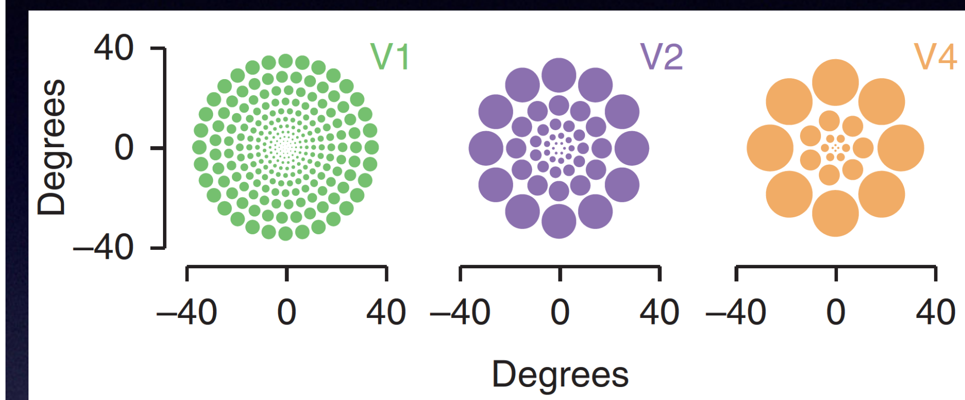
Adapted from Freeman, Ziemba, Heeger, Simoncelli, & Movshon, Nature Neuro (2013)

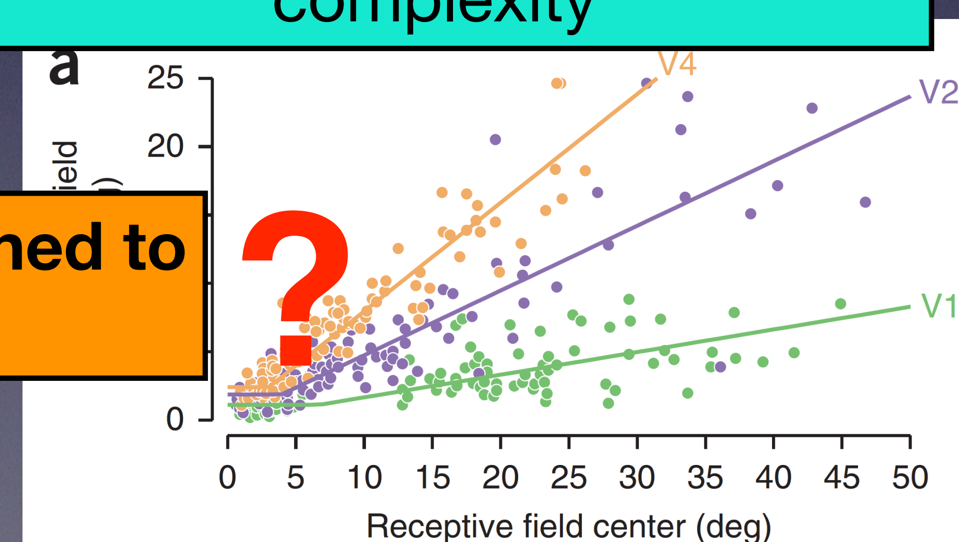
Primary Visual Cortex - V4



V1: Edges
V2: Curved Lines (texture like)
V4: Complex Gratings

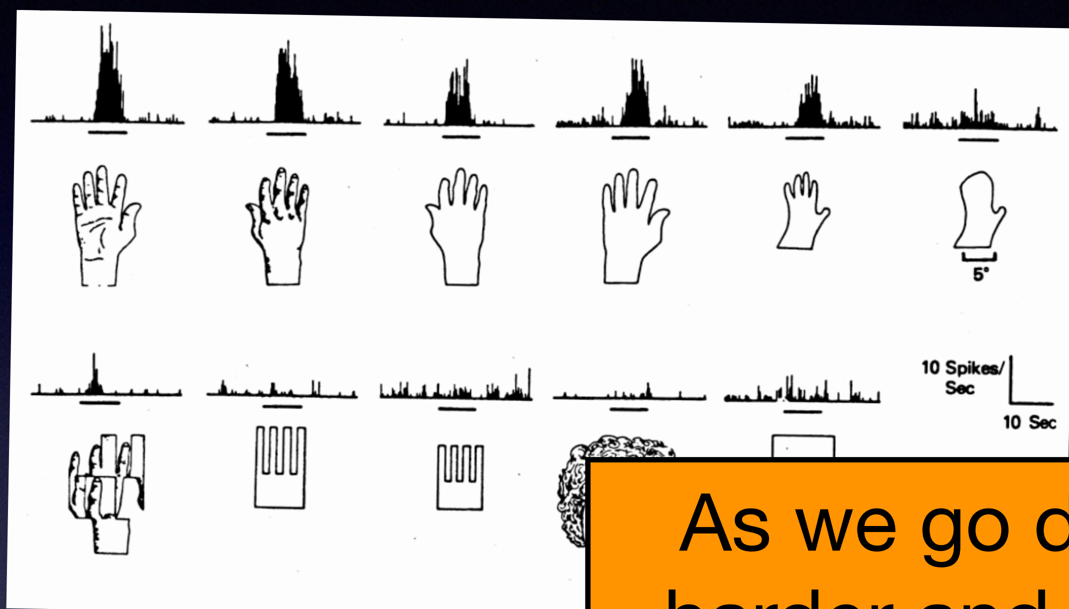
Ventral Stream: Captures visual regularities of increasing complexity





Are these regularities tuned to behavior or not?

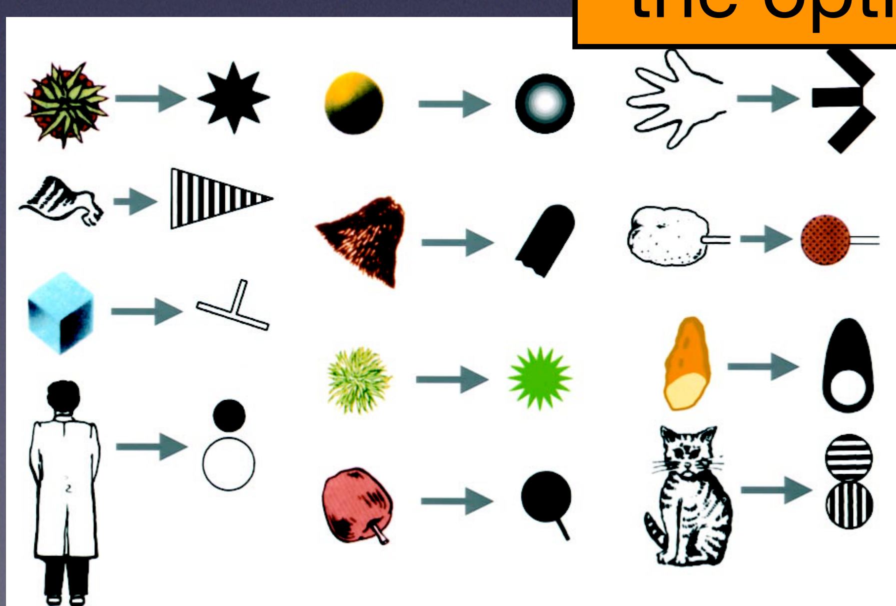
Primary Visual Cortex - IT



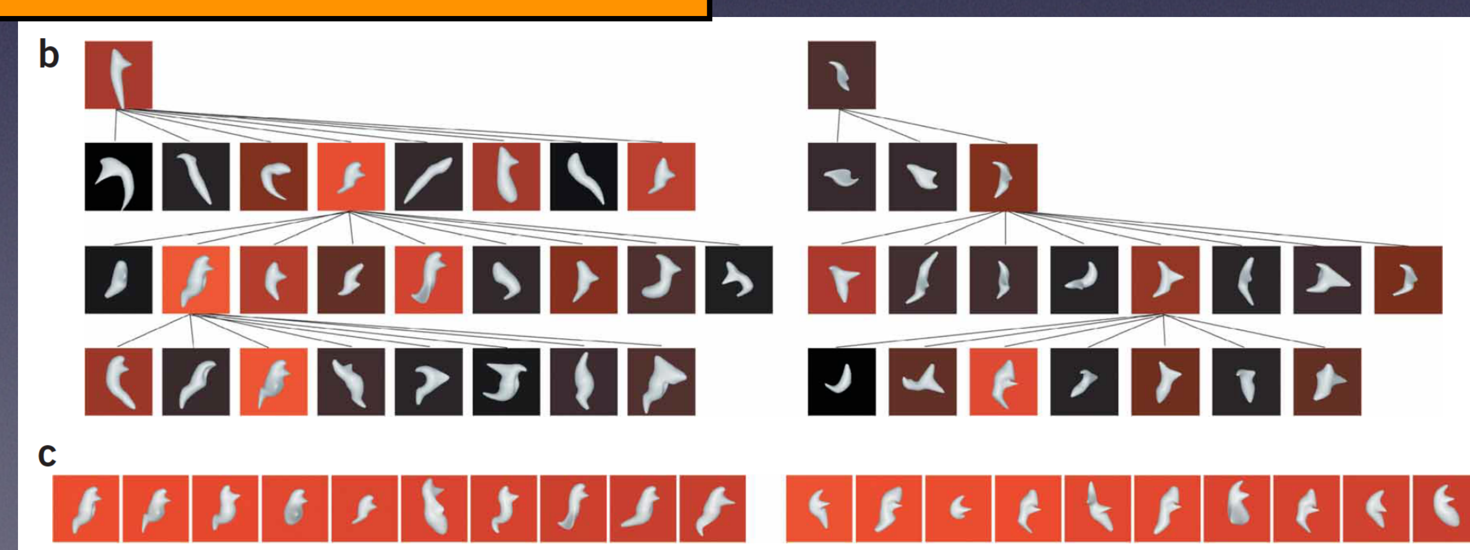
Adapted from Desimone et al. (1984)

IT neurons can be tuned to specific combinations of features (high “selectivity”)

As we go deeper it becomes harder and harder to discover the optimal stimuli for neurons.



Adapted from Tanaka

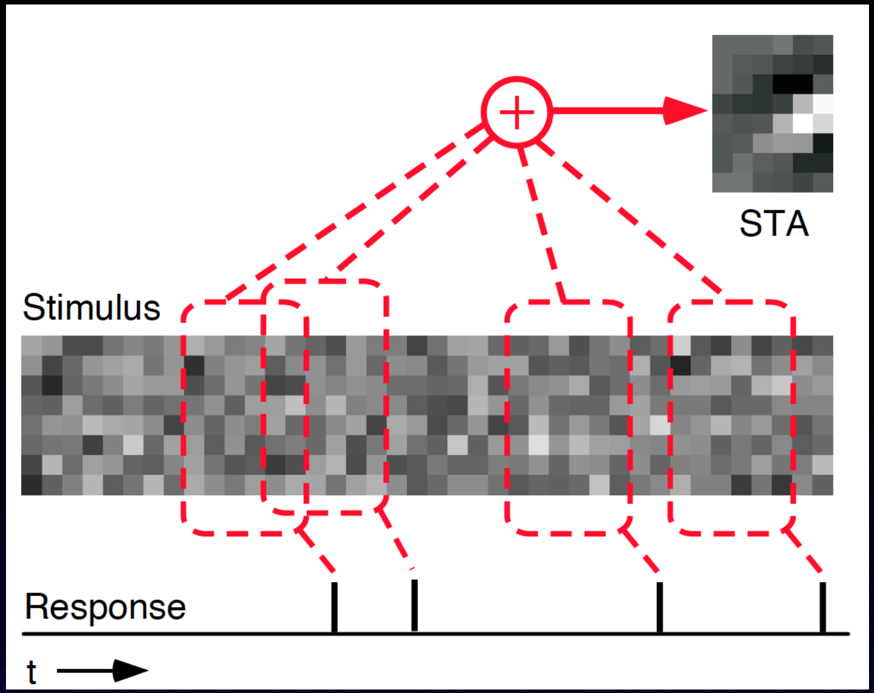


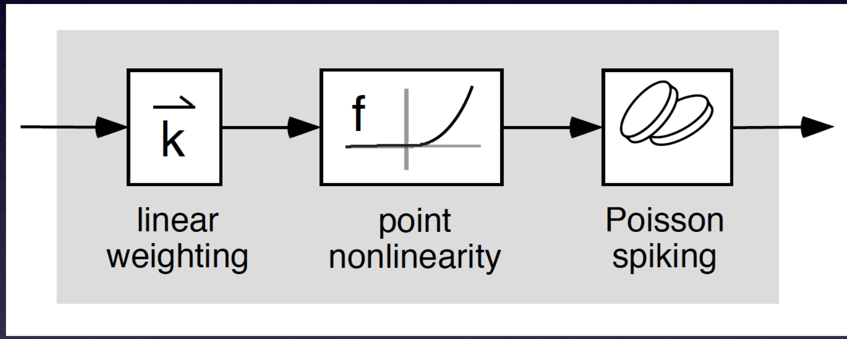
Adapted from Yamane, Connor Nature (2008)

Models of Vision

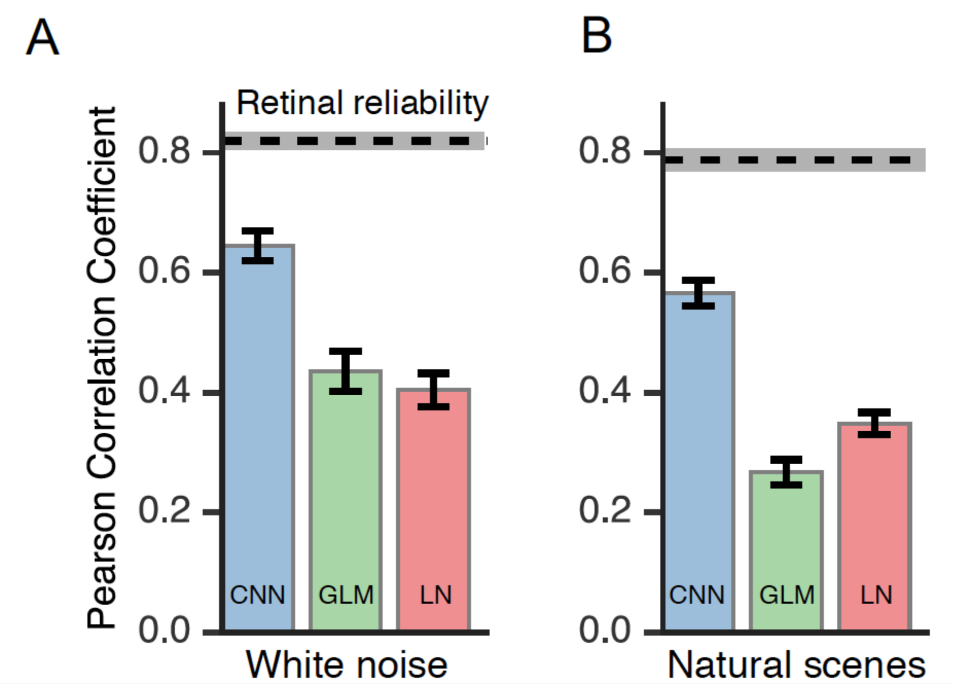
Vision Models
Retina

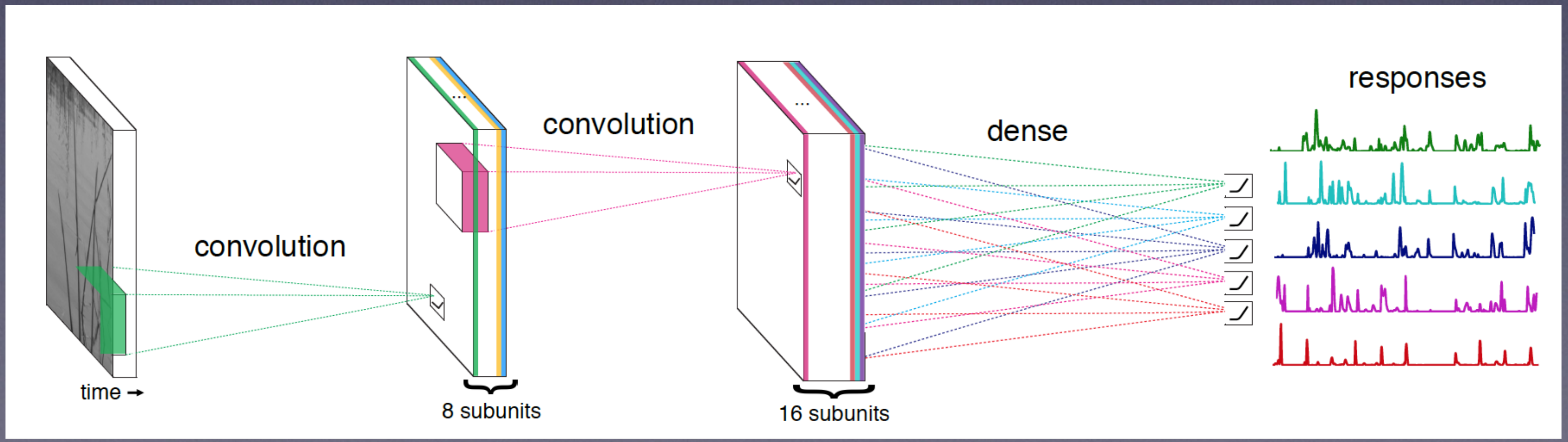
- Spike-Triggered Averaging (STA)
- Linear-Nonlinear Method (LN)
- Generalized Linear Models (GLM)
- Convolutional models (CNN)





Adapted from Simoncelli et al. (2004)

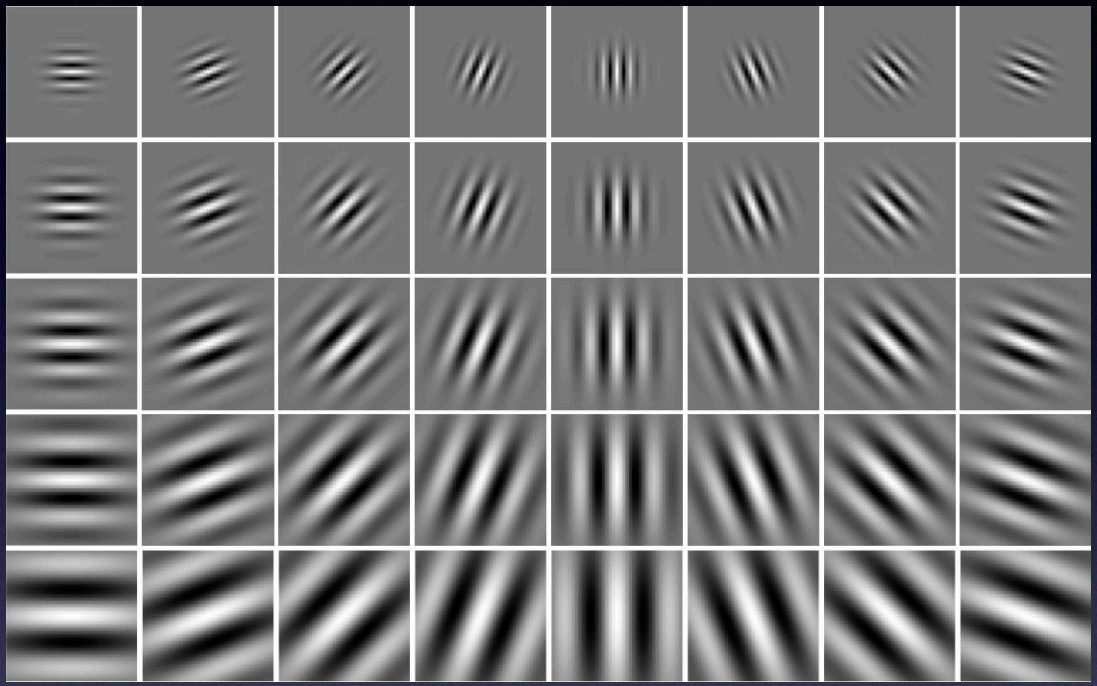


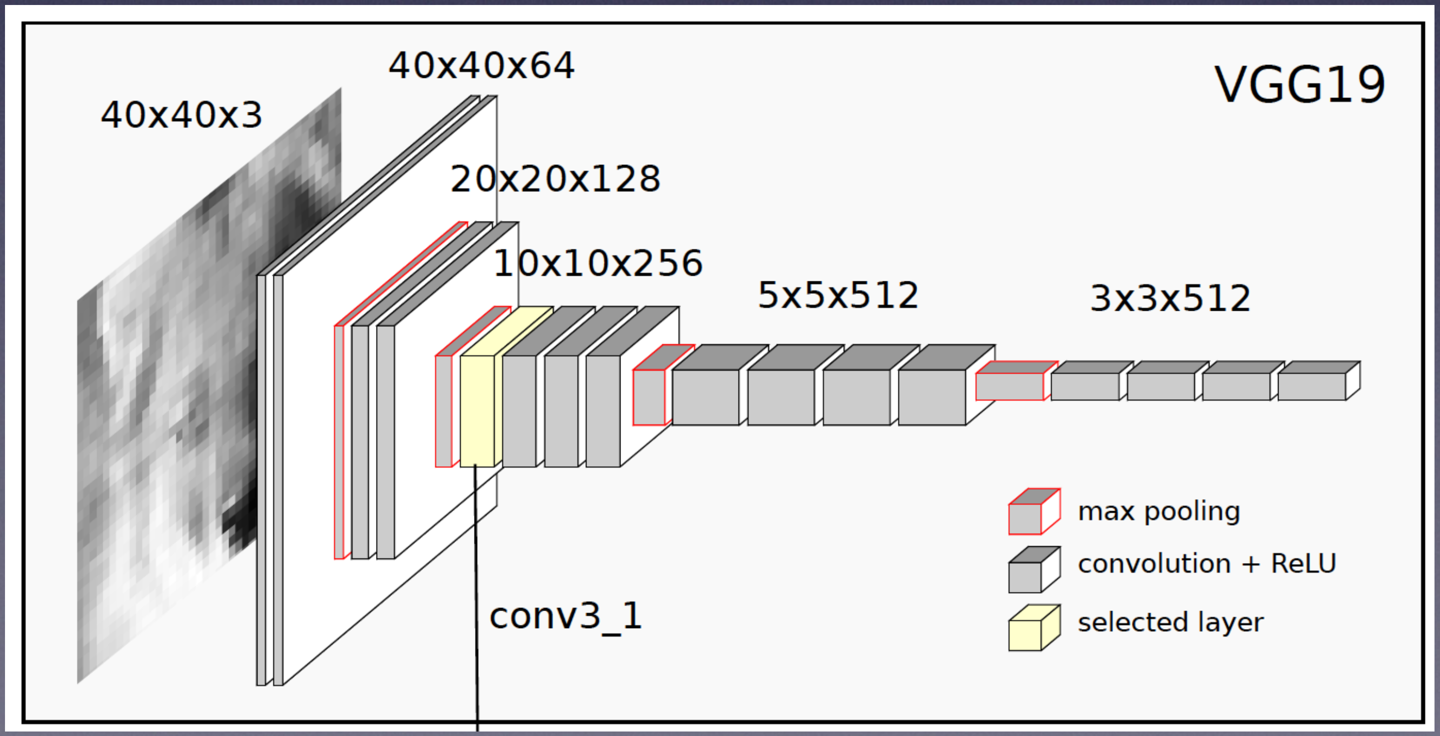


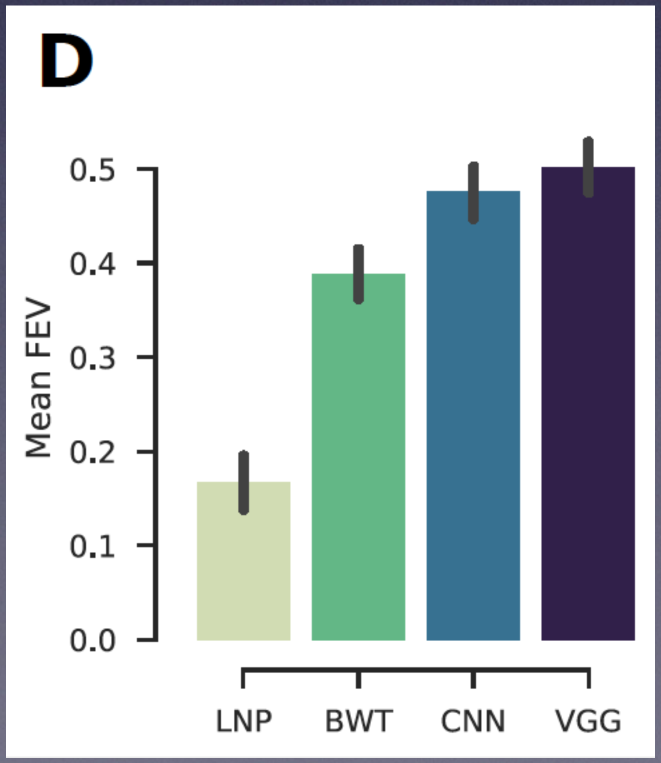
Adapted from McIntosh et al. (2017)

Vision Models - V1

- Gabor filter banks - Wavelet transforms
- LN-LN Cascade model
- CNNs



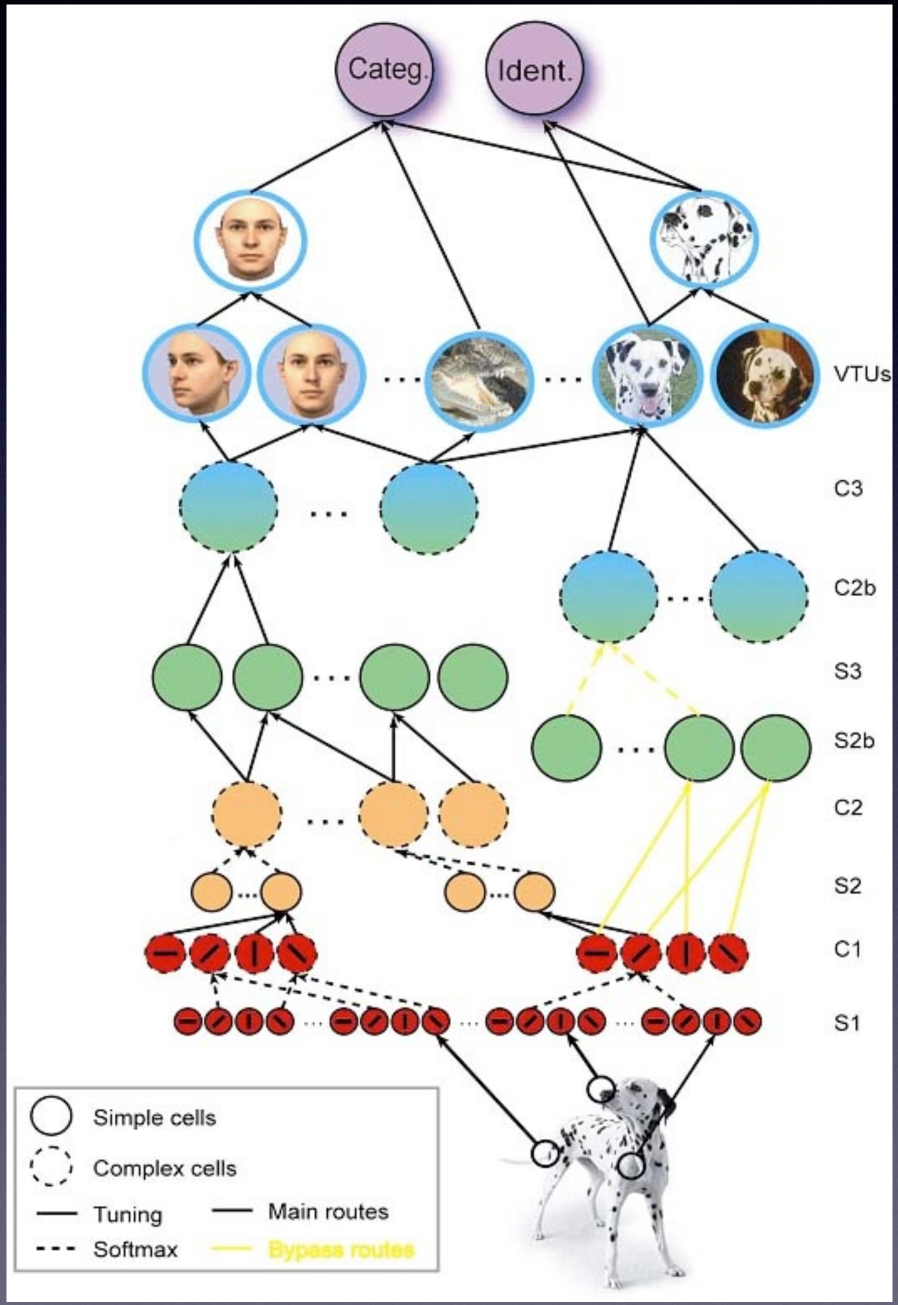




Adapted from Cadena et al. 2017

Vision Models - V2

- Hierarchy
- Spatially local filters
- Convolution
- Normalization
- …



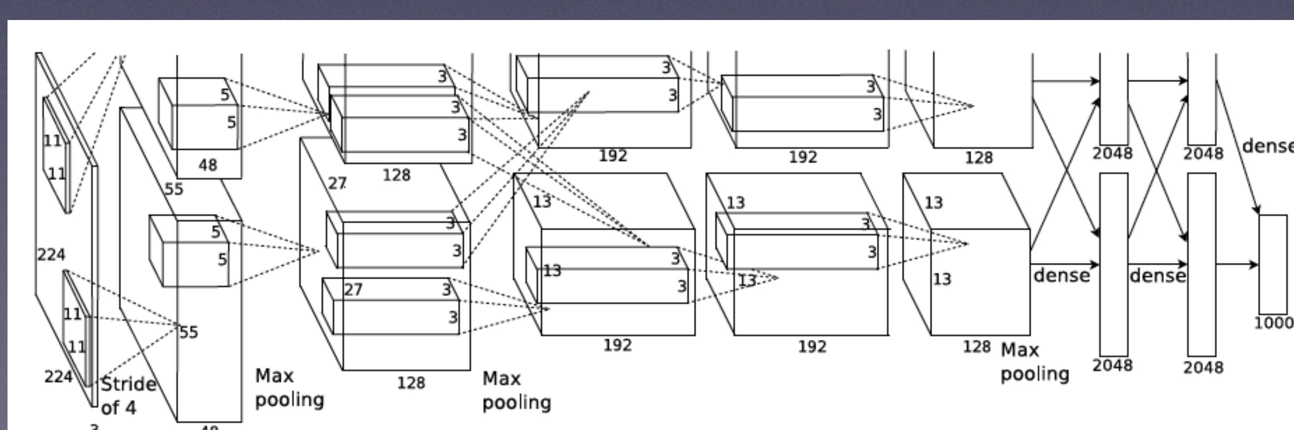
HMAX Model

Serre, Kouh, Cadieu, Knoblich, Kreiman & Poggio 2005

Vision Models - CNNs

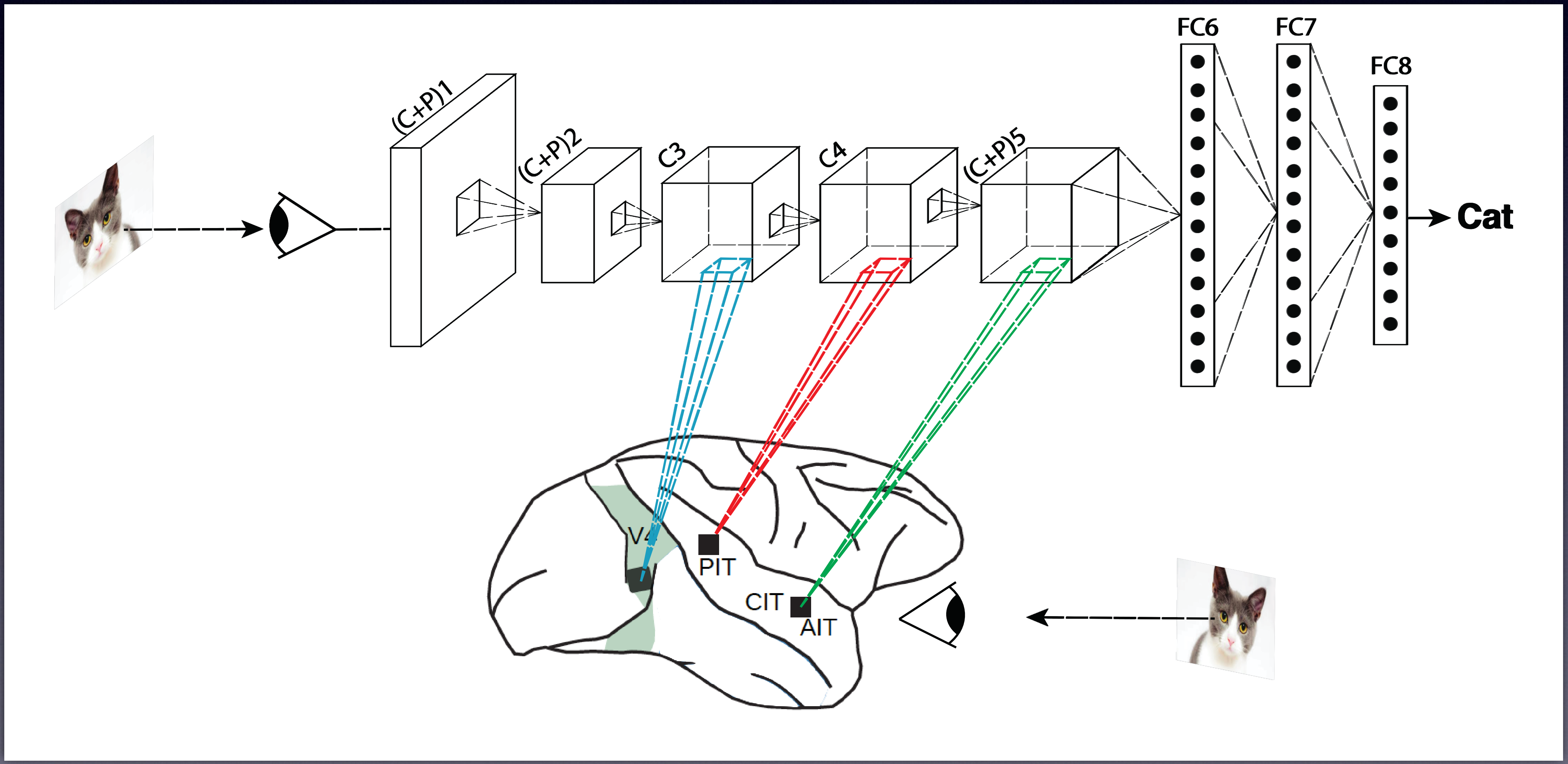
- Stack of Convolutions and Max-Pooling Layers with nonlinearities and normalization

- Parameters tuned on 1.28 M <u>labeled</u> images to minimize the <u>error on object classification</u> task.

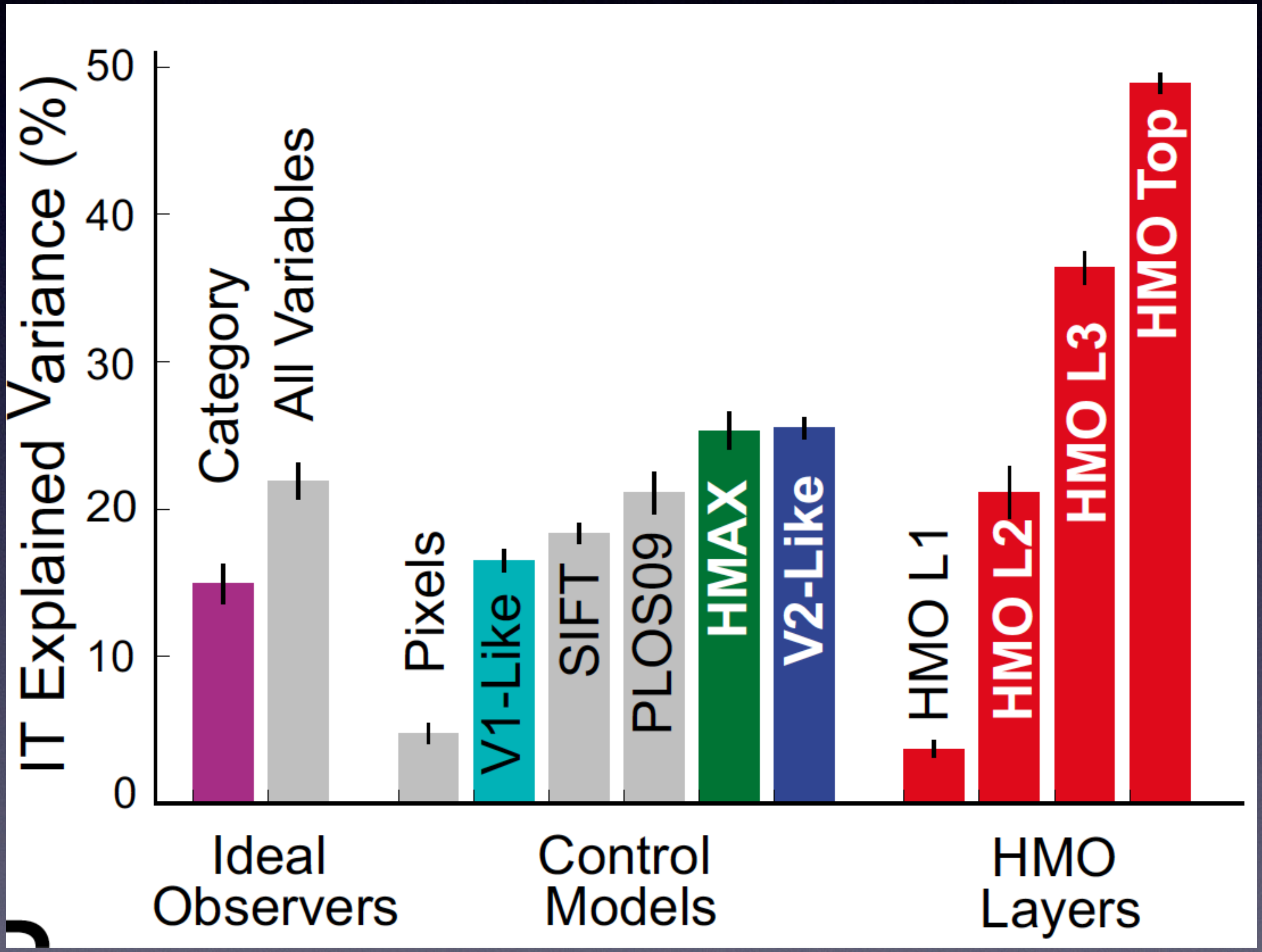


Adapted from Krizhevsky et al. NIPS 2012

Models of Higher Visual Areas
CNNs



Vision Models - CNNs

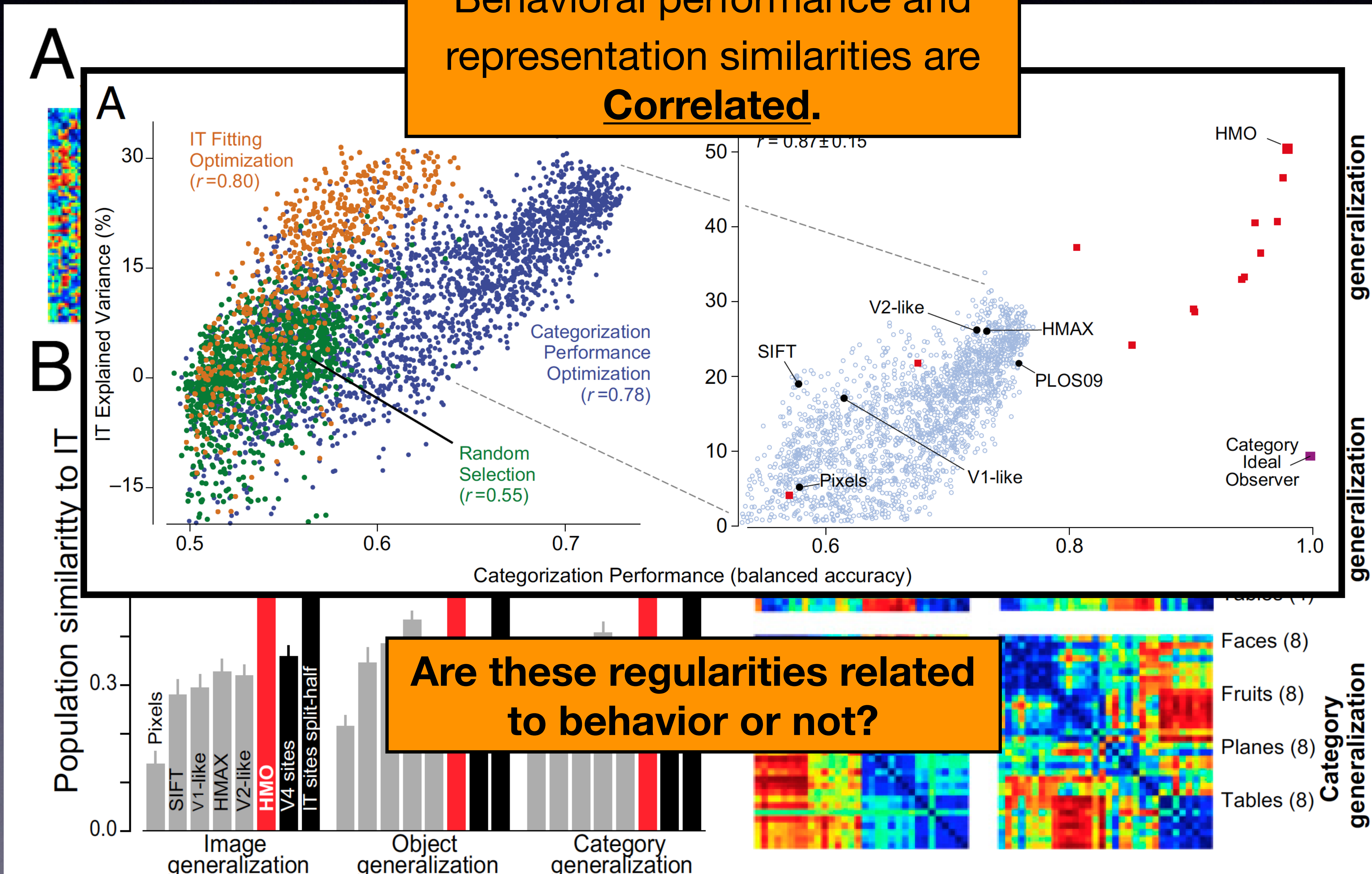


Higher layers of CNNs explain **50-60%** of the explainable variance in IT neural activity.

Adapted from Yamins, DiCarlo PNAS 2014

Vision Models - CNNs

Behavioral performance and representation similarities are **Correlated**.



Are these regularities related to behavior or not?

Adapted from Yamins, DiCarlo PNAS 2014

Applications-Automation

- Face recognition
- Self-driving Cars
- Security
- Many forms of intelligence…

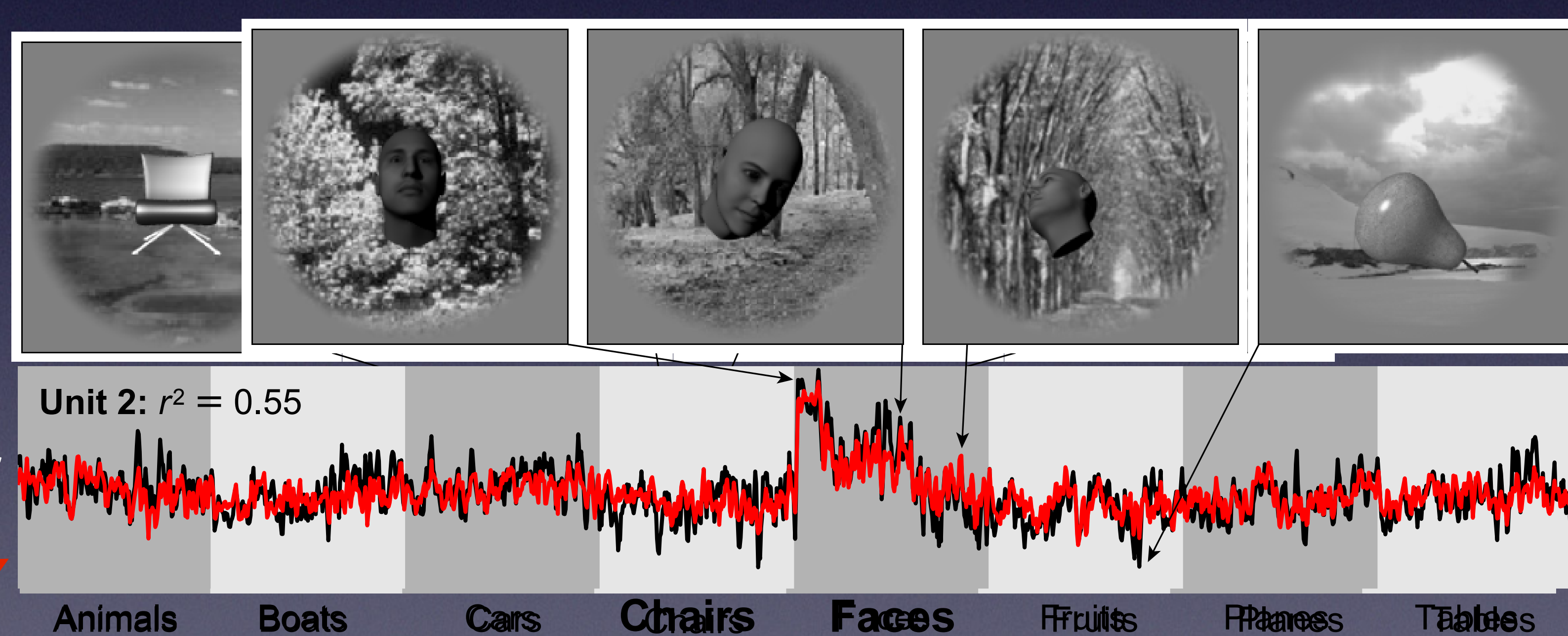
What about in neuroscience?

Vision is necessary to make sense of the visual world.

Neuroscience Applications
Prediction

Predictions of single site IT responses from layer 4 of HMO 1.0 model

*These are **predictions:** All of these objects and images were never previously seen by the HMO model*



Response* of IT neural site

Prediction of HMO model

(* mean rate 70-170 ms after image onset)

Yamins, Hong, Solomon, Seibert and DiCarlo ***PNAS (2014)***

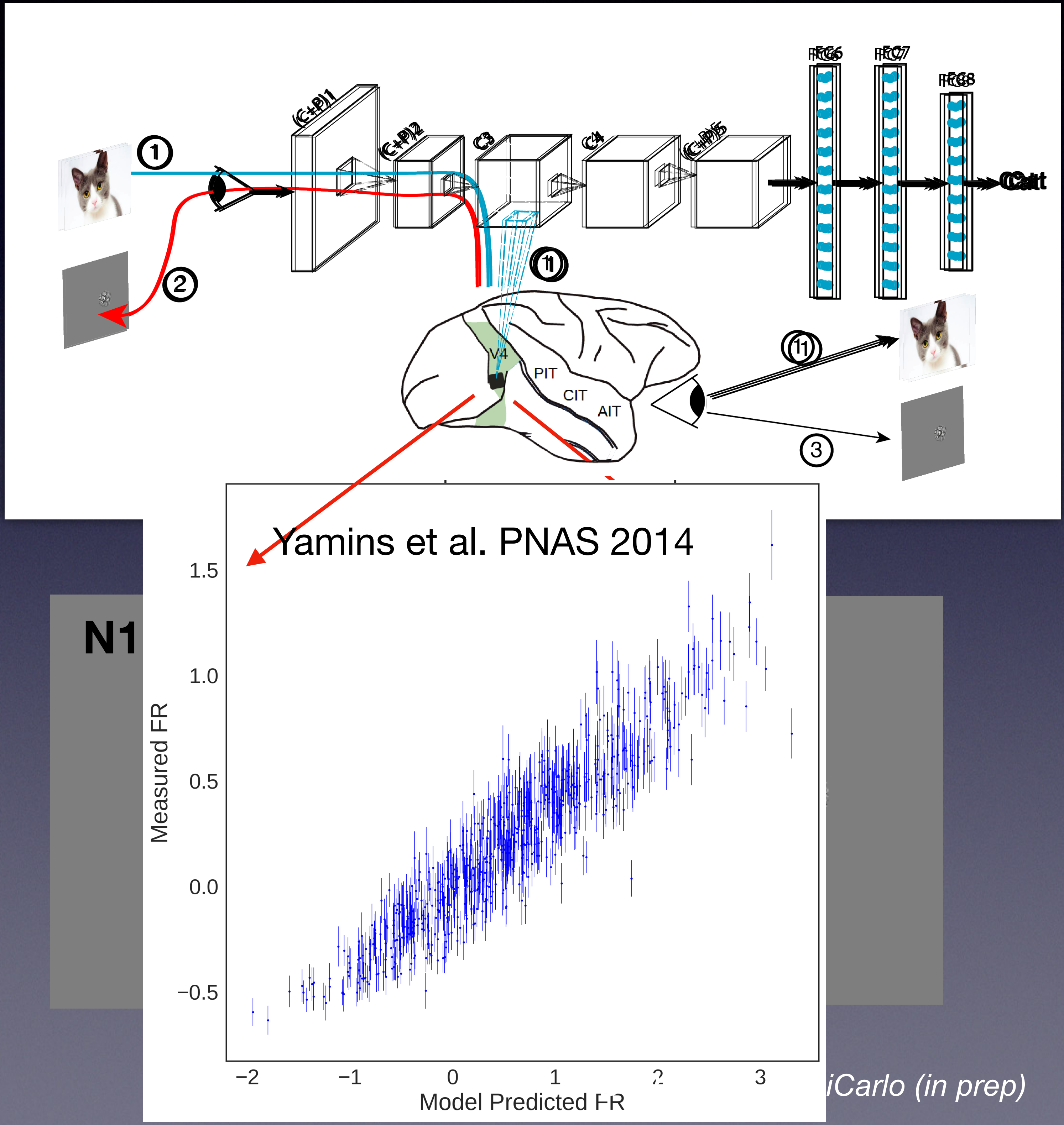
Neuroscience Applications
Control

- As a stronger model test, we here asked <u>can we use a model to generate new visual images to drive a recorded neural population response into any desired state?</u>

- We started with two specific goals:

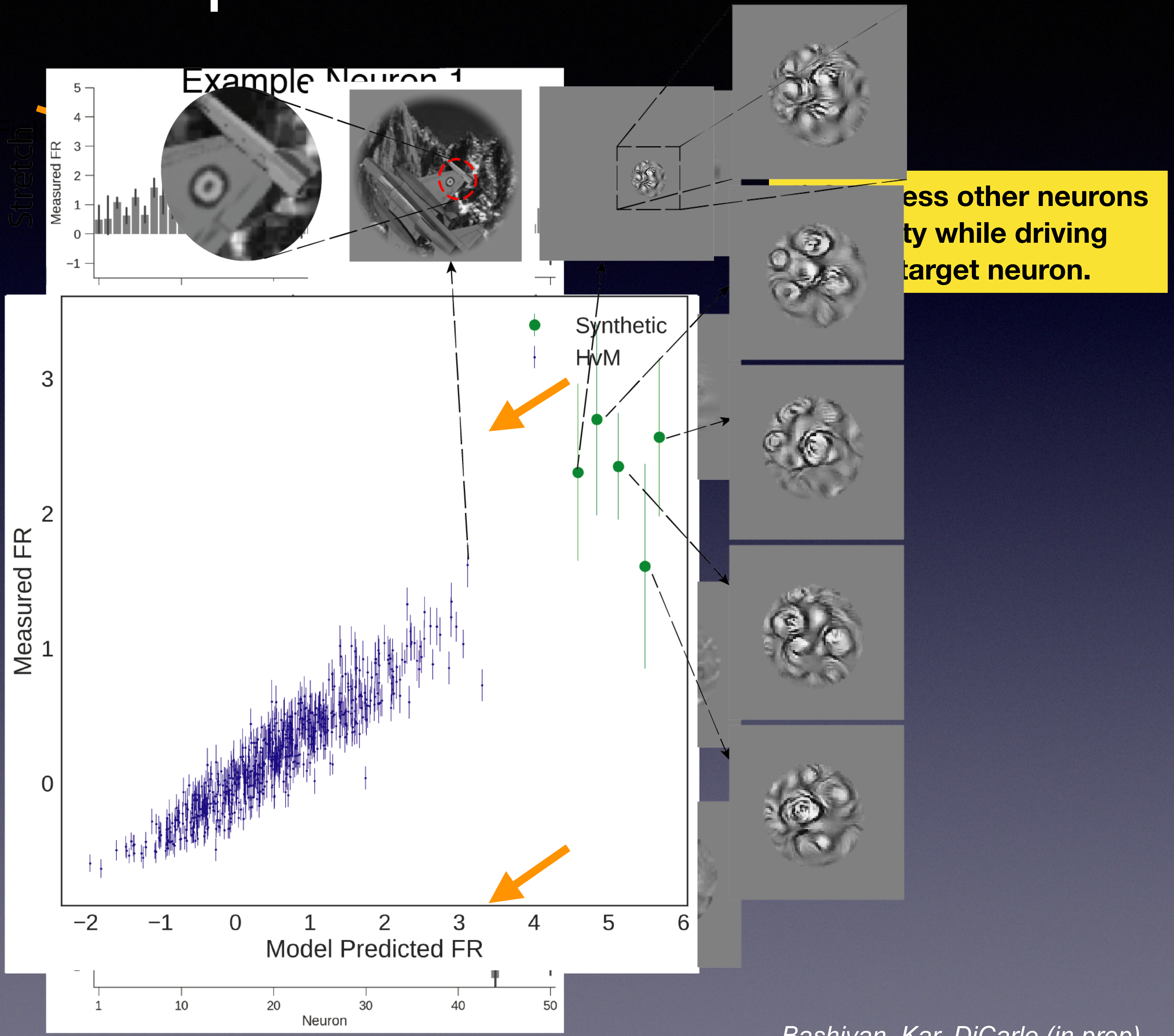
    1. **“Maximal drive” (Stretch)**: Drive any <u>single</u> neuron’s response beyond the <u>maximum response</u> observed thus far.

    2. **“One hot population” (OHP)**: Drive any given neuron’s response up while holding the responses of all other (recorded) neurons at baseline (a test of <u>independent control</u>).



Neural Population Control

1) Stretch the firing rates beyond previously observed



Bashivan, Kar, DiCarlo (in prep)

Questions

“The only stupid question is the one that is never asked.”

Ramon Bautista