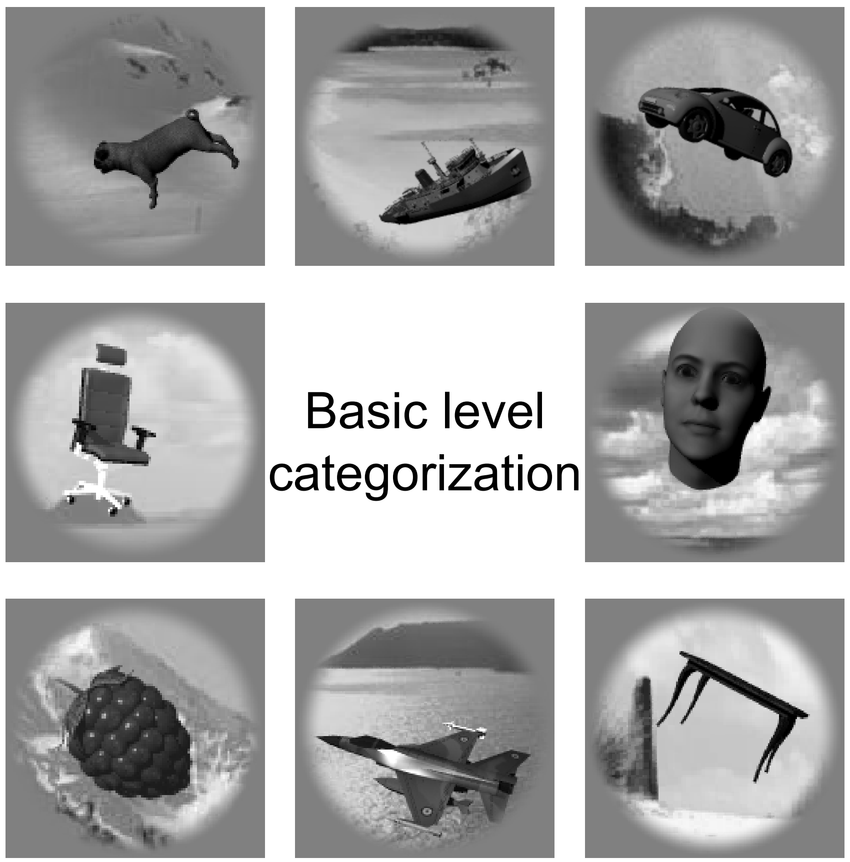
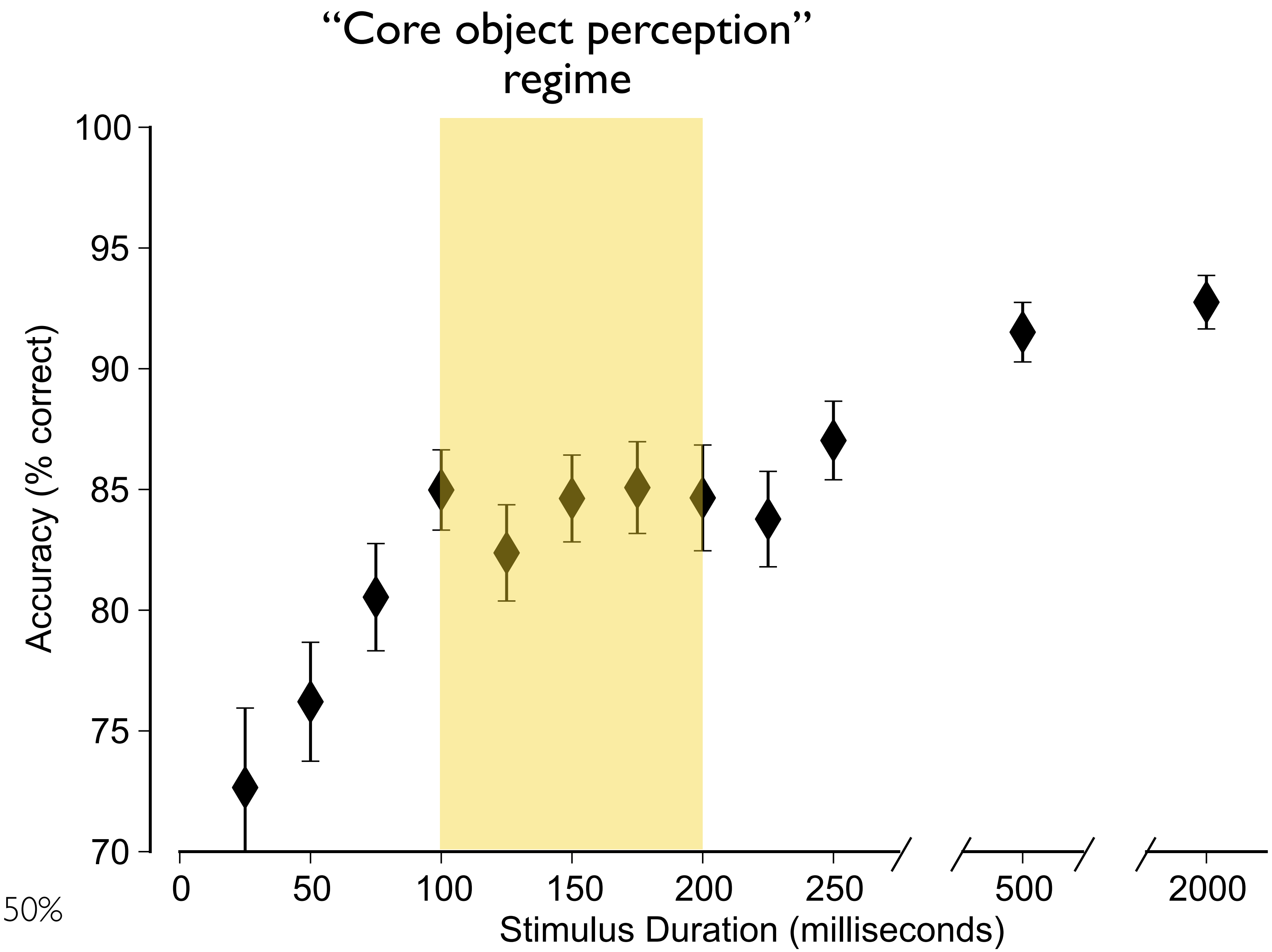
CS375 / Psych 249:
Large-Scale Neural Network Models for Neuroscience

Lecture 3: Deep CNNs and the Ventral Visual stream
2025.01.12

Daniel Yamins
Departments of Computer Science and of Psychology
Stanford Neuroscience and Artificial Intelligence Laboratory
Wu Tsai Neurosciences Institute
Stanford University

Problem: Entity Extraction

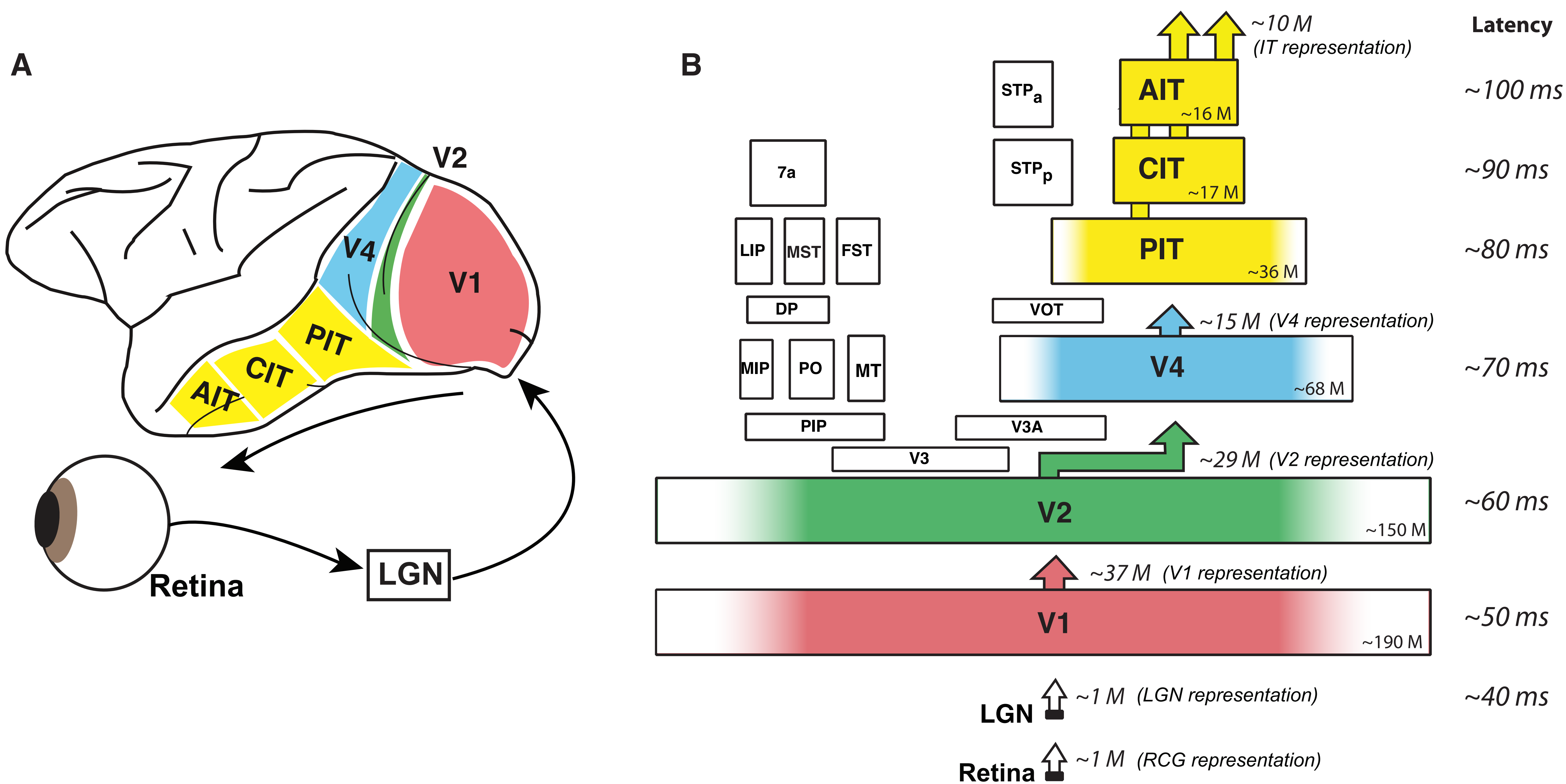




Chance is 50%

All the data I will show you today

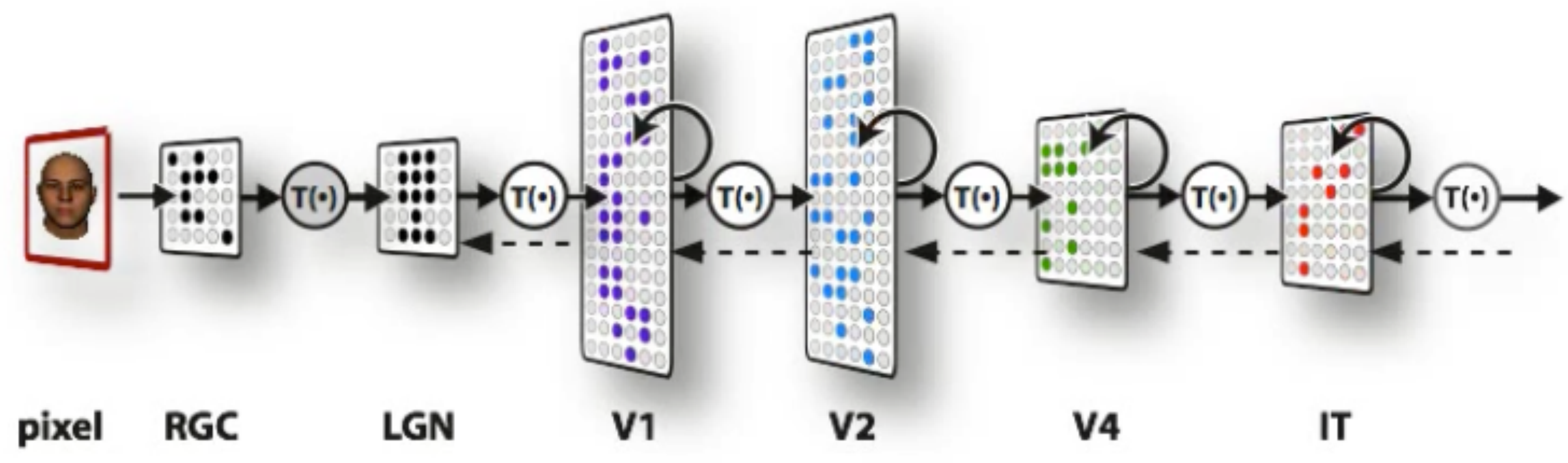
Typical primate fixation duration during natural viewing



Adapted from DiCarlo et al. 2012

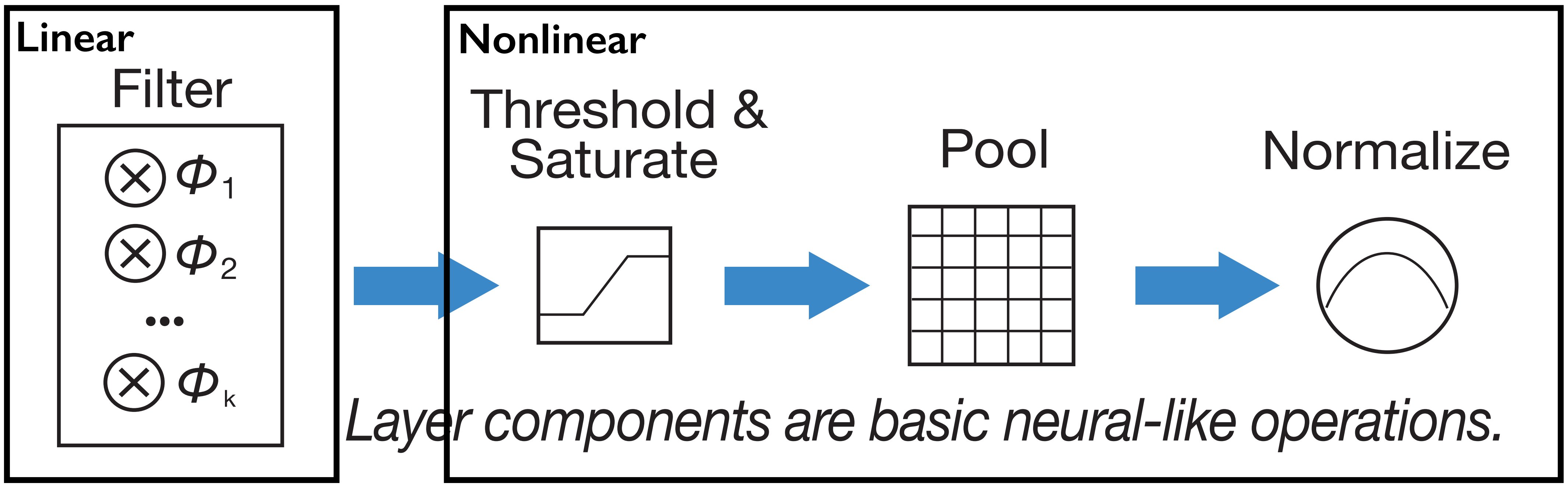
Problem: Entity Extraction

GOAL: Predictive model of single-neuron responses throughout the ventral stream to arbitrary image stimuli.



What We Learned from V1

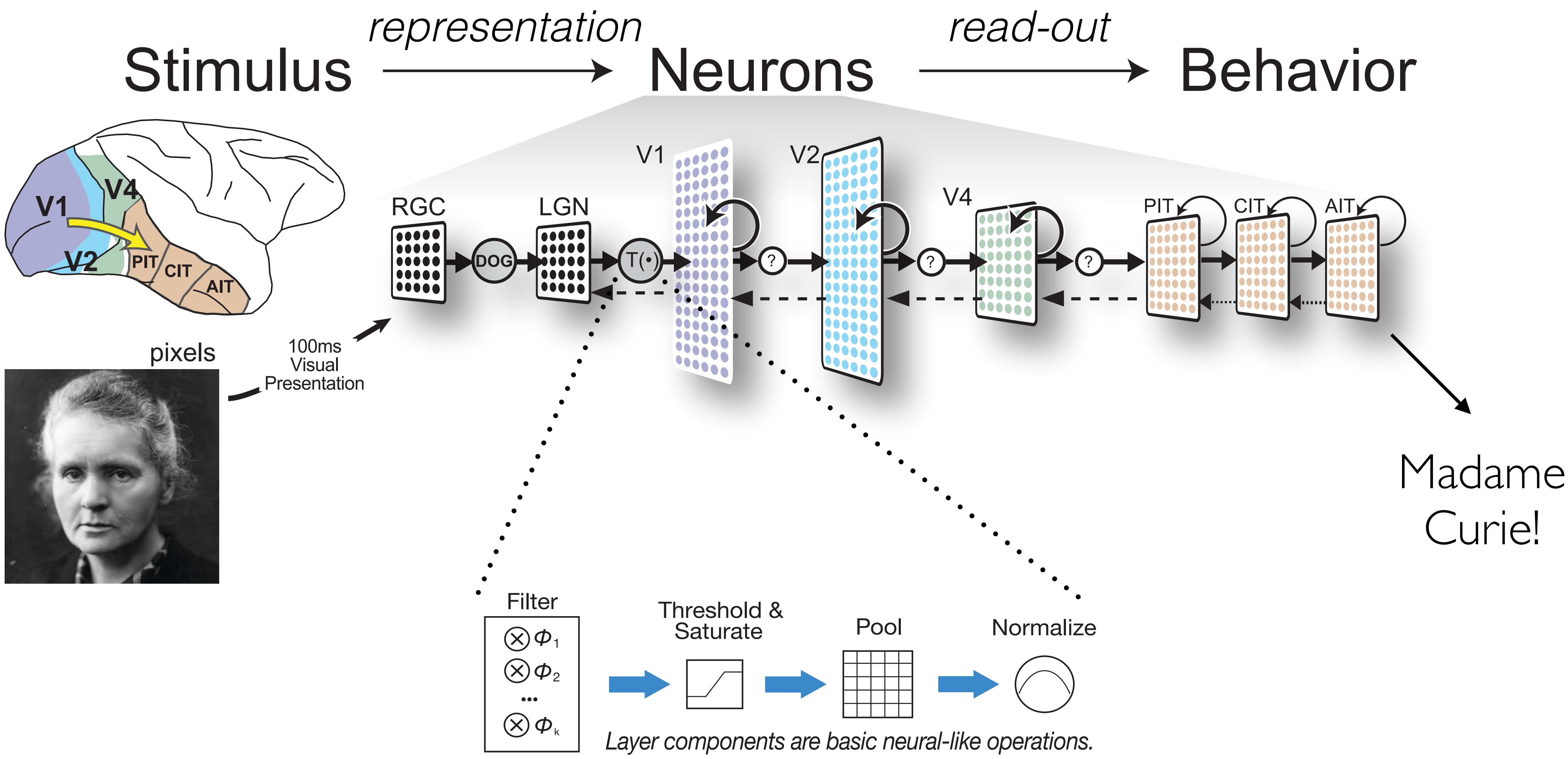
- Linear-Nonlinear neurally-plausible **basic operations** within layer



neuro:	synaptic weights patterns	single-unit activations	complex cells	competitive inhibition
data:	untangling through dimension expansion	"AND" operation by limiting dynamic range	adding robustness by dimension reduction	put results back into standard range

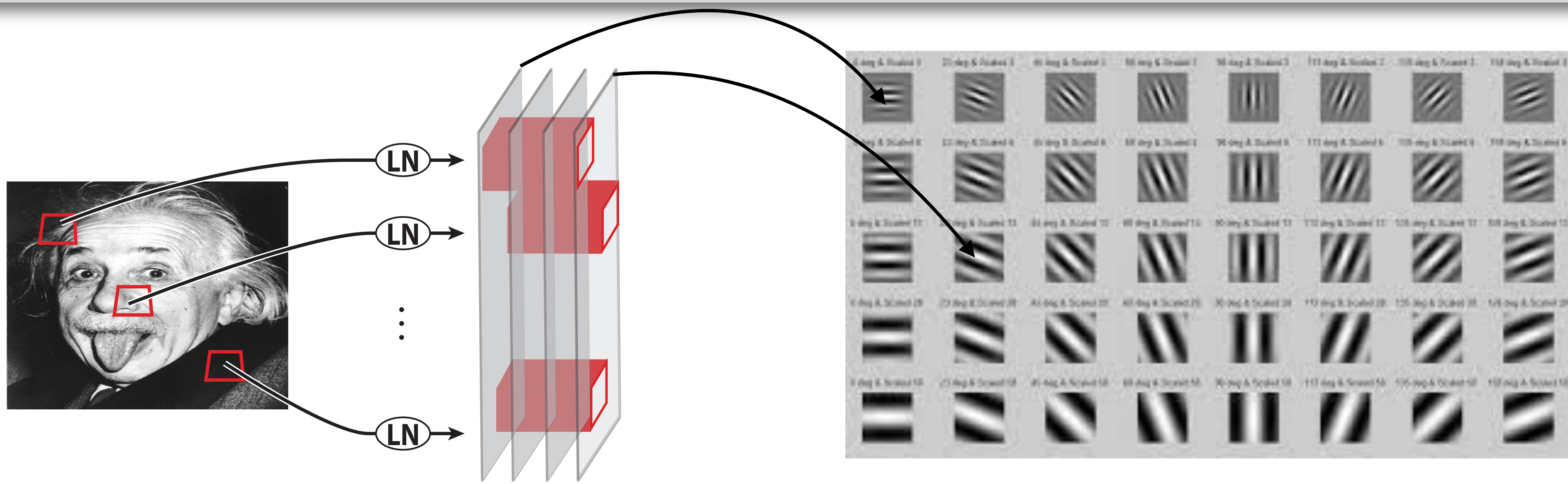
Hubel and Wiesel (1965-1975) Lecun (2004), Carandini et. al (2005), Lennie & Movshon (2005) , DiCarlo (2012)

What We Learned from V1

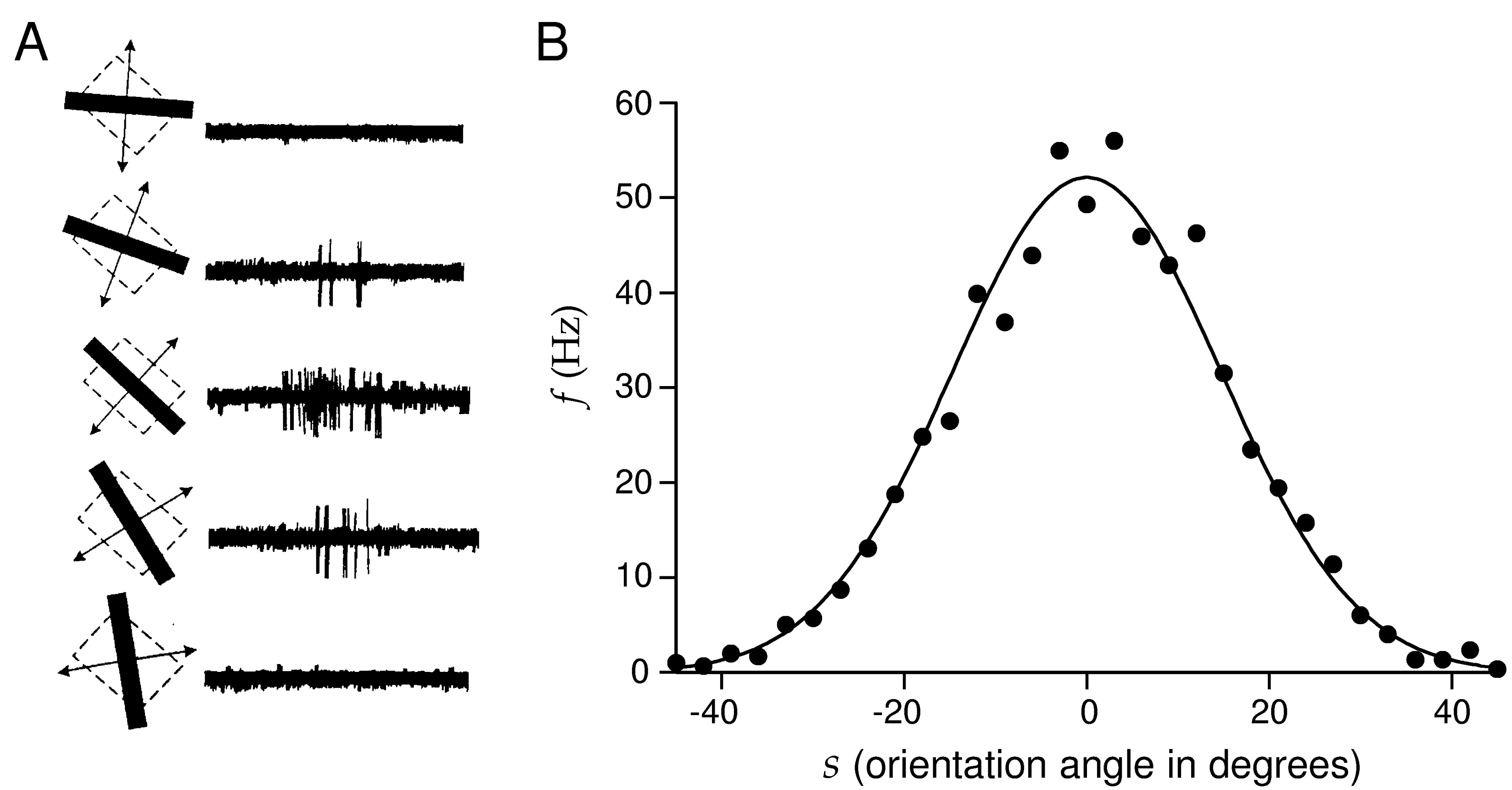


Linear-Nonlinear neurally-plausible **basic operations** within layer

What We Learned from V1



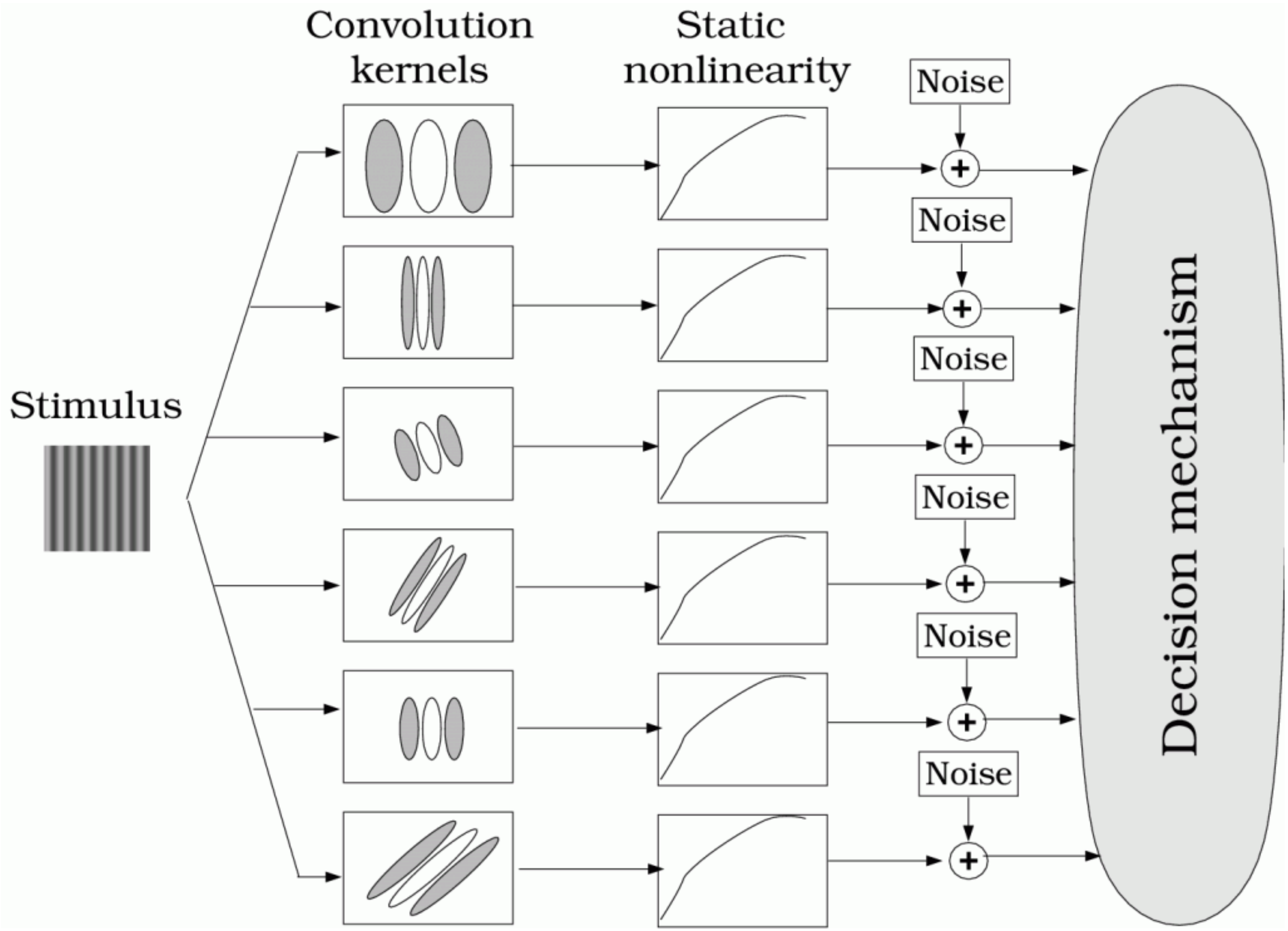
Gaussian tuning curve of V1 simple cell



adapted from Adrienne Fairhall

"Hubel and Wiesel's Intuition"
~1970s and formalized later
via Gabor wavelets

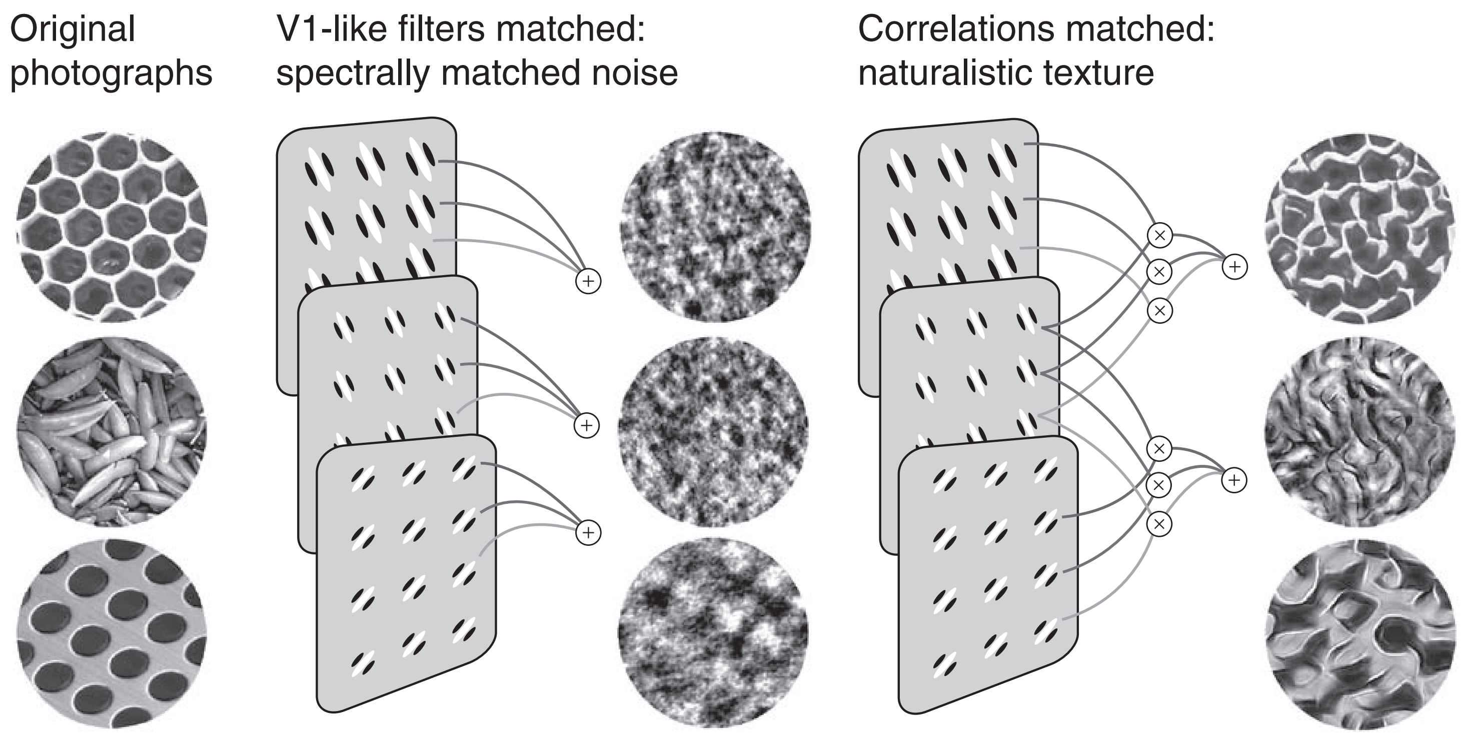
What We Learned from V1



from Wandell 1996

What We Learned from V2 and V4?

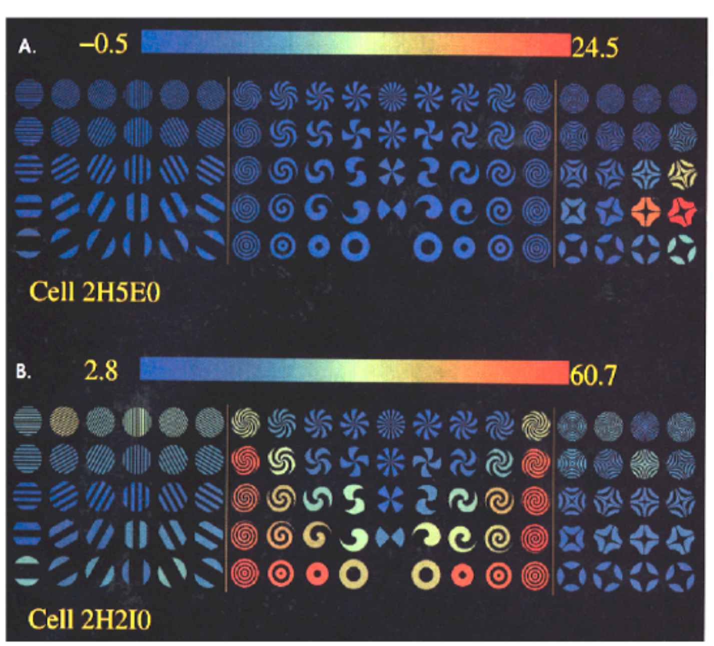
V2:

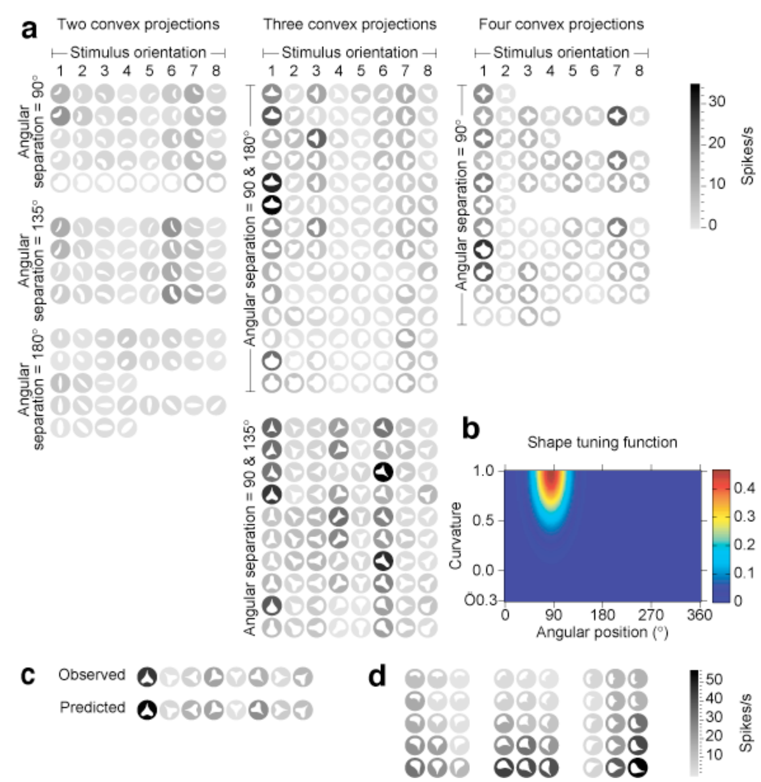


So, maybe a <u>hierarchically-built</u> sparse auto-encoding in a 2-layer model with max pooling?? … but doesn't really work well in practice.

V4:

V4 Responses to Non-Cartesian Gratings
Gallant et al. 1996

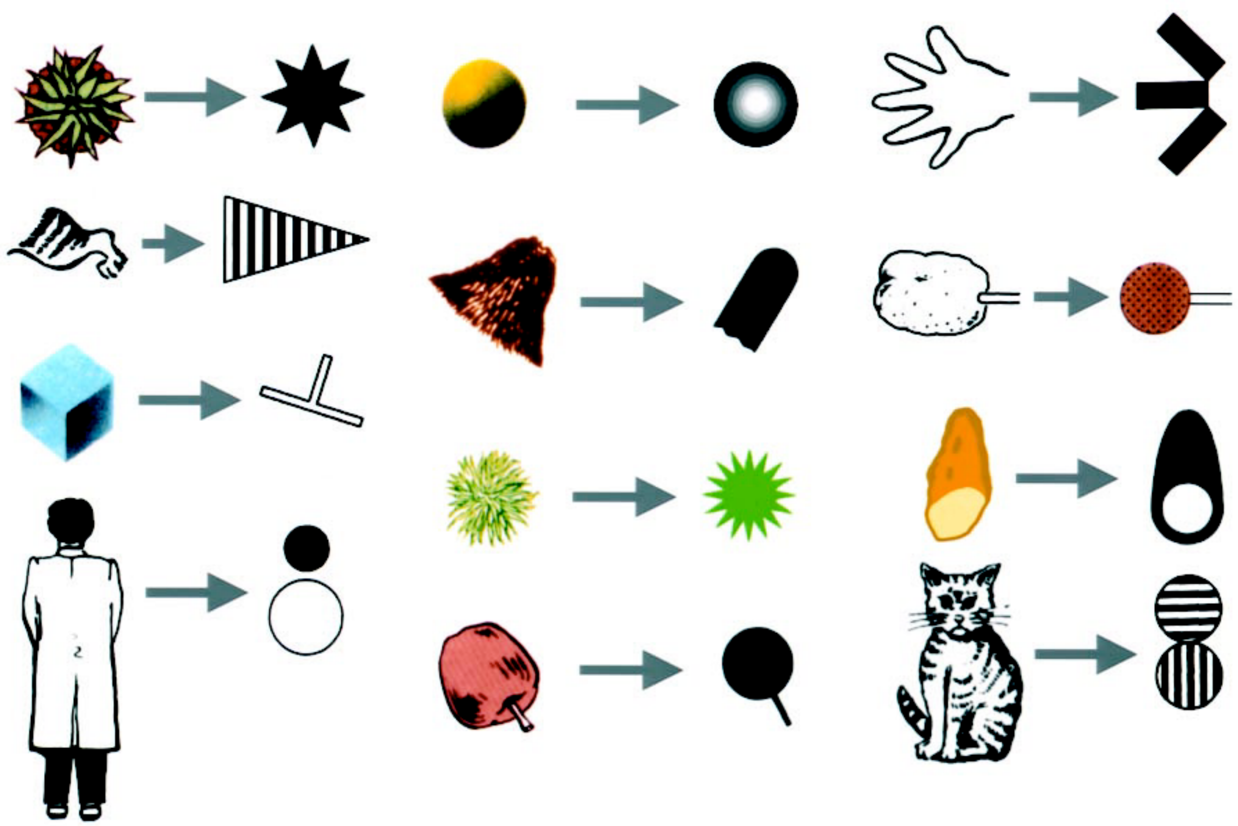


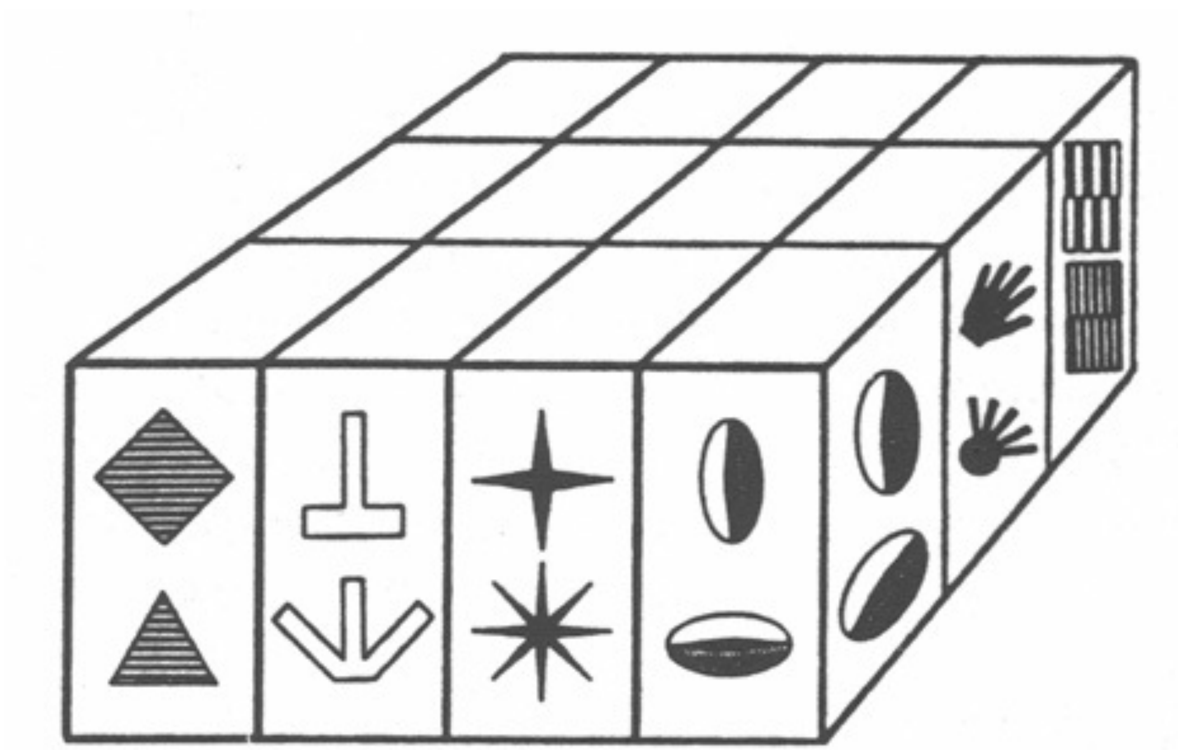


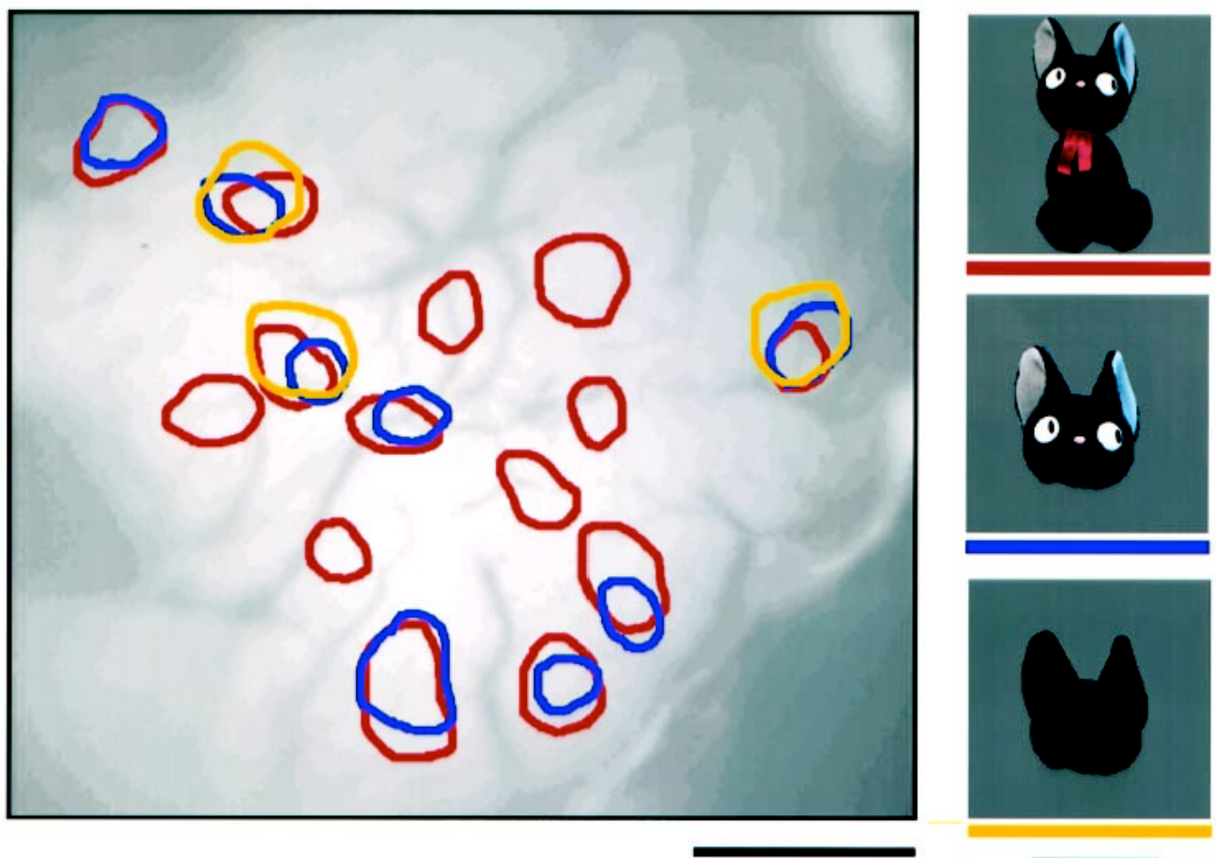
Problem:
No predictions for any other images.
i.e.
is not an "image-computable" model

What We Learned from IT?

IT:



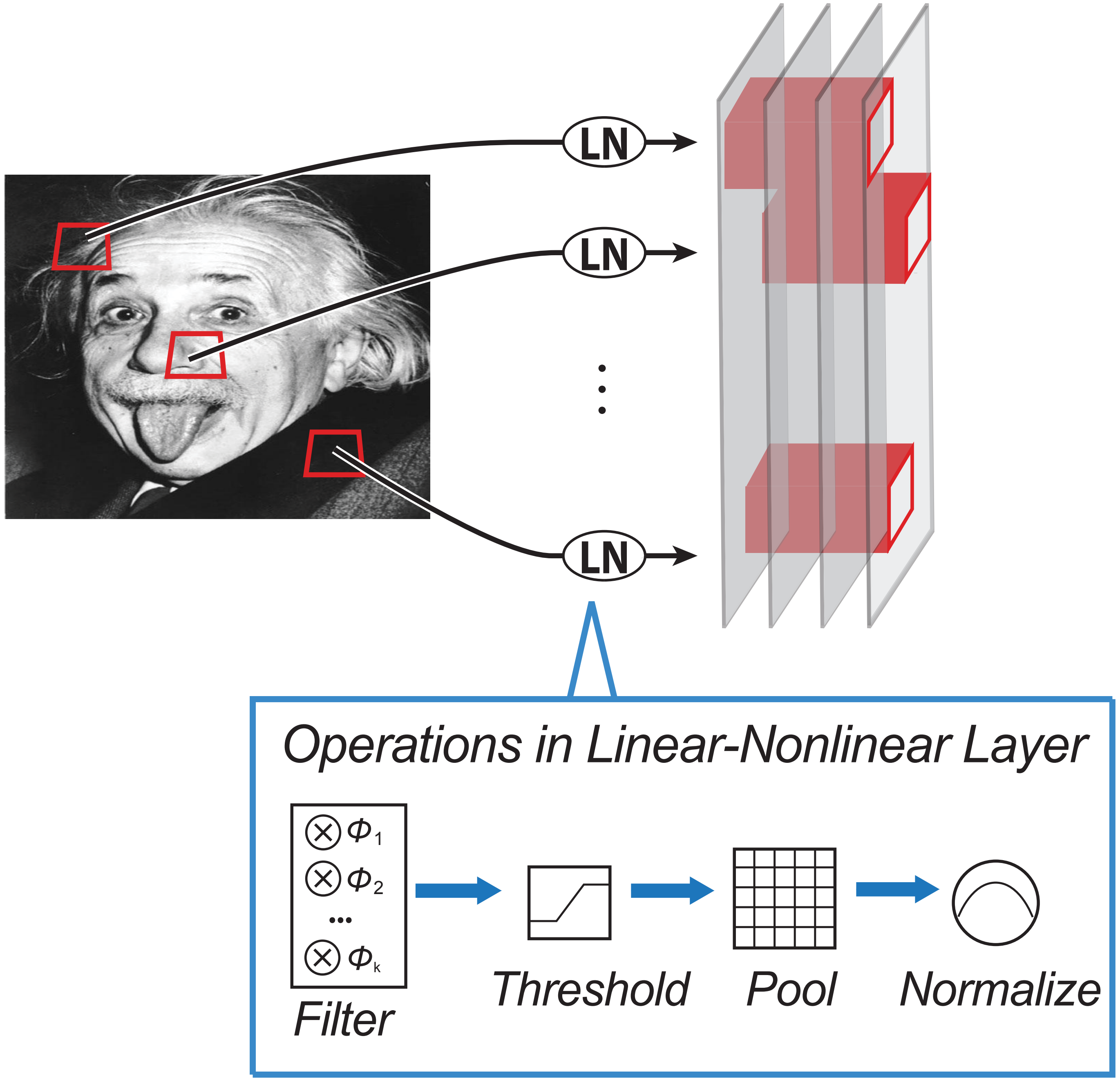




Tsunoda et al.

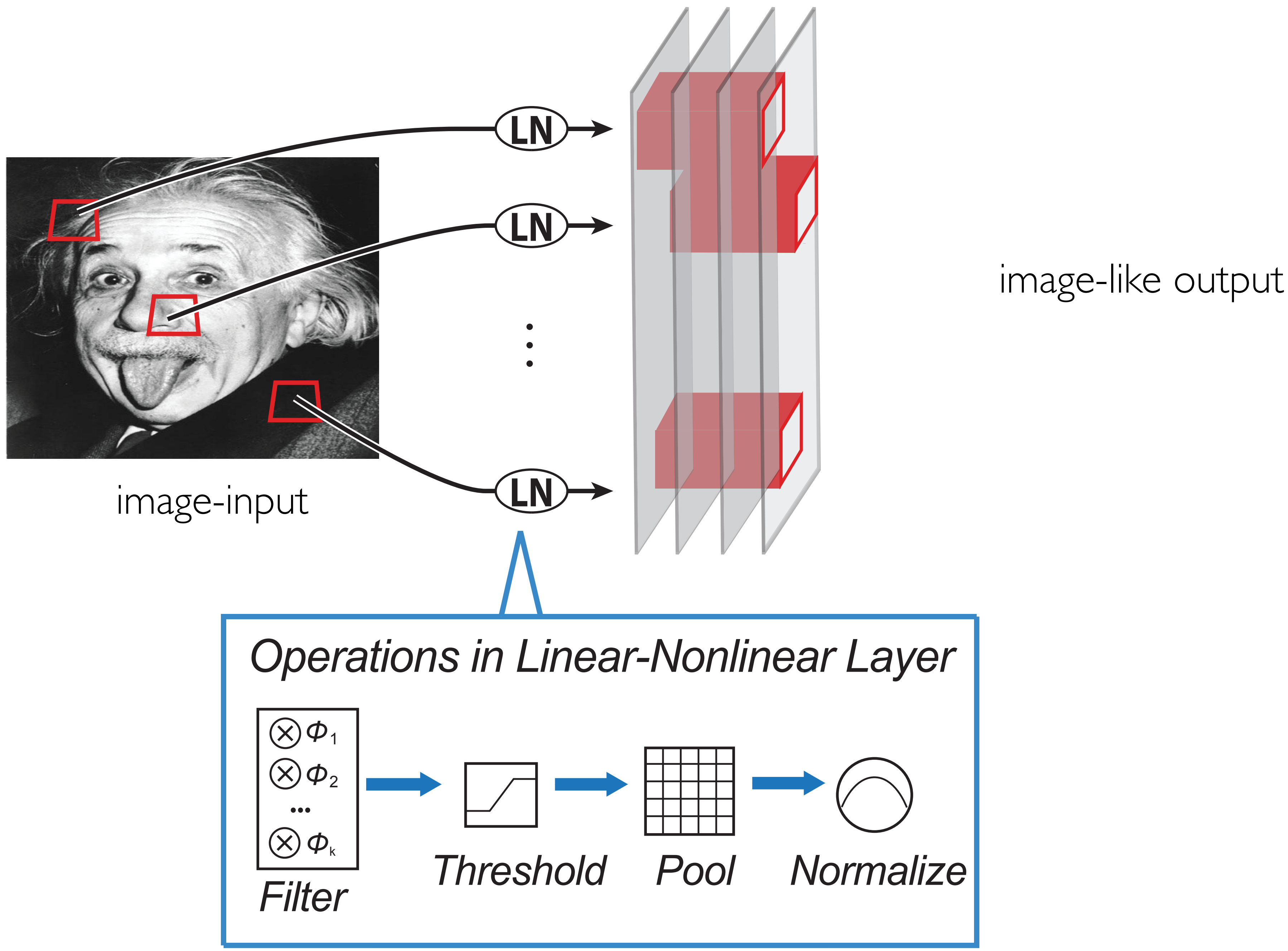
Hierarchical Convolutional Neural Networks

- Individual layers of neurally-plausible **basic operations**
- Applied **convolutionally** — same at all locations: approx. retinopy



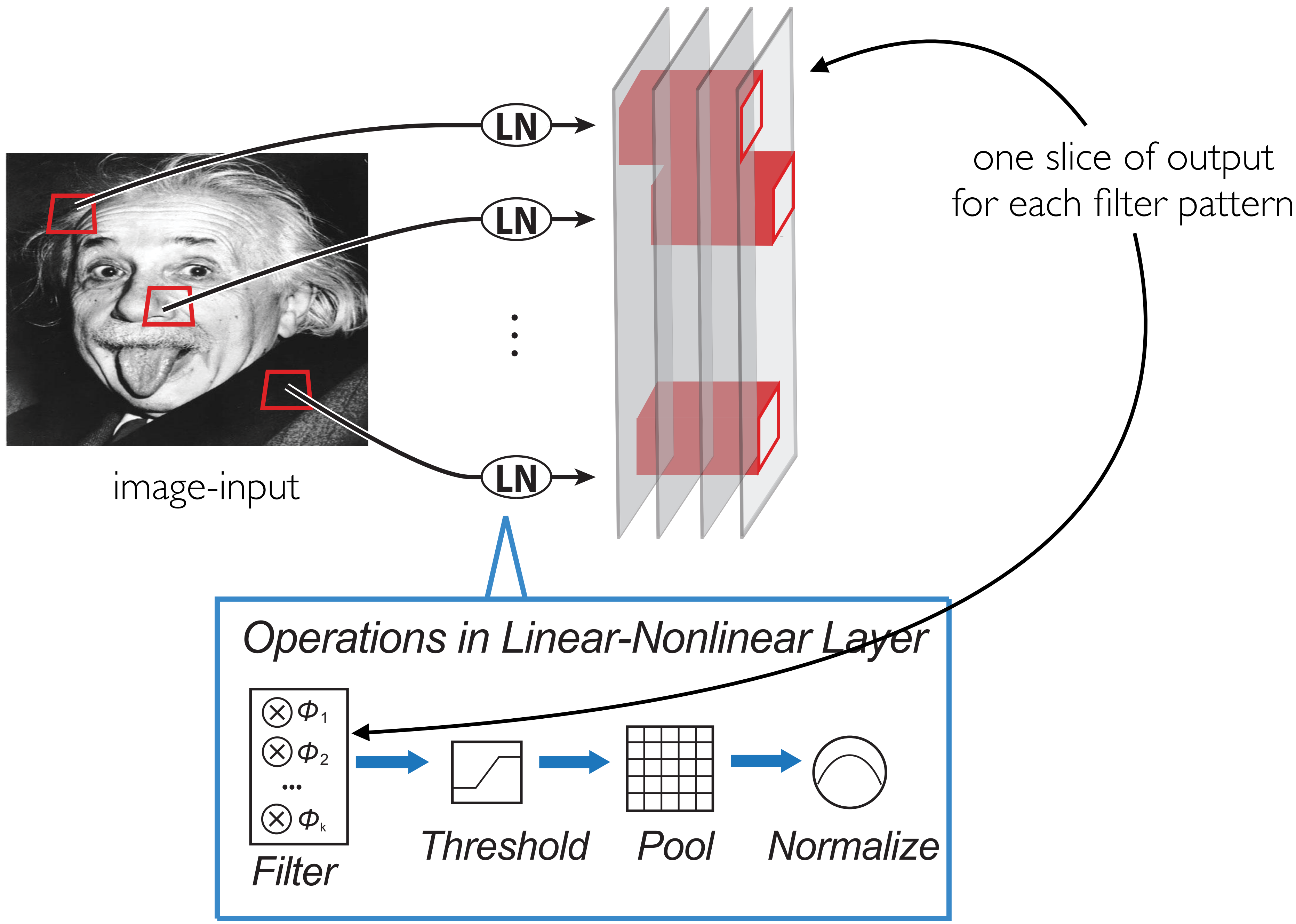
Hierarchical Convolutional Neural Networks

- Individual layers of neurally-plausible **basic operations**
- Applied **convolutionally** — same at all locations: approx. retinopy



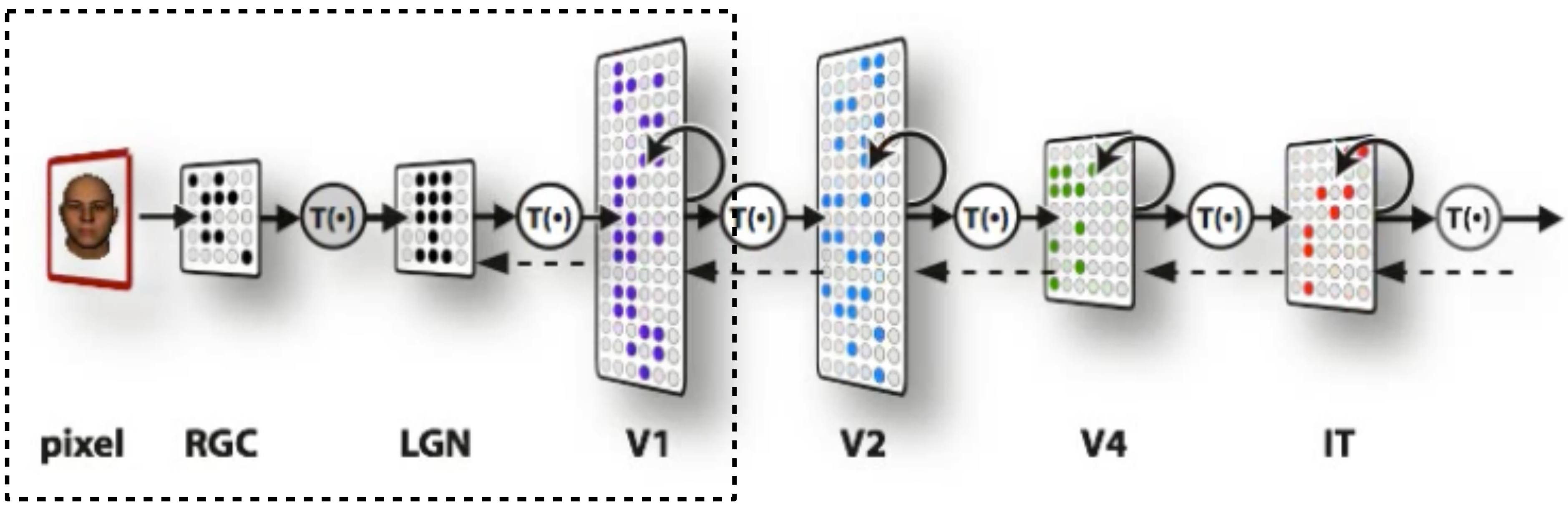
Hierarchical Convolutional Neural Networks

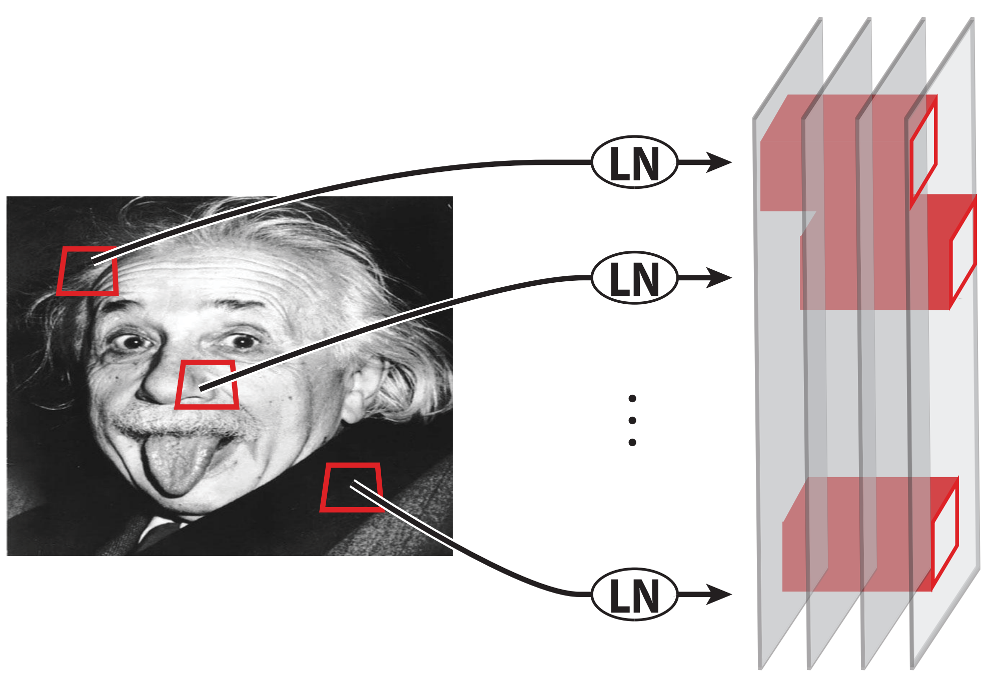
- Individual layers of neurally-plausible **basic operations**
- Applied **convolutionally** — same at all locations: approx. retinopy



Hierarchical Convolutional Neural Networks

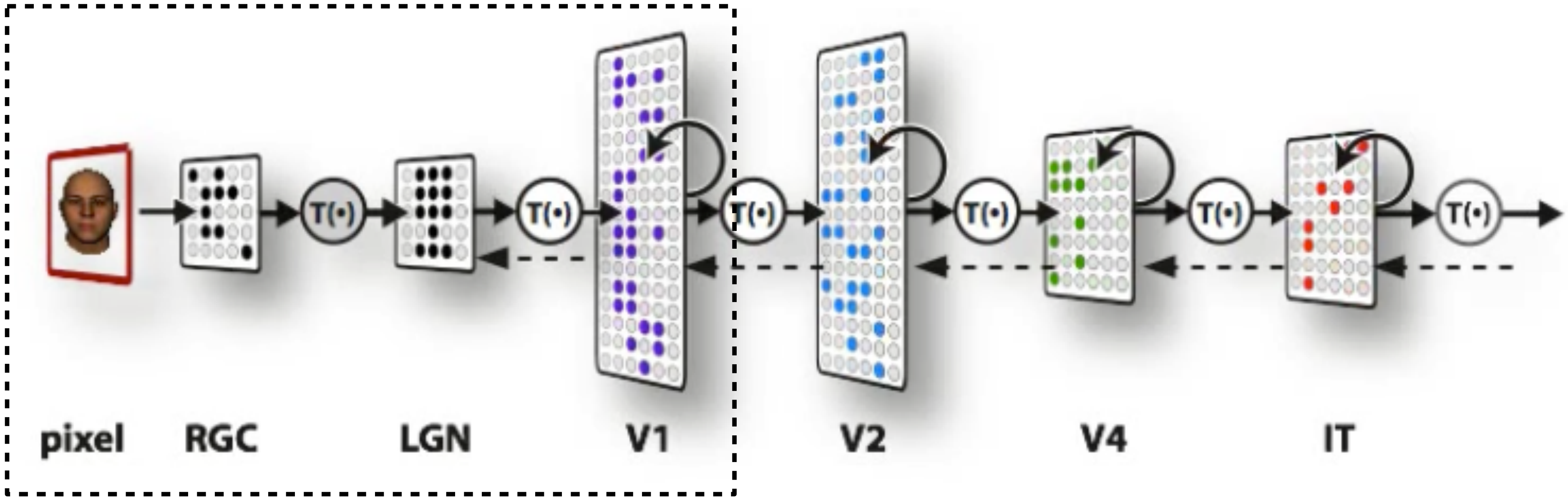
Lower areas, (RGC, LGN,V1) have been reasonably captured by single-layer models: <u>~40% of variance explained</u>. Carandini et. al (2005), Lennie & Movshon (2005)

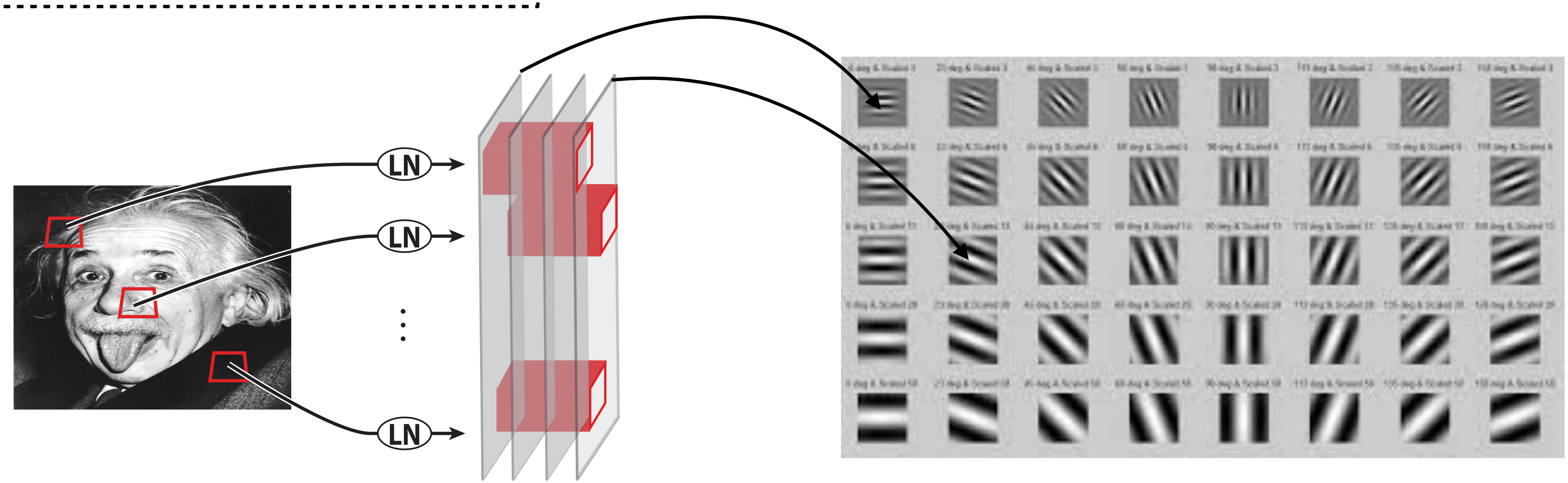




Hierarchical Convolutional Neural Networks

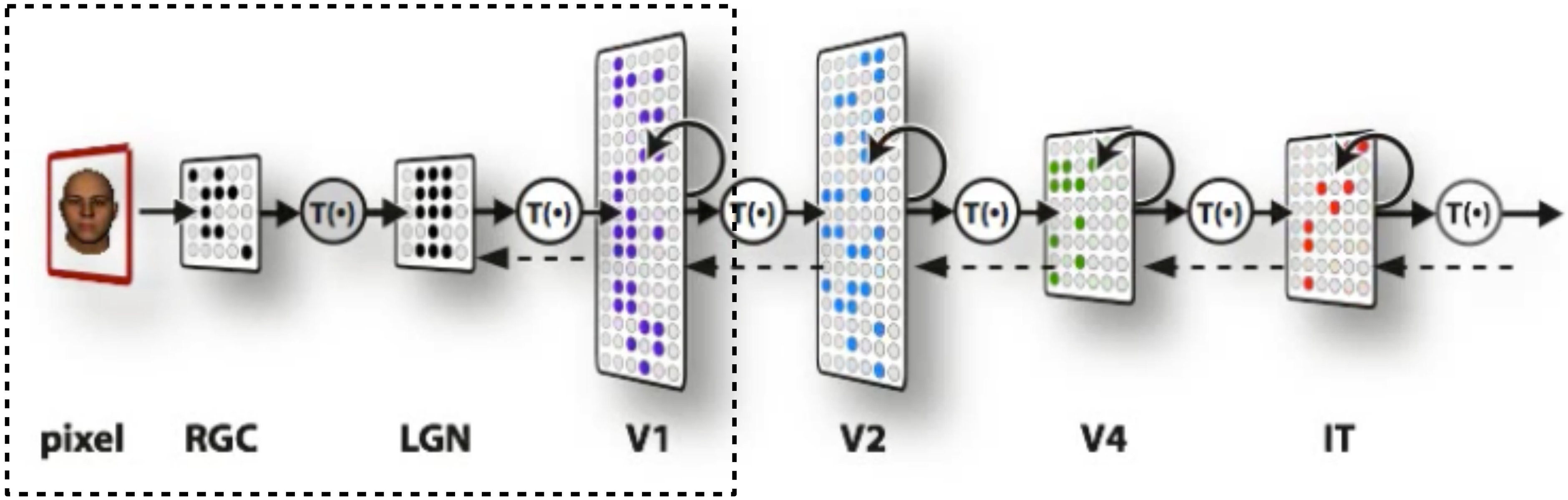
Lower areas, (RGC, LGN, V1) have been reasonably captured by single-layer models: ~50% of variance explained. Carandini et. al (2005), Lennie & Movshon (2005)

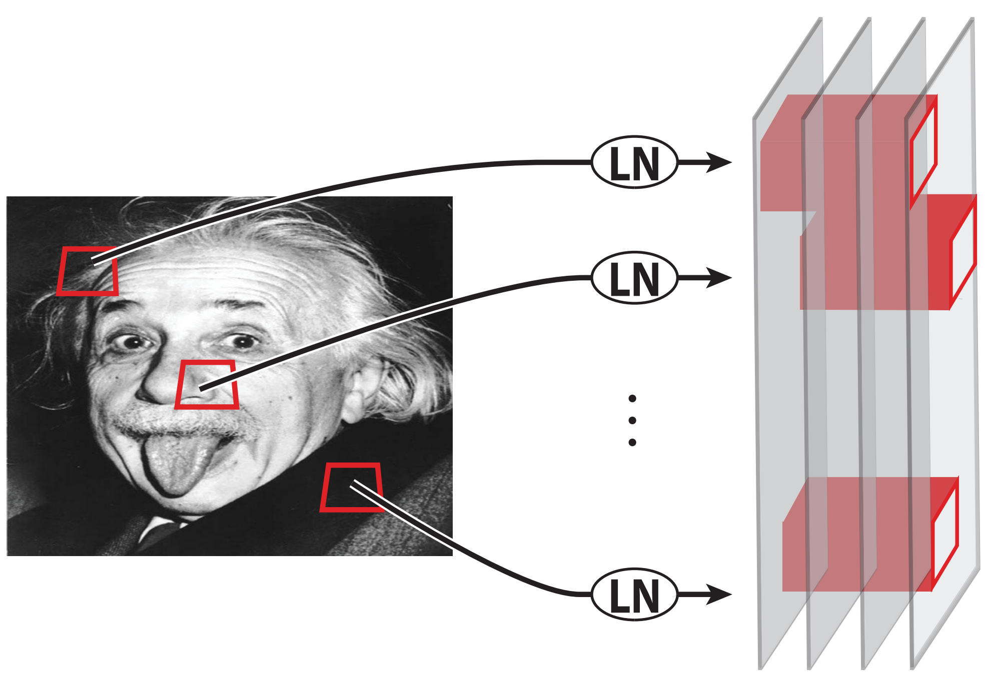




Hierarchical Convolutional Neural Networks

Lower areas, (RGC, LGN, V1) have been reasonably captured by single-layer models: ~50% of variance explained. Carandini et. al (2005), Lennie & Movshon (2005)





Push up the ventral stream?

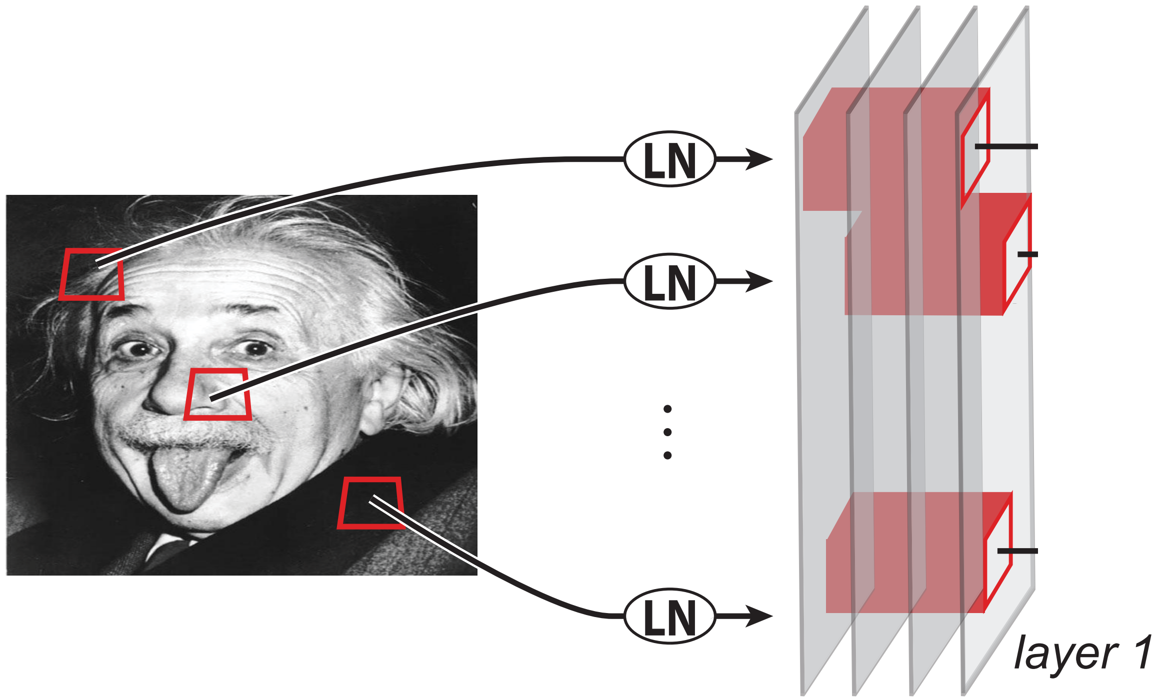
Hierarchical CNNs

Hierarchical Convolutional Neural Networks

- Individual layers of neurally-plausible **basic operations**

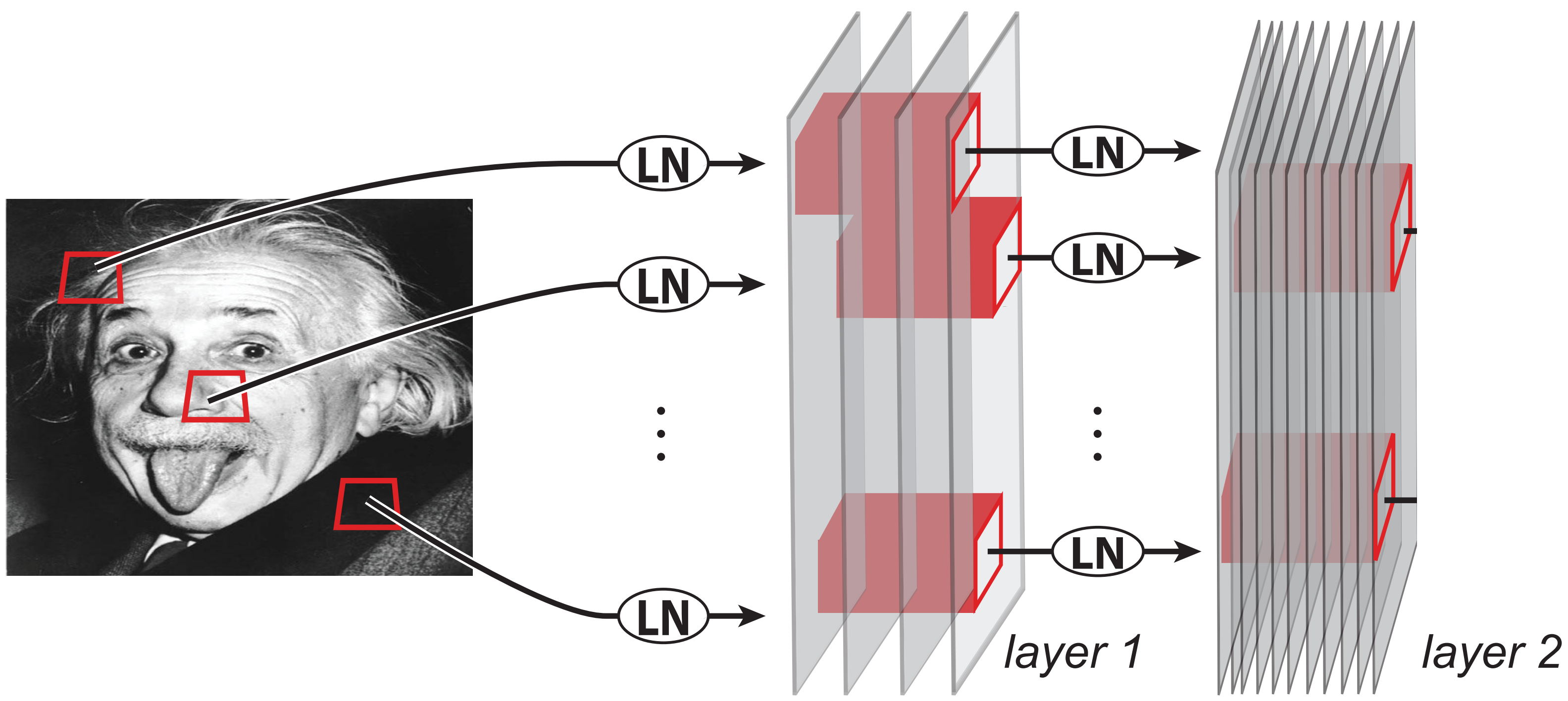
- Applied **convolutionally** — same at all locations

- Stacked **hierarchically** to produce more complex operations



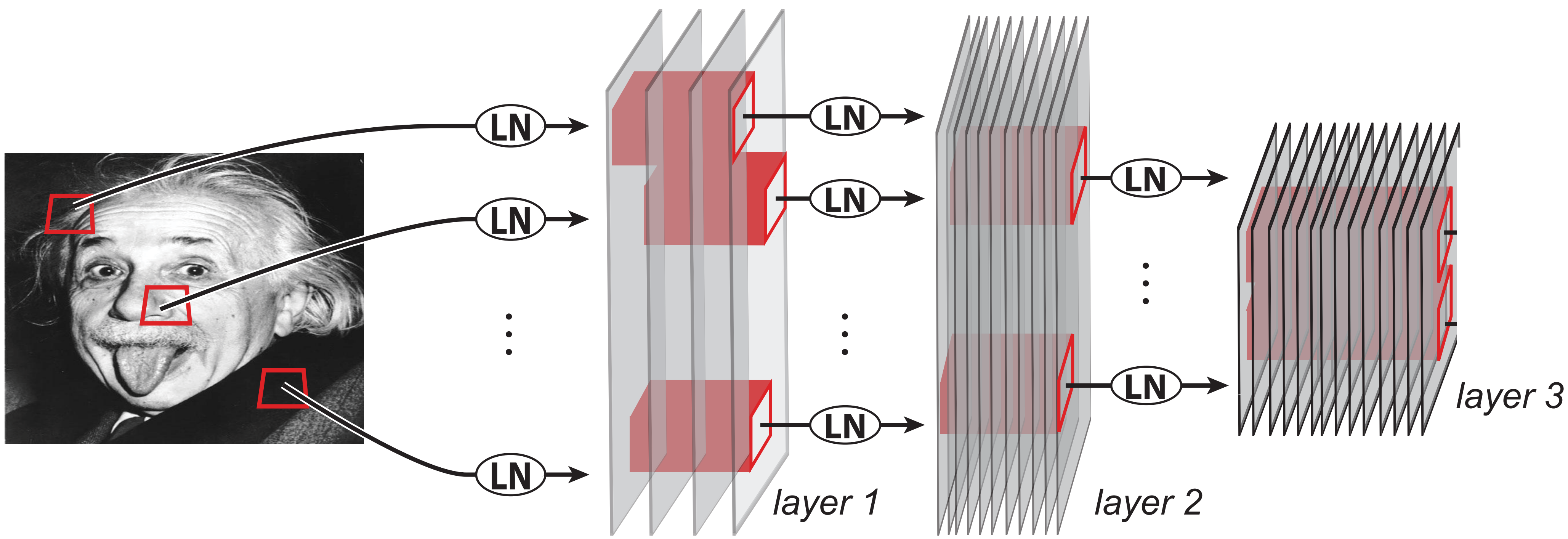
Hierarchical Convolutional Neural Networks

- Individual layers of neurally-plausible **basic operations**
- Applied **convolutionally** — same at all locations
- Stacked **hierarchically** to produce more complex operations



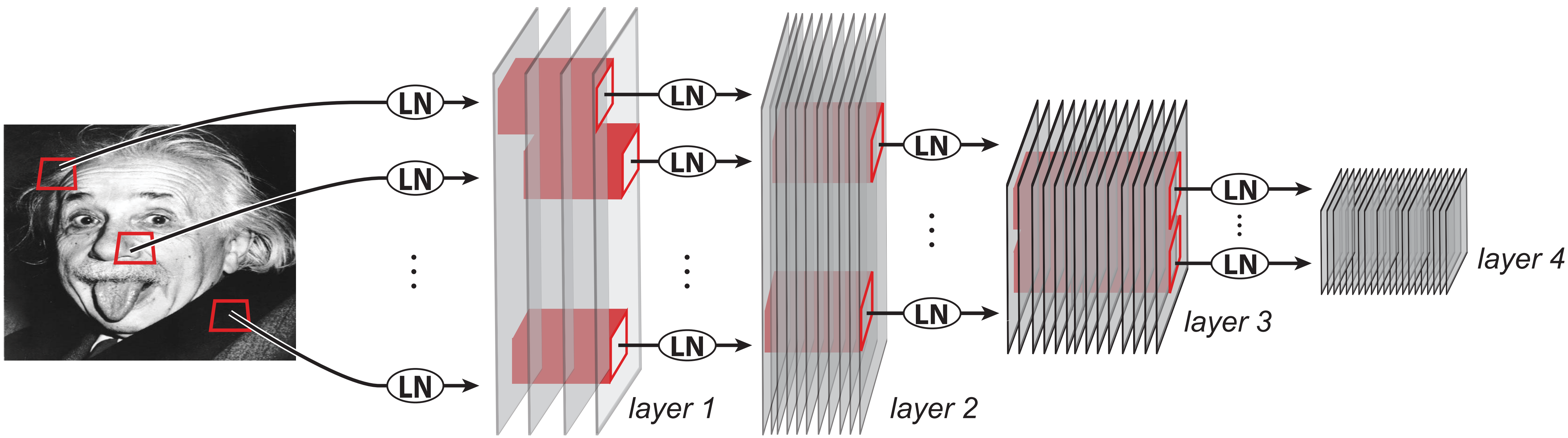
Hierarchical Convolutional Neural Networks

- Individual layers of neurally-plausible **basic operations**
- Applied **convolutionally** — same at all locations
- Stacked **hierarchically** to produce more complex operations



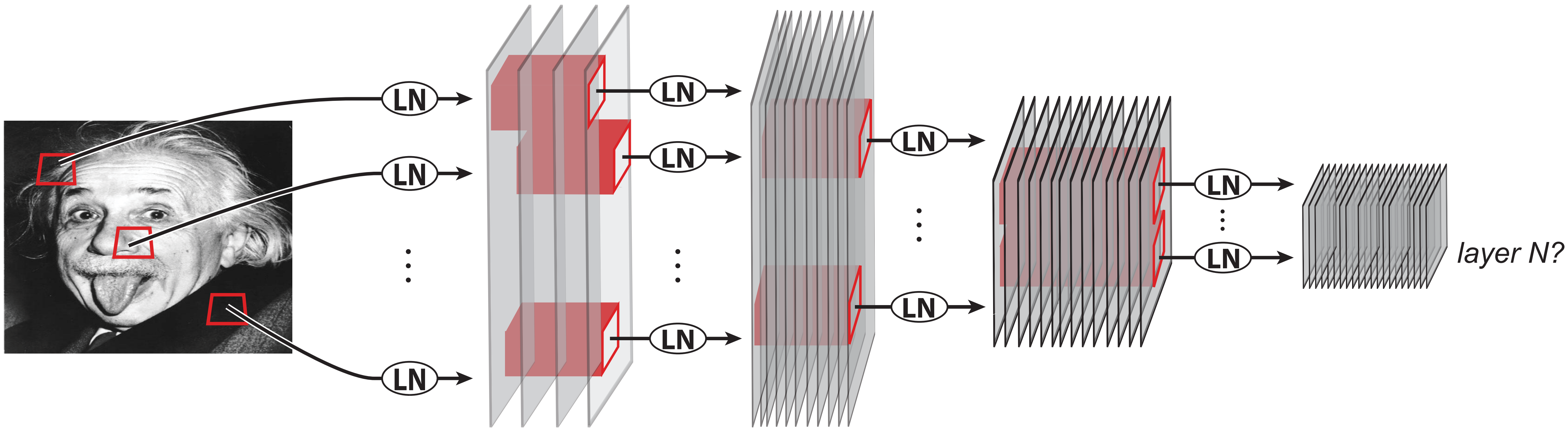
Hierarchical Convolutional Neural Networks

- Individual layers of neurally-plausible **basic operations**
- Applied **convolutionally** — same at all locations
- Stacked **hierarchically** to produce more complex operations



Hierarchical Convolutional Neural Networks

Tensor dimensionality:



$$(s_0, s_0, c_0) \longmapsto (s_1, s_1, c_1)$$

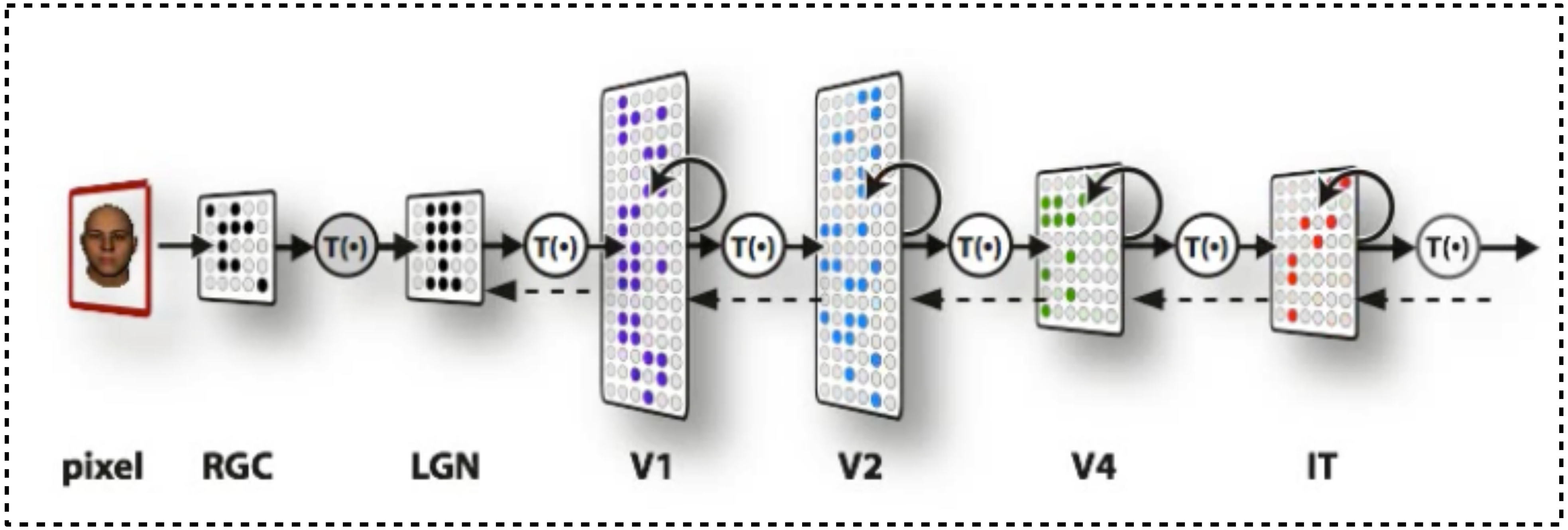
$$(k_x, k_y, c_0, c_1)$$ filter is 4-tensor

s0 = input dimension size; s1 = output dimension

c0 = number of input channels; c1 = number of output channels

kx, ky = kernel size

Hierarchical Convolutional Neural Networks



→ Convolutional Neural Networks (CNNs) Fukushima, 1980; Lecun, 1995

CNNs condense rough neuroanatomy of the ventral stream b:

1) being **hierarchical**

2) being **retinotopic** (spatially tiled)

Hierarchical Convolutional Neural Networks

Fukushima, 1978

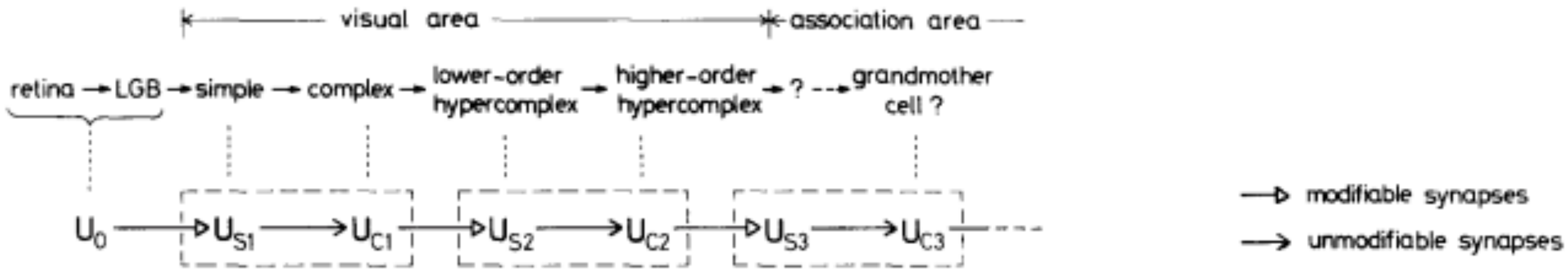


Fig. 1. Correspondence between the hierarchy model by Hubel and Wiesel, and the neural network of the neocognitron

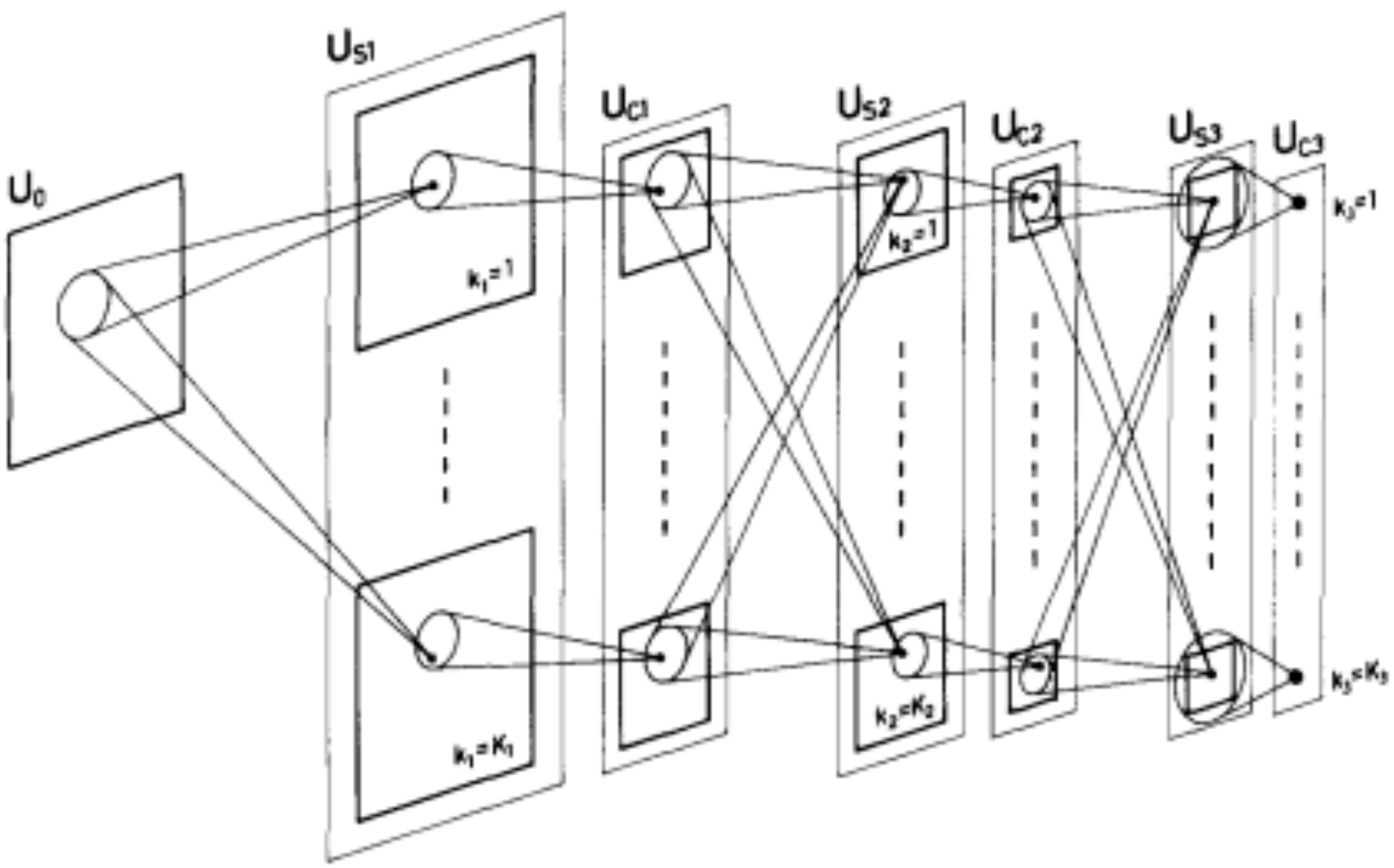


Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

Hierarchical Convolutional Neural Networks

Kunihiko Fukushima!

Tokyo, November 2015

Hierarchical Convolutional Neural Networks

Tokyo, November 2015

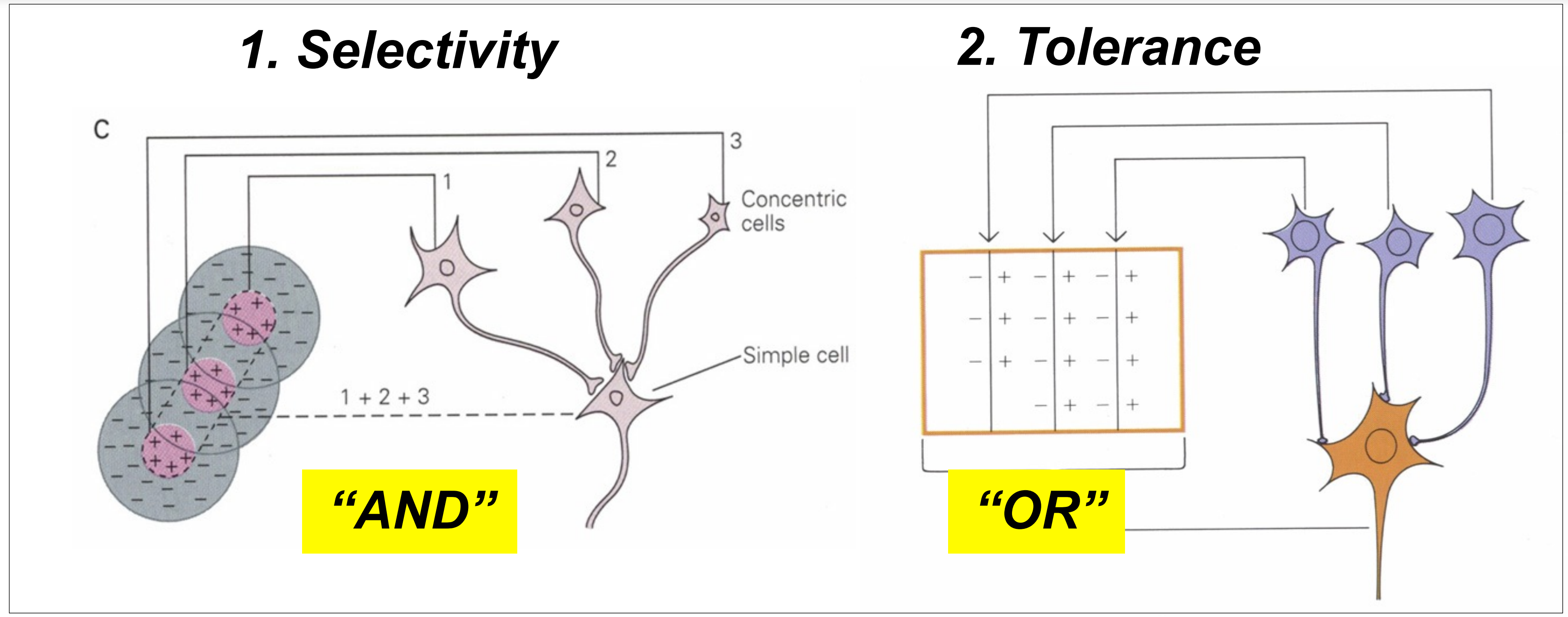
Kunihiko Fukushima!

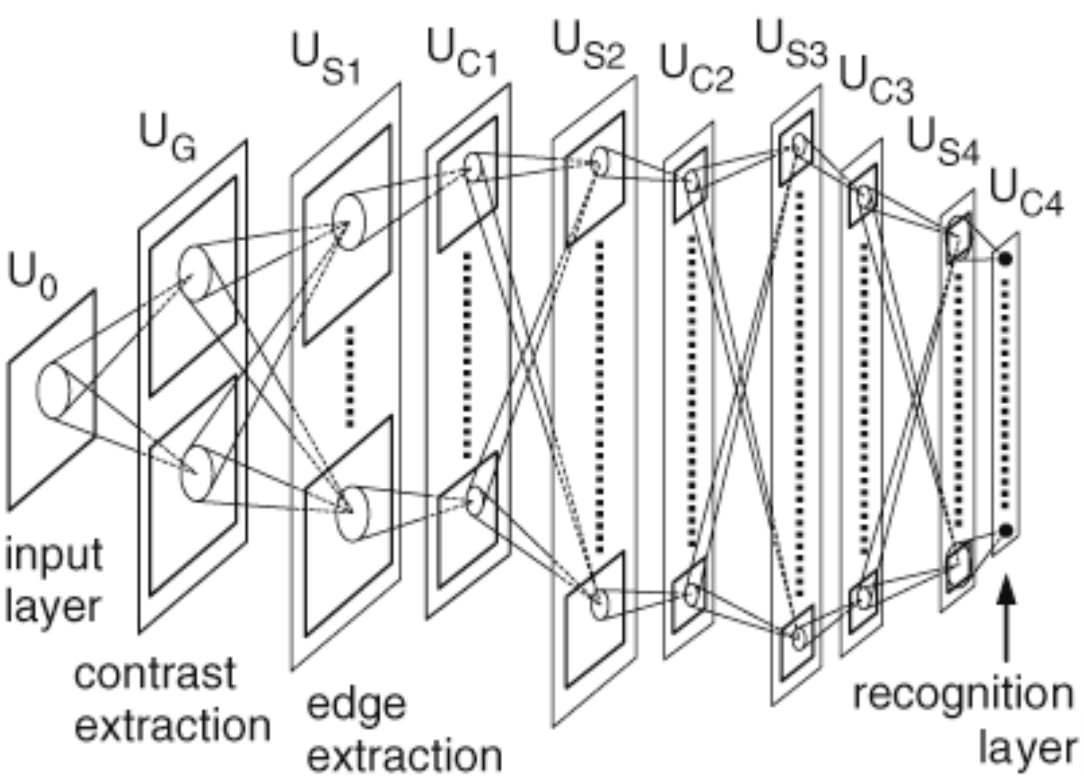
Developed first convnet
in the late 1970s
while Japan Broadcasting
Corporation (NHK)
… office directly next
door to Keiji Tanaka's

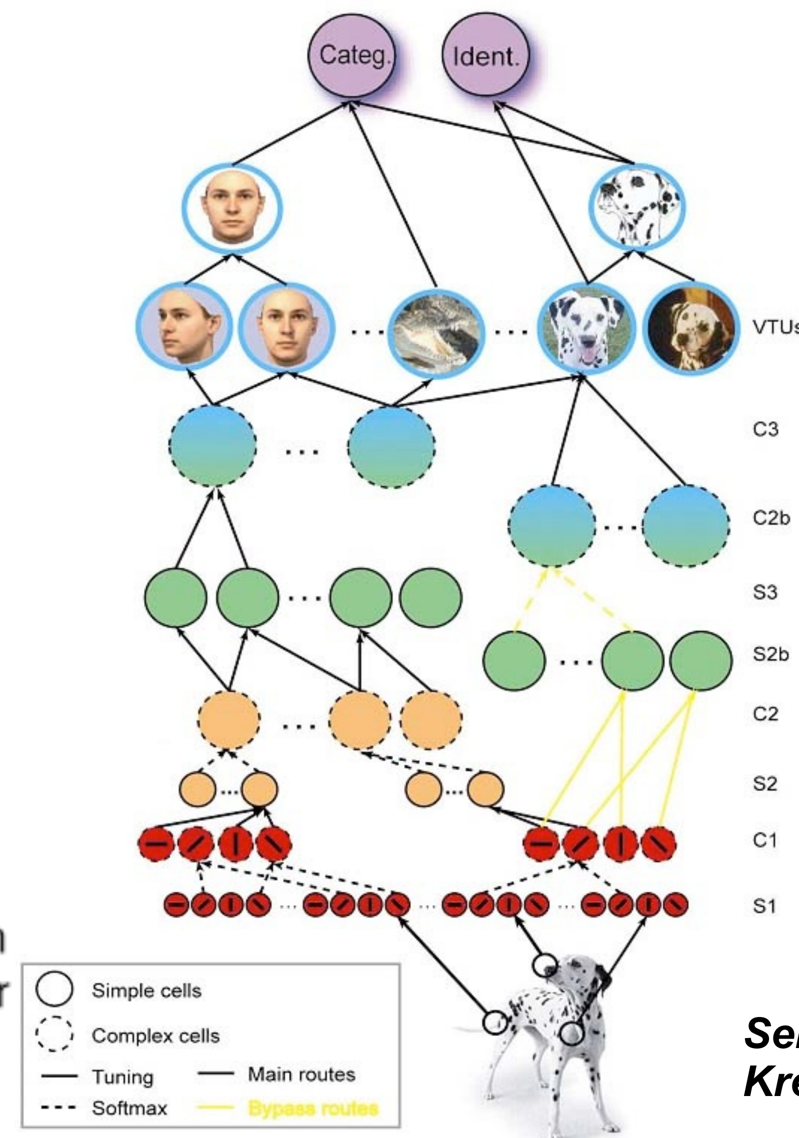
Various attempts at models ….

Examples:

- *Hubel & Wiesel (1962)*
- *Fukushima (1980)*
- *Perrett & Oram (1993)*
- *Olshausen & Field (1996)*
- *Wallis & Rolls (1997)*
- *LeCun et al. (1998)*
- *Riesenhuber & Poggio (1999)*
- *Serre, Kouh, et al. (2005)*



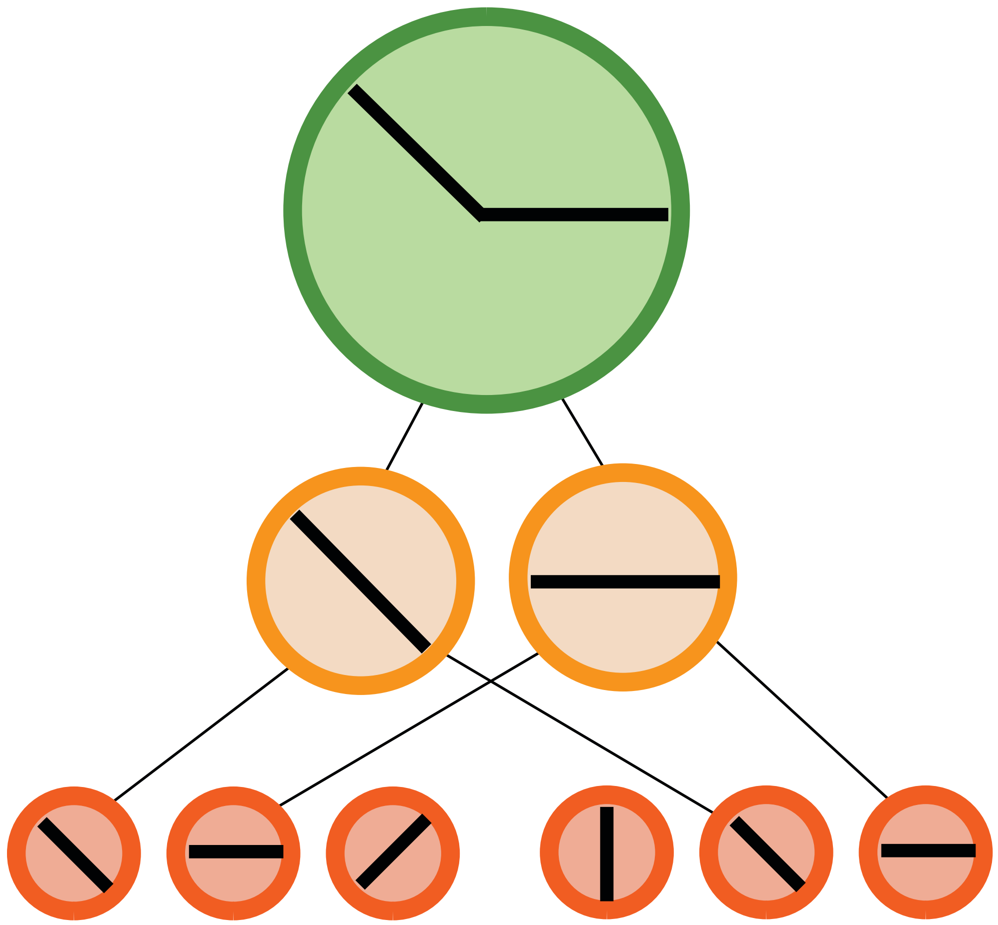




Serre, Kouh, Cadieu, Knoblich, Kreiman & Poggio 2005

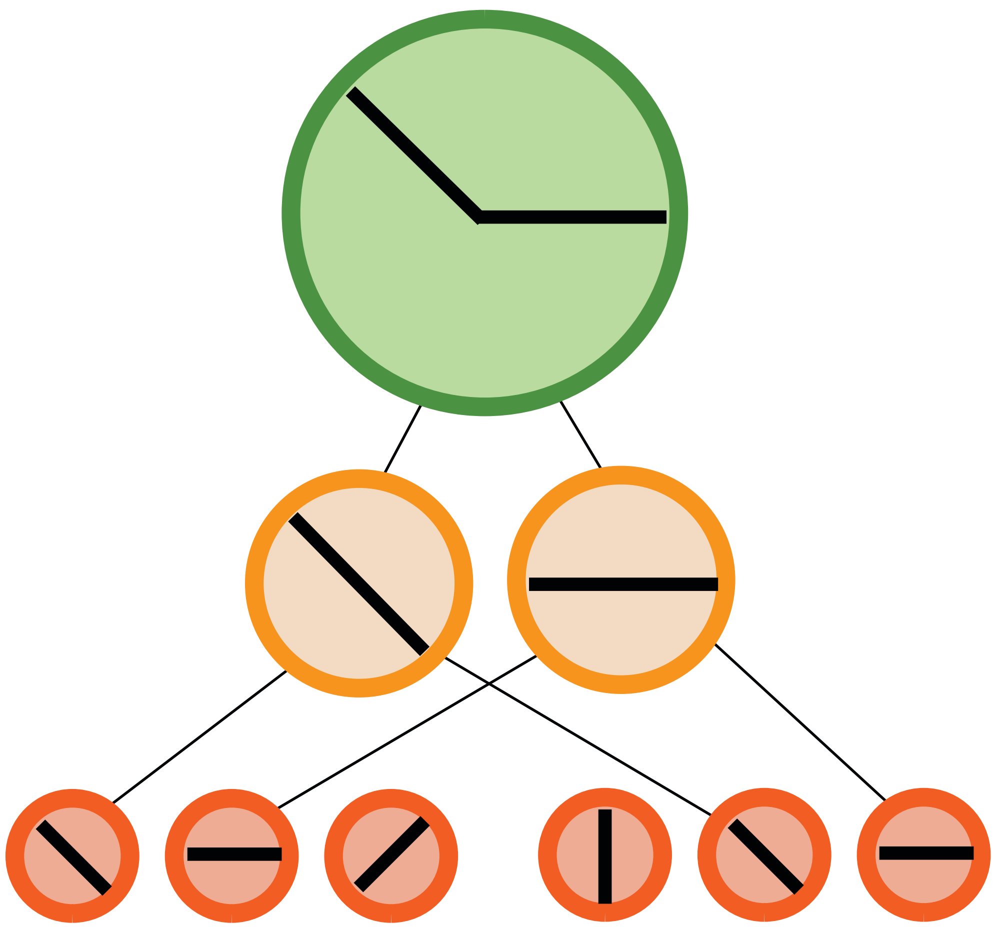
"Intuitive idea" of hierarchical processing

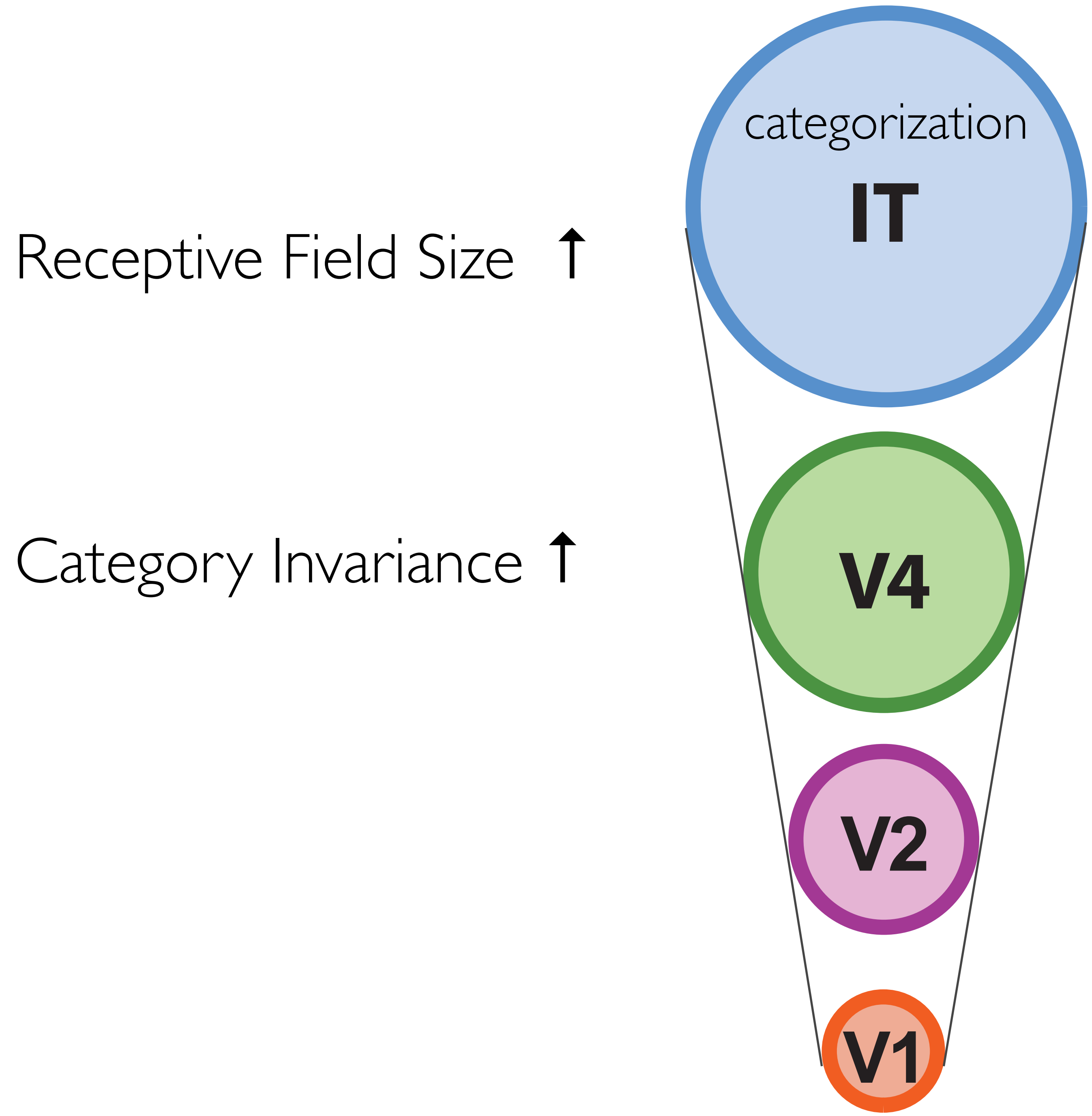
Aggregation over identity-preserving transformations, e.g. translation.



Beyond categorization

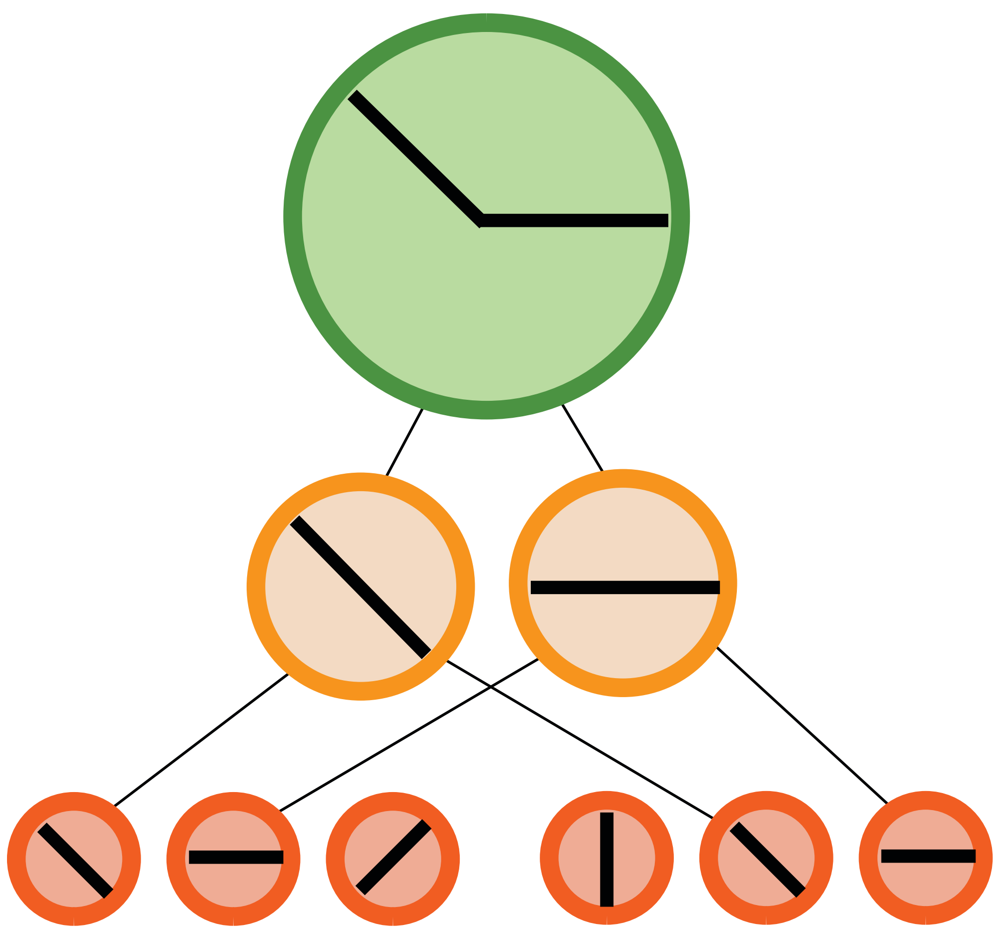
Aggregation over identity-preserving transformations, e.g. translation.





Beyond categorization

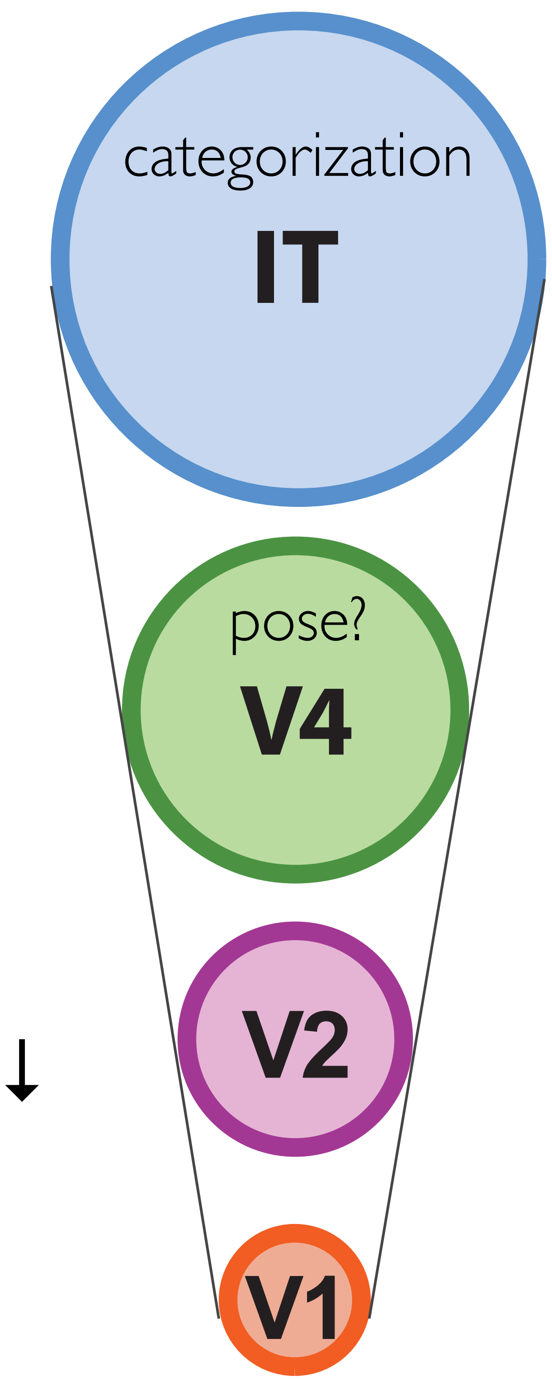
Aggregation over identity-preserving transformations, e.g. translation.



Receptive Field Size ↑

Category Invariance ↑

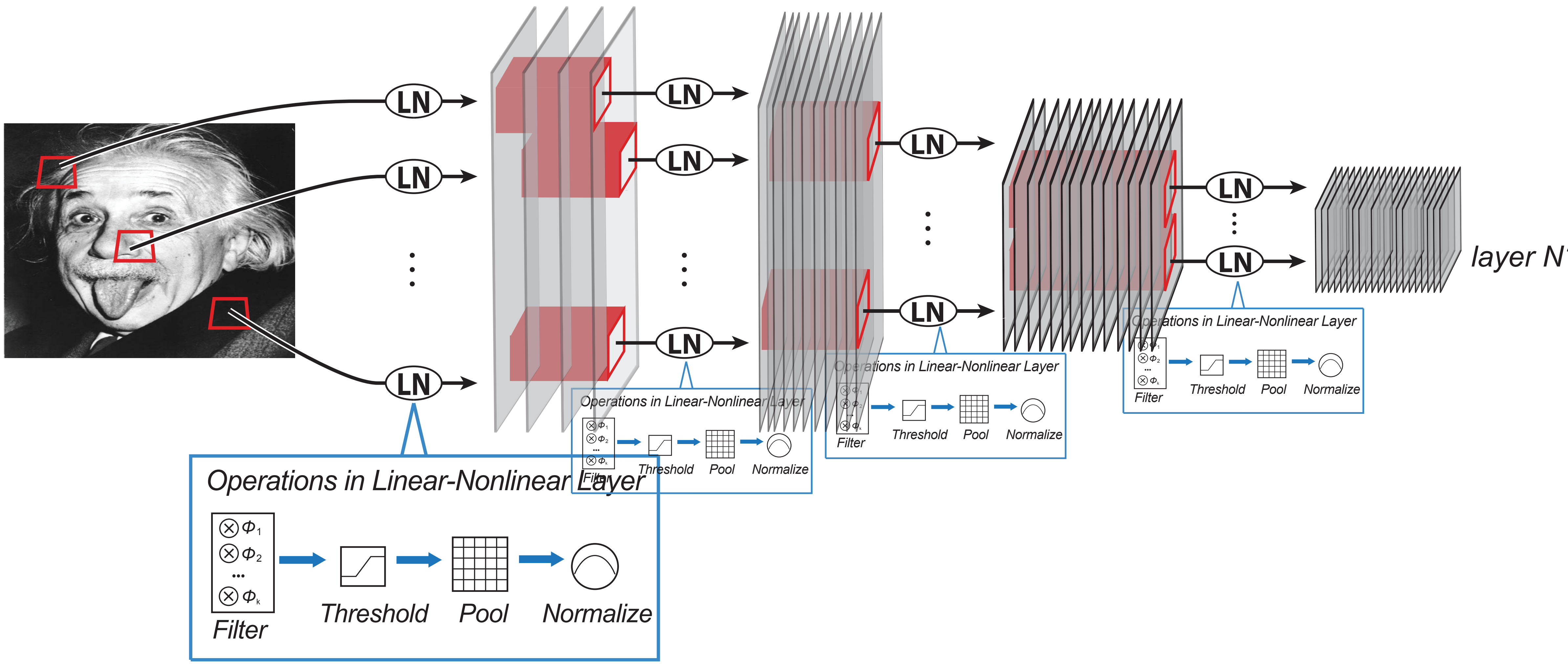
(e.g.) Position Sensitivity ↓



position / size estimation

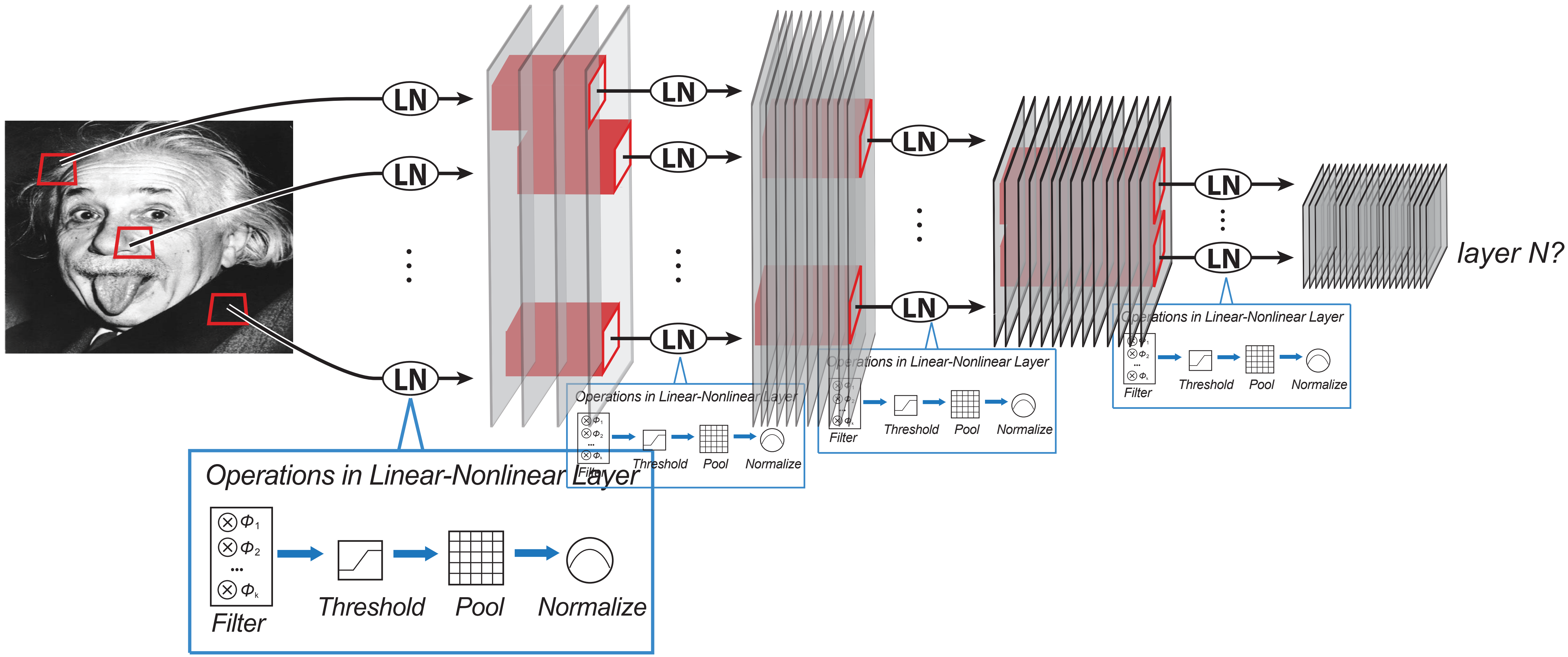
Hierarchical Convolutional Neural Networks

Huge number of parameters consistent with HCNN concept.



Hierarchical Convolutional Neural Networks

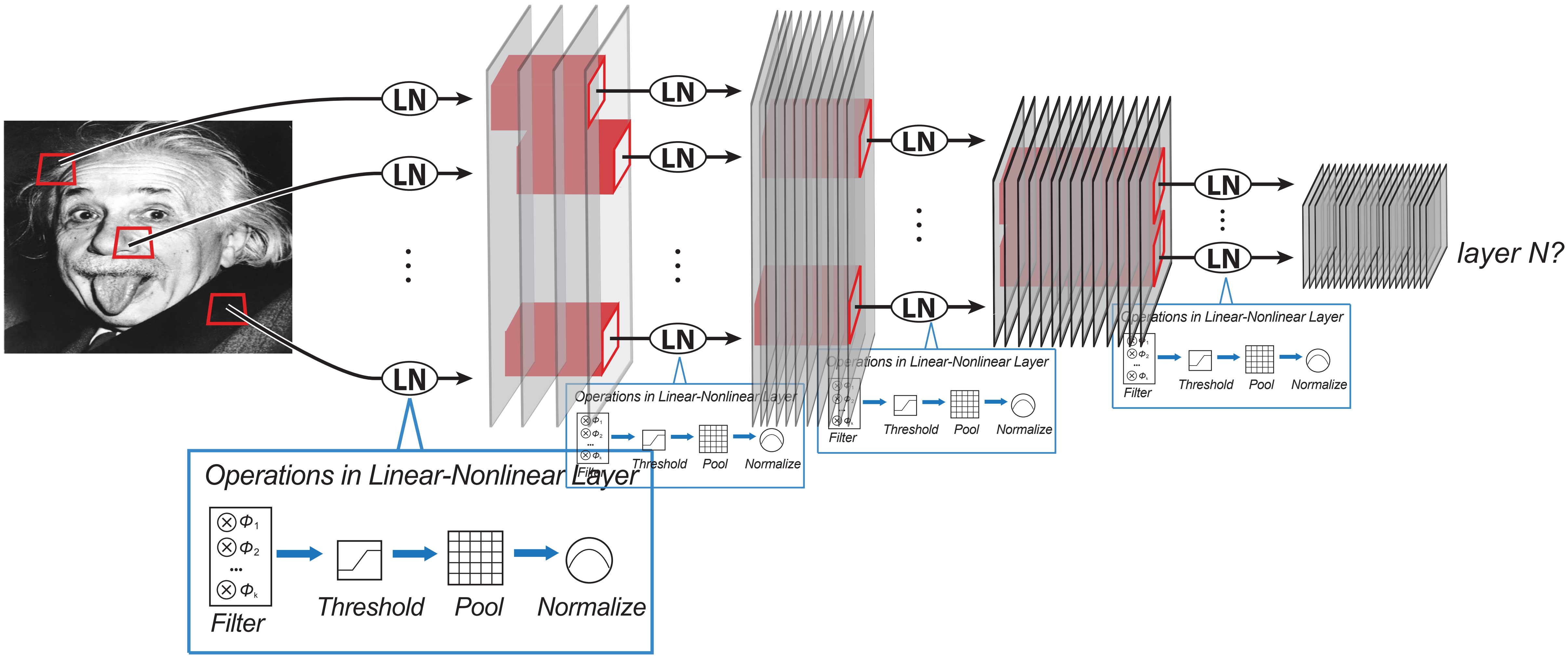
Huge number of parameters consistent with HCNN concept.



i. **architectural** params: (# layers, # filters, receptive field sizes, &c) — "network structure"

Hierarchical Convolutional Neural Networks

Huge number of parameters consistent with HCNN concept.

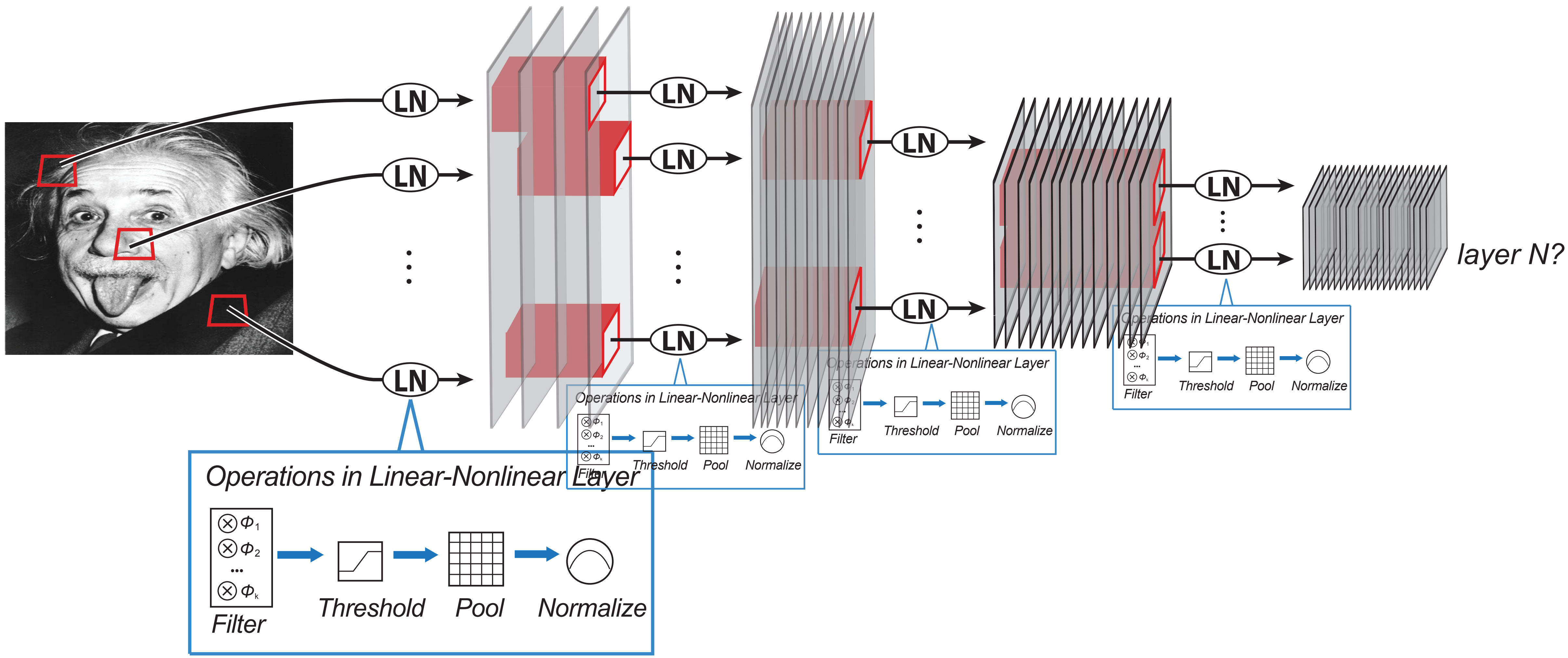


i. **architectural** params: (# layers, # filters, receptive field sizes, &c) — "network structure"

ii. **filter** parameters: continuous valued pattern templates — "network contents"

Hierarchical Convolutional Neural Networks

Huge number of parameters consistent with HCNN concept.



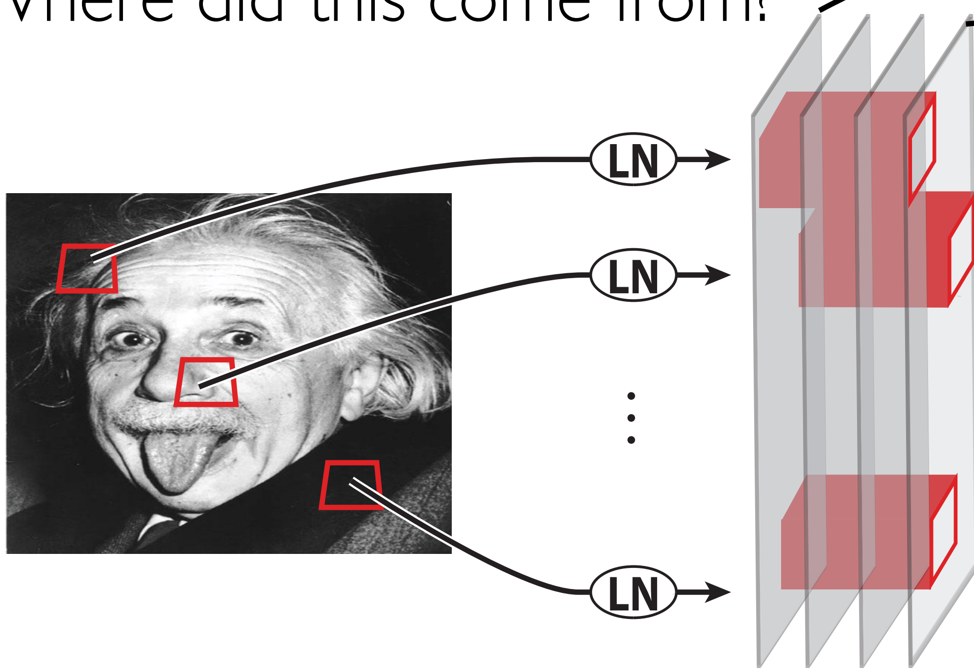
i. **architectural** params: (# layers, # filters, receptive field sizes, &c) — "network structure"

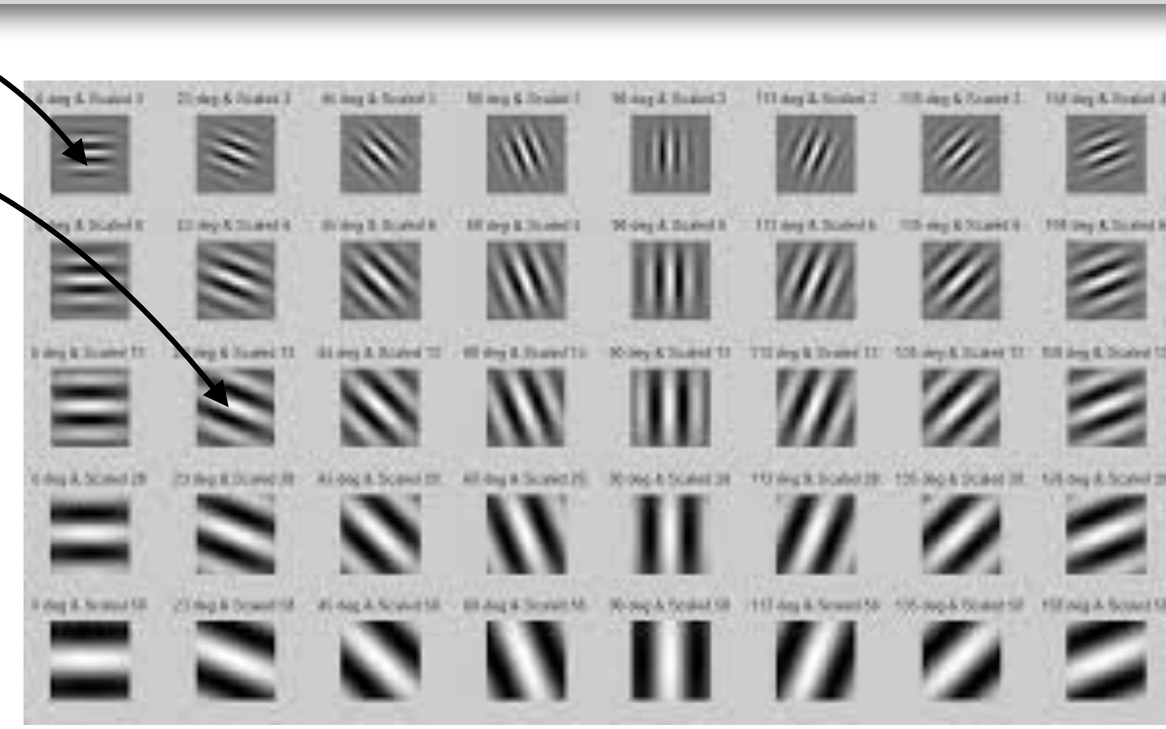
ii. **filter** parameters: continuous valued pattern templates — "network contents"

*Q: How to discover the **"right"** parameters to understand real cortex?*

Recall from V1

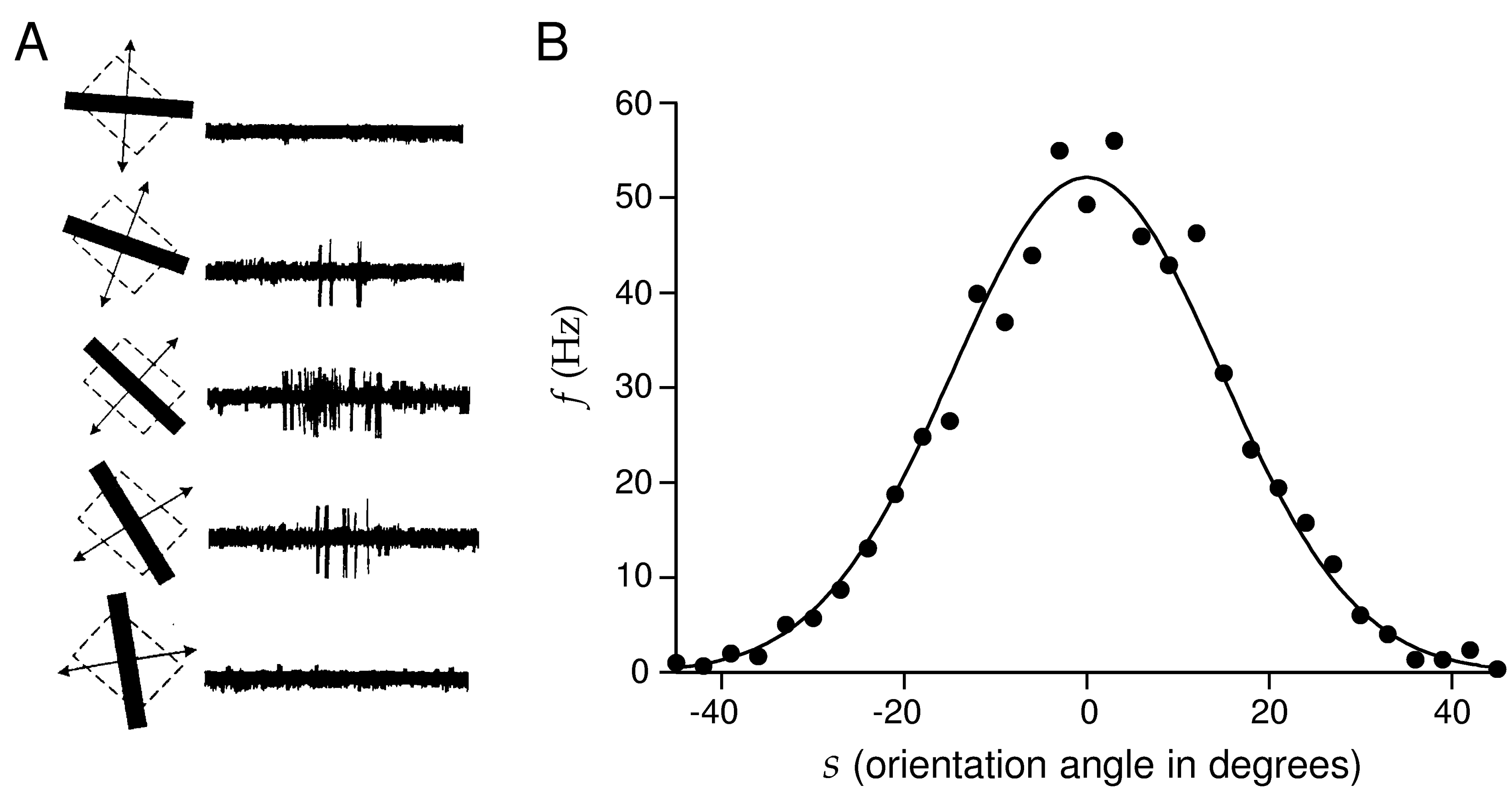
Where did this come from?





"Hubel and Wiesel's Intuition"
~1970s and formalized later
via Gabor wavelets

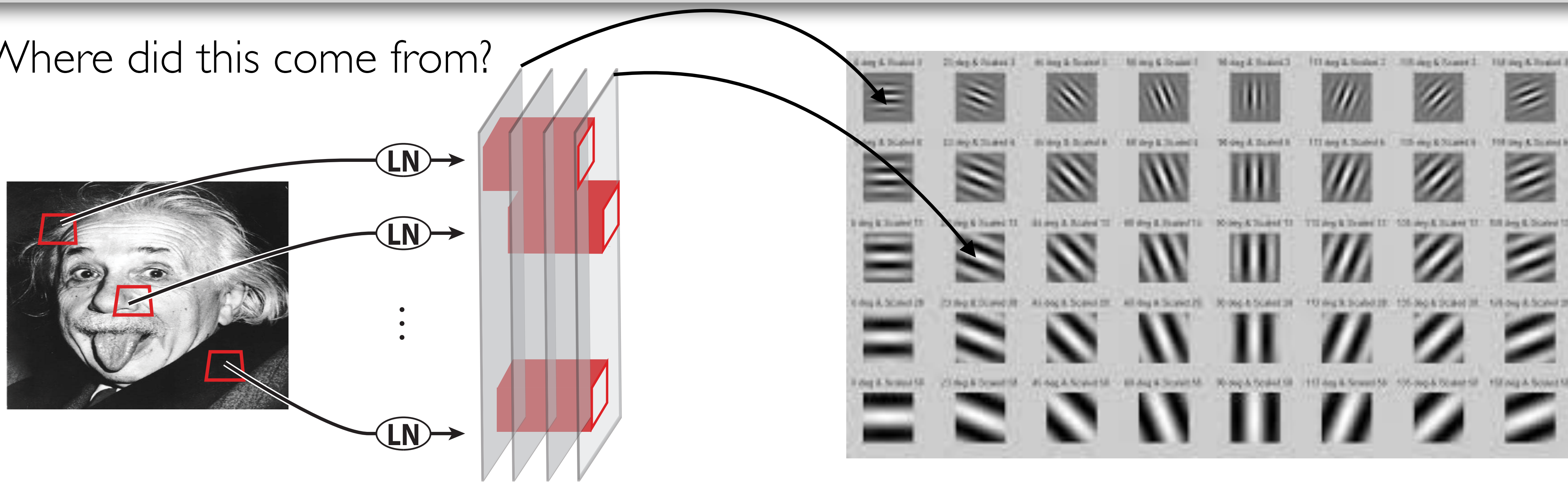
Gaussian tuning curve of V1 simple cell



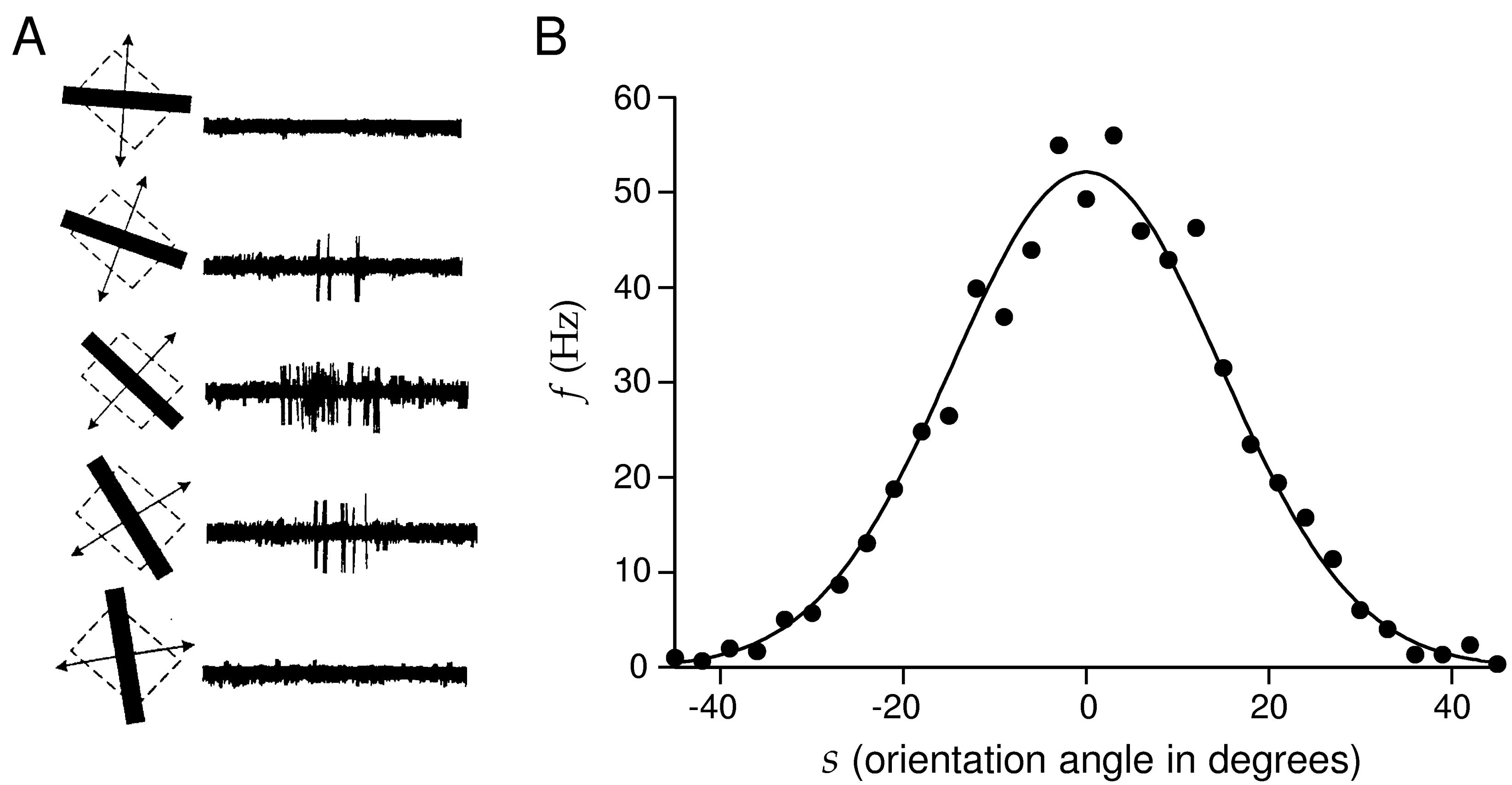
adapted from Adrienne Fairhall

Recall from V1

Where did this come from?



Gaussian tuning curve of V1 simple cell



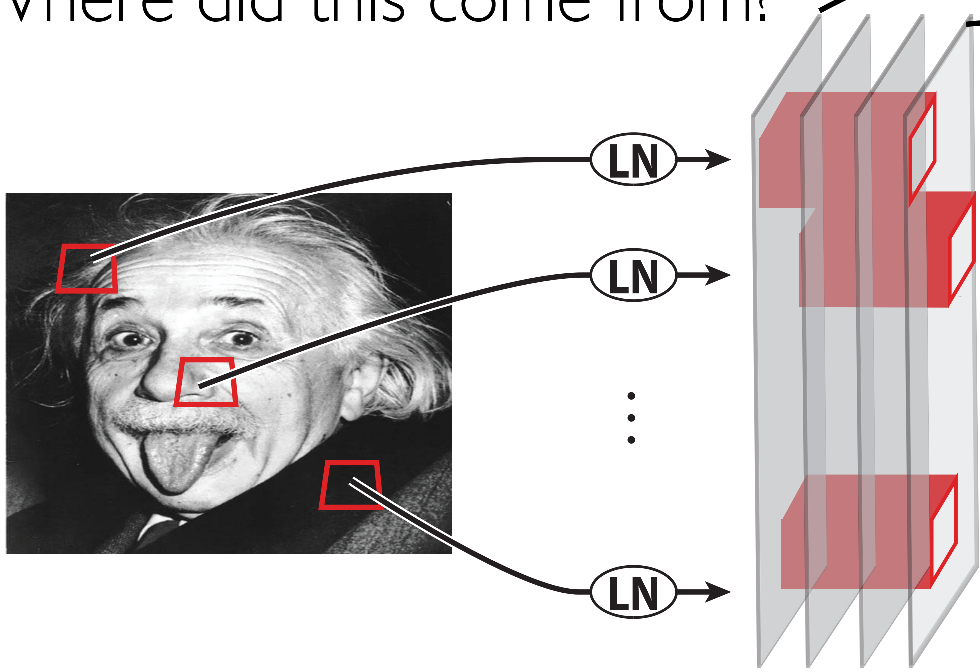
"Hubel and Wiesel's Intuition"
~1970s and formalized later via Gabor wavelets

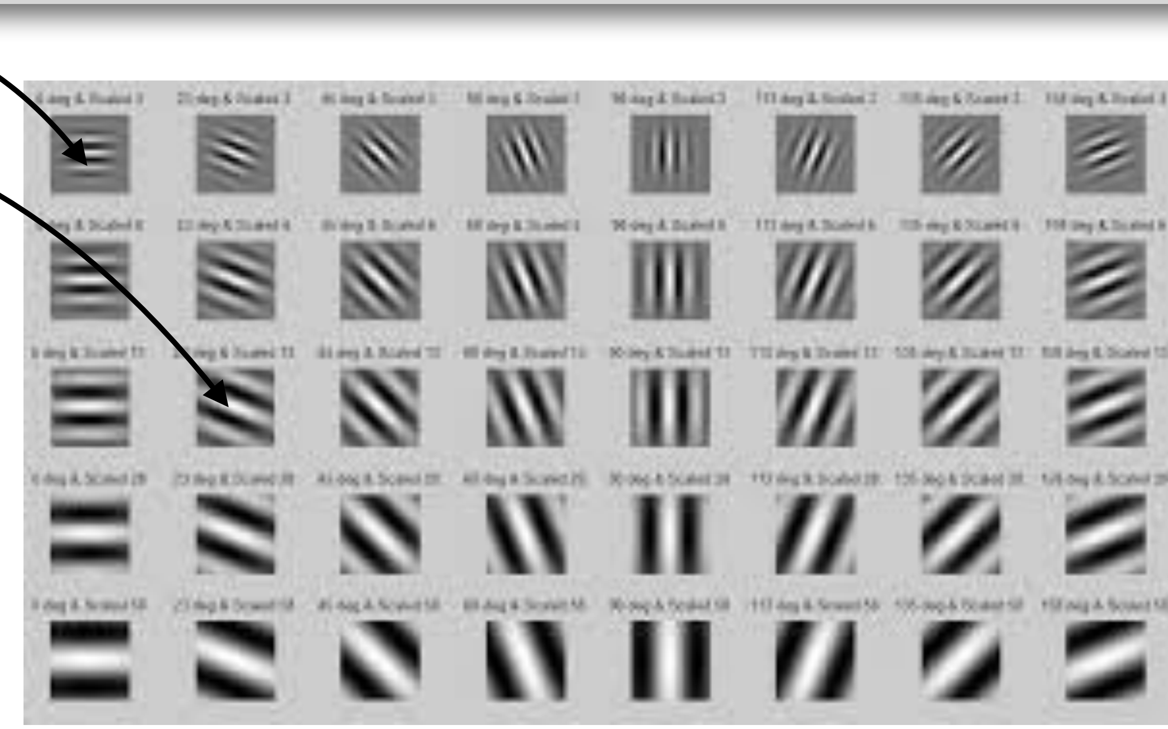
→ can just "see" what the right axes for measuring good tuning curves are, if we're smart enough

adapted from Adrienne Fairhall

Recall from V1

Where did this come from?

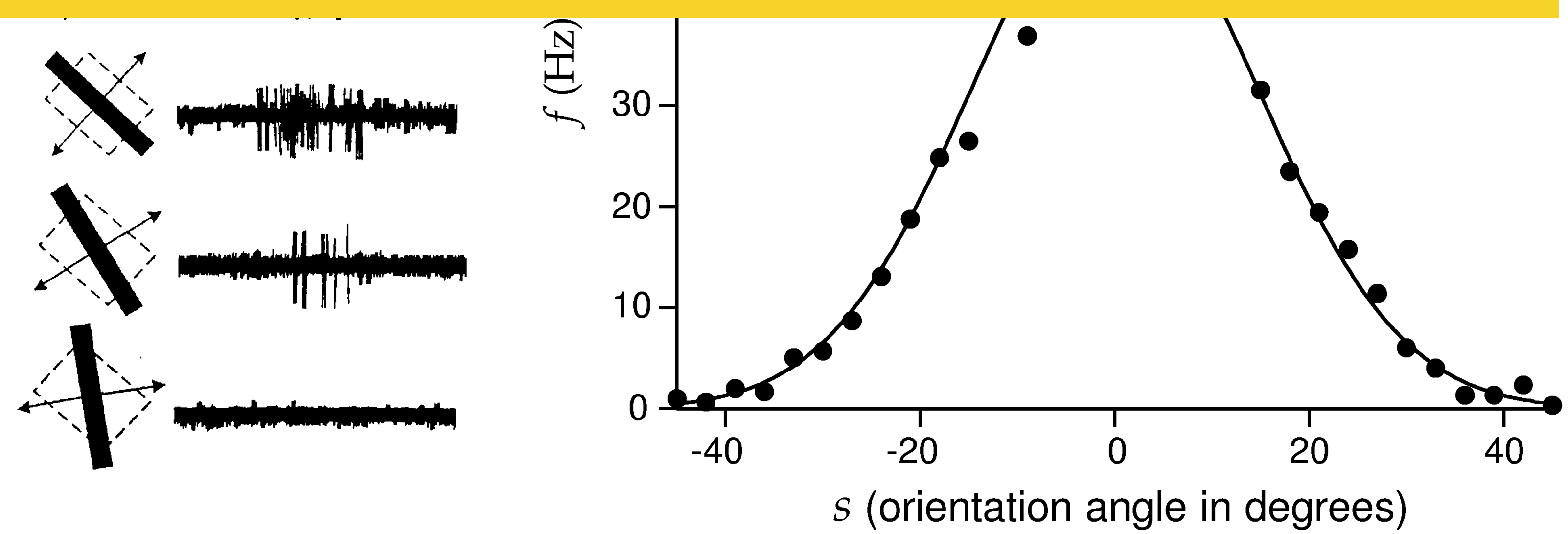




Gaussian tuning curve of V1 simple cell

REALLY HARD TO GENERALIZE TO MULTI-LAYER NETWORKS

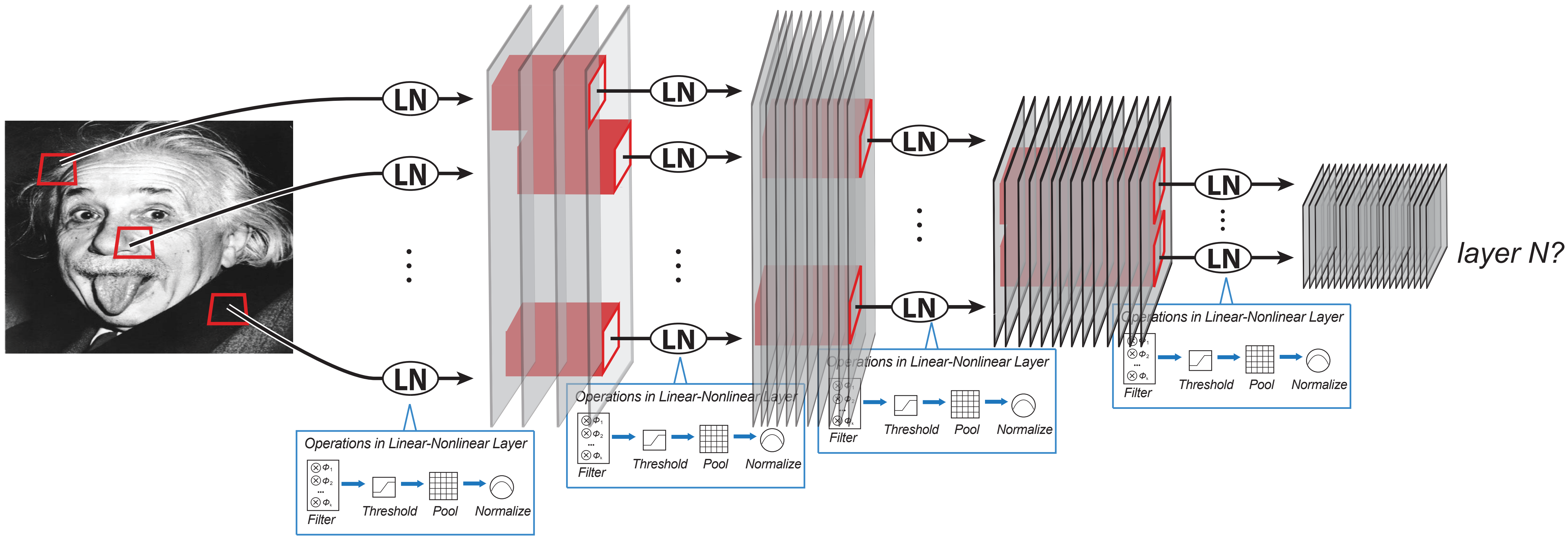
→ can just "see" what the right axes for measuring good tuning curves are, if we're smart enough

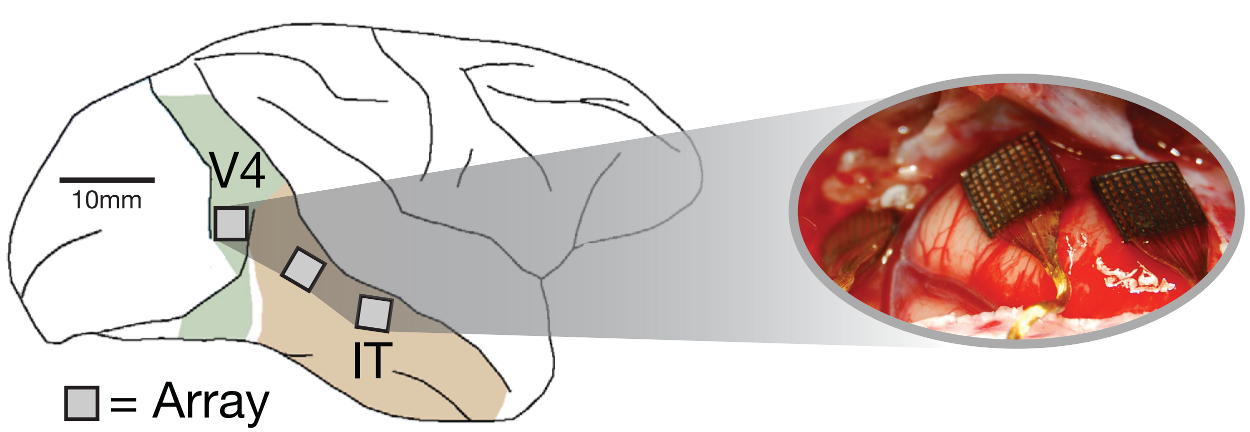


adapted from Adrienne Fairhall

Neural Fitting Strategy?

Huge number of parameters consistent with HCNN concept.

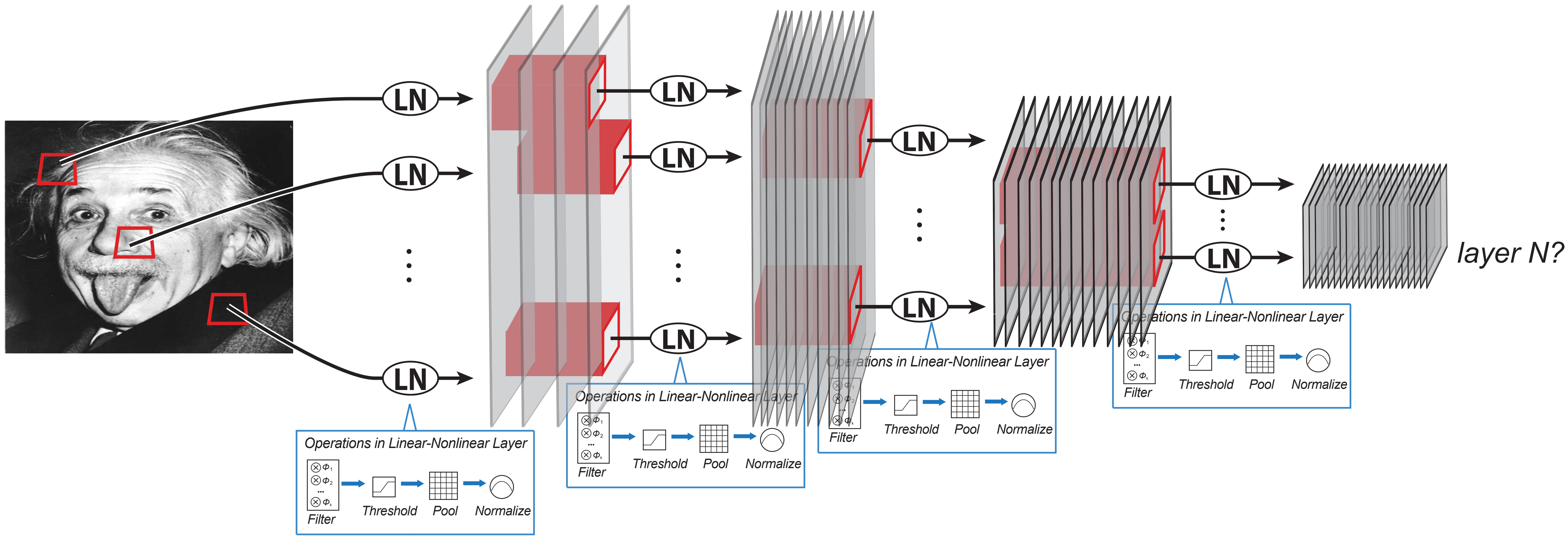


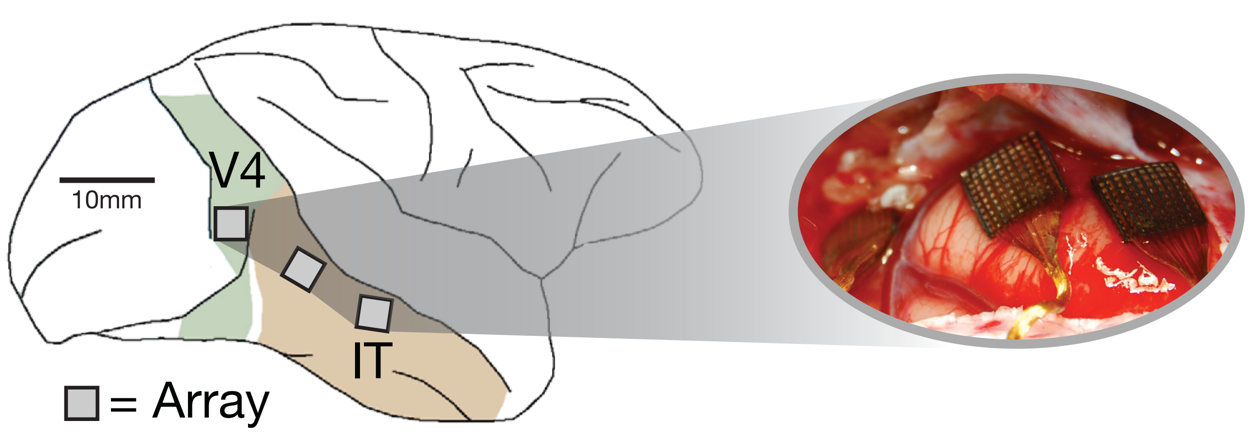


Obvious alternative strategy: fit parameters to neural data.

Neural Fitting Strategy?

Huge number of parameters consistent with HCNN concept.

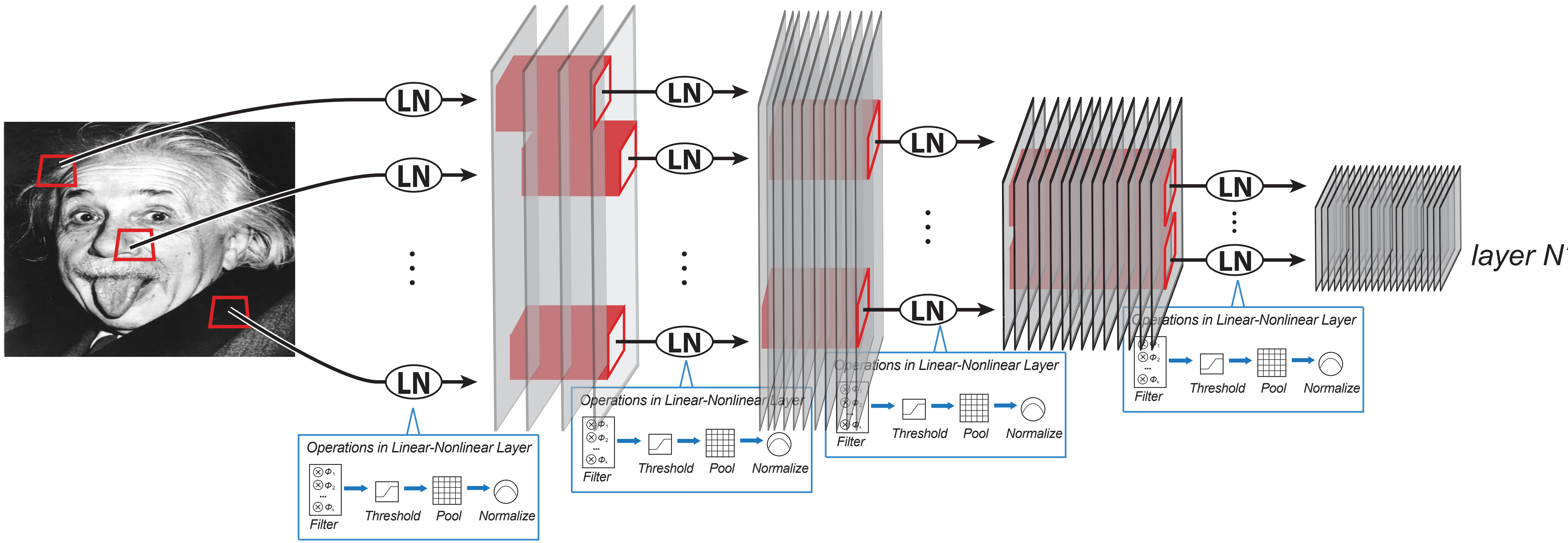


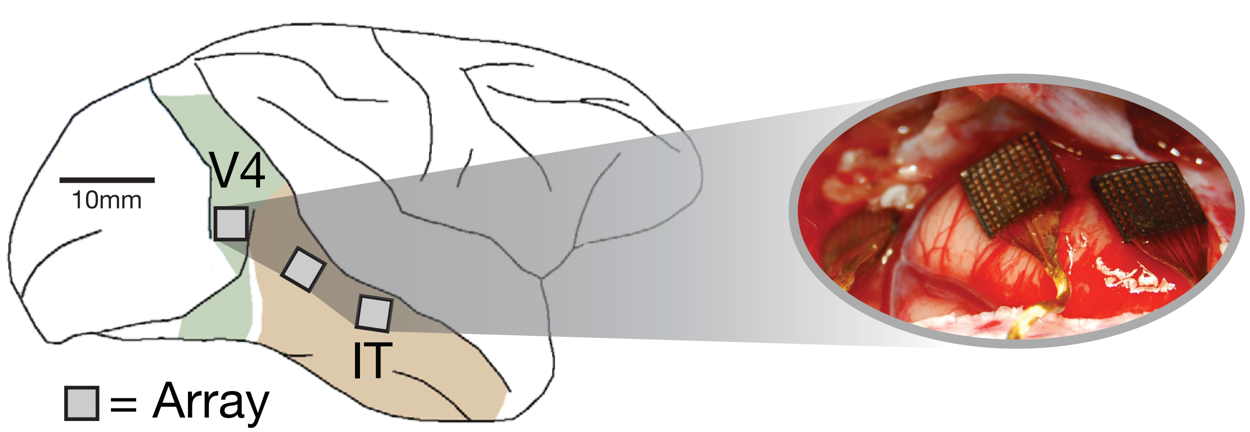


…not enough neural data to constrain model class. Gallant (2007); Rust & Movshon (2006)

Neural Fitting Strategy?

Huge number of parameters consistent with HCNN concept.



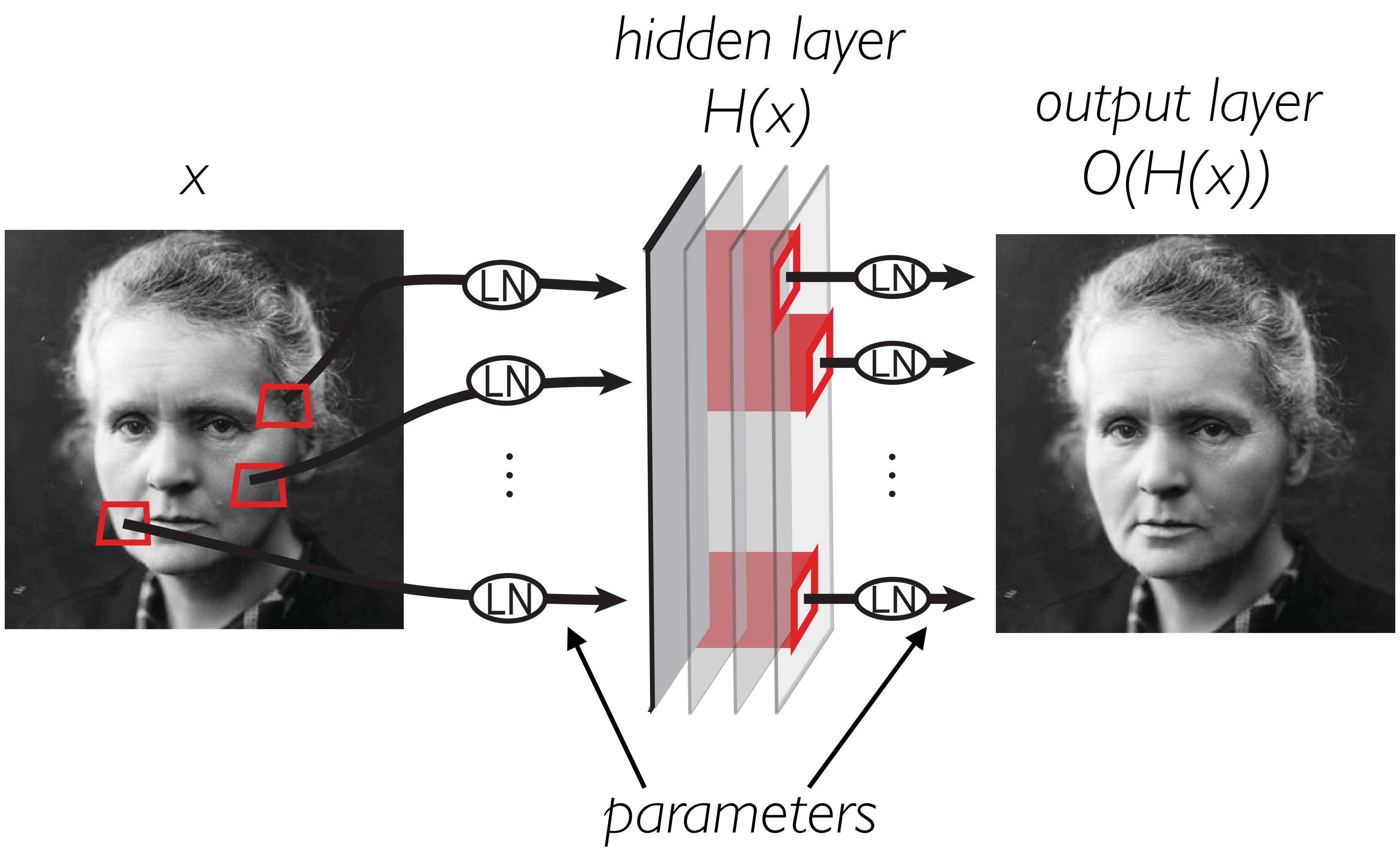


…not enough neural data to constrain model class. Gallant (2007); Rust & Movshon (2006)

↓

Overfitting.

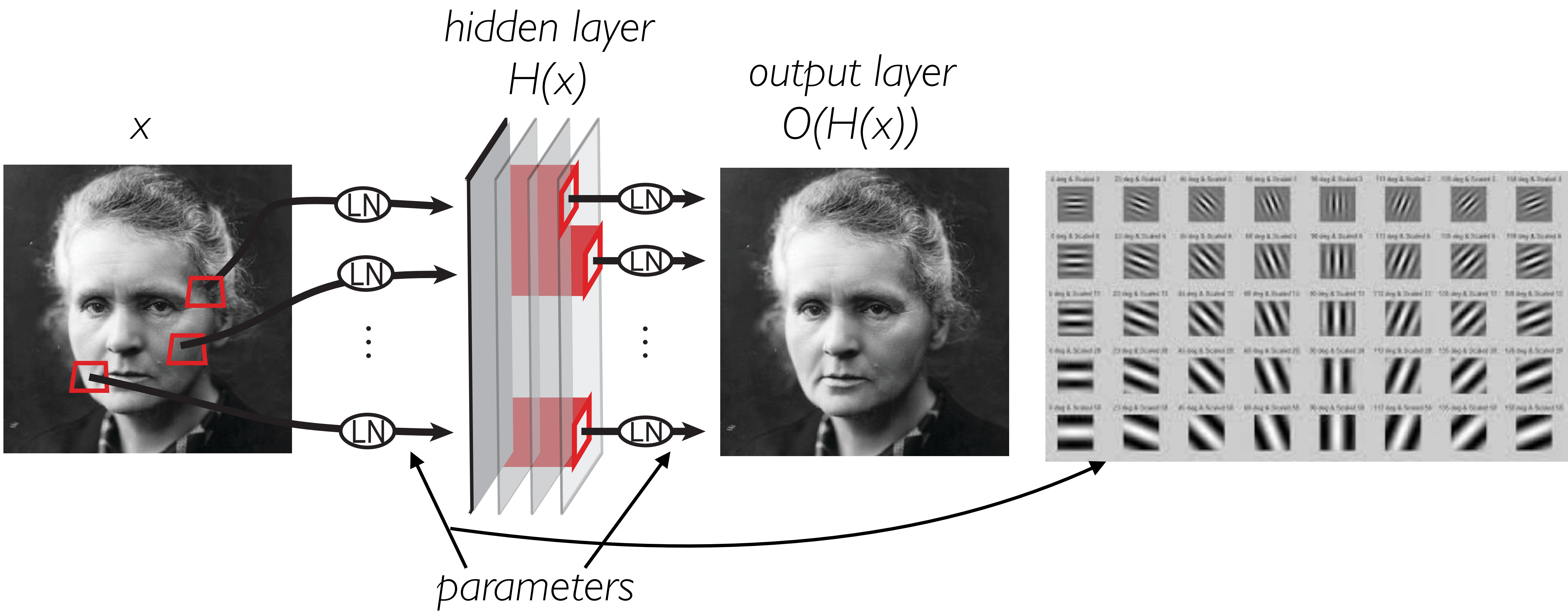
V1 as Sparse Autoencoder



$$L(x) = |x - O(H(x))|^2 + \lambda \cdot |H(x)|$$

Sparse Coding Foldiak, Olshausen,
mid 1990s

V1 as Sparse Autoencoder

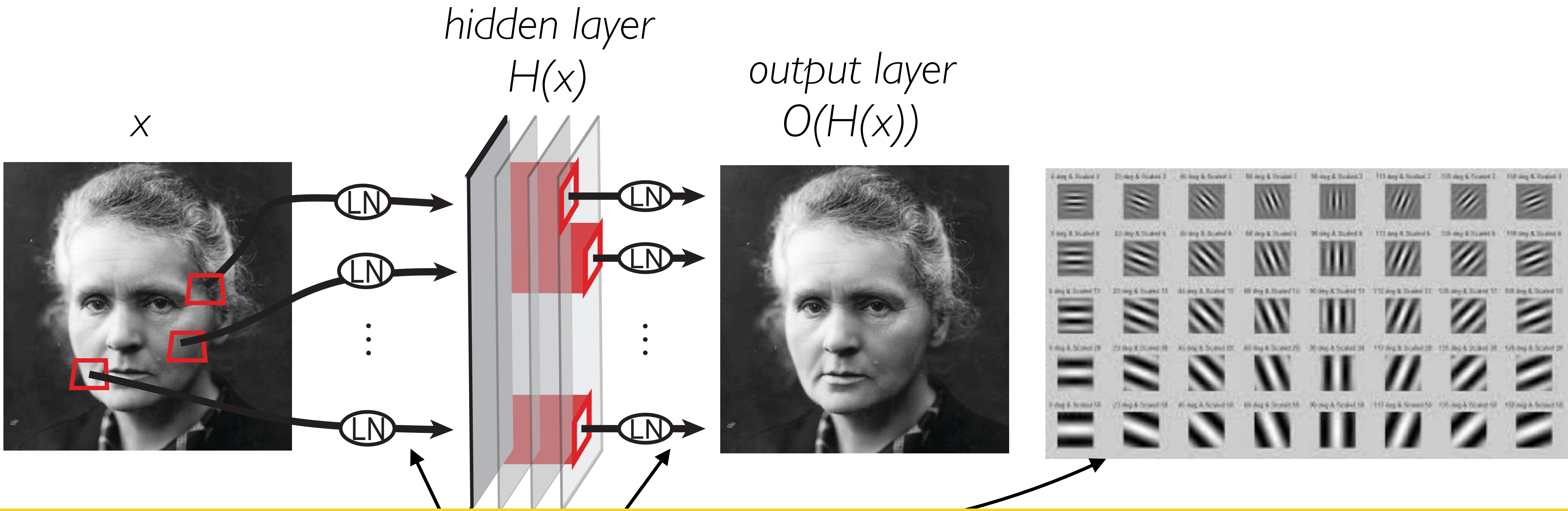


$$L(x) = |x - O(H(x))|^2 + \lambda \cdot |H(x)|$$

Sparse Coding Foldiak, Olshausen,
mid 1990s

→neurons have to represent their environment, as efficiently as possible

V1 as Sparse Autoencoder

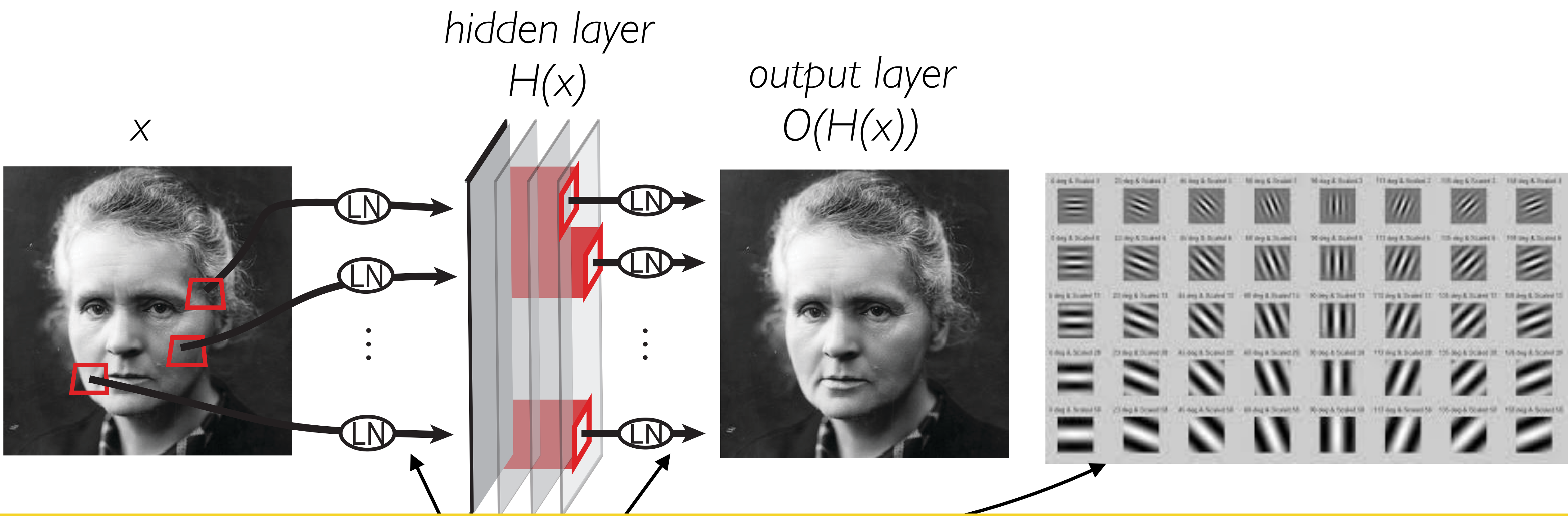


**Also turns out not to generalize to multi-layer networks very well . . .
at least not directly.**

Sparse Coding Foldiak, Olshausen, mid 1990s

→neurons have to represent their environment, as efficiently as possible

V1 as Sparse Autoencoder



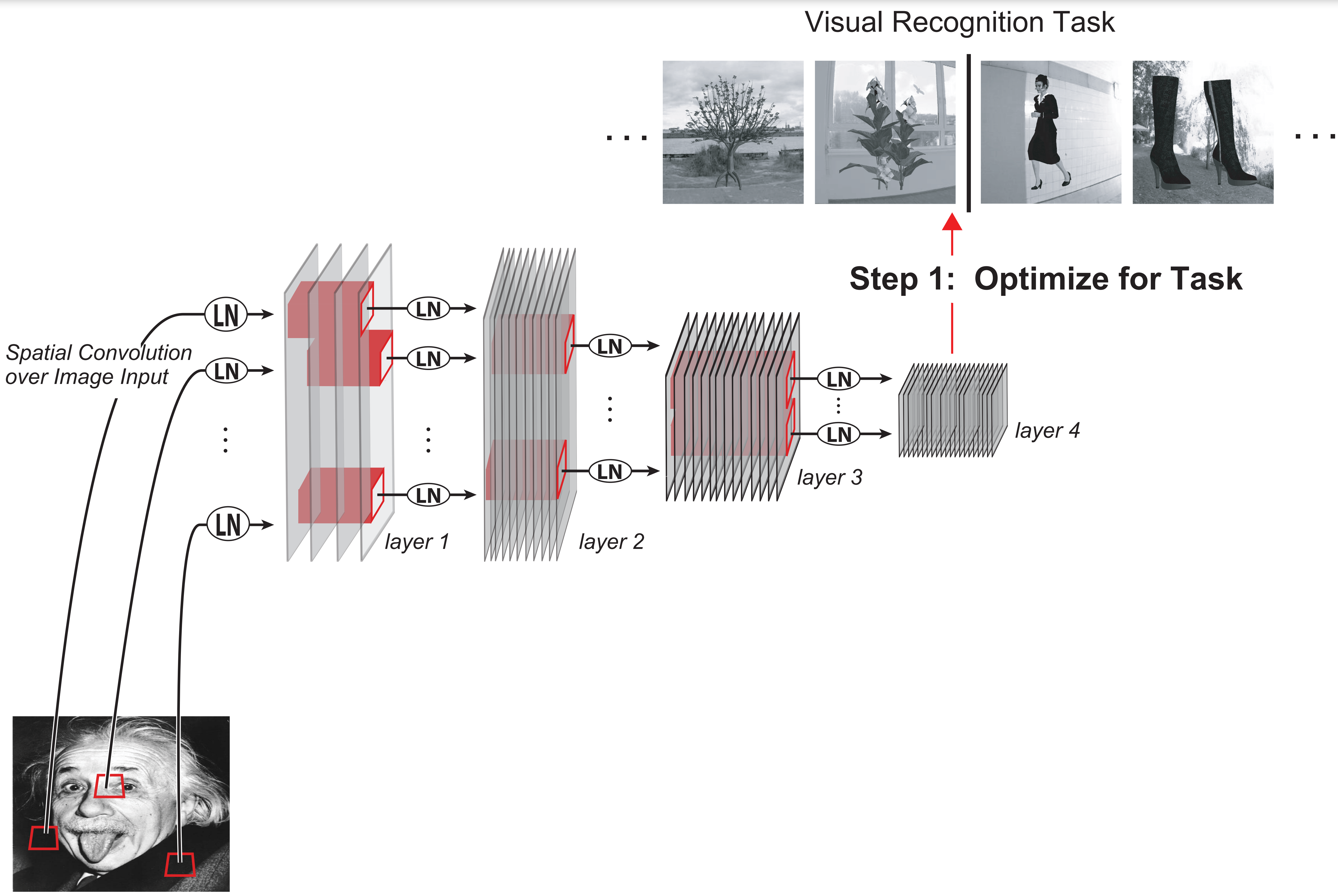
**Also turns out not to generalize to multi-layer networks very well . . .
at least not directly.**

but we will return to this point when we study self-supervised learning

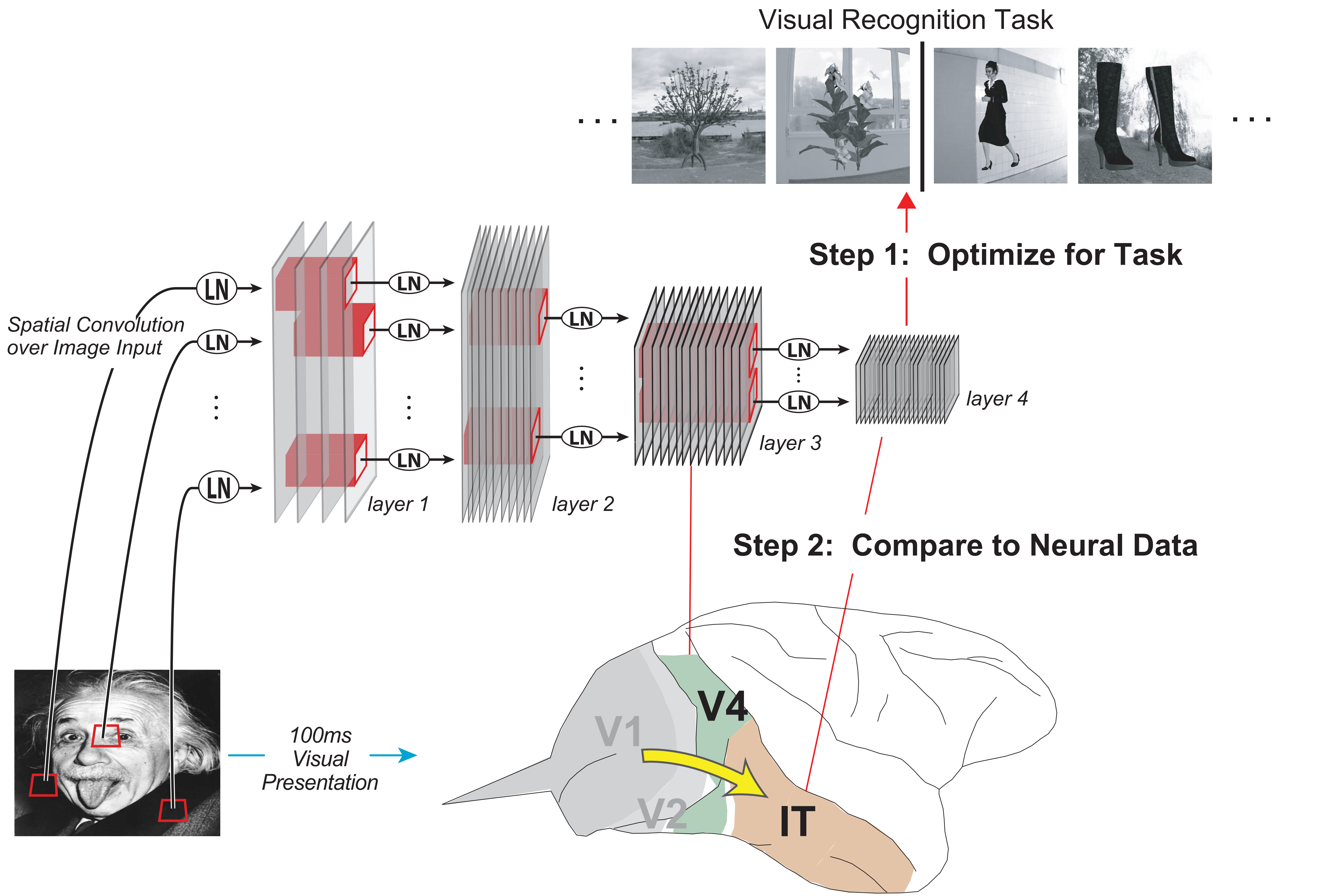
Sparse Coding Foldiak, Olshausen, mid 1990s

→neurons have to represent their environment, as efficiently as possible

Optimize for Performance, Test Against Neurons



Optimize for Performance, Test Against Neurons



Optimize for Performance, Test Against Neurons

1. **Performance**: accuracy on a challenging, high-variation visual object categorization task.

2. **Neural predictivity**: the ability of model to predict each individual neural site's activity.

Optimize for Performance, Test Against Neurons

1. **Performance**: accuracy on a challenging, high-variation* visual object categorization task.

2. **Neural predictivity**: the ability of model to predict each individual neural site's activity.

challenging for neural network engineers, not the animal

Optimize for Performance, Test Against Neurons

1. **Performance**: accuracy on a challenging, high-variation* visual object categorization task.

2. **Neural predictivity**: the ability of model to predict each individual neural site's activity.

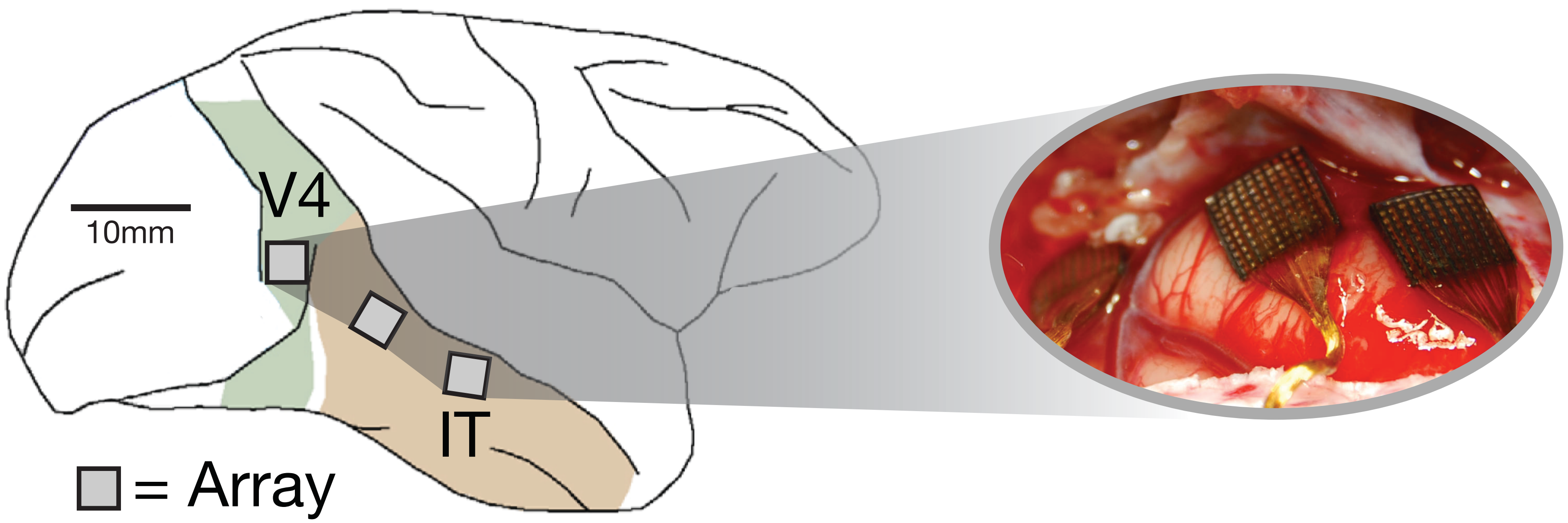
Our hypothesis: Performance (1) → neural predictivity (2).

***challenging for neural network engineers, not the animal**

Multi-array Electrophysiology Experiment

Multi-array electrophysiology in macaque V4 and IT.

(somewhere between single and multi-unit recording)



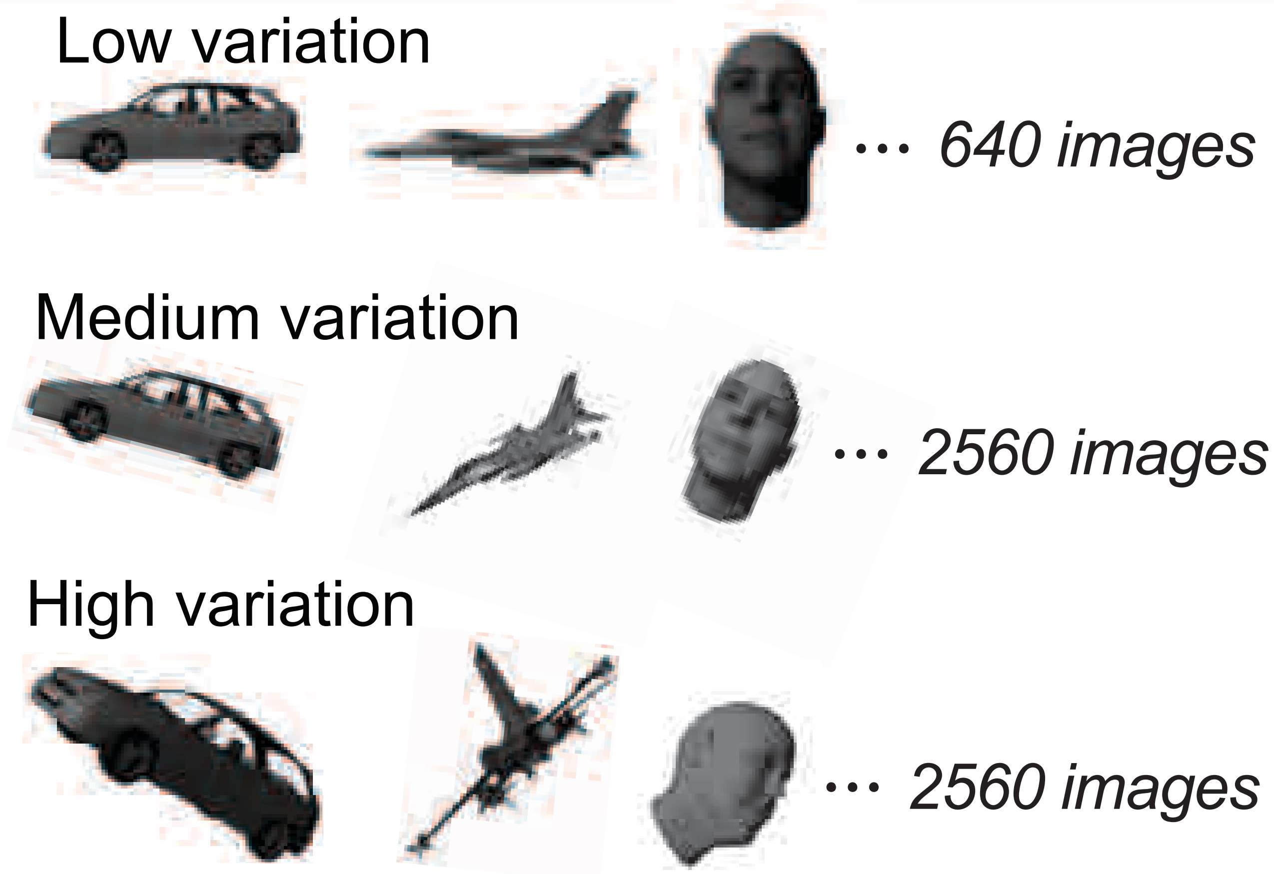
Multi-array Electrophysiology Experiment

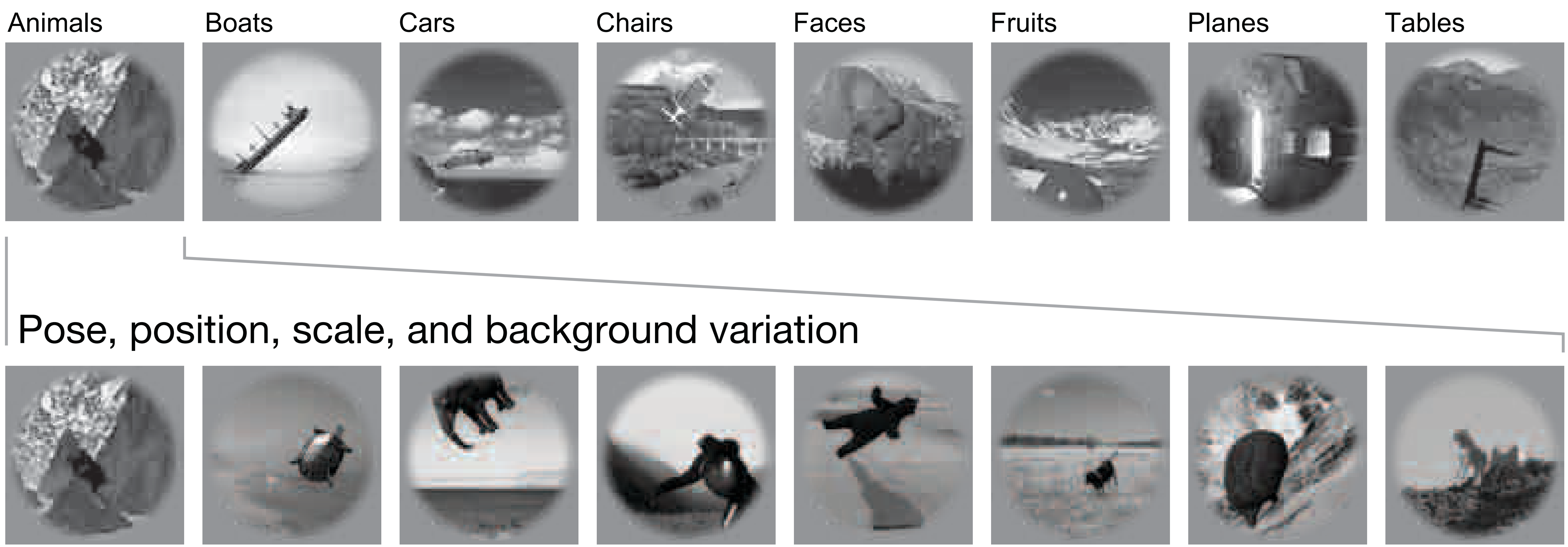
5760 images

64 objects

8 categories

uncorrelated photo backgrounds

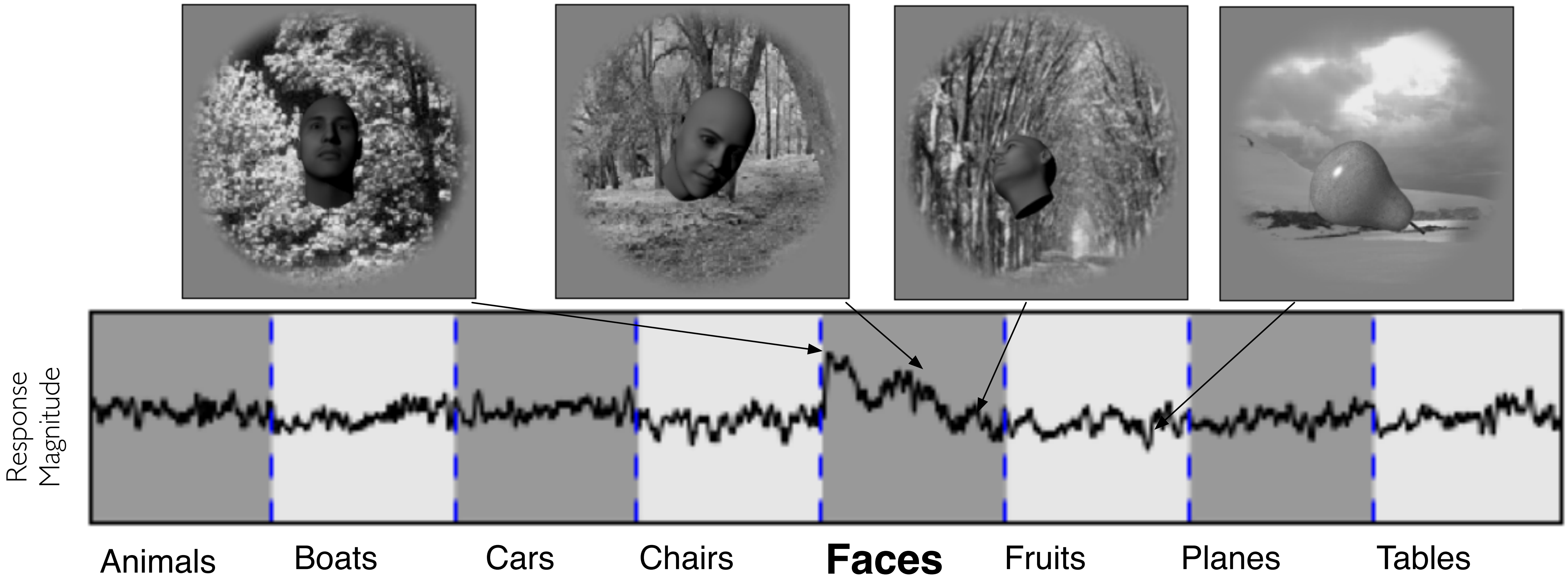




Multi-array Electrophysiology Experiment

Responses to 1600 test images of two example units

IT unit 53



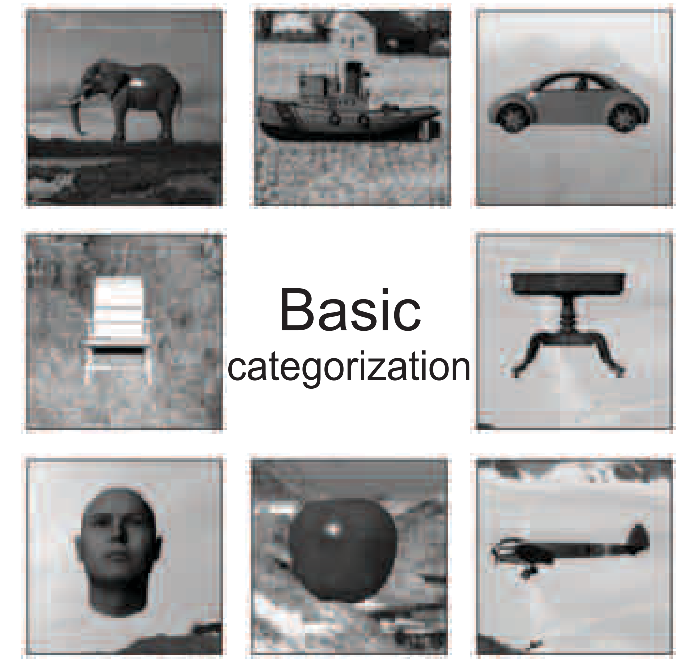
Images sorted first by **category**, then **variation level**.

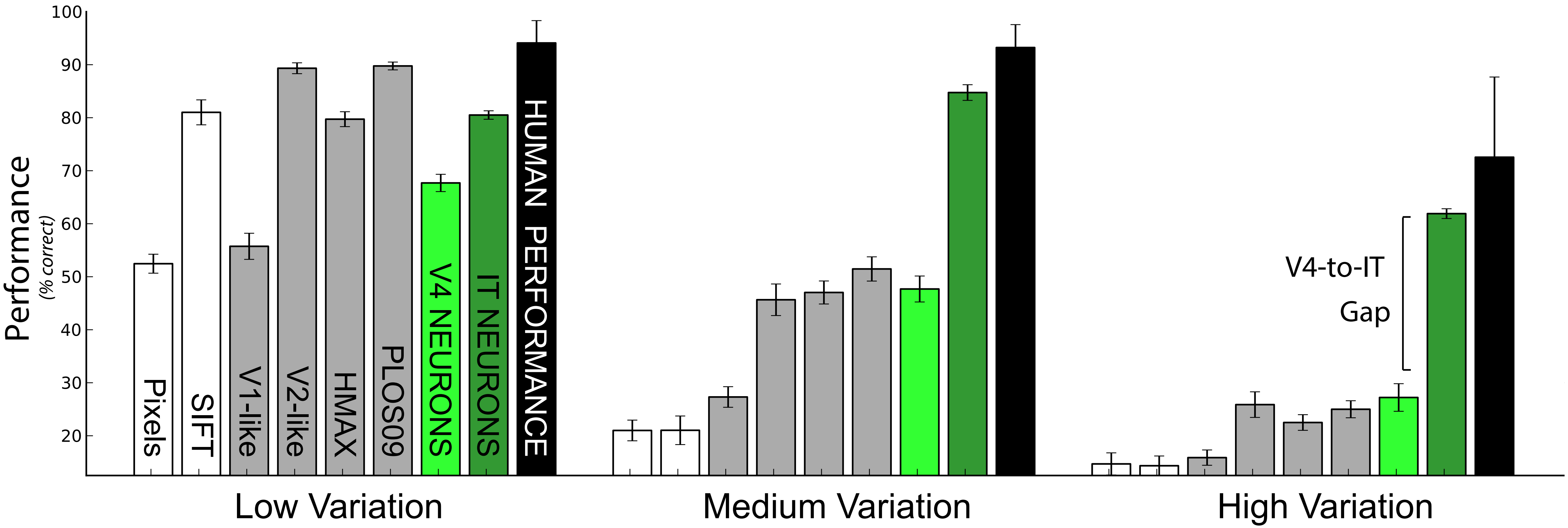
IT Neurons Track Human Performance

V4 loses out at higher variation:

… but humans are much less affected.

… as is the IT neural population.

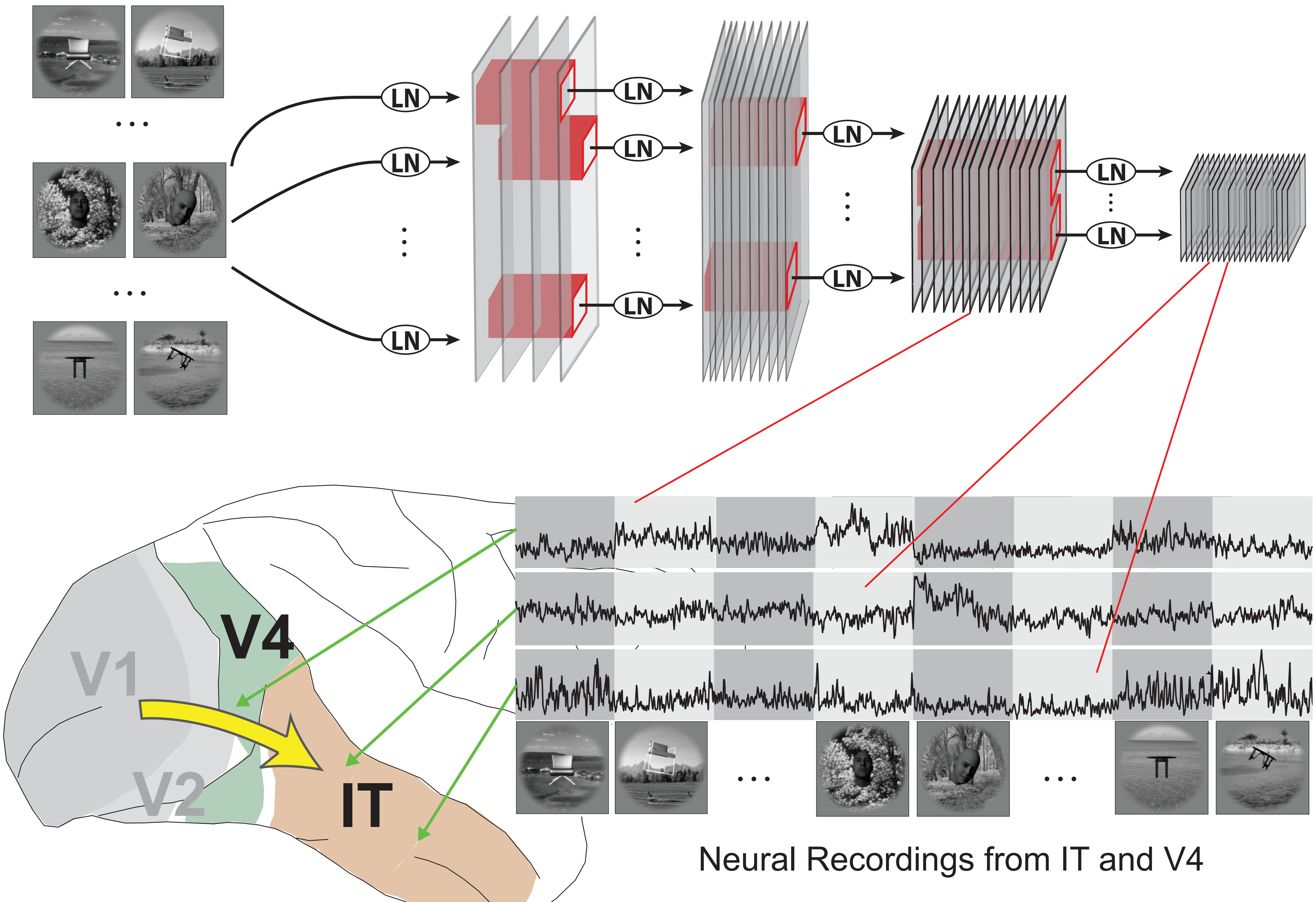




Yamins and Hong* et. al.* ***PNAS*** *(2014)*

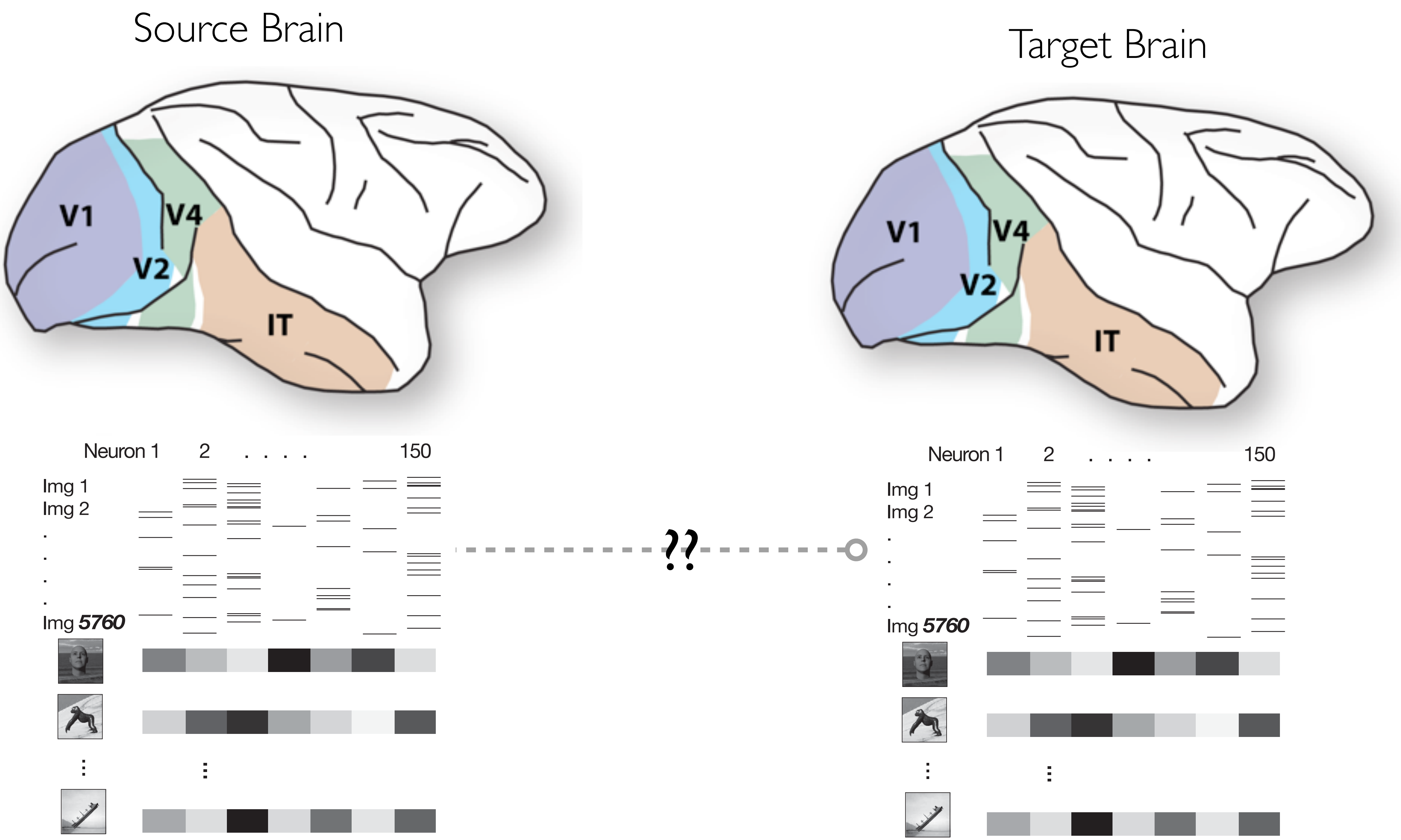
At <u>high variation levels</u>, IT much better than V4 and existing models.

Neural predictivity: the ability of model to predict each individual neural site's activity.



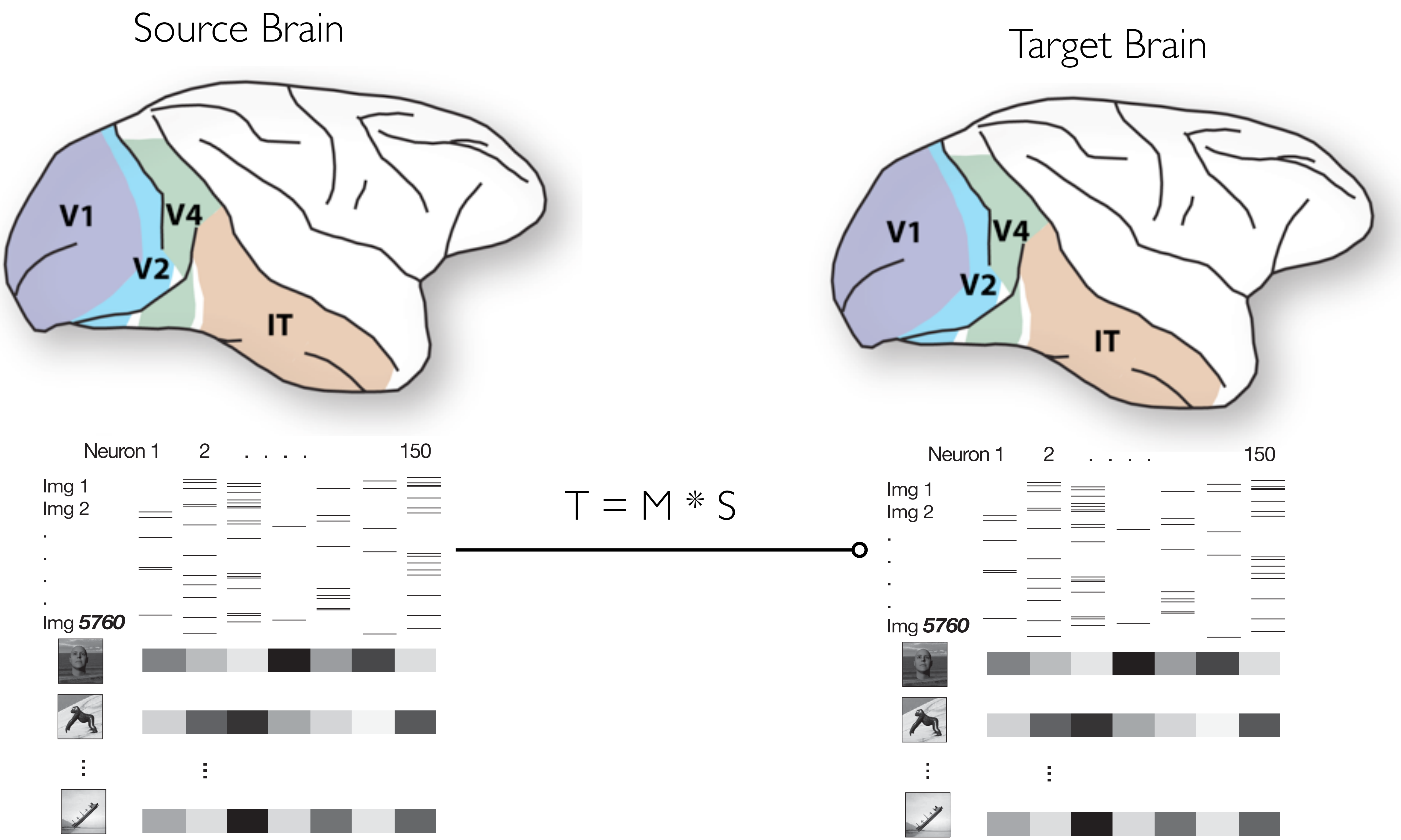
Neural Response Prediction

Some kind of mapping is necessary.



Neural Response Prediction

Here, we use linear regression.



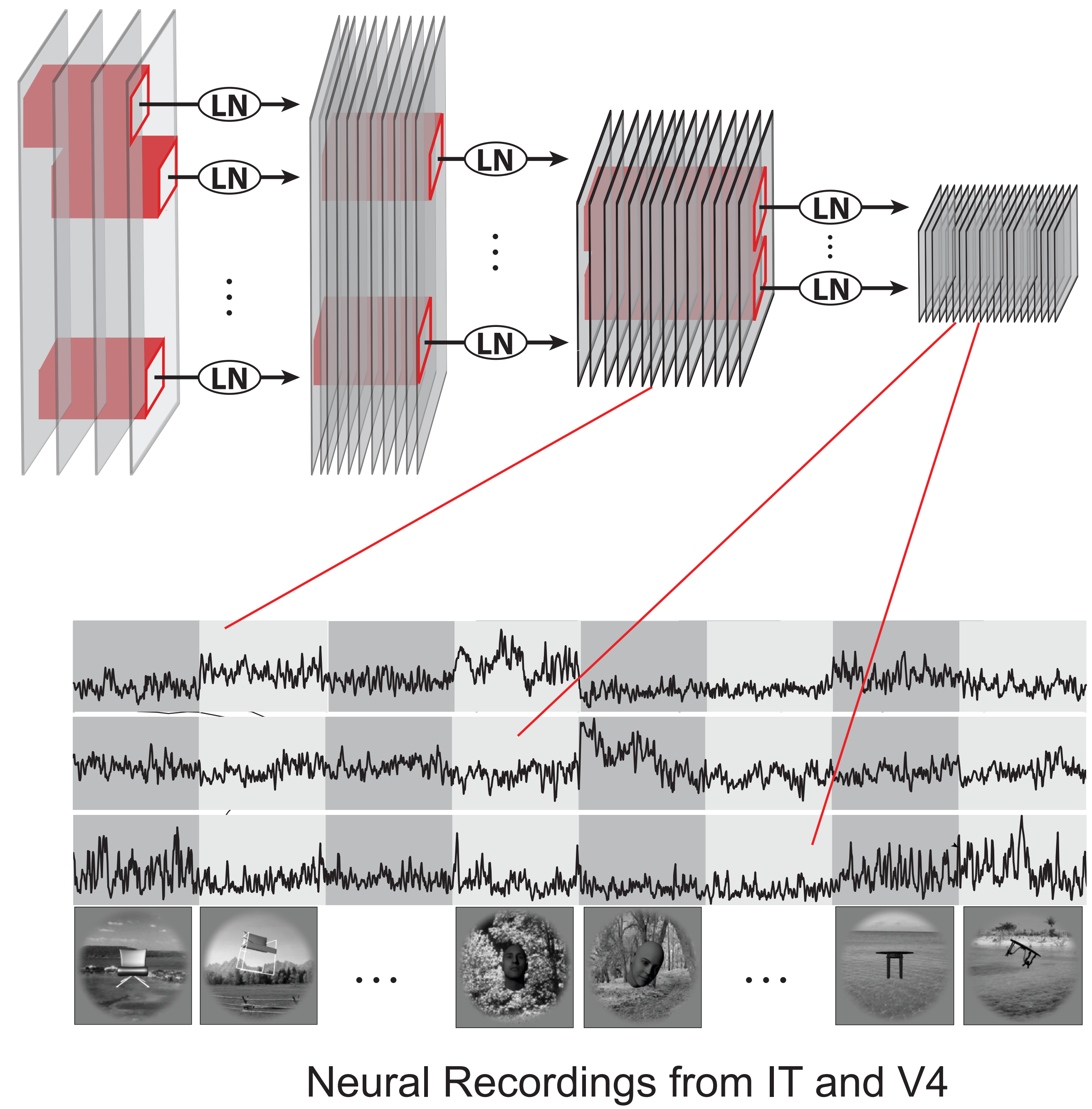
Neural predictivity: the ability of model to predict each individual neural site's activity.

Neural site unit ~ sparse linear combination of model units

Linear regression with fixed training images.

Accuracy = goodness-of-fit on held-out testing images (Cross validated)

Neural predictivity = median accuracy over all units.

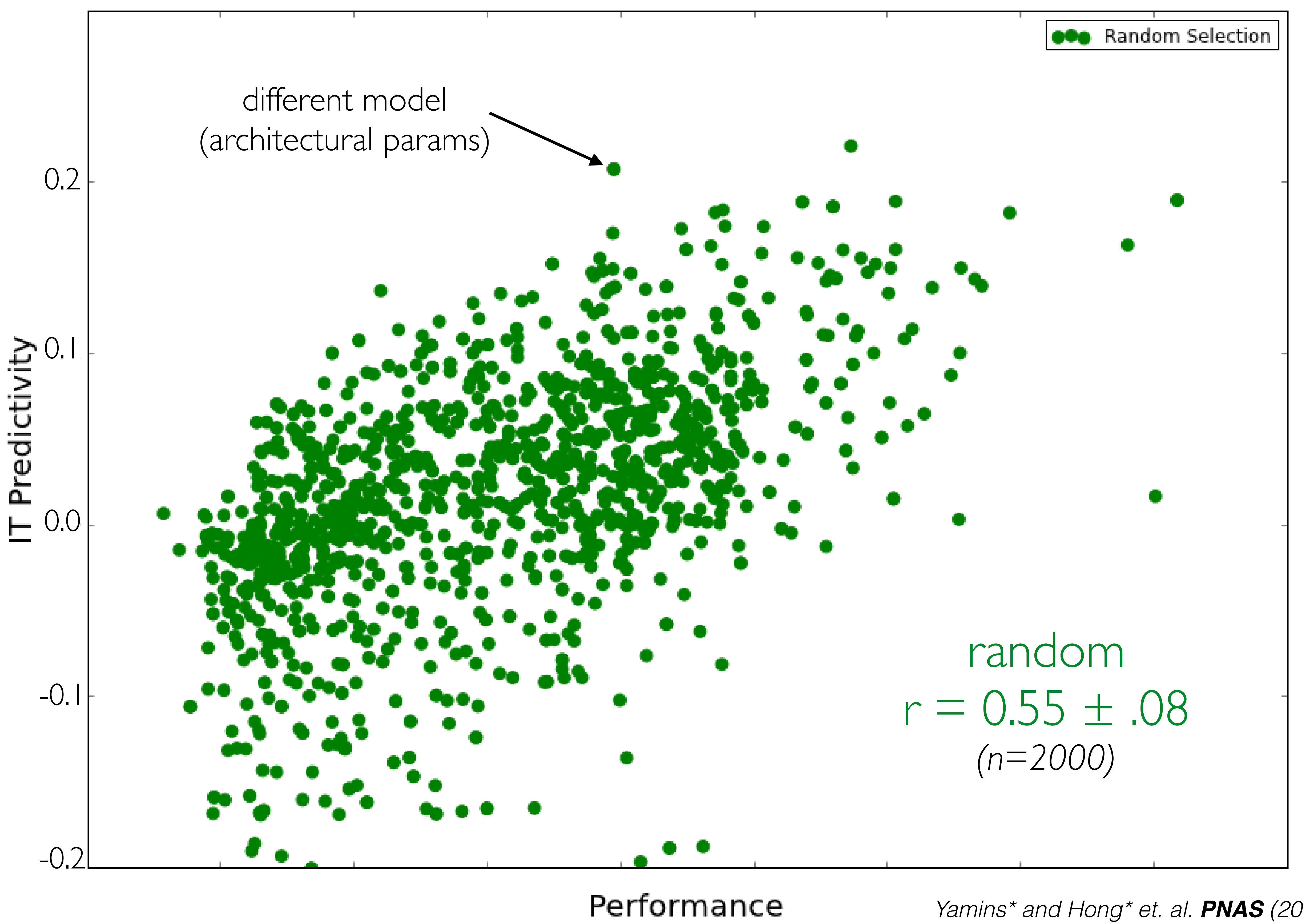


Initial Validation of Idea

High-throughput experiments to directly test the relationship between performance and IT neural predictivity.

- Random selection of model parameters; measure performance and neural predictivity Pinto et. al (2008, 2009)

Initial Validation of Idea



Yamins and Hong* et. al.* ***PNAS*** *(2014)*

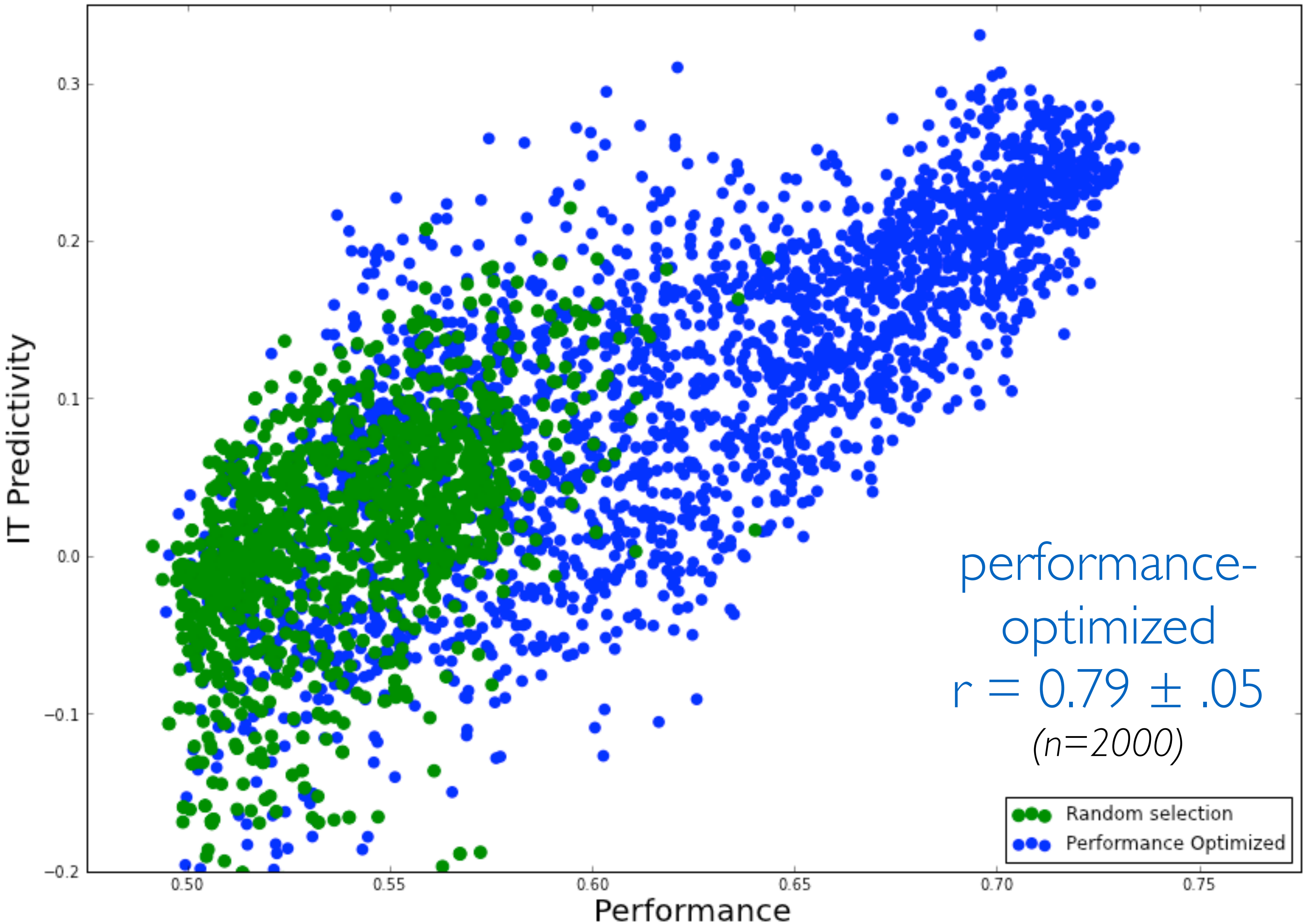
Initial Validation of Idea

High-throughput experiments to directly test the relationship between neural predictivity and performance.

- Random selection of model parameters; measure performance and neural predictivity   Pinto et. al (2008, 2009)

- Optimize parameters for performance; measure neural predictivity. optimization techniques: Bergstra Yamins & Cox (2013)

Initial Validation of Idea



Yamins and Hong* et. al.* ***PNAS*** *(2014)*

Initial Validation of Idea

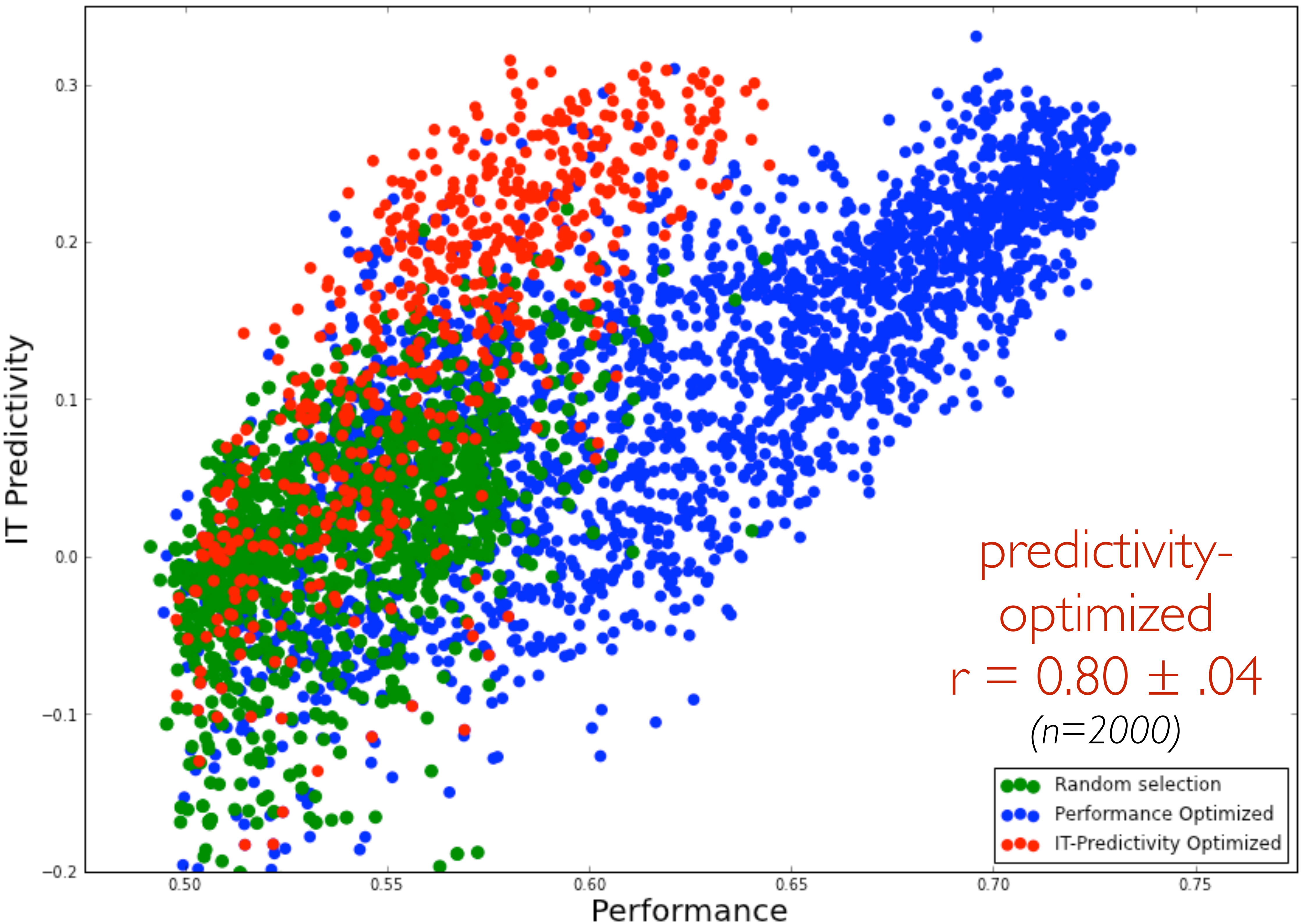
High-throughput experiments to directly test the relationship between neural predictivity and performance.

- Random selection of model parameters; measure performance and neural predictivity Pinto et. al (2008, 2009)

- Optimize parameters for performance; measure neural predictivity optimization techniques: Bergstra Yamins & Cox (2013)

- Optimize parameters for neural predictivity; measure performance

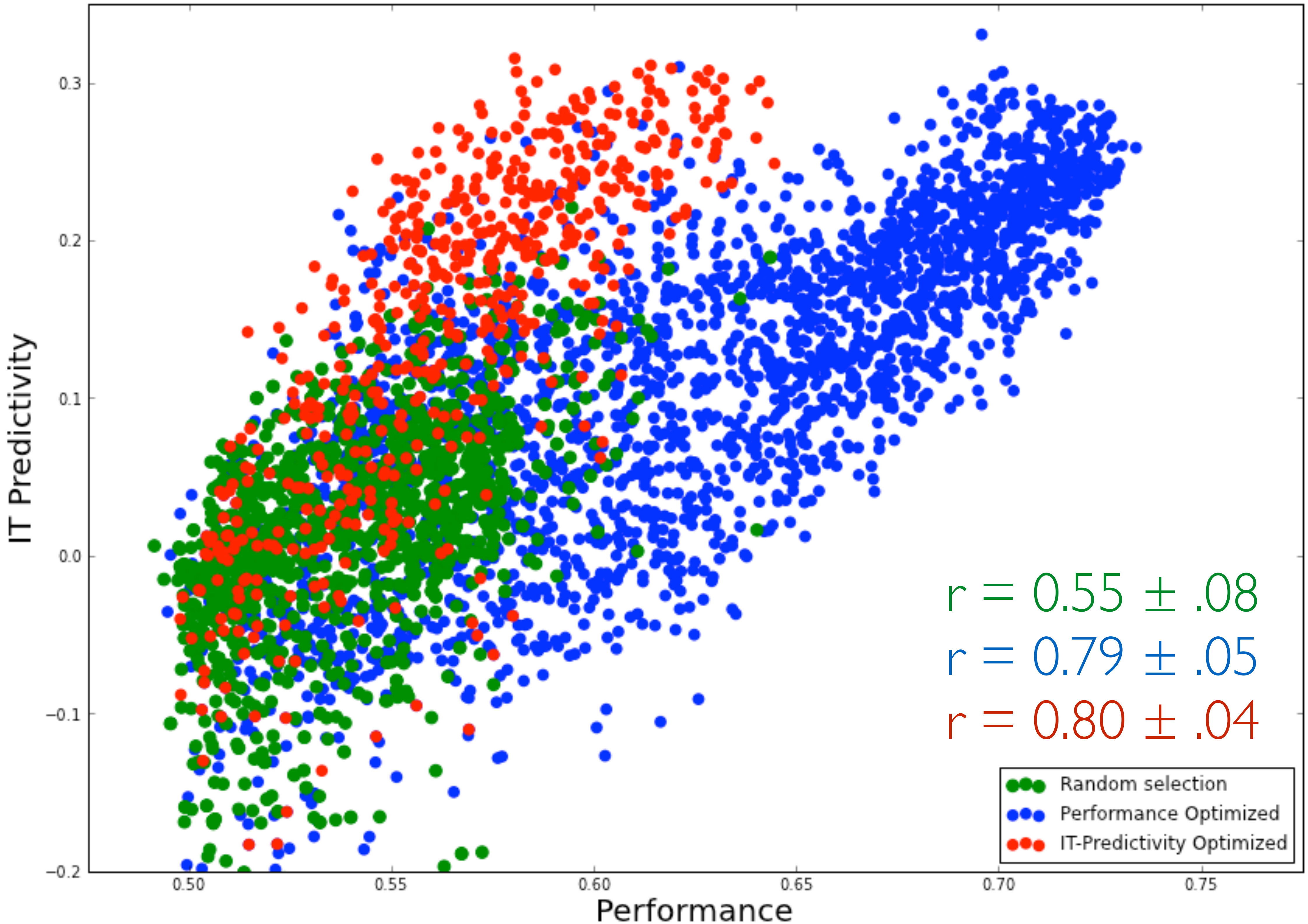
Performance vs IT predictivity: Predictivity-Optimized



Yamins and Hong* et. al.* ***PNAS*** *(2014)*

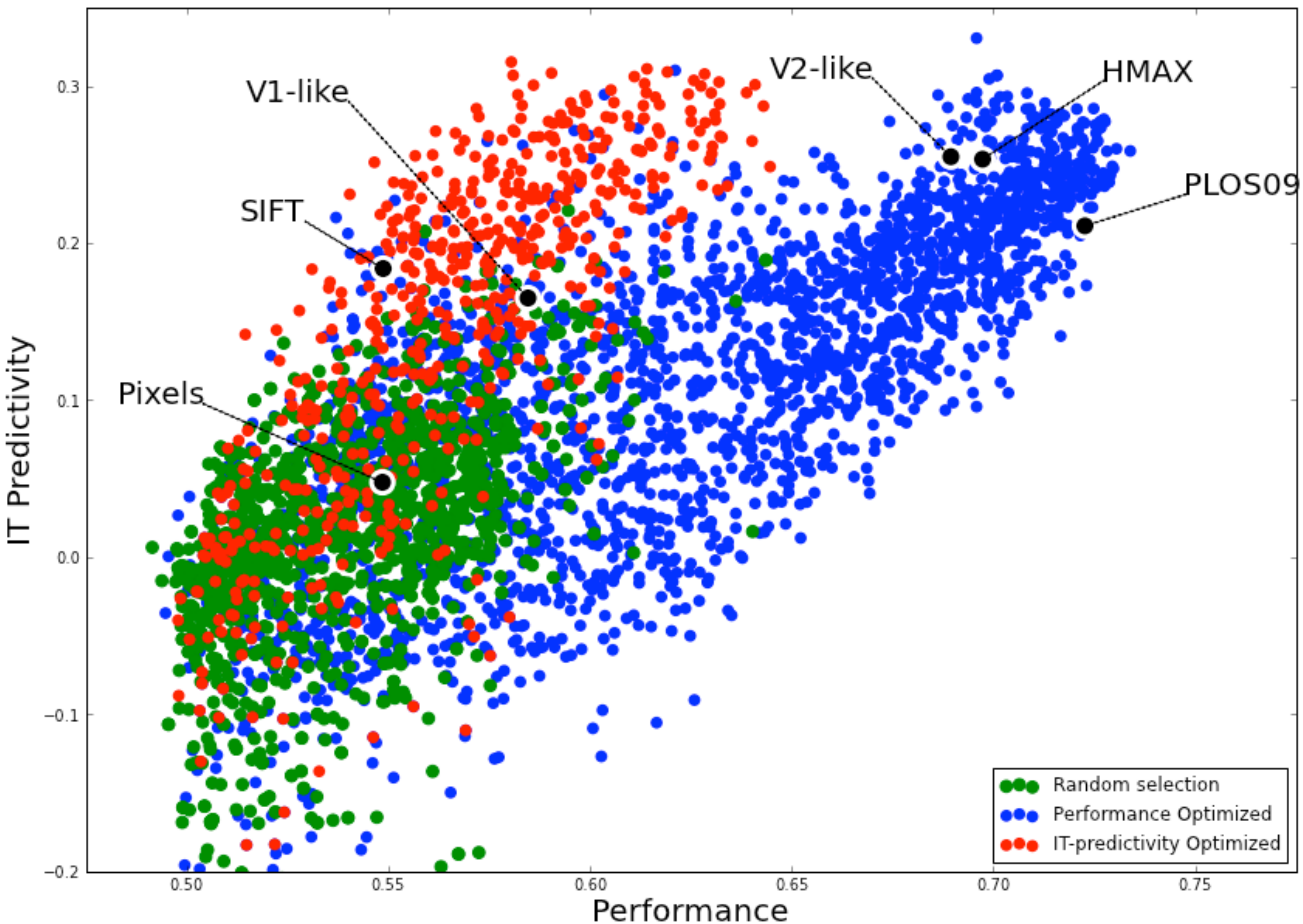
Performance vs IT predictivity: Predictivity-Optimized

Performance is a potentially very good driver of neural prediction.



Yamins and Hong* et. al.* ***PNAS*** *(2014)*

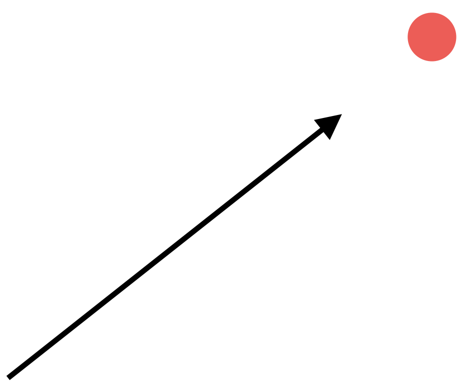
Performance vs IT predictivity

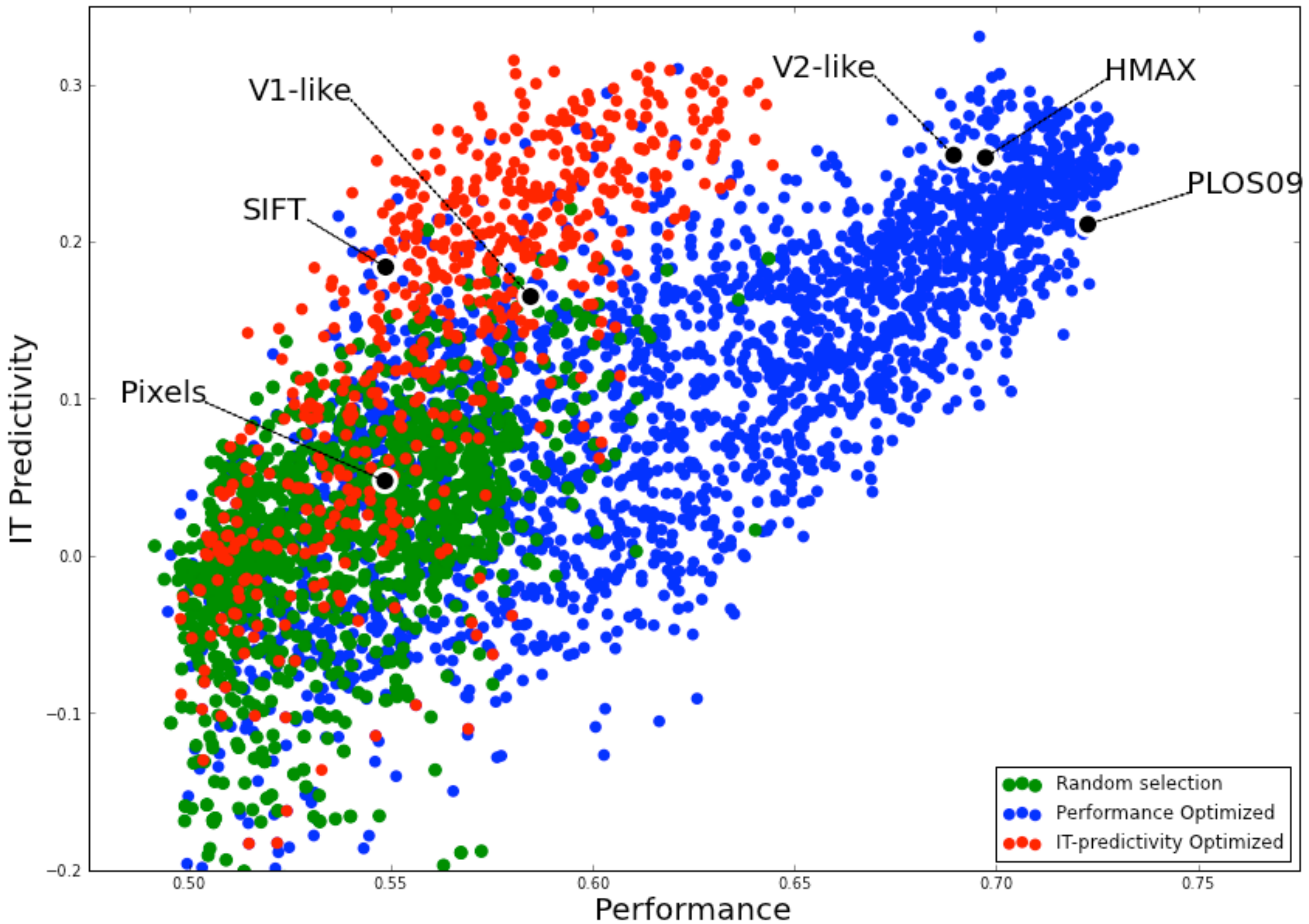


Yamins and Hong* et. al.* ***PNAS*** *(2014)*

Performance vs IT predictivity

But, not doing that well. Really want to be here:





Optimization Strategy

i. **architectural** params: (# layers, # filters, receptive field sizes, &c) — "network structure"

→ Automated meta-parameter optimization in high-dimensional discrete parameter spaces
Bergstra Yamins & Cox (2013)

→ Ensembles of models chosen through modified boosting Yamins et. al (2013, 2014)

Optimization Strategy

i. **architectural** params: (# layers, # filters, receptive field sizes, &c) — "network structure"

→ Automated meta-parameter optimization in high-dimensional discrete parameter spaces
Bergstra Yamins & Cox (2013)

→ Ensembles of models chosen through modified boosting Yamins et. al (2013, 2014)

ii. **filter** parameters: continuous valued pattern templates — "network contents"

→GPU-accelerated stochastic gradient descent Pinto et. al., (2009), Krizhevsky et. al. (2012)

Gradient descent eq: $$\frac{dp}{dt} = -\lambda(t) \cdot \frac{\partial L}{\partial P}$$

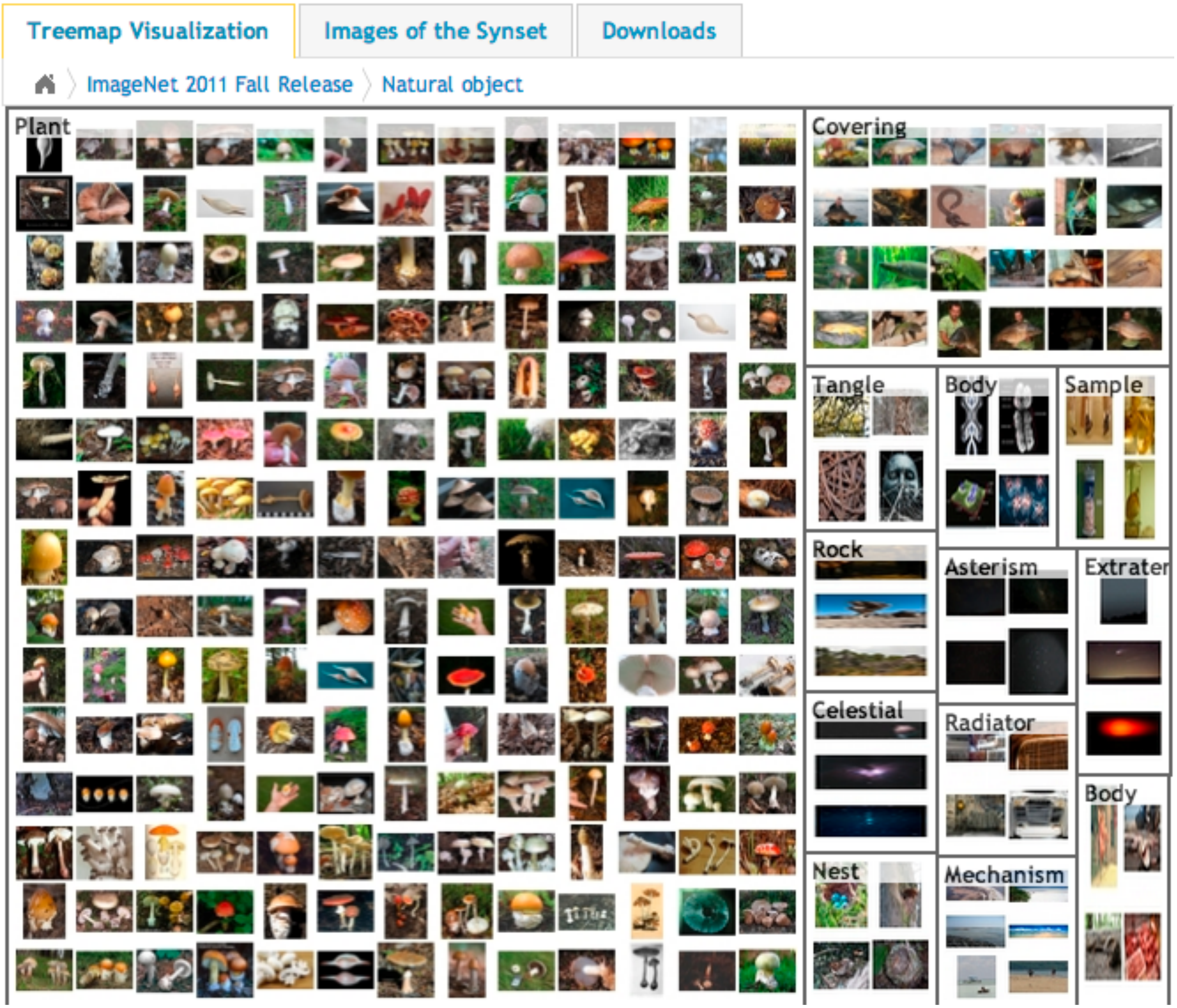
L = loss function
λ = learning rate

In current practice:

L = loss computed from **large numbers of externally-provided object category labels.**

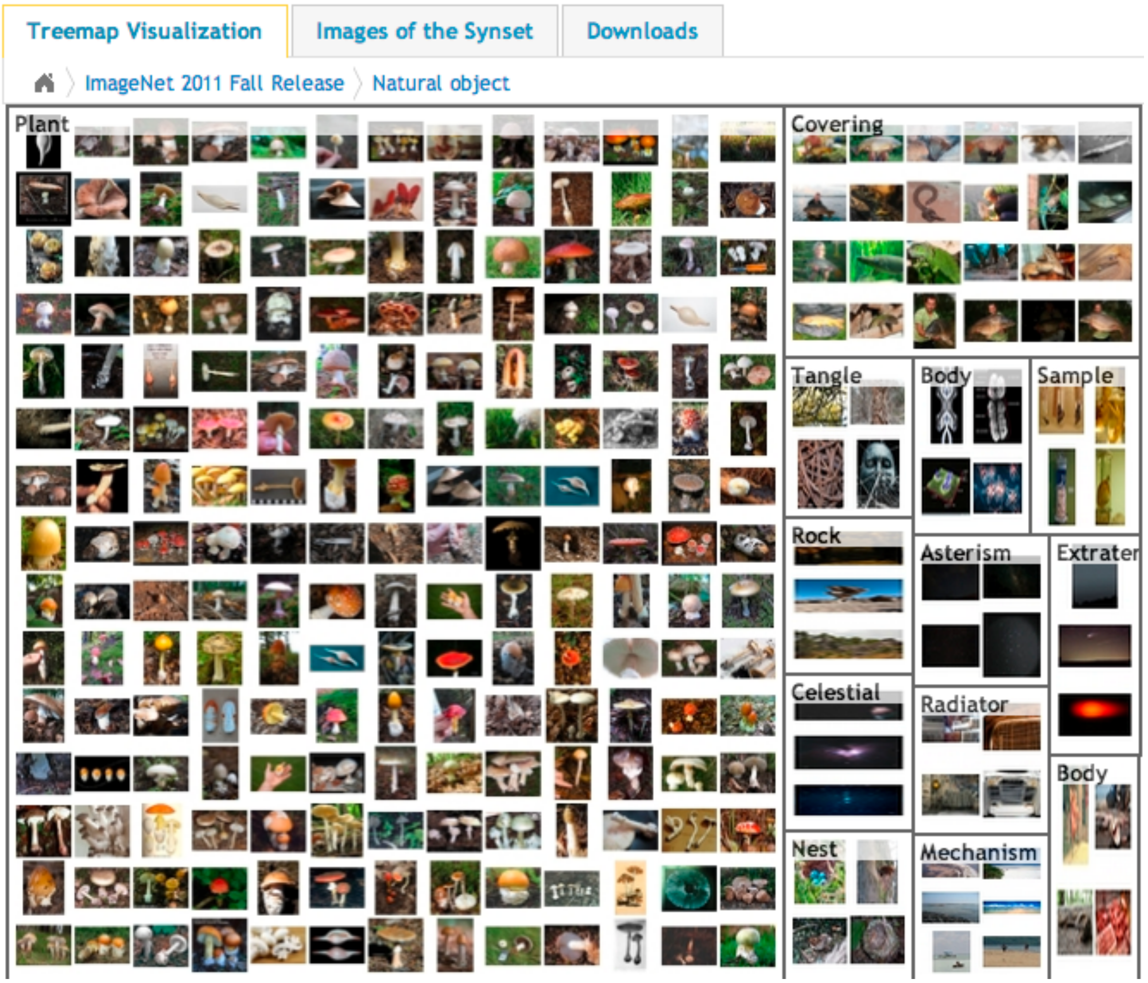
Model Training Regimen

ImageNet (2012). Thousands of images in thousands of categories.



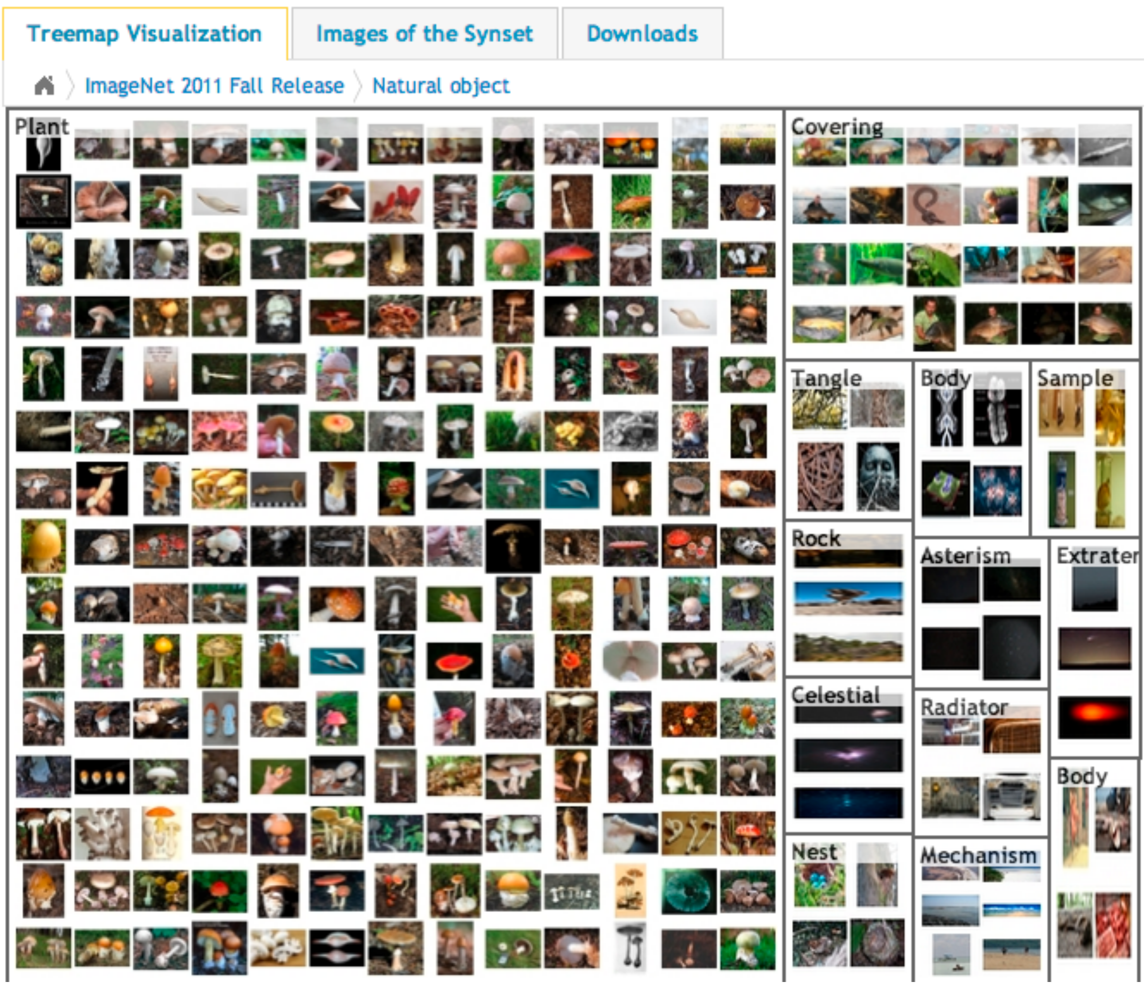
Model Training Regimen

train: real photos



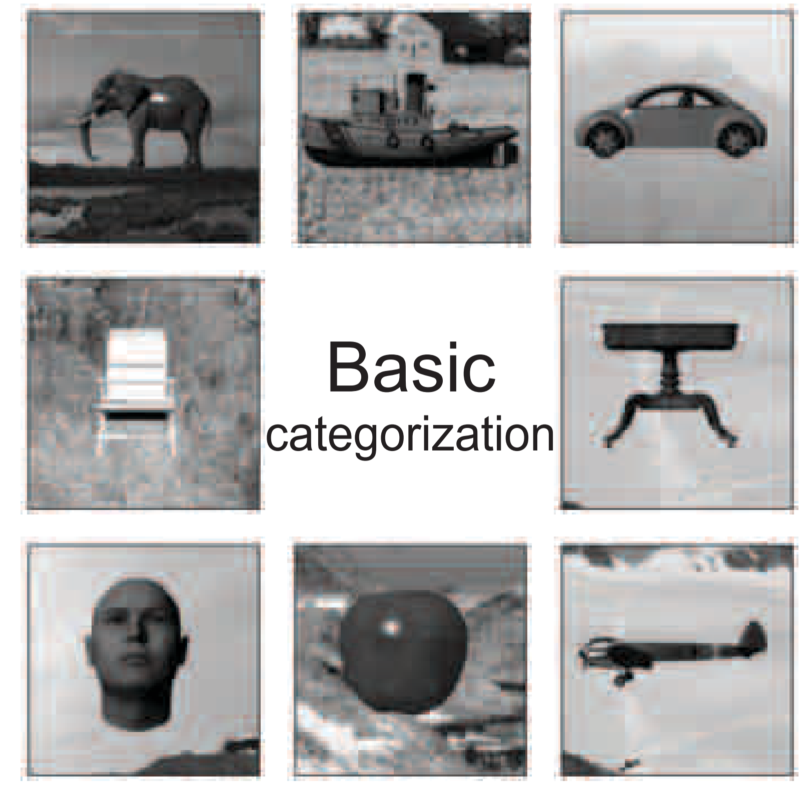
Model Training Regimen

train: real photos



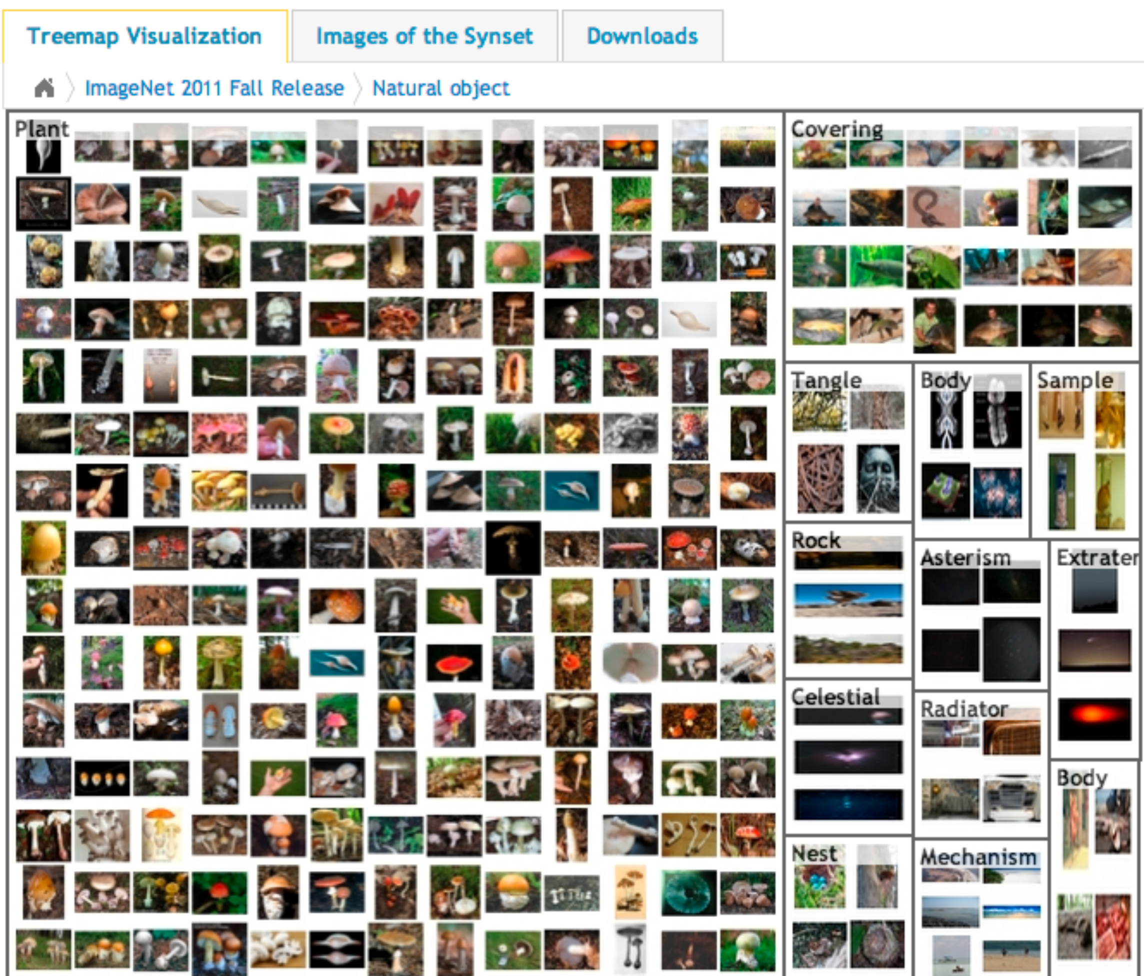
generalize?

test: neural stimuli



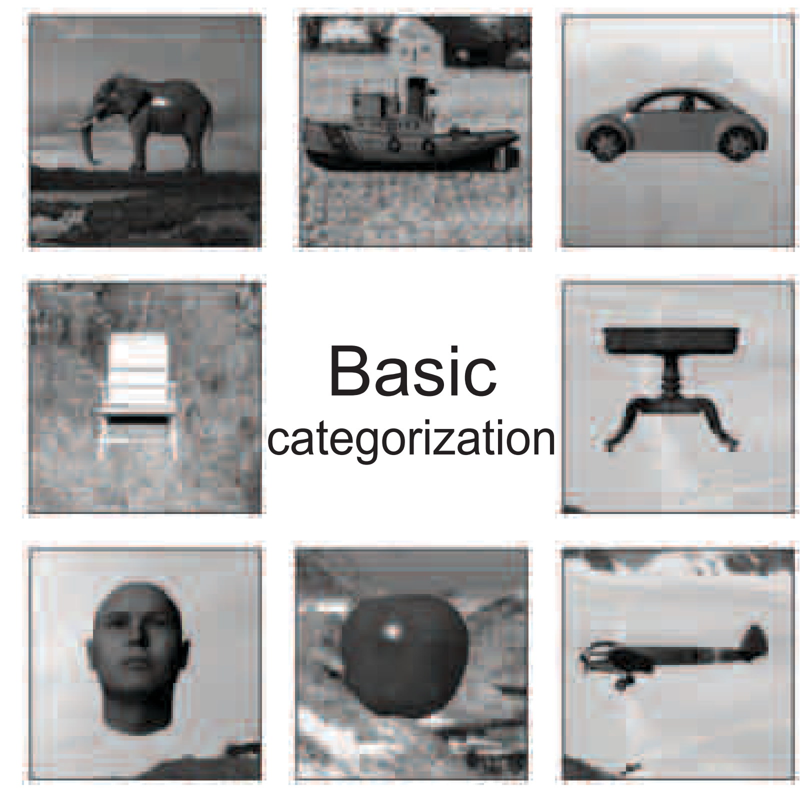
Model Training Regimen

train: real photos



generalize?

test: neural stimuli

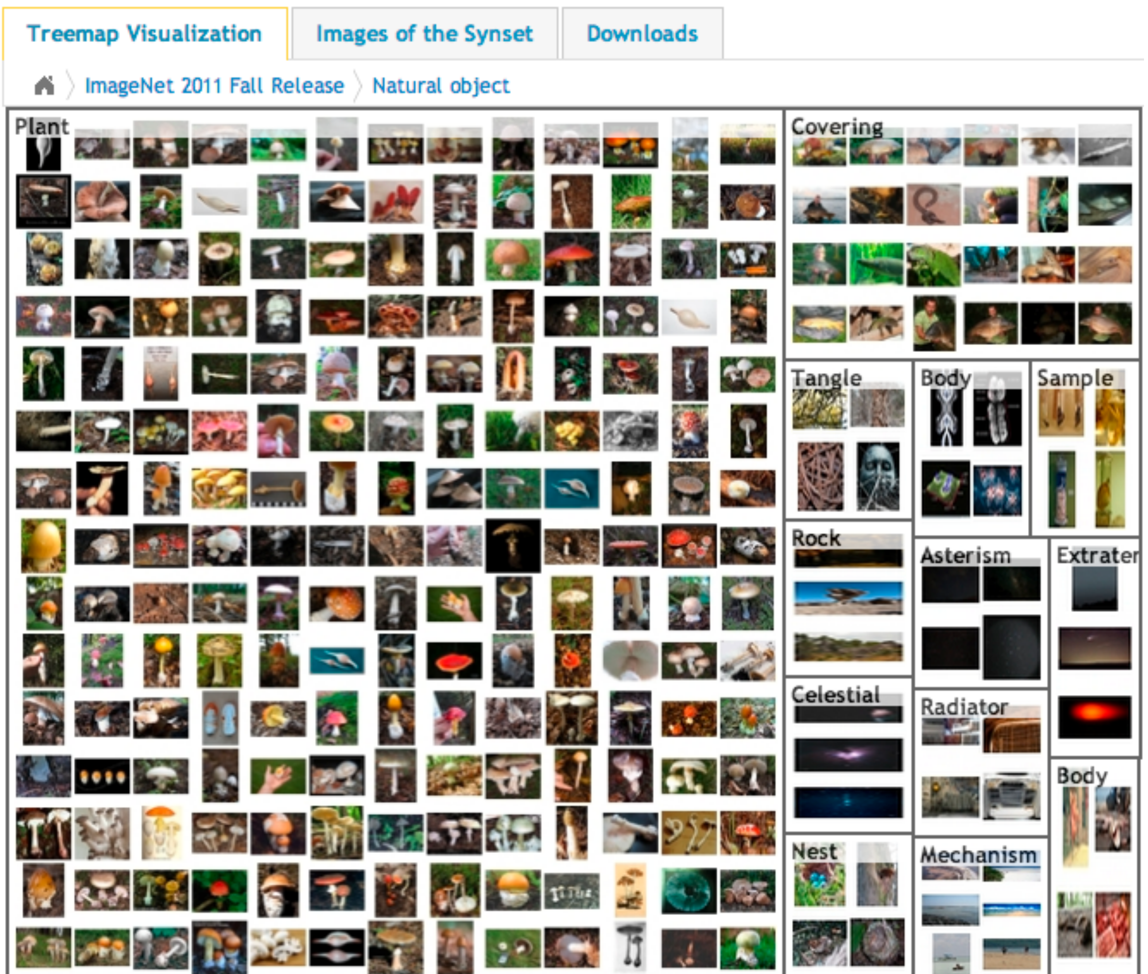


removed categories of photos that appeared in the test stimuli

(animals, boats, cars, chairs, faces, fruits, planes, tables)

Model Training Regimen

train: real photos



test: neural stimuli

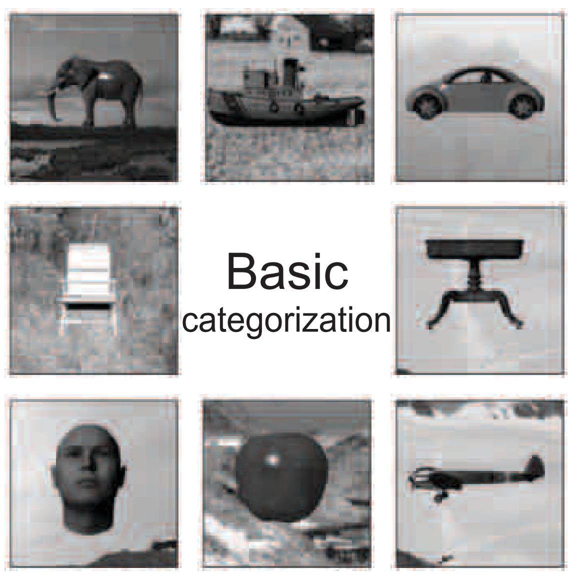
generalize? →

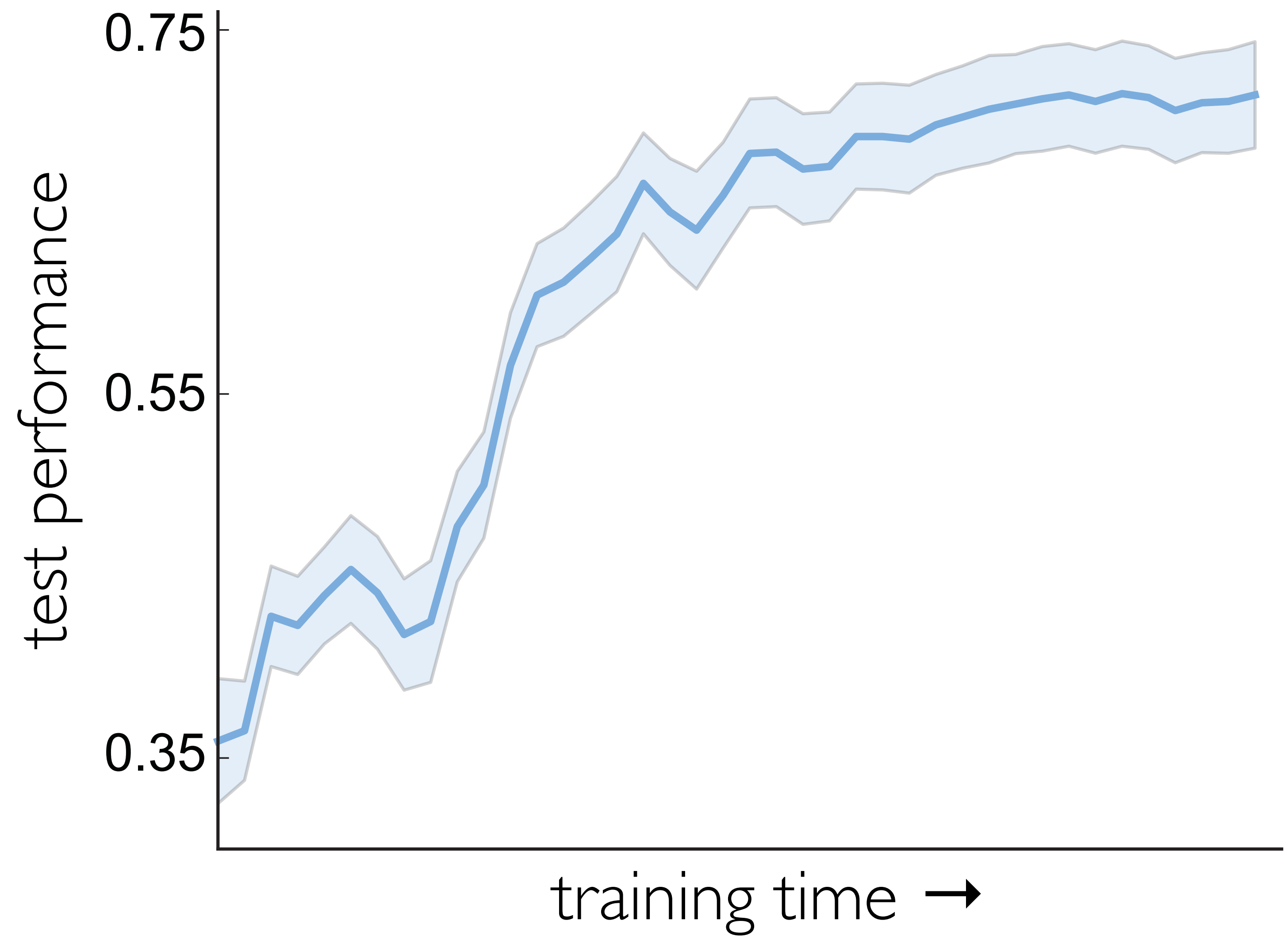
removed categories of photos that appeared in the test stimuli

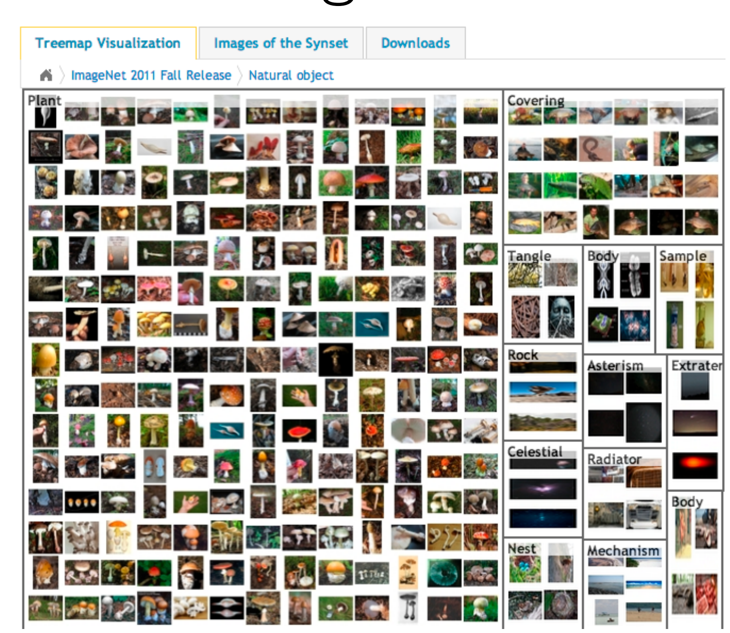
(animals, boats, cars, chairs, faces, fruits, planes, tables)

→ Specific 4-layer model that achieved high recognition performance.

Performance Generalization

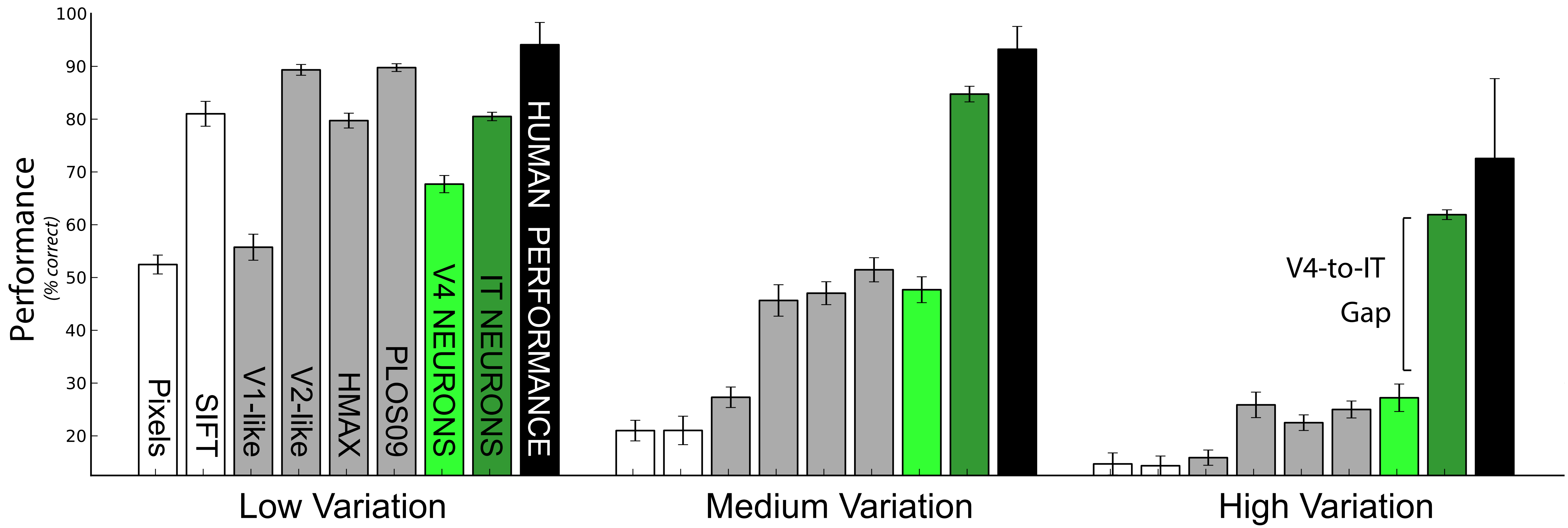






Performance Comparison

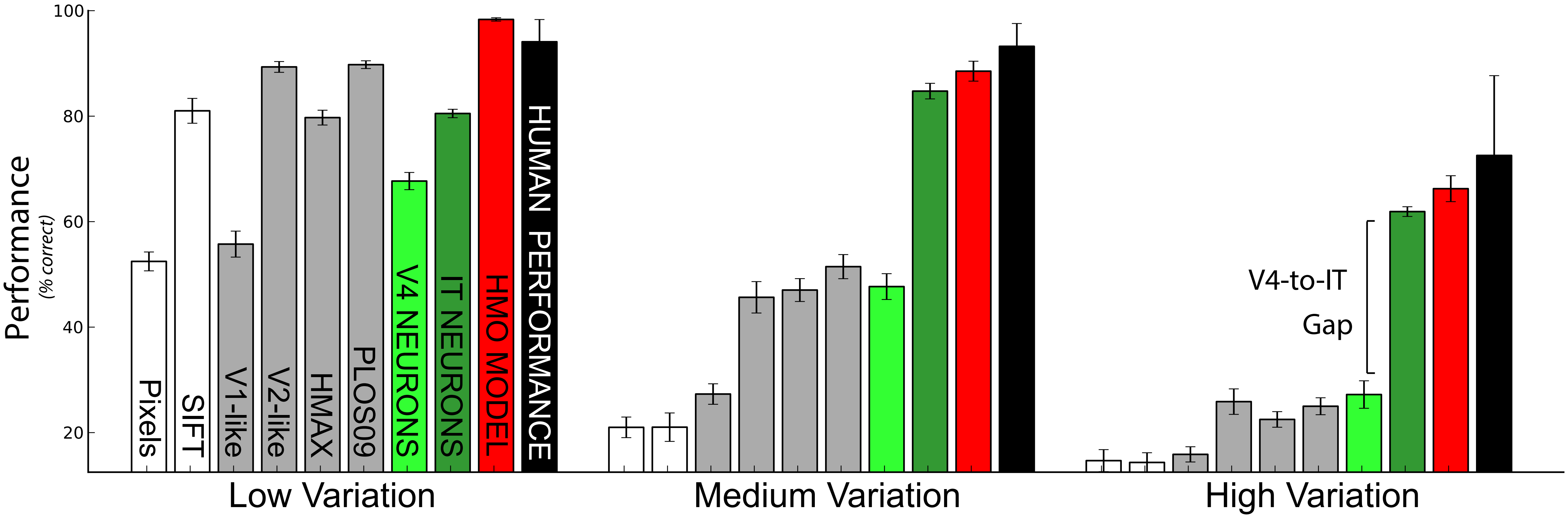
At <u>high variation levels</u>, IT much better than V4 and existing models



Yamins and Hong* et. al.* ***PNAS*** *(2014)*

Performance Comparison

At <u>high variation levels</u>, IT much better than V4 and existing models

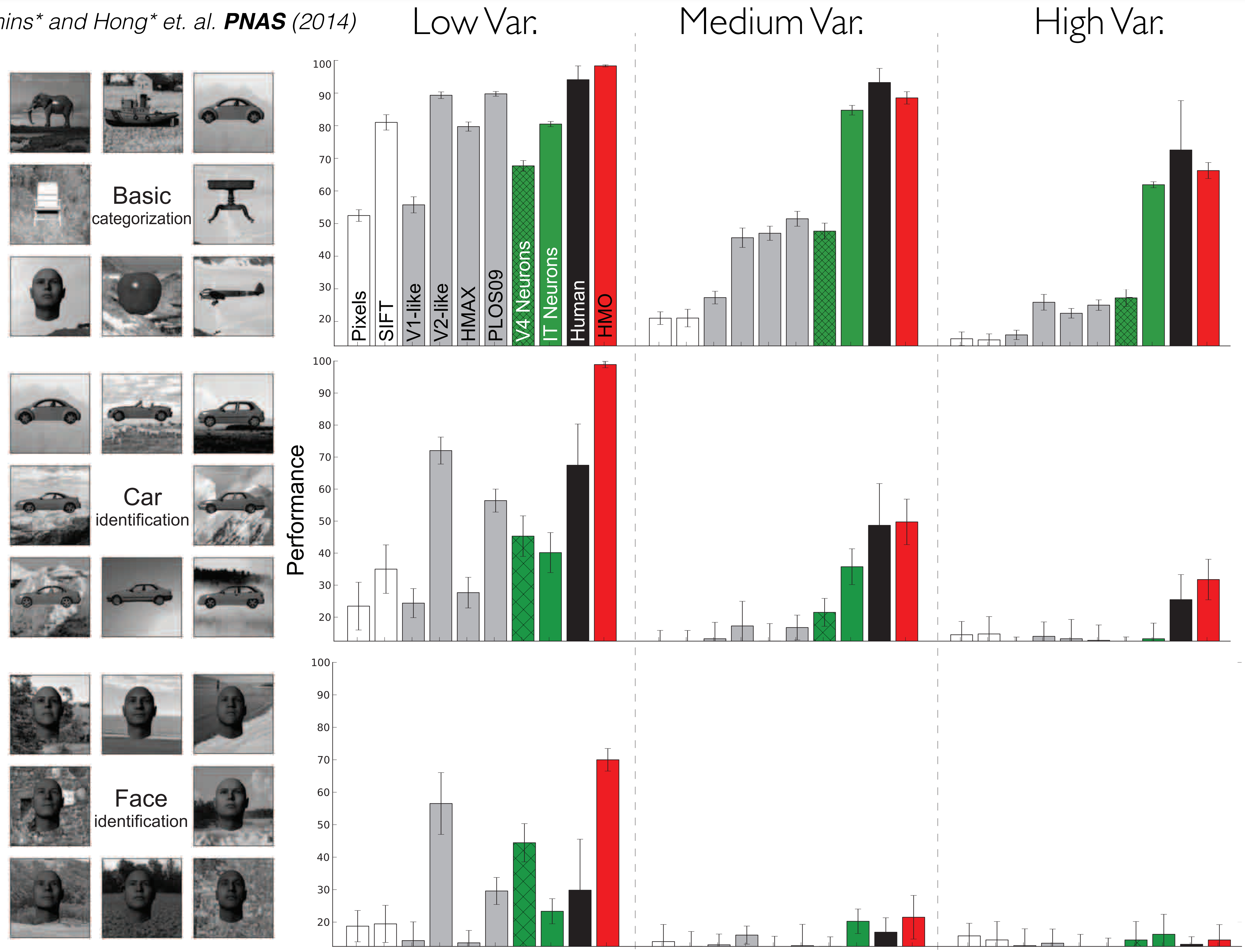


Yamins and Hong* et. al.* ***PNAS*** *(2014)*

New model comparable to IT / human performance levels.

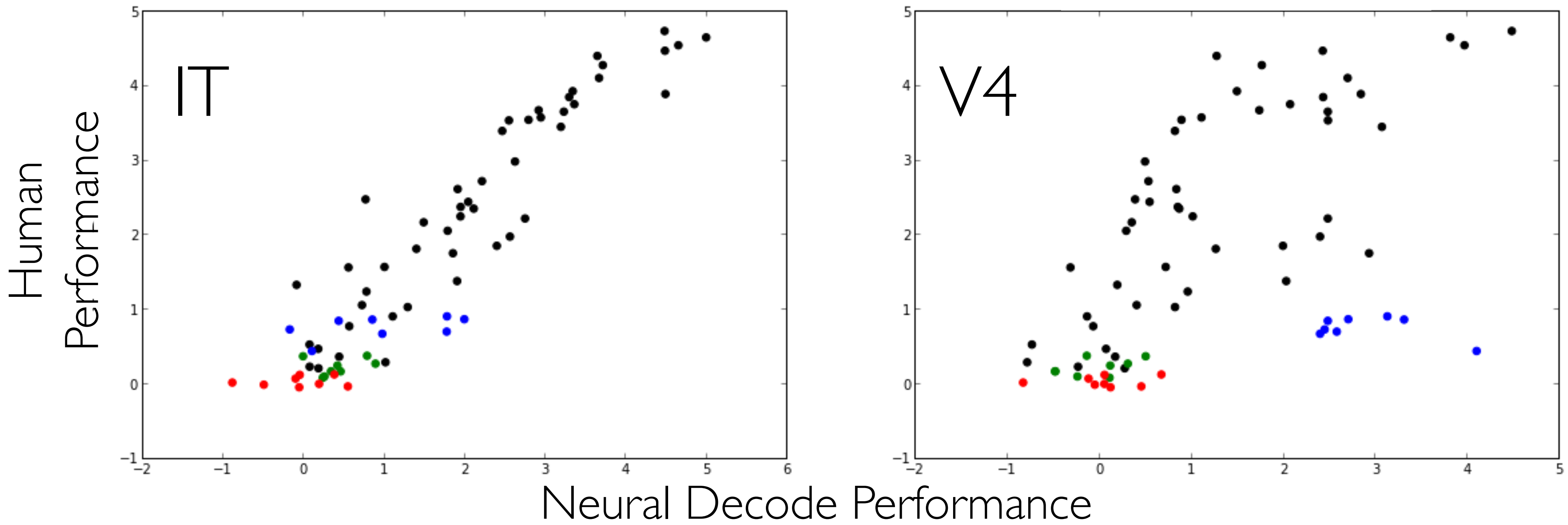
Performance Comparison

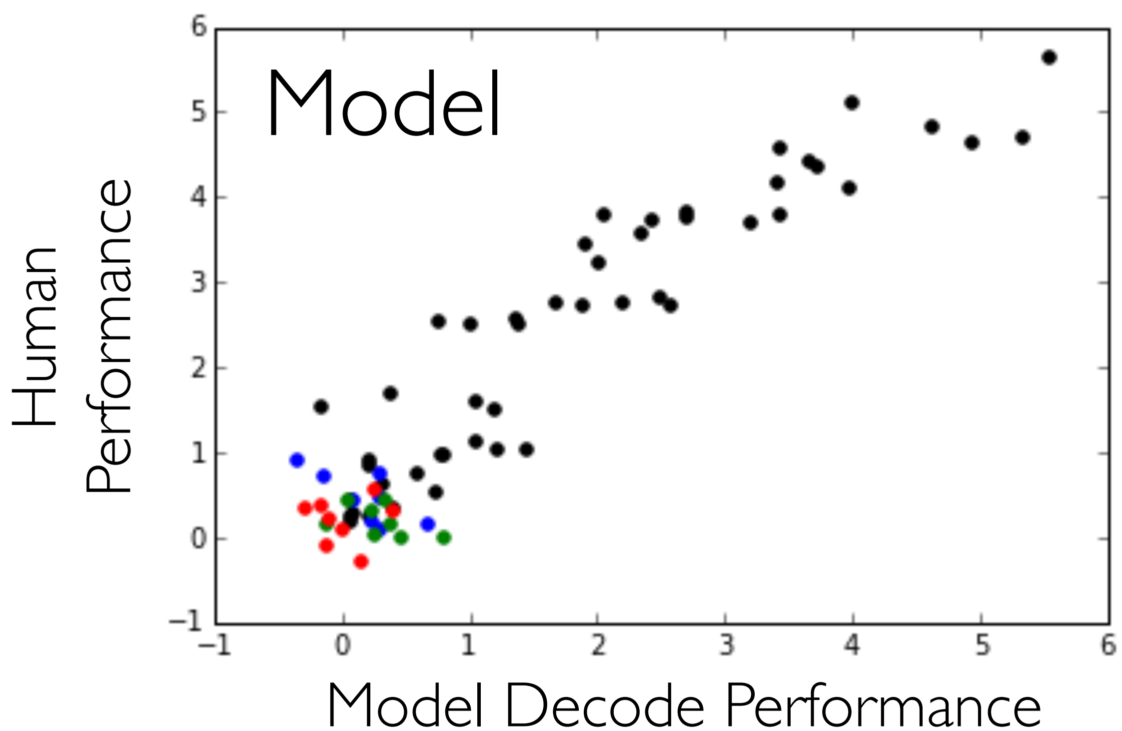
Yamins and Hong* et. al.* ***PNAS*** *(2014)*



Performance Comparison

Behavioral match between models and data at category confusion level is pretty good …



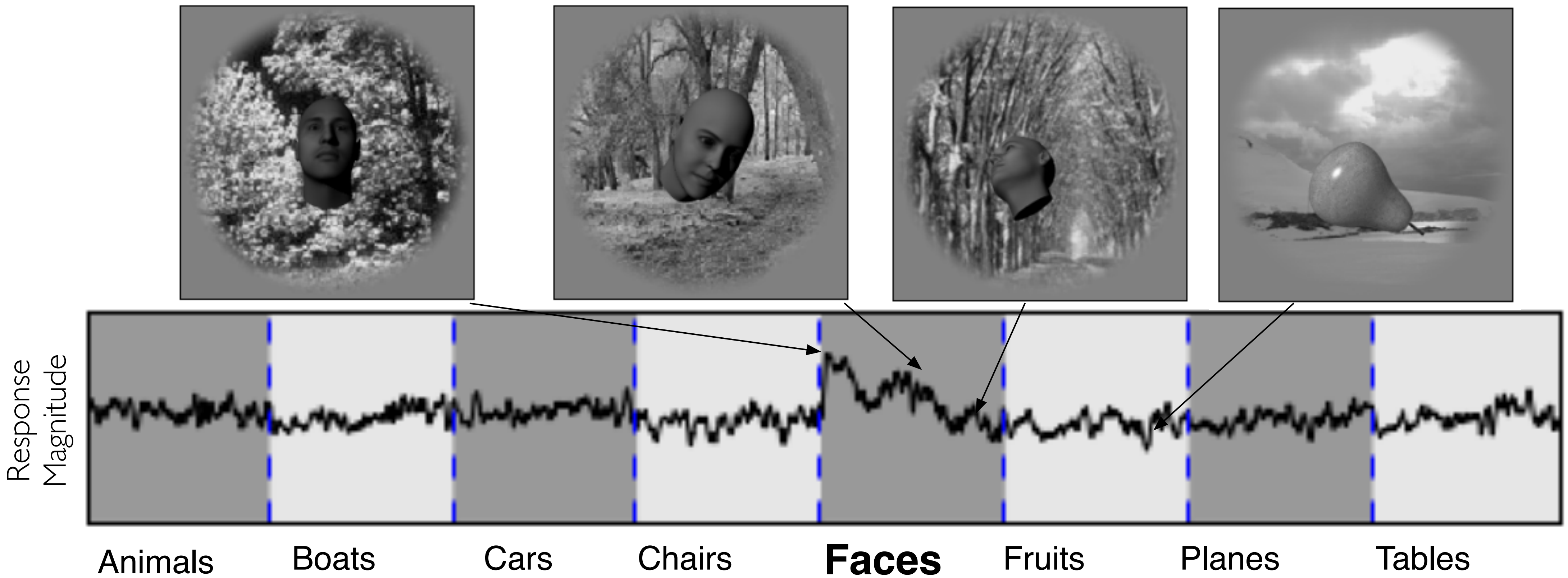


Does it predict neurons better?

Does it predict neurons better?

Yamins and Hong* et. al.* ***PNAS*** *(2014)*

unit 53



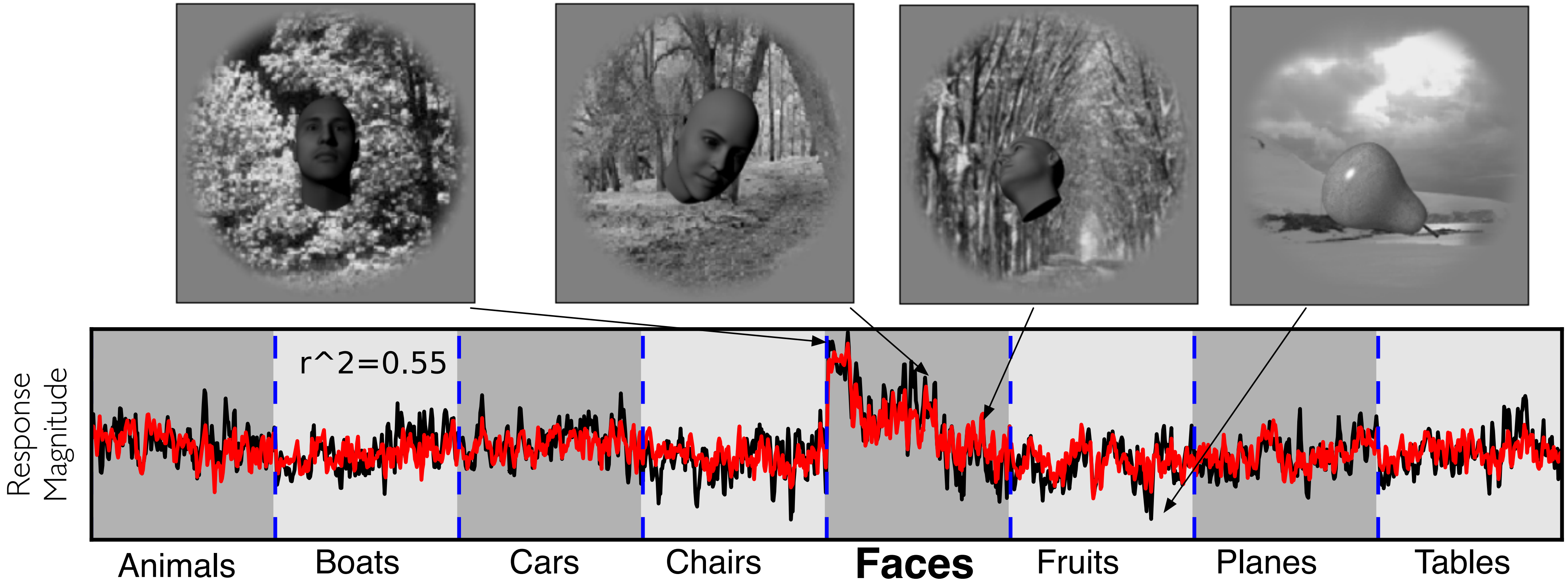
Images sorted first by **category**, then **variation level**.

Neural data

Does it predict neurons better?

unit 53

Yamins* and Hong* et. al. ***PNAS*** *(2014)*

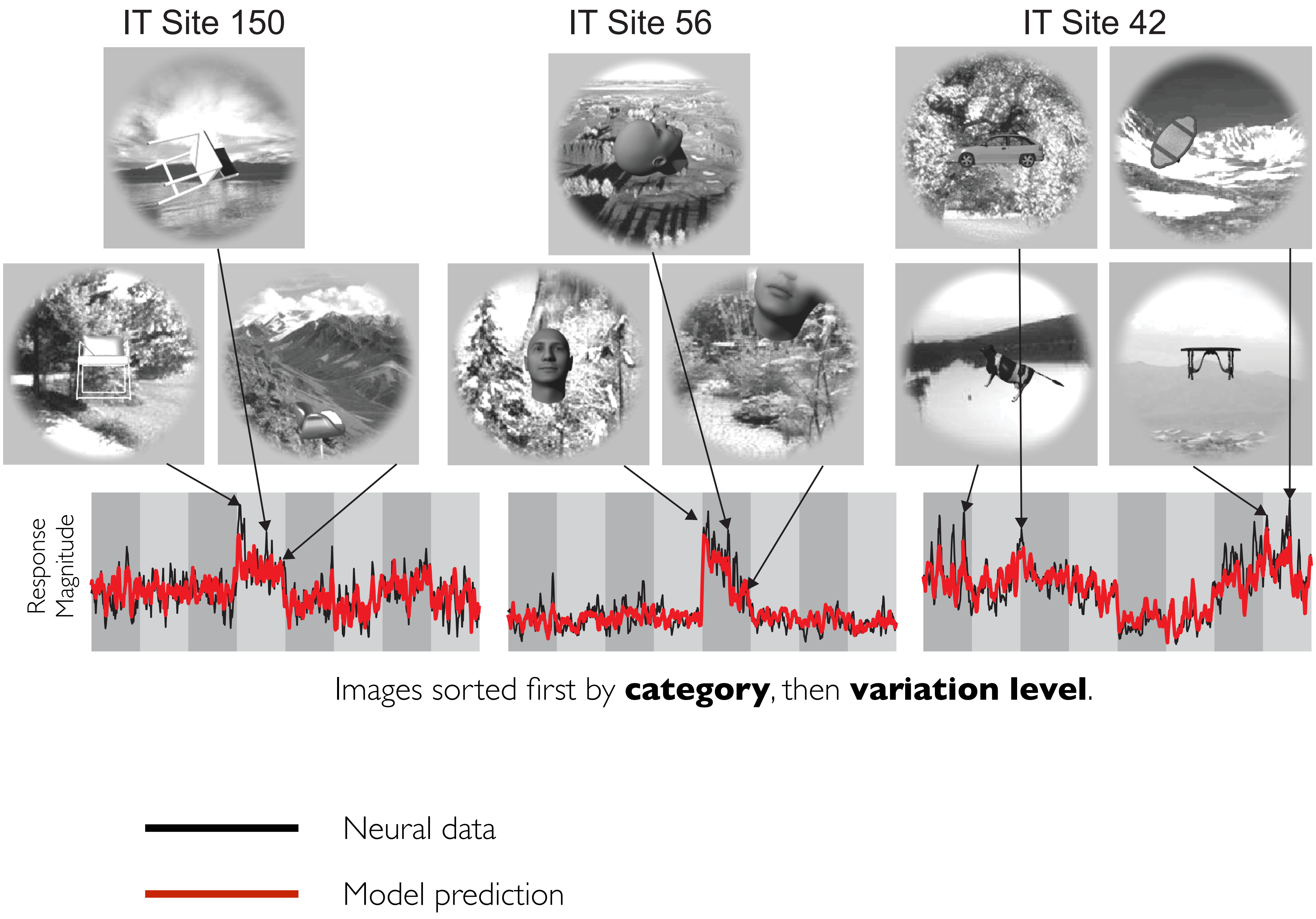


Images sorted first by **category**, then **variation level**.

Neural data

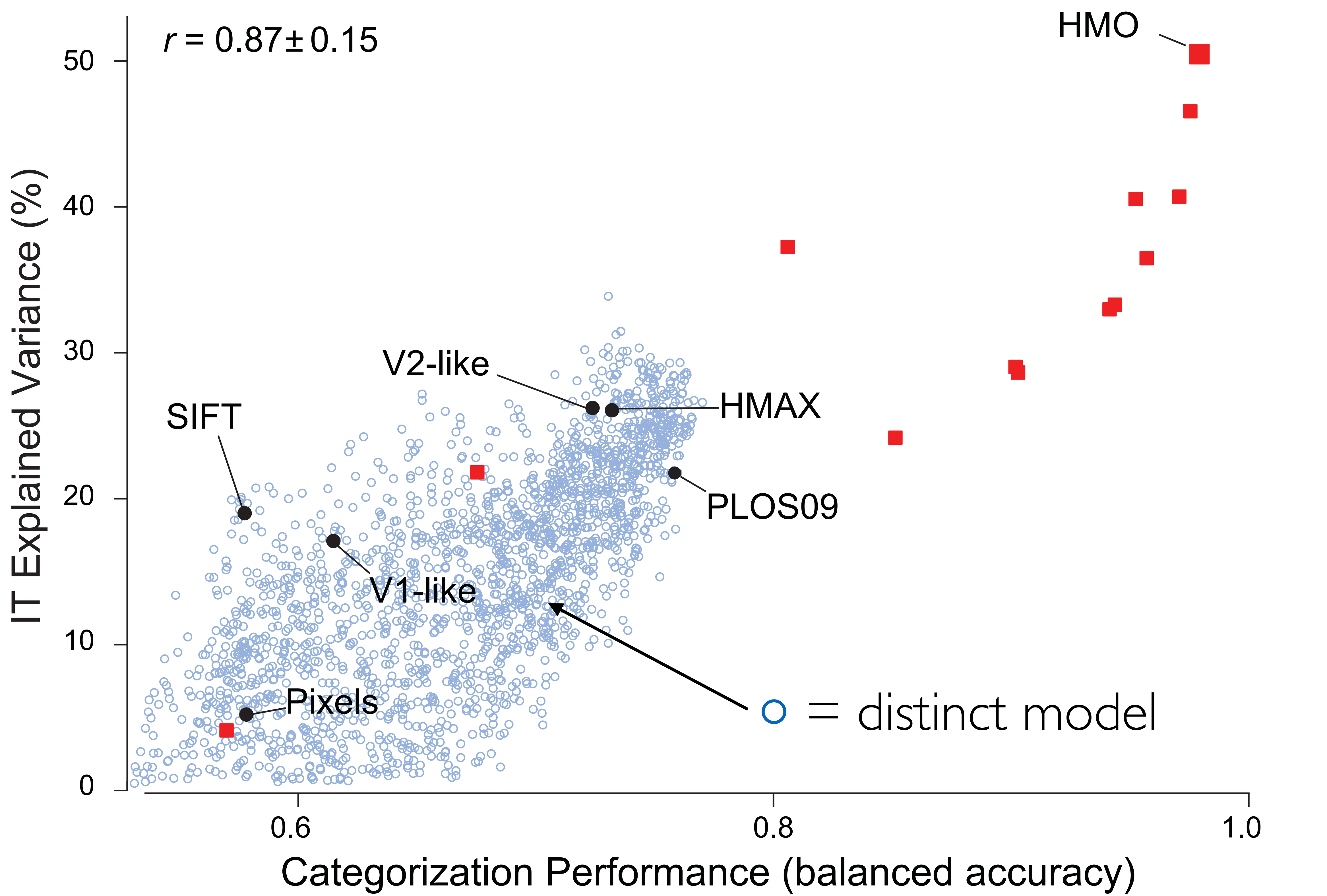
Model prediction

Predicting IT Neural Responses



Key Underlying Principle

Yamins and Hong* et. al.* ***PNAS*** *(2014)*



Predicting IT Neural Responses

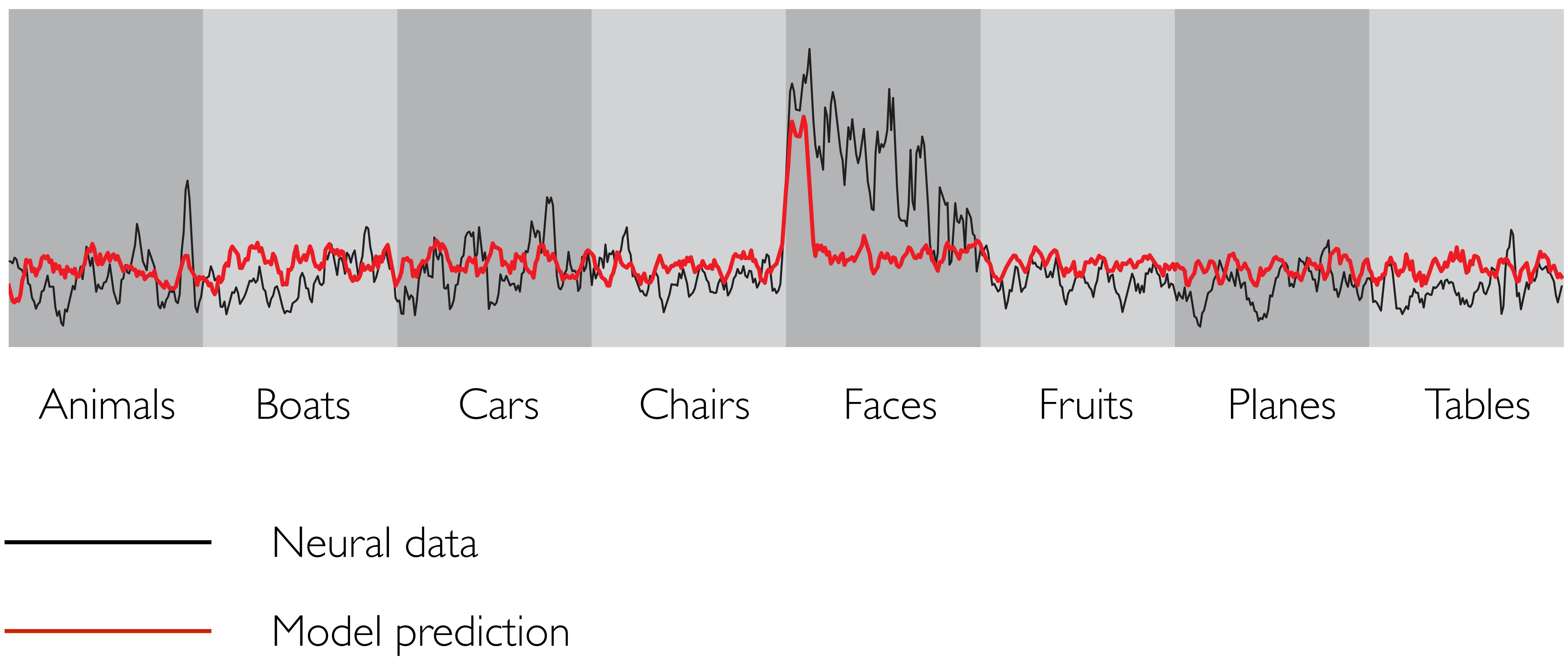
What about intermediate layers?

i. compare intermediate model layers to IT neural data

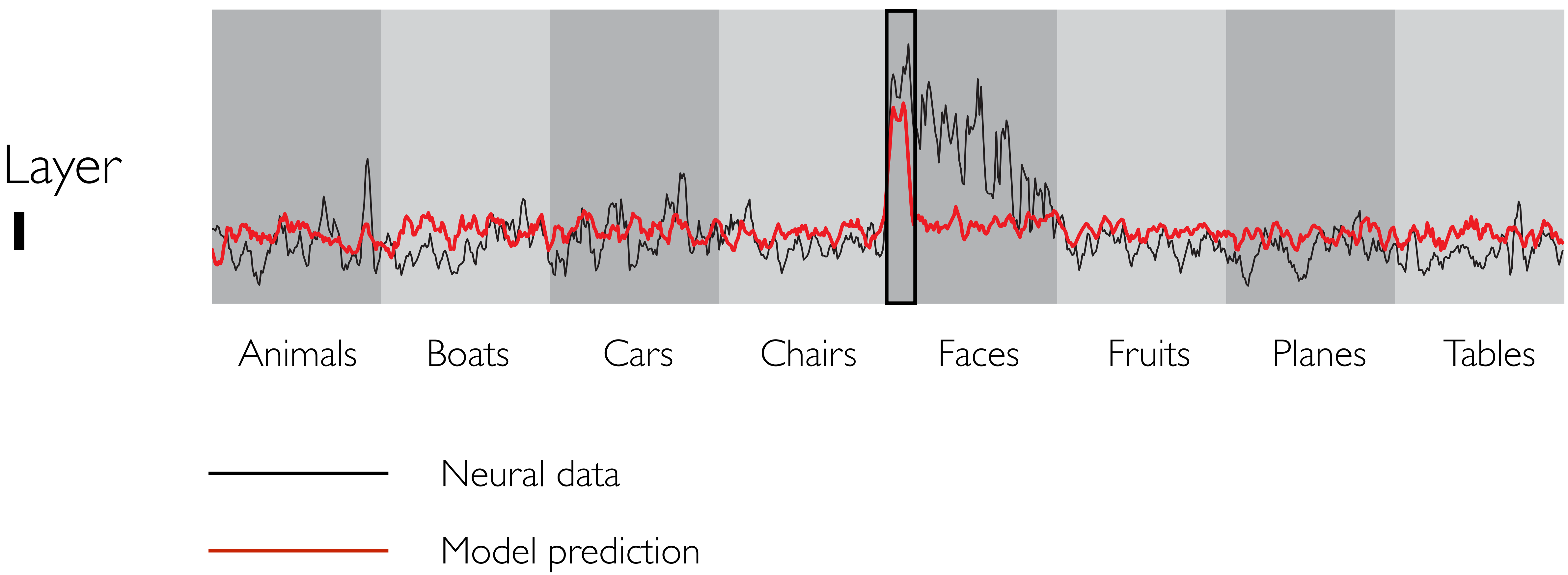
ii. compare all model layers to intermediate visual areas (V4)

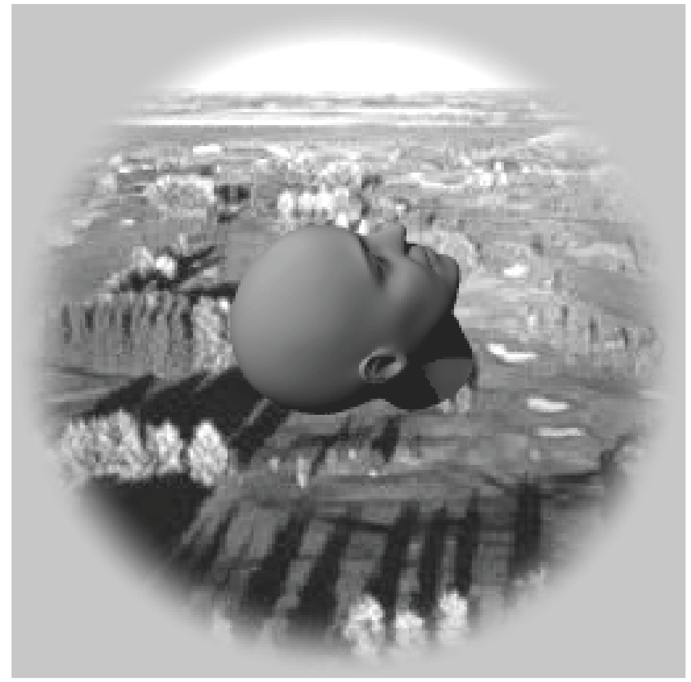
Captures low variation image response patterns …

Layer 1



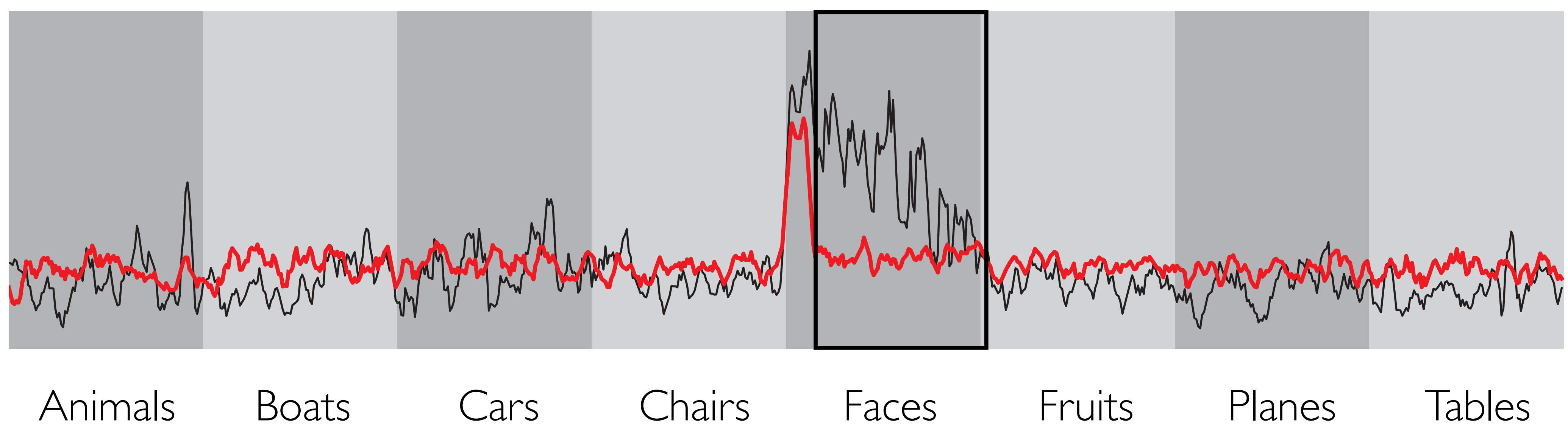
Captures low variation image response patterns …





Layer

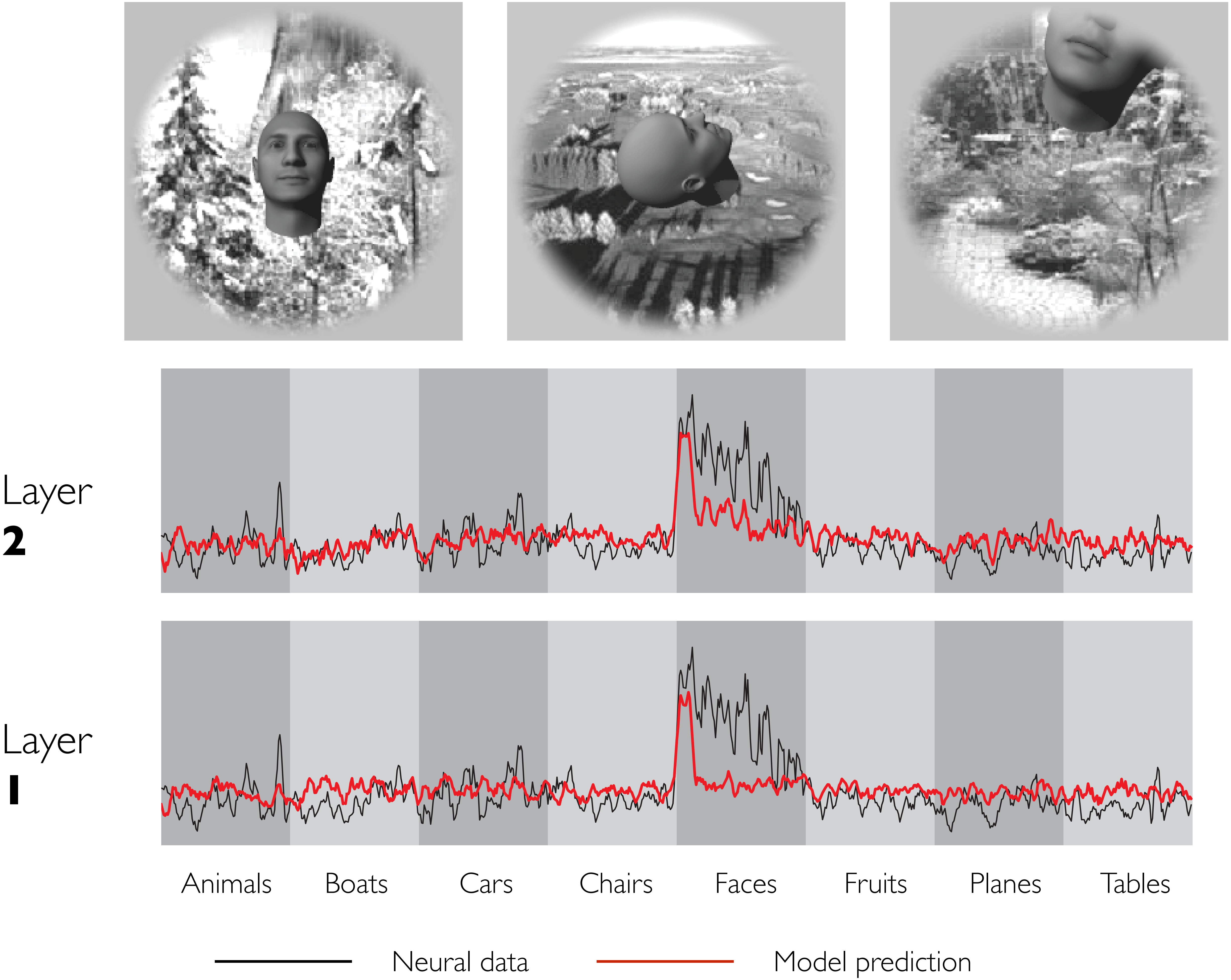
I

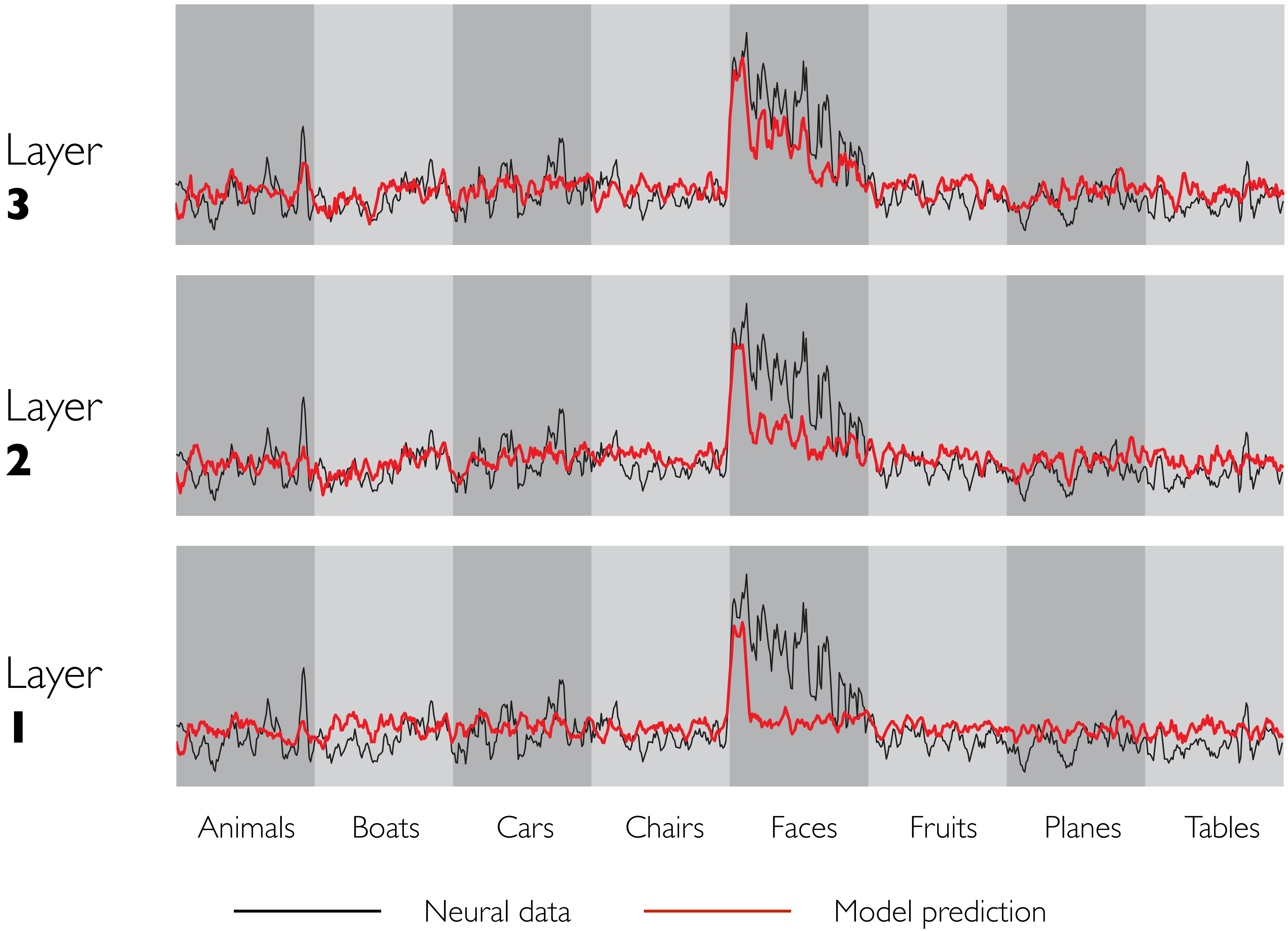


Neural data

Model prediction

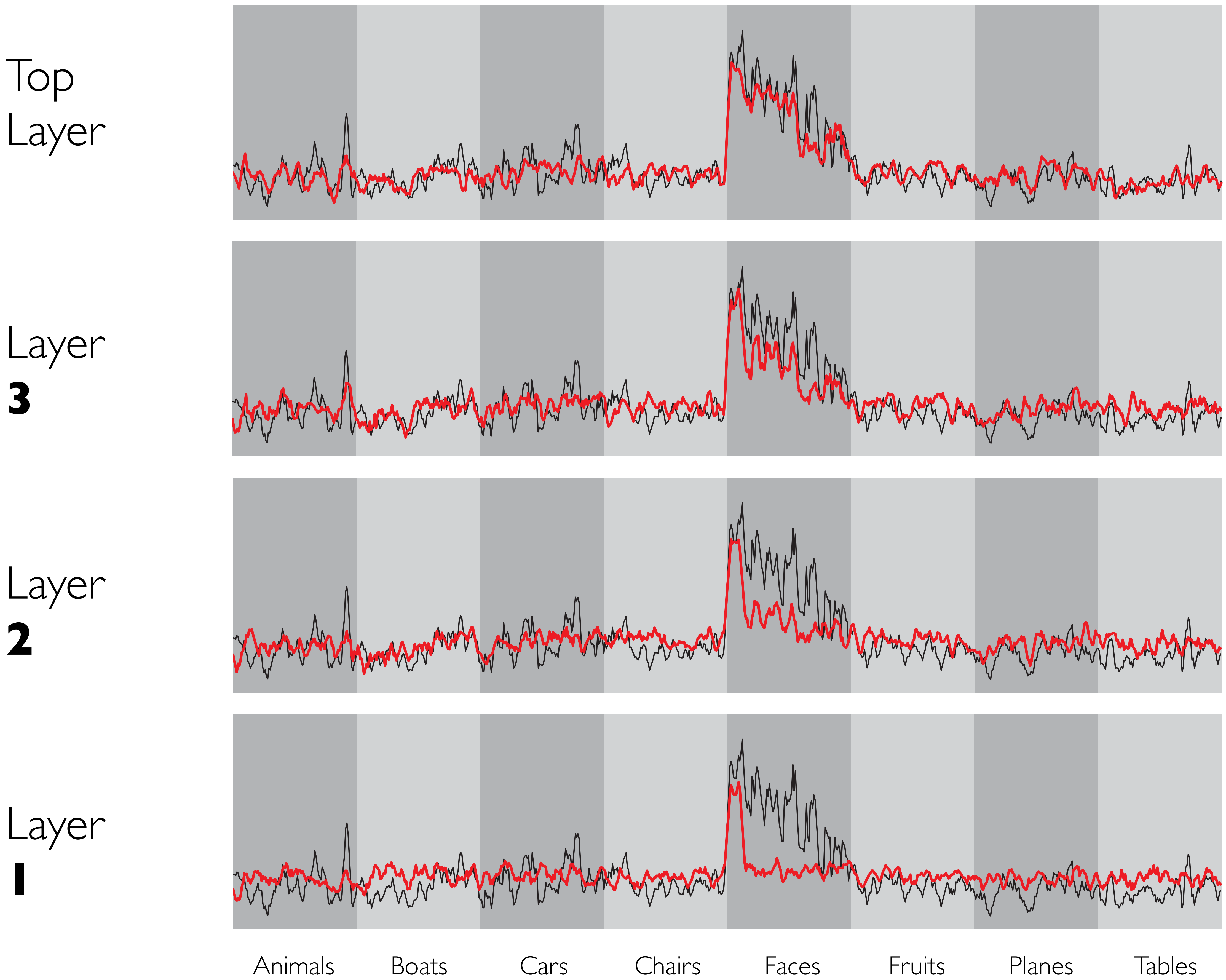
… but fails to capture higher variation response patterns.


Layer 2
Layer 1
Animals
Boats
Cars
Chairs
Faces
Fruits
Planes
Tables
Neural data
Model prediction



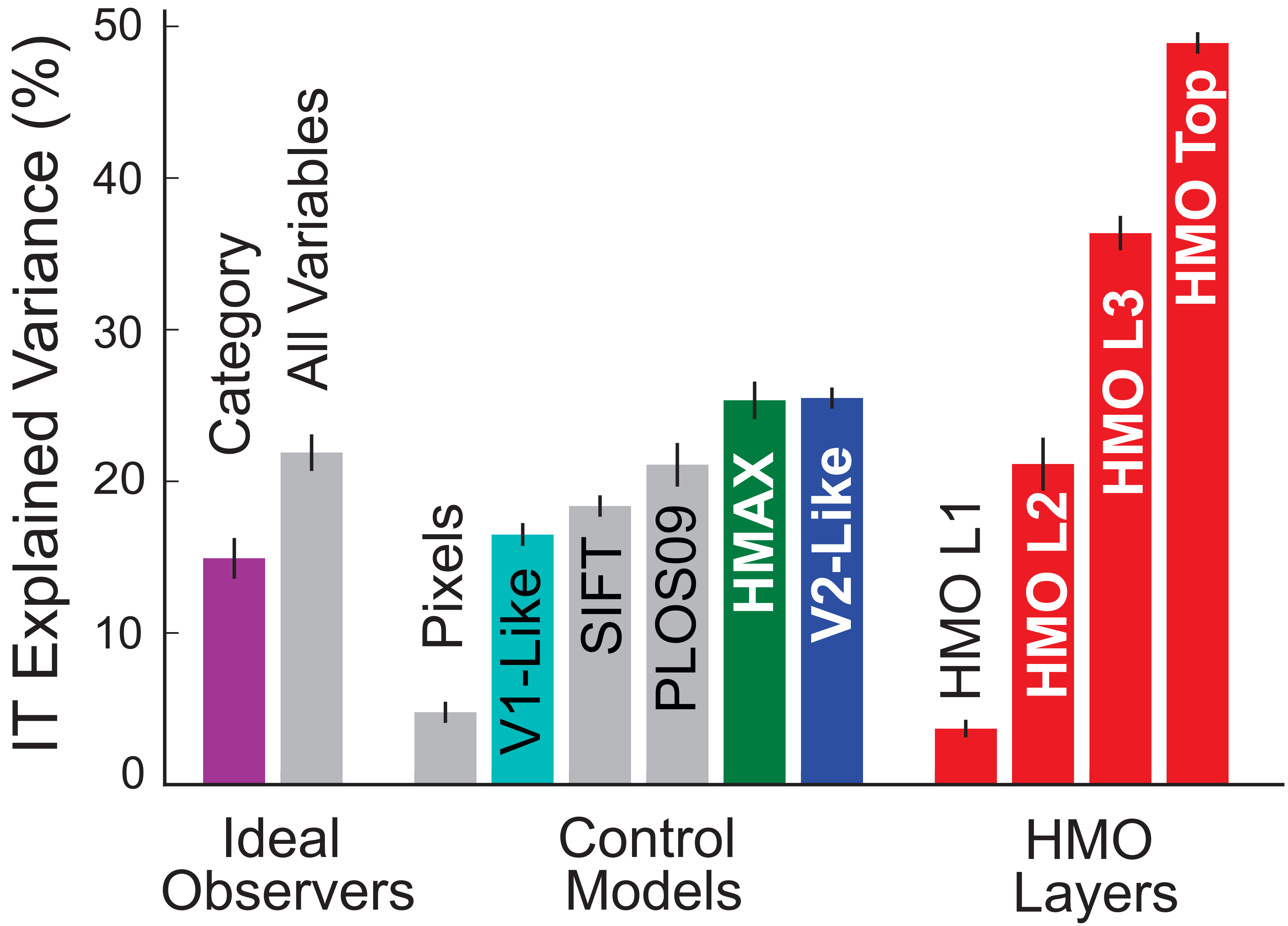
Layer 3
Layer 2
Layer 1
Animals
Boats
Cars
Chairs
Faces
Fruits
Planes
Tables
Neural data
Model prediction

Building tolerance while maintaining selectivity



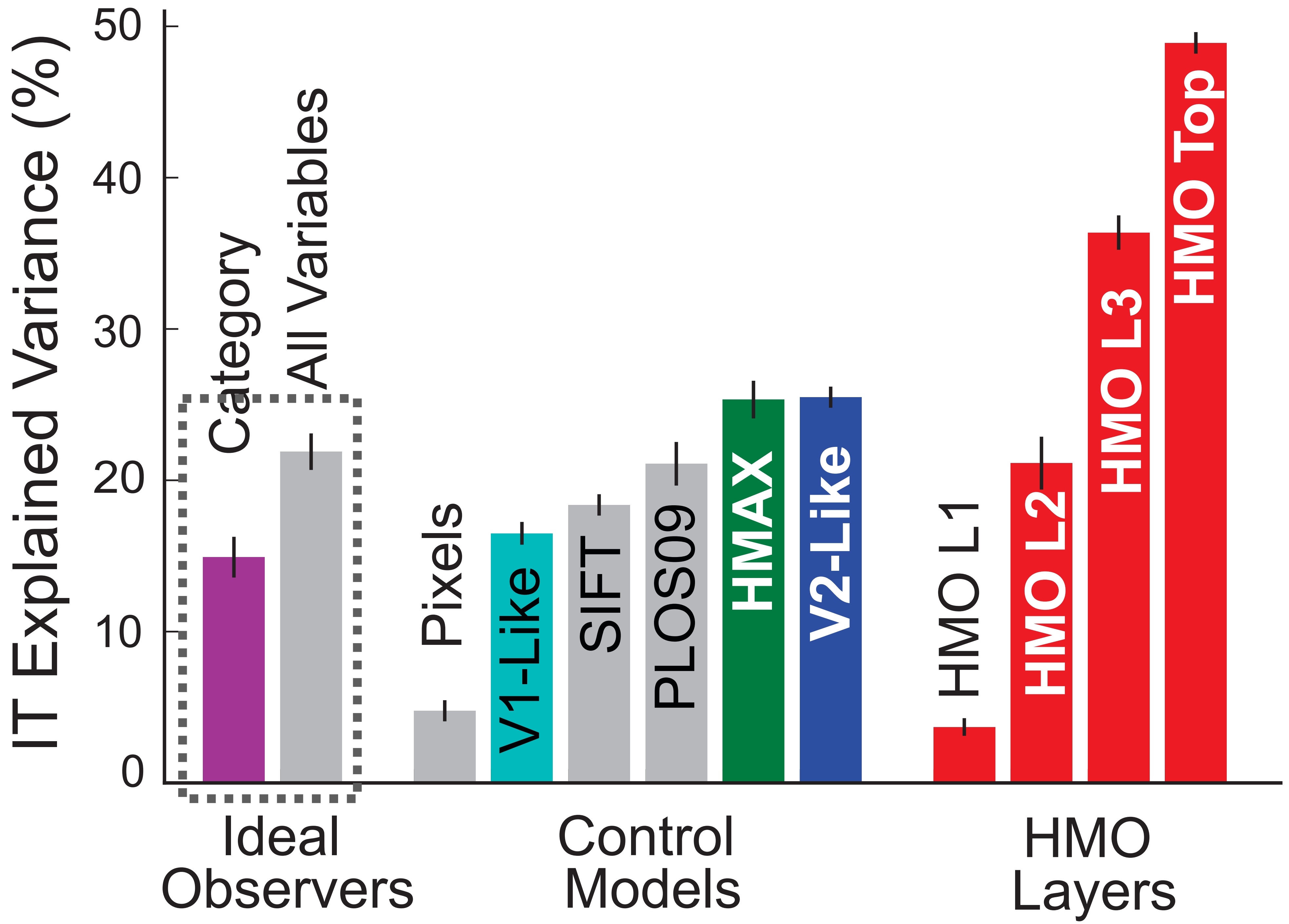
Predicting IT Neural Responses

Yamins and Hong* et. al.* ***PNAS*** *(2014)*



Predicting IT Neural Responses

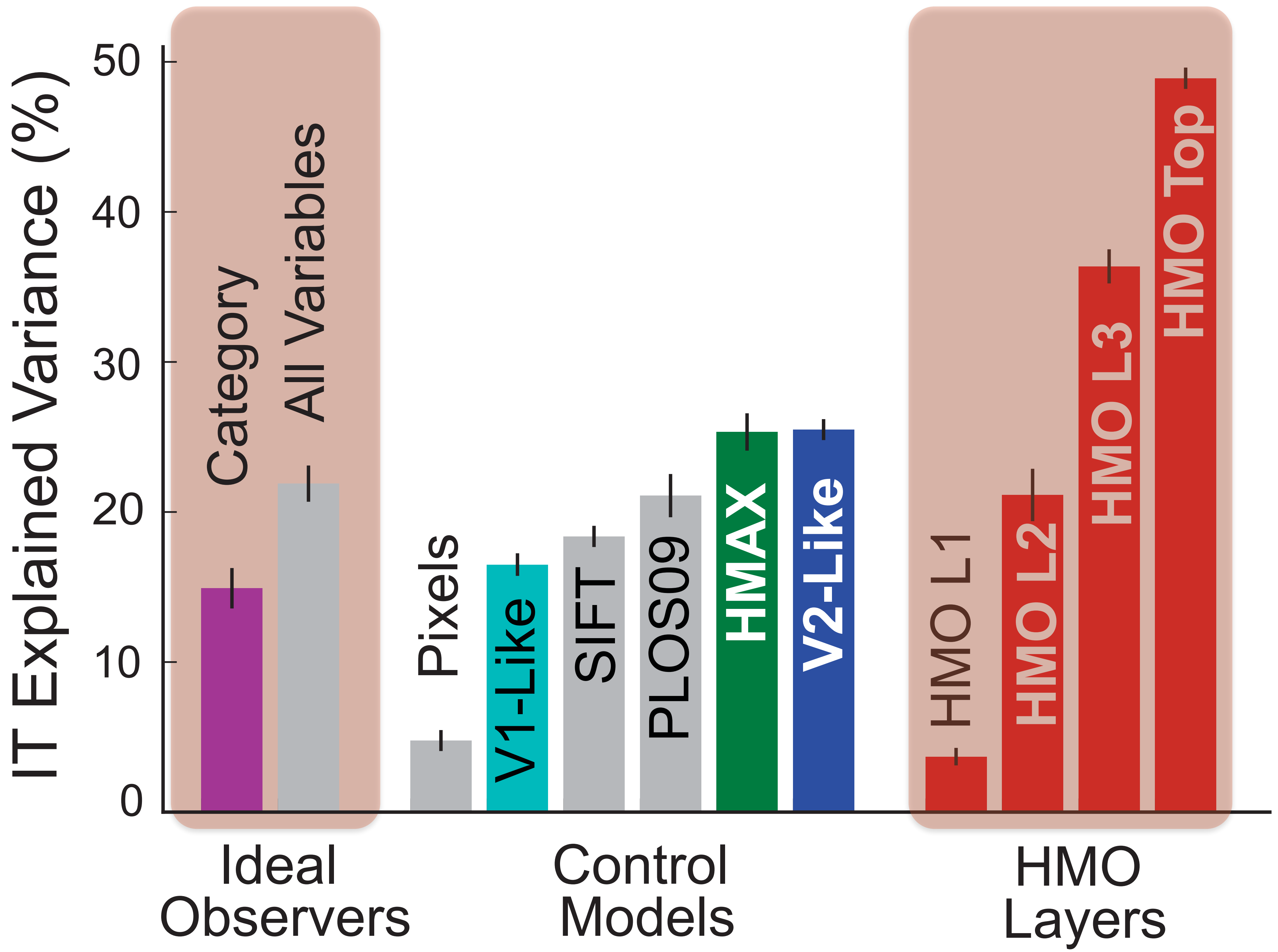
Yamins and Hong* et. al.* ***PNAS*** *(2014)*



Predicting IT Neural Responses

Yamins and Hong* et. al.* ***PNAS*** *(2014)*

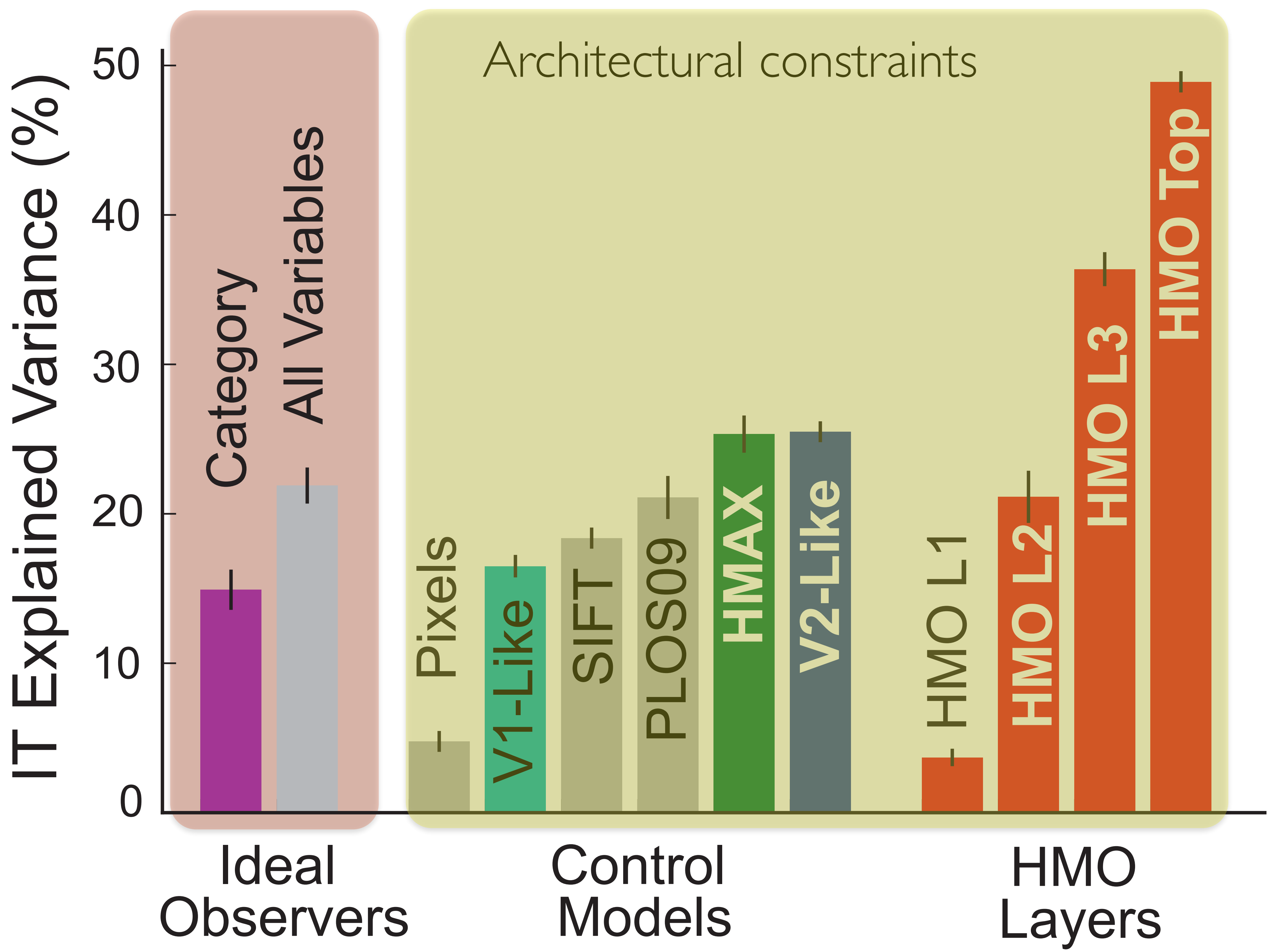
Performance constraints



Predicting IT Neural Responses

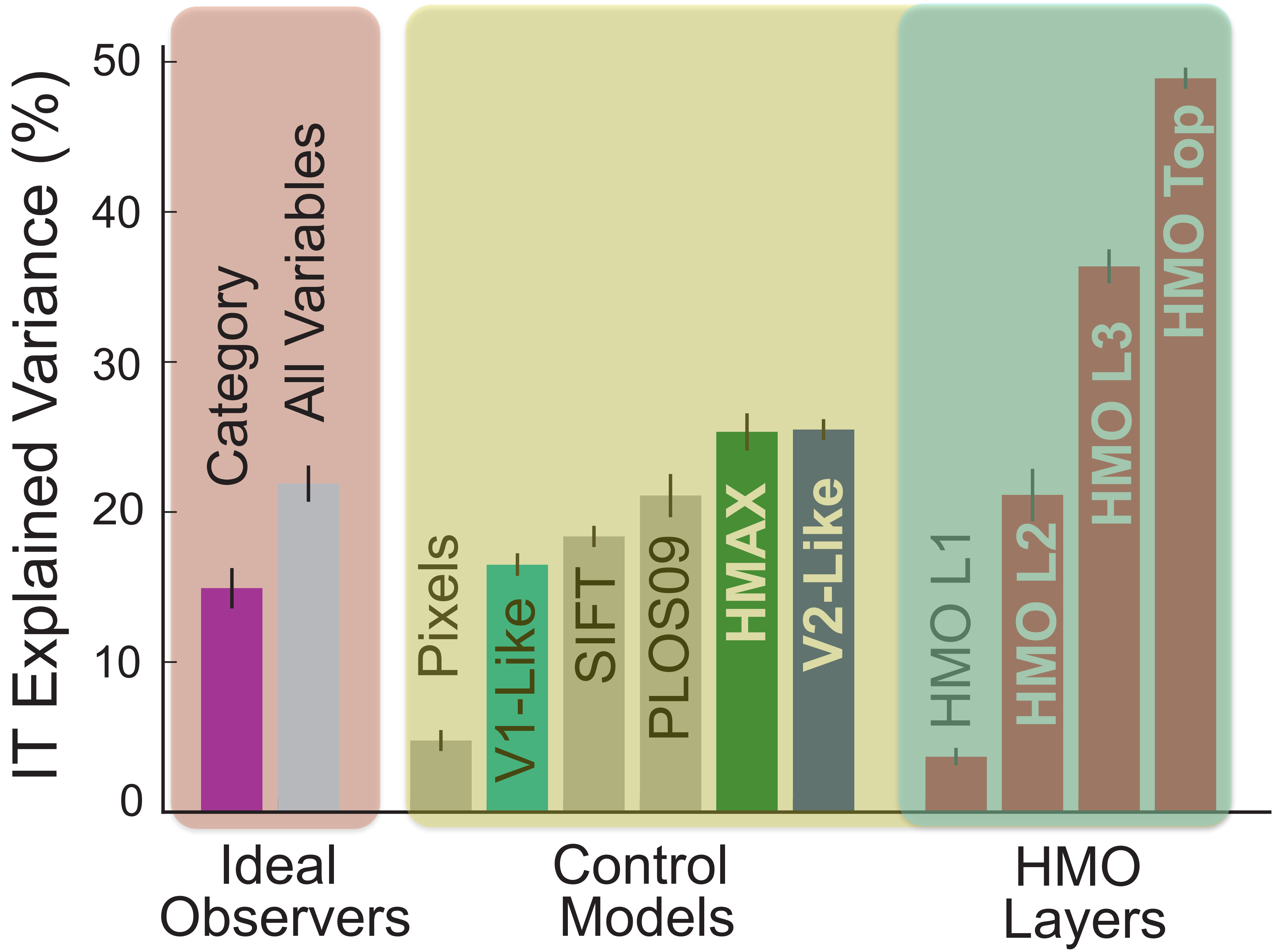
Yamins and Hong* et. al.* ***PNAS*** *(2014)*

Performance constraints



Predicting IT Neural Responses

Performance constraints + architectural constraints → better neural prediction



Predicting IT Neural Responses

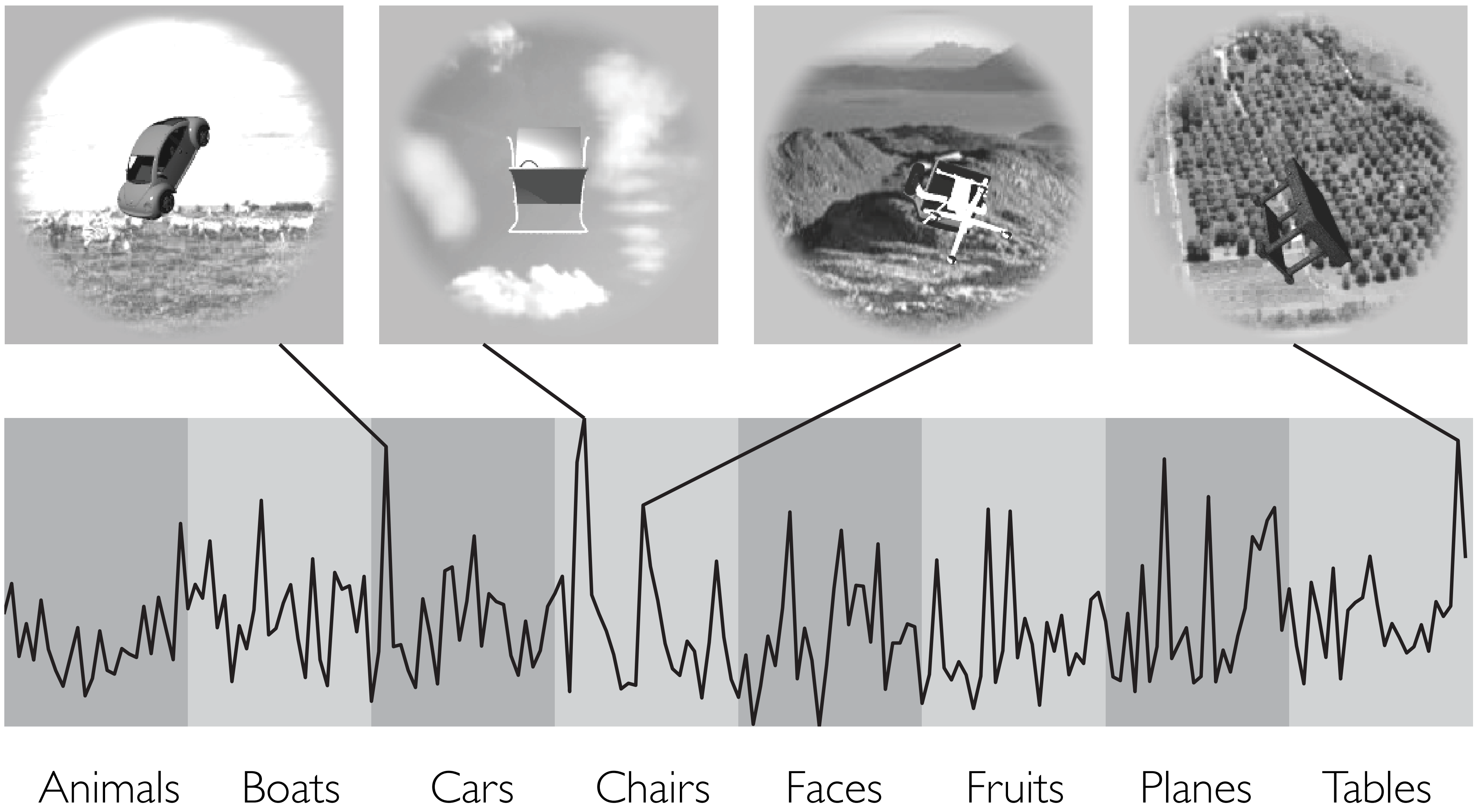
What about intermediate layers?

i. compare intermediate model layers to IT neural data

ii. compare all model layers to intermediate visual areas (V4)

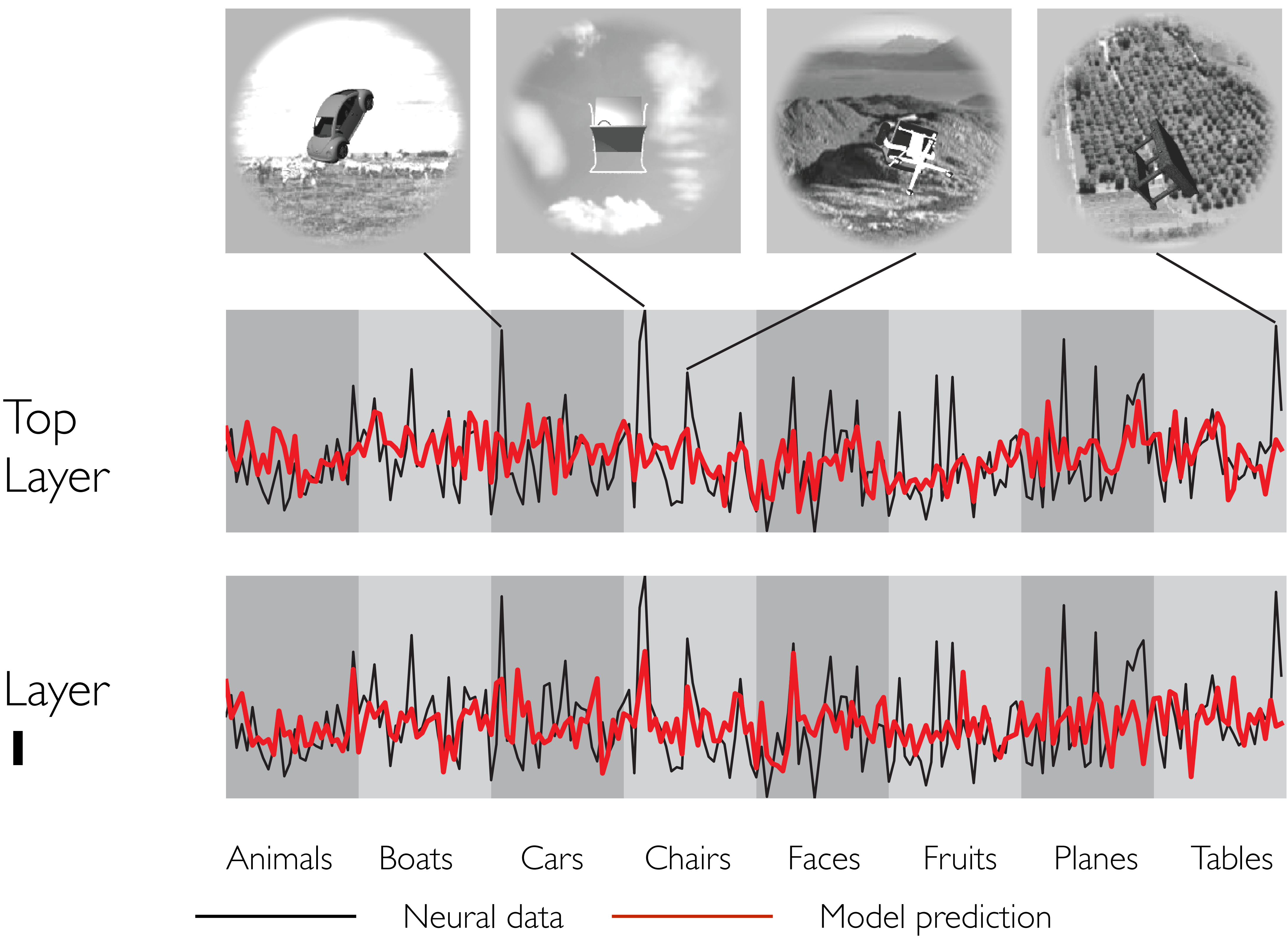
Predicting V4 Neural Responses

V4 unit 60

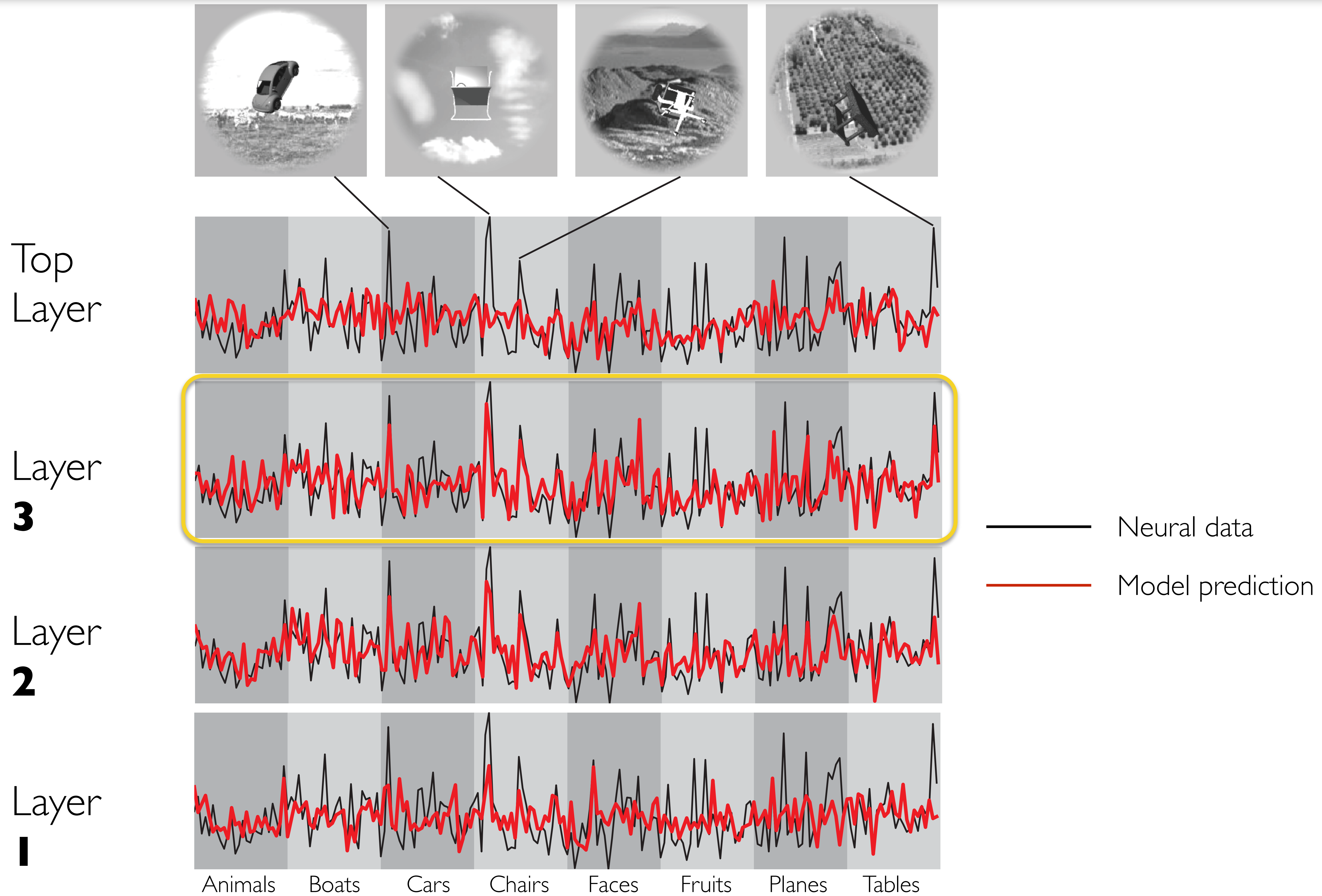


Predicting V4 Neural Responses

V4 unit 60

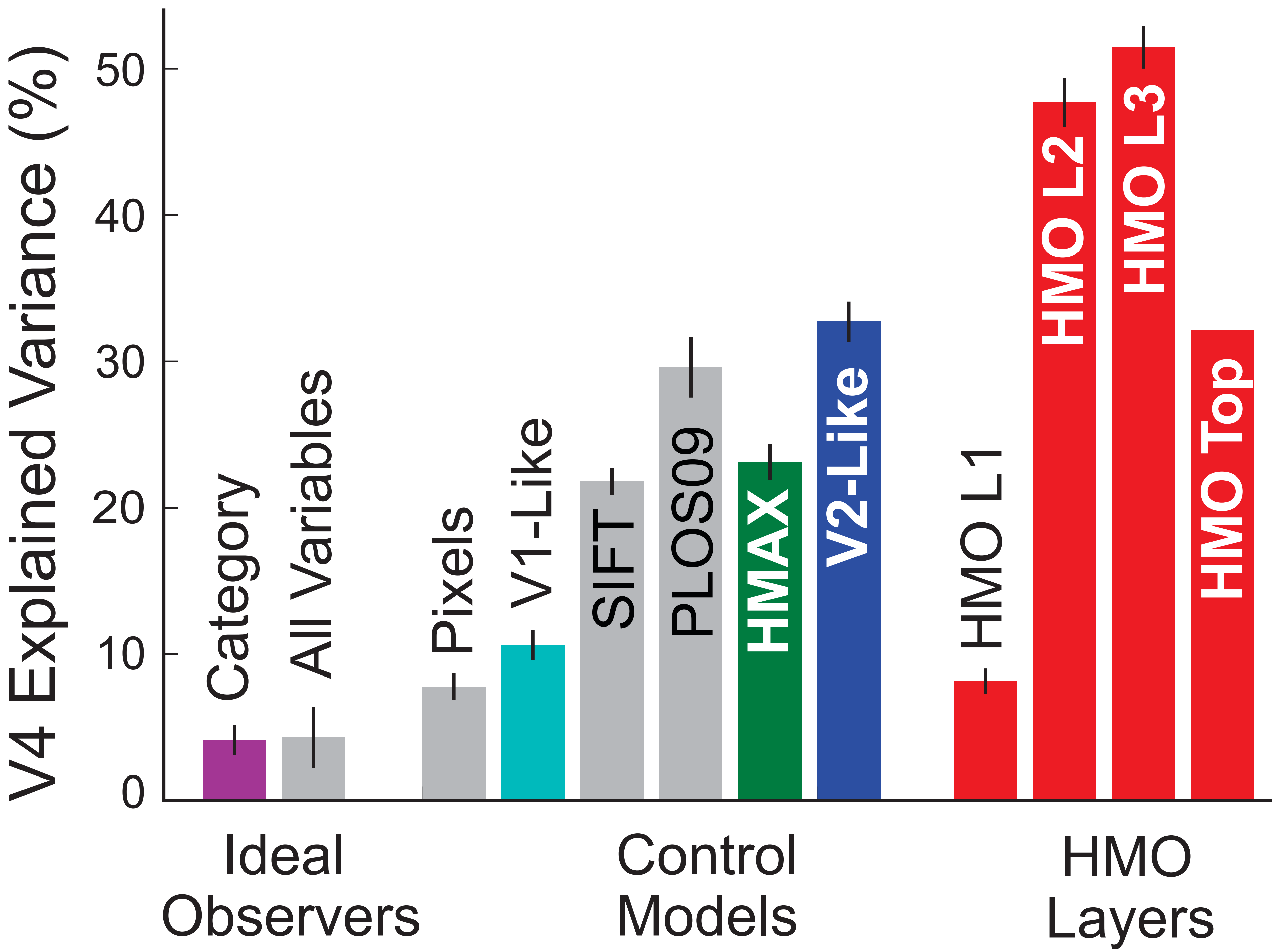


Predicting V4 Neural Responses



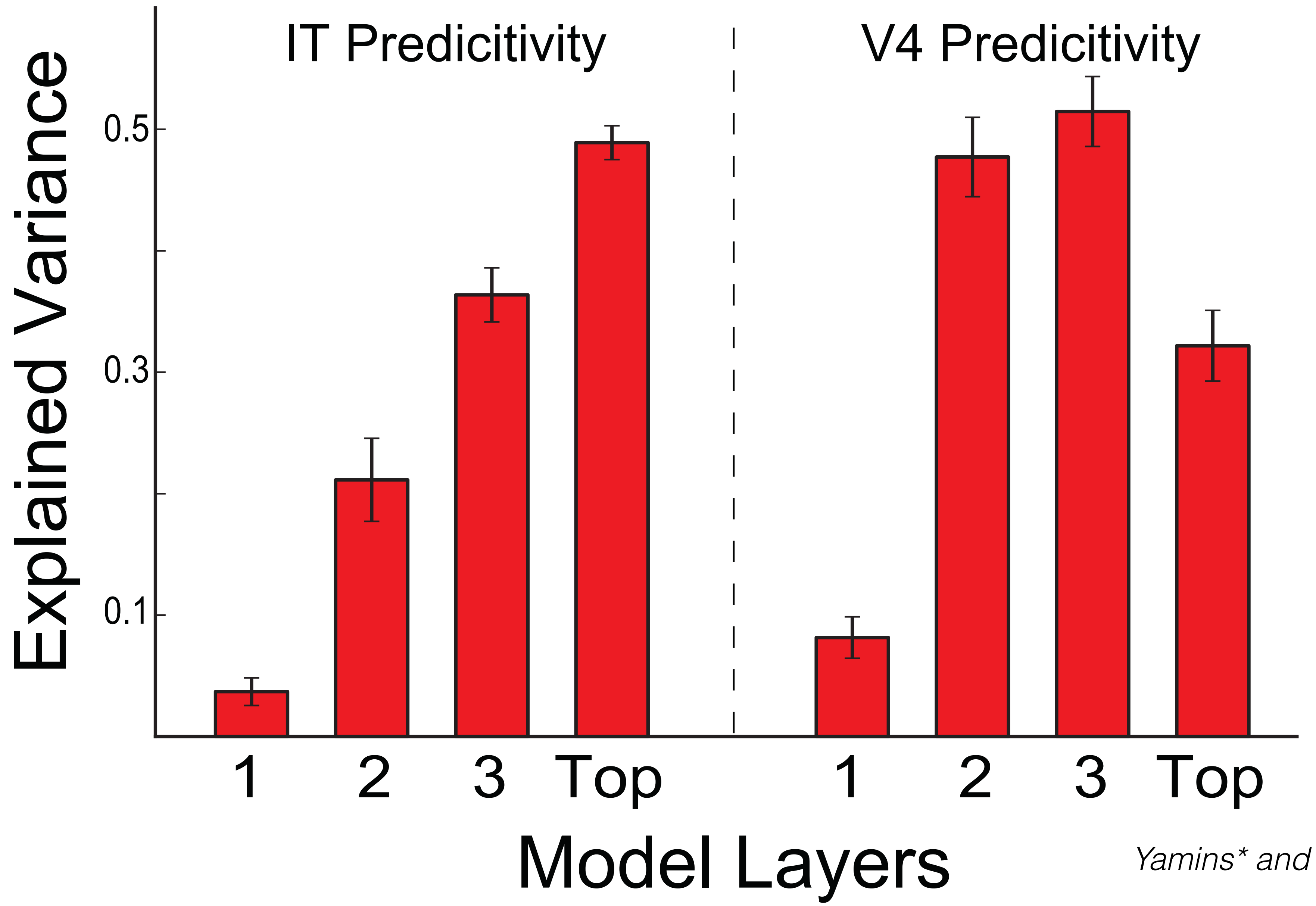
Predicting V4 Neural Responses

Yamins and Hong* et. al.* ***PNAS*** *(2014)*



Layer-area correspondence

Investigating fits as a function of model layer:

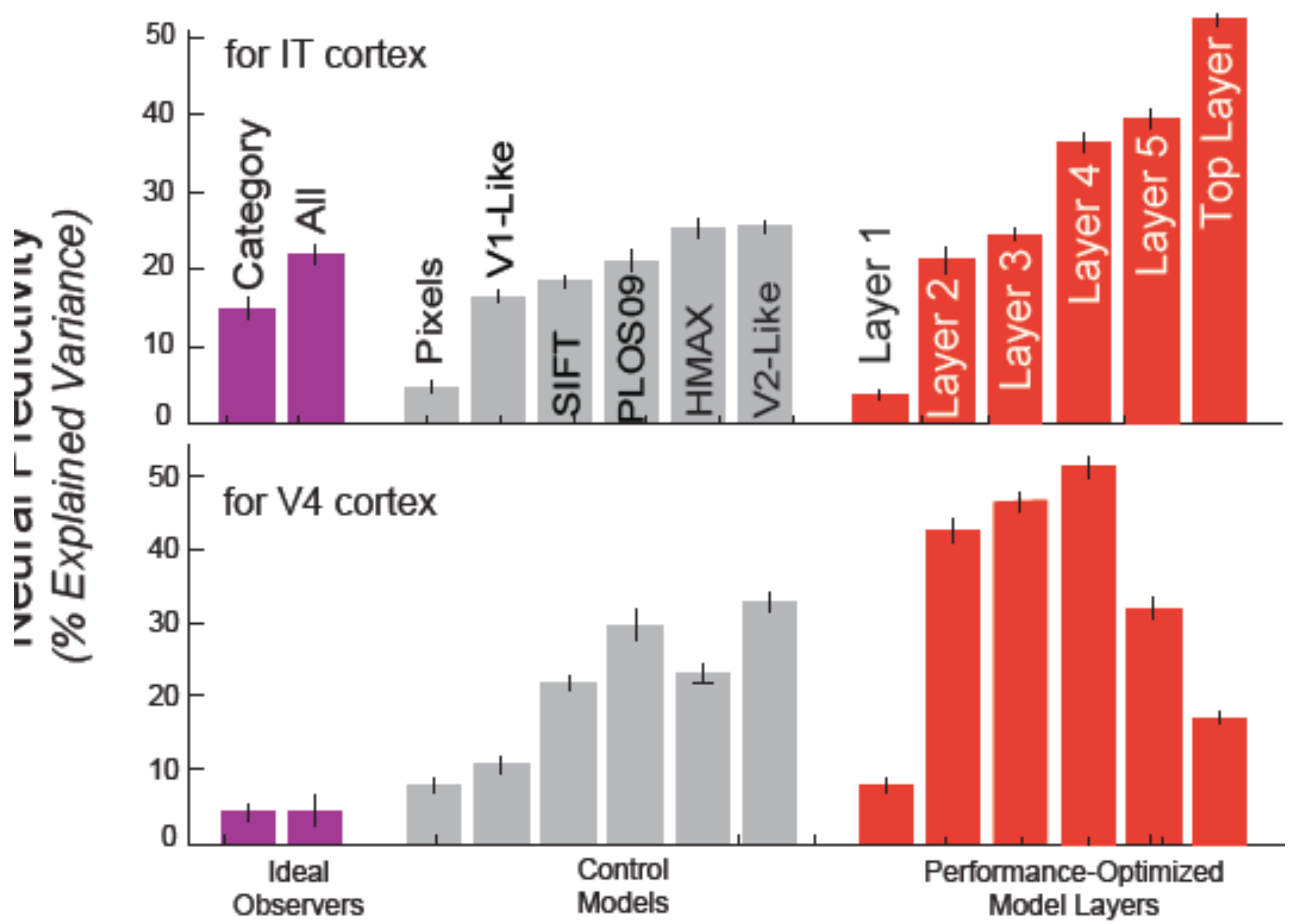


Yamins and Hong* et. al.* ***PNAS*** *(2014)*

IT fit increases at each layer. In contrast, V4 fit peaks and then goes down.

Layer-area correspondence

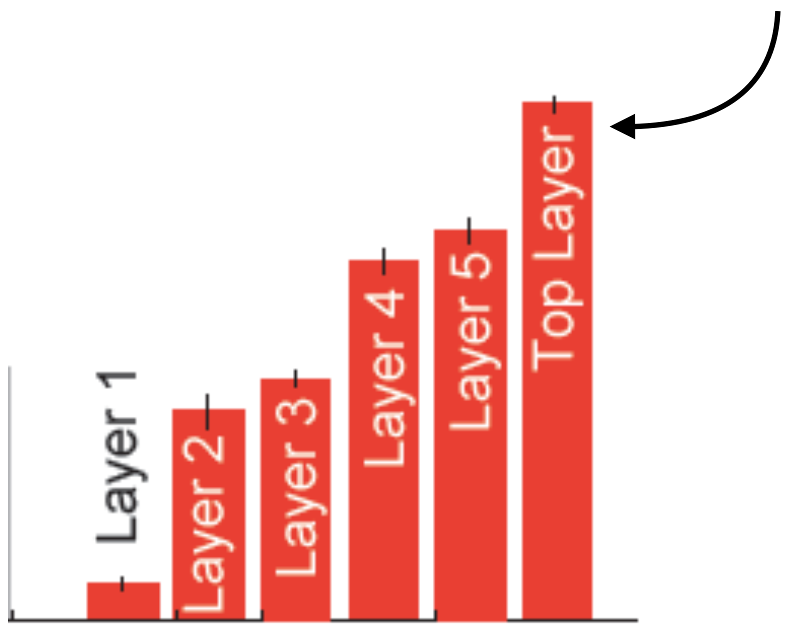
Nothing special about 4 layers — deeper models can be better:



Hong and Yamins*et. al.*
Nature Neuroscience
(2016)

Layer-area correspondence

Top **hidden** layer (***not*** explicit categorization layer)

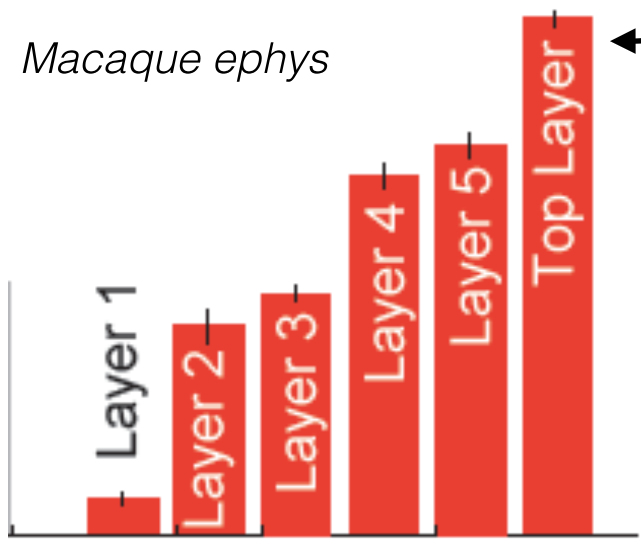


Hong and Yamins*et. al.*
Nature Neuroscience
(2016)

Layer-area correspondence

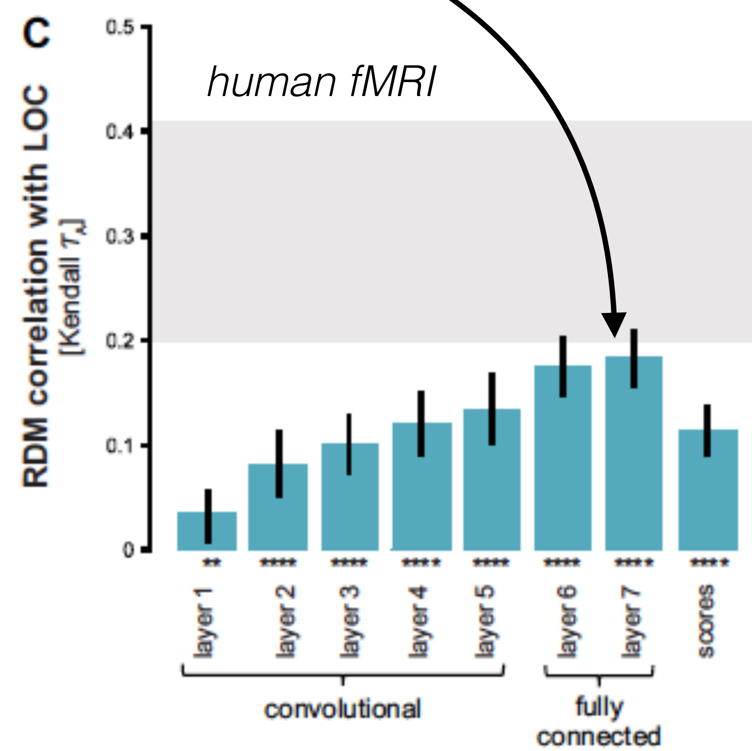
Top **hidden** layer (***not*** explicit categorization layer)

Macaque ephys



Hong and Yamins*et. al.*
Nature Neuroscience
(2016)

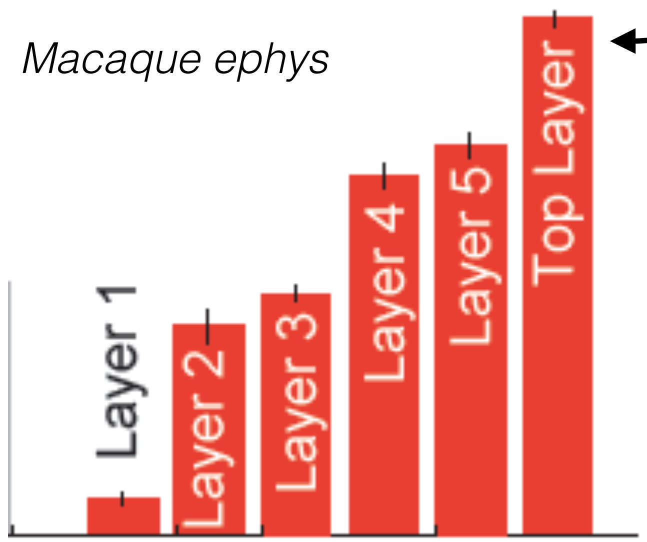
human fMRI



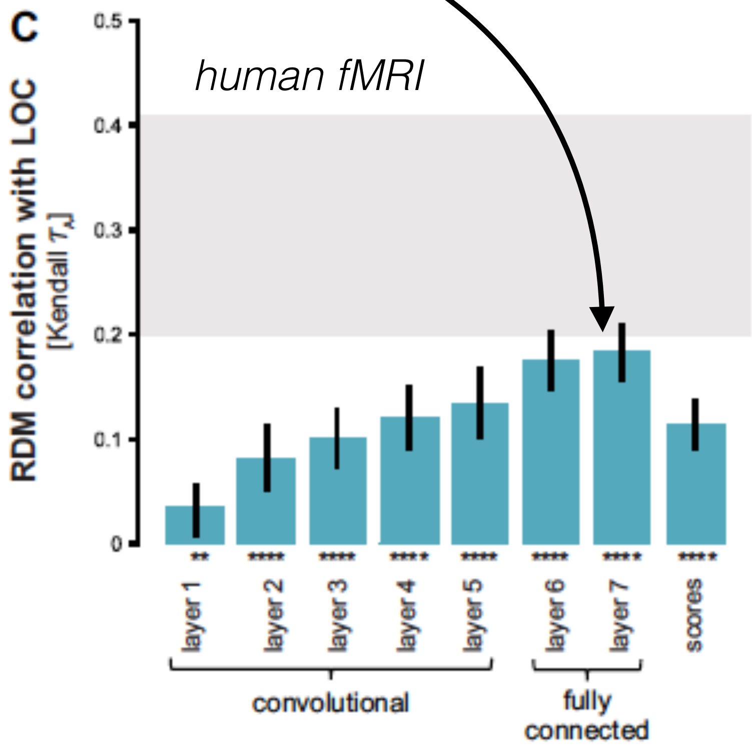
Khaligh-Razavi & Kriegestkorte (2014)

Layer-area correspondence

Top **hidden** layer (***not*** explicit categorization layer)



Hong and Yamins*et. al.*
Nature Neuroscience
(2016)

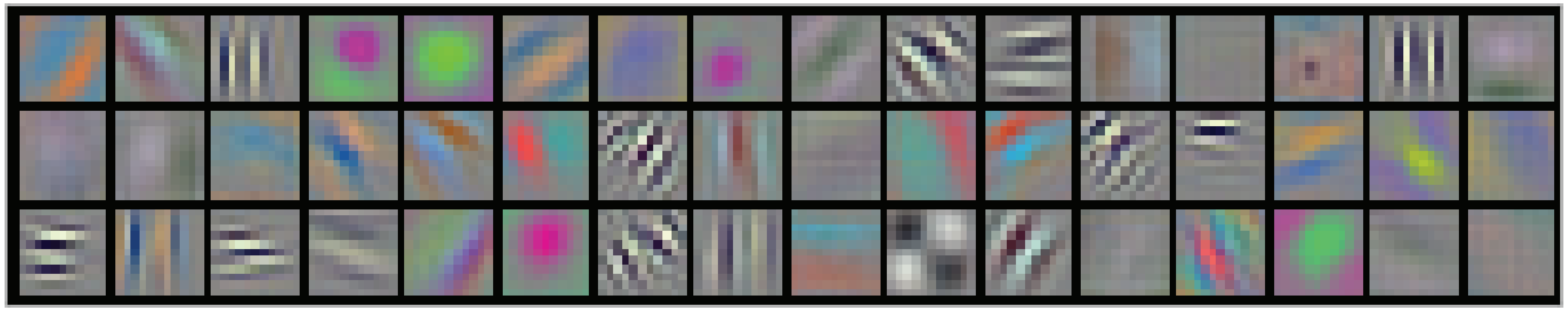


Khaligh-Razavi & Kriegestkorte (2014)

*Best recent models: ~**13** layers deep, with IT best predicted around ~**80%** of the way through (e.g. 10 layers)*

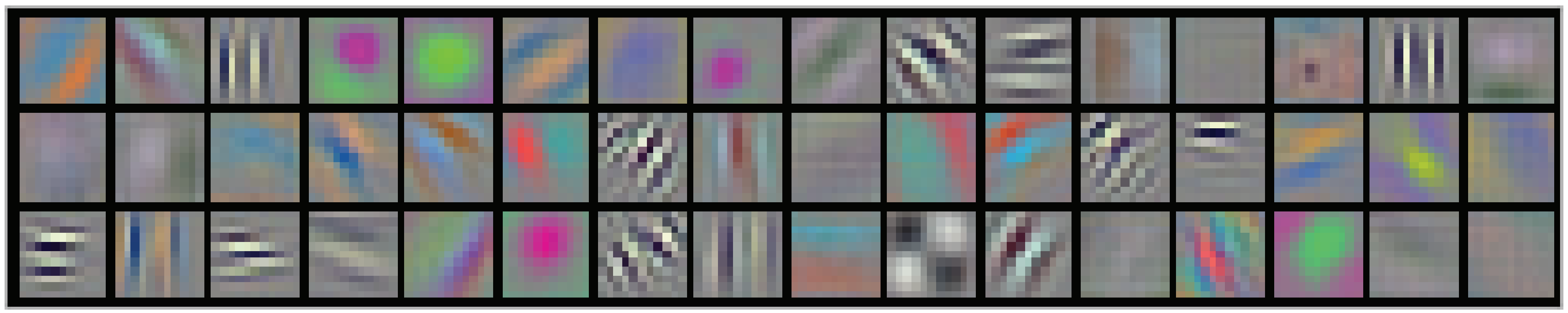
Layer-area correspondence

Emergently, AlexNet filters at lowest layer resemble Gabor wavelets:

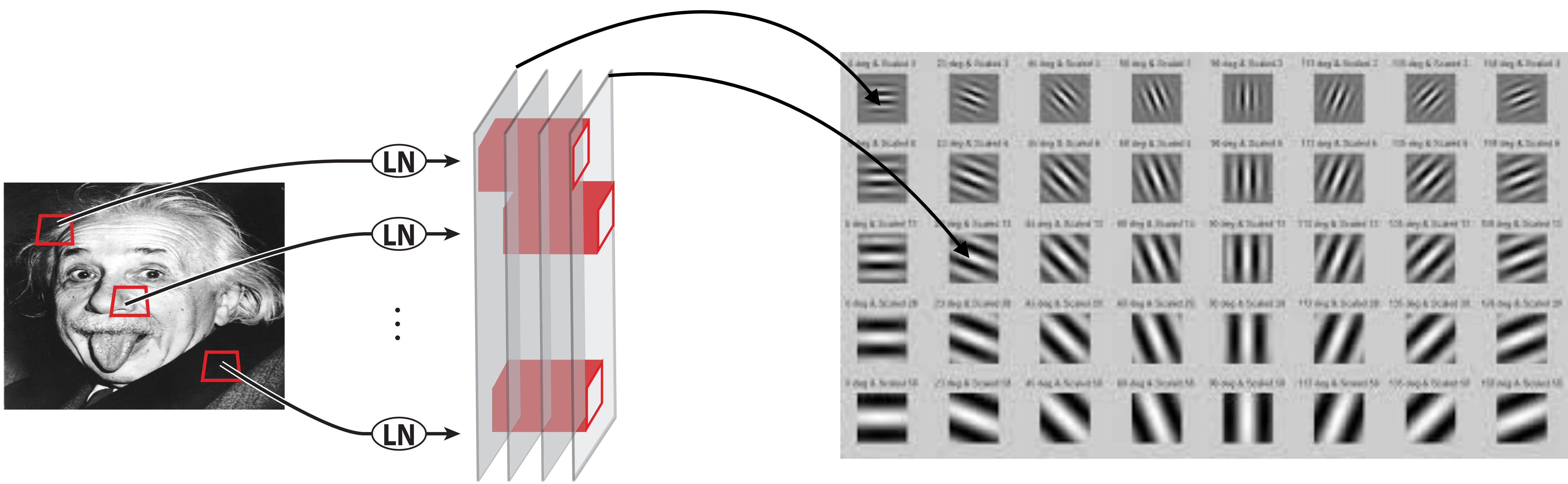


Layer-area correspondence

Emergently, AlexNet filters at lowest layer resemble Gabor wavelets:



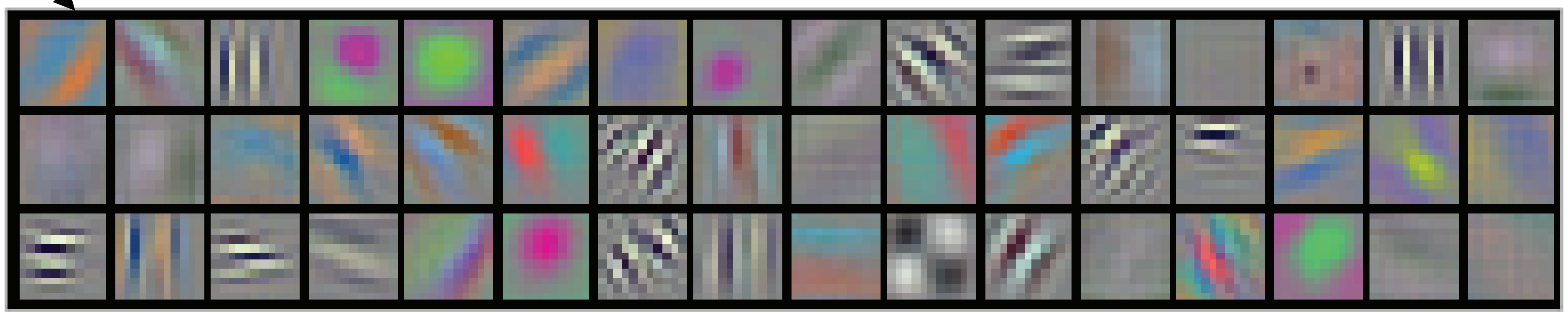
Compare to:



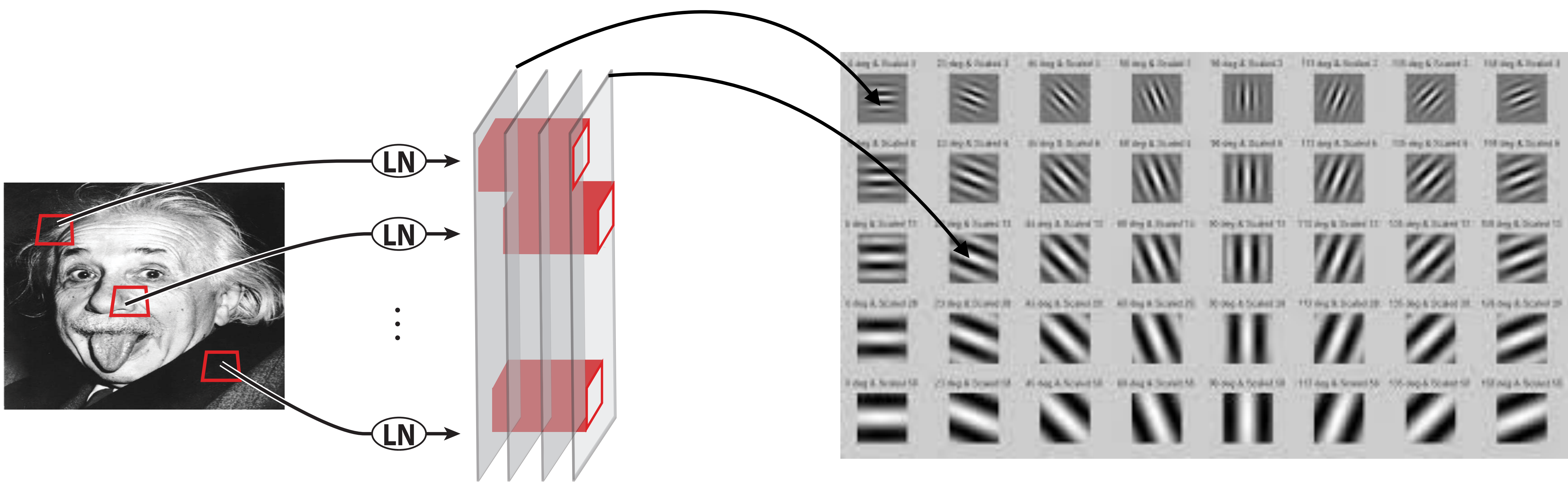
Layer-area correspondence

Emergently, AlexNet filters at lowest layer resemble Gabor wavelets:

actually, this is "better" than Gabor model b/c it naturally has "color opponency"

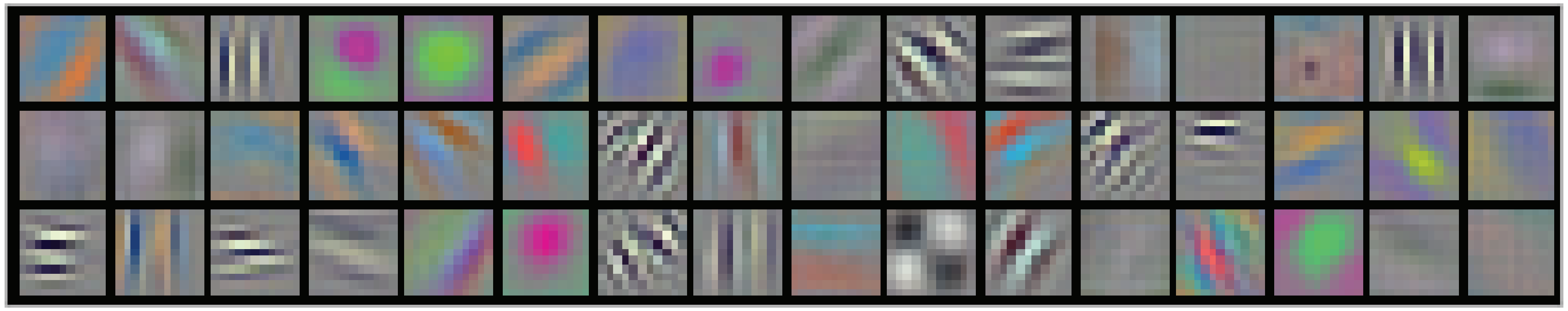


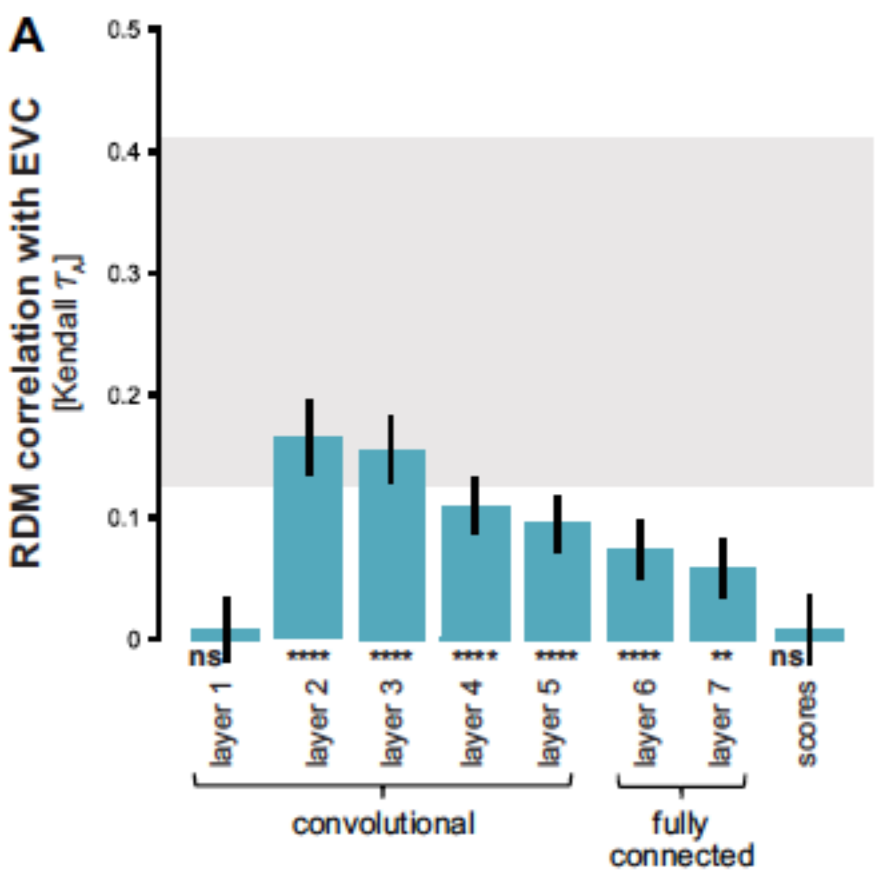
Compare to:



Layer-area correspondence

Emergently, AlexNet filters at lowest layer resemble Gabor wavelets:





Model early layers are best explanation of fMRI data in V1. (*with Darren Seibert and Justin Gardner)*

Kaligh-Razavi and Kriegeskorte (2014)

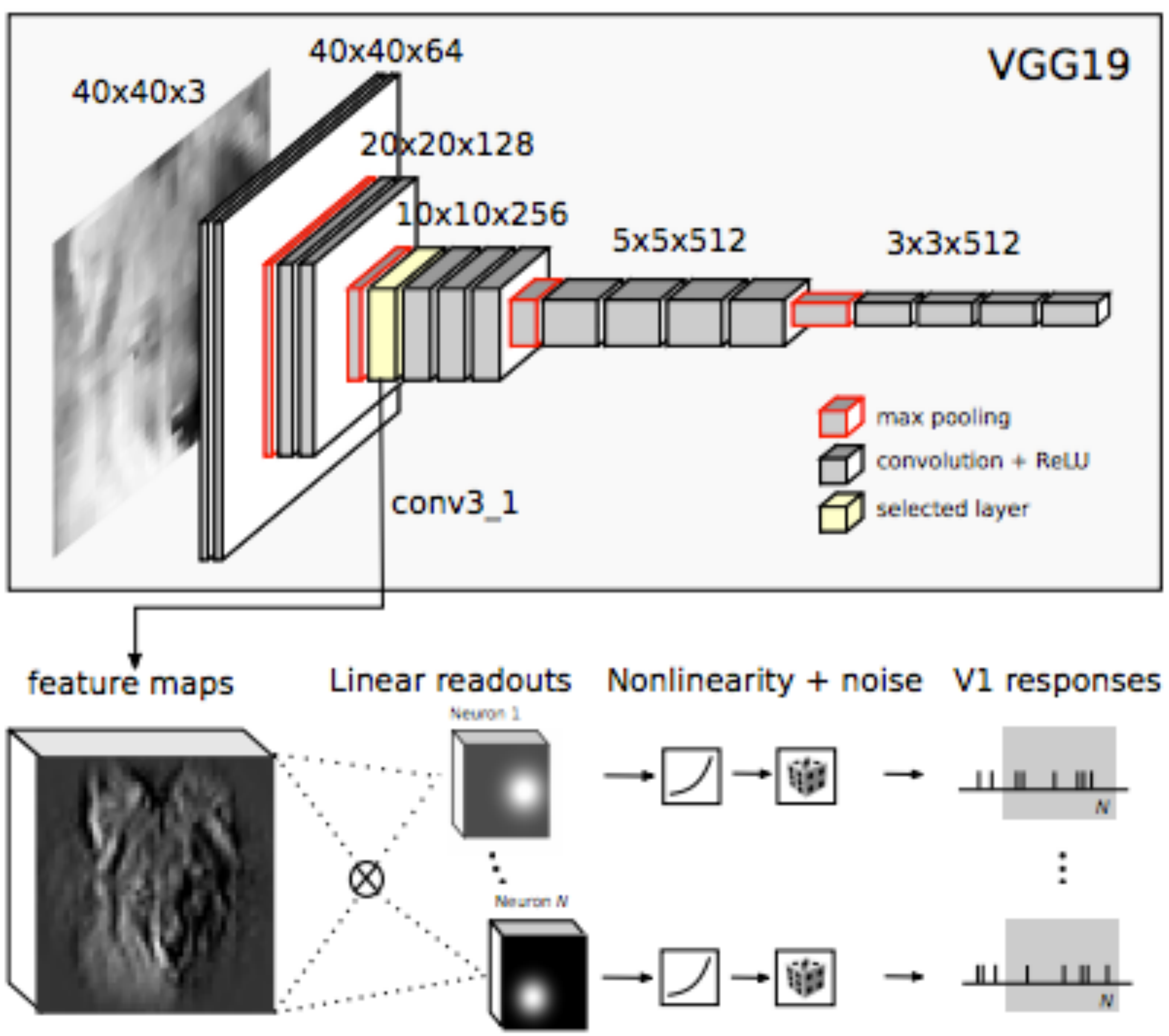
Similar result: Guclu & Van Gerven (2015)

Layer-area correspondence

Deep convolutional models improve predictions of macaque V1 responses to natural images

Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, Alexander S Ecker

doi: https://doi.org/ 10.1101/201764

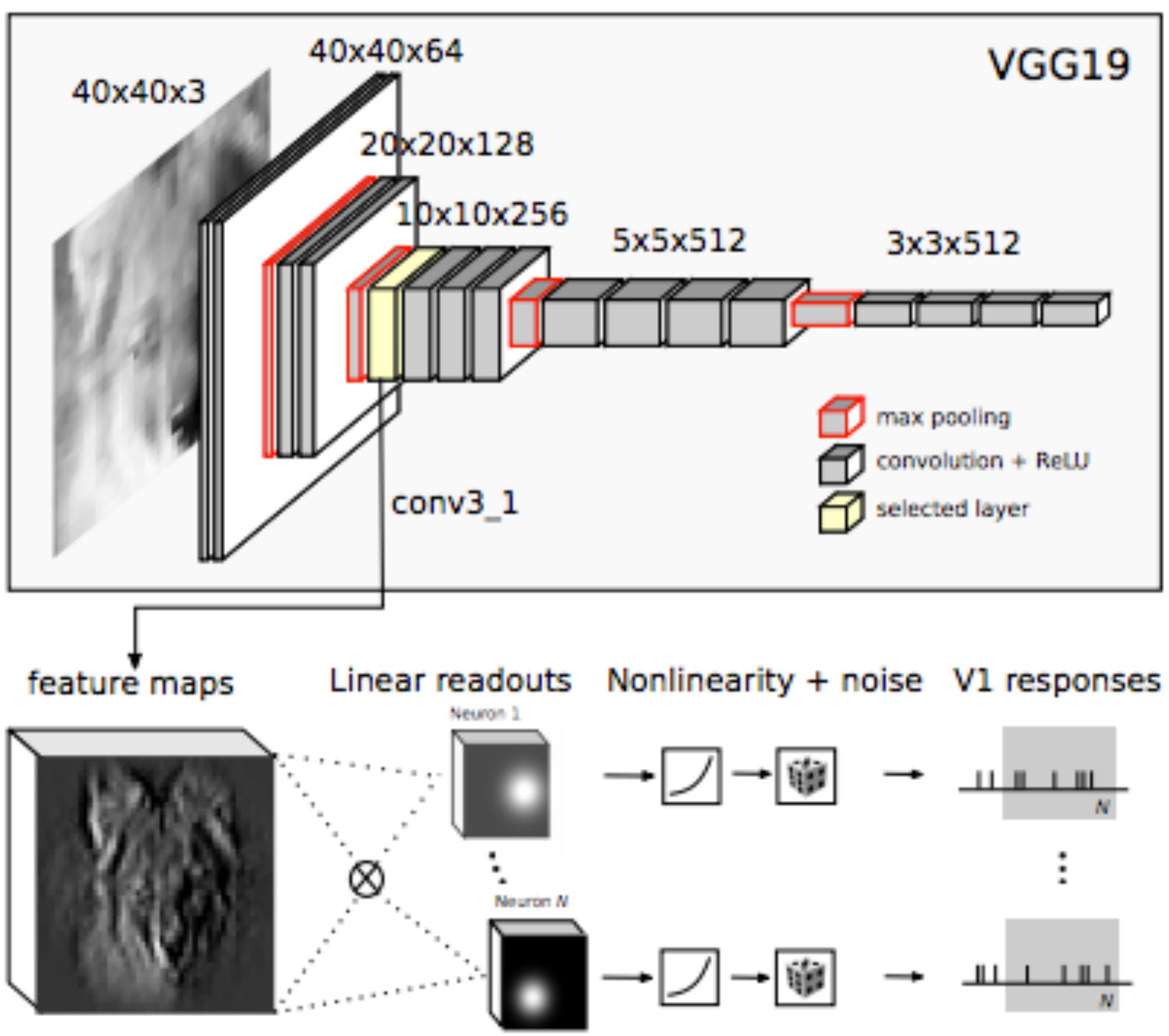


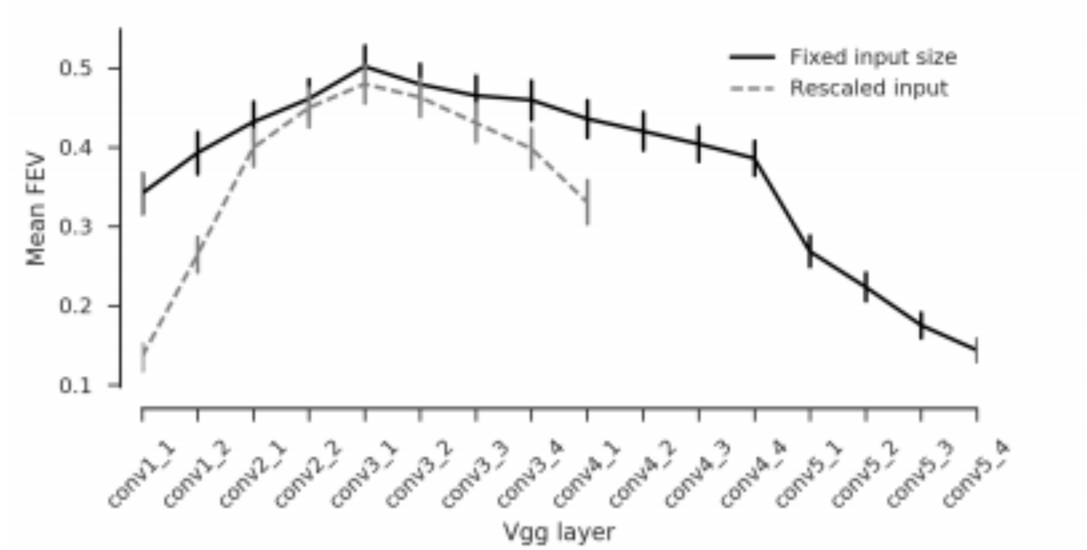
Layer-area correspondence

Deep convolutional models improve predictions of macaque V1 responses to natural images

Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, Alexander S Ecker

doi: https://doi.org/ 10.1101/201764



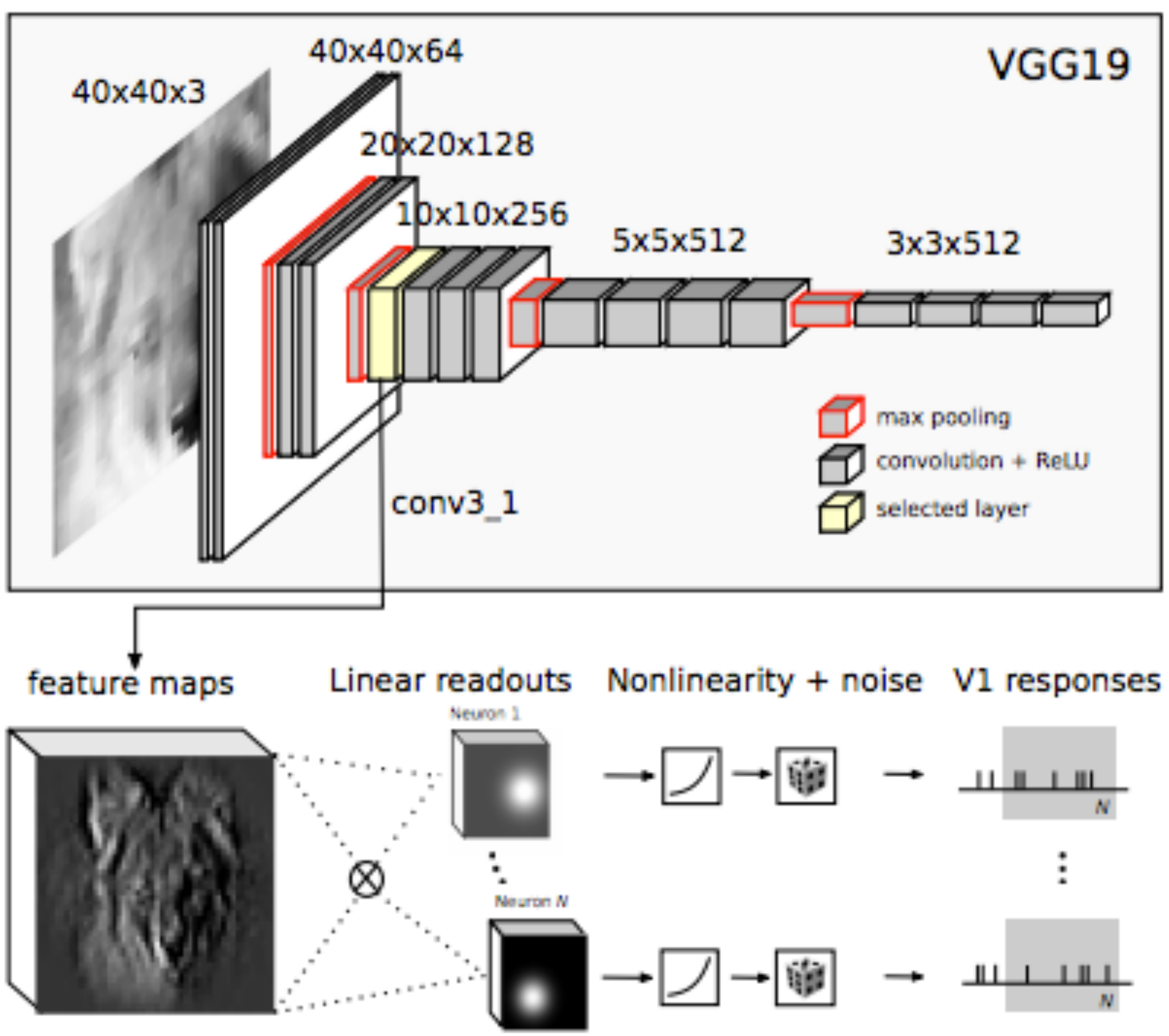


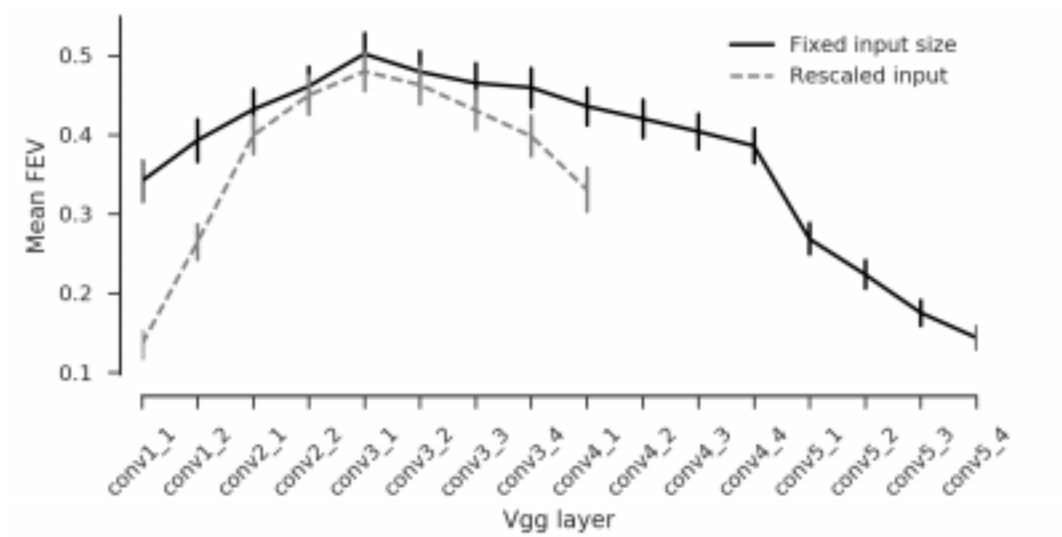
Layer-area correspondence

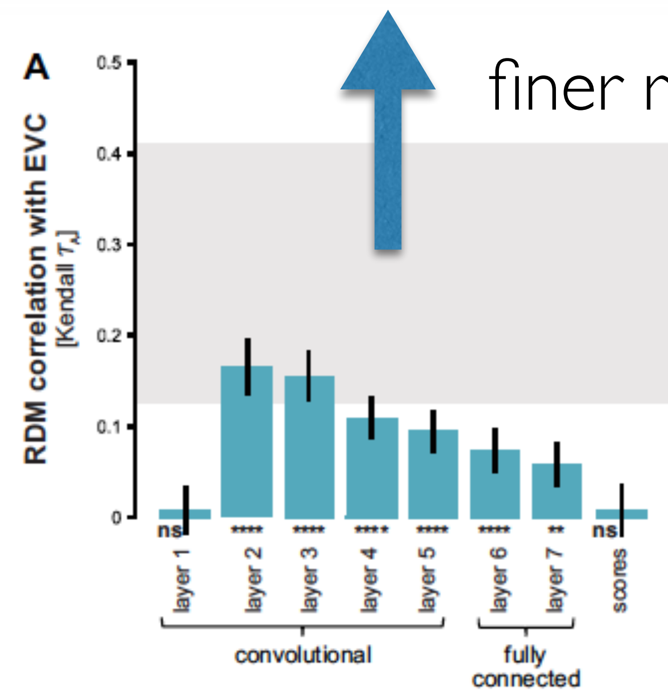
Deep convolutional models improve predictions of macaque V1 responses to natural images

Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, Alexander S Ecker

doi: https://doi.org/ 10.1101/201764







finer resolution

Layer-area correspondence

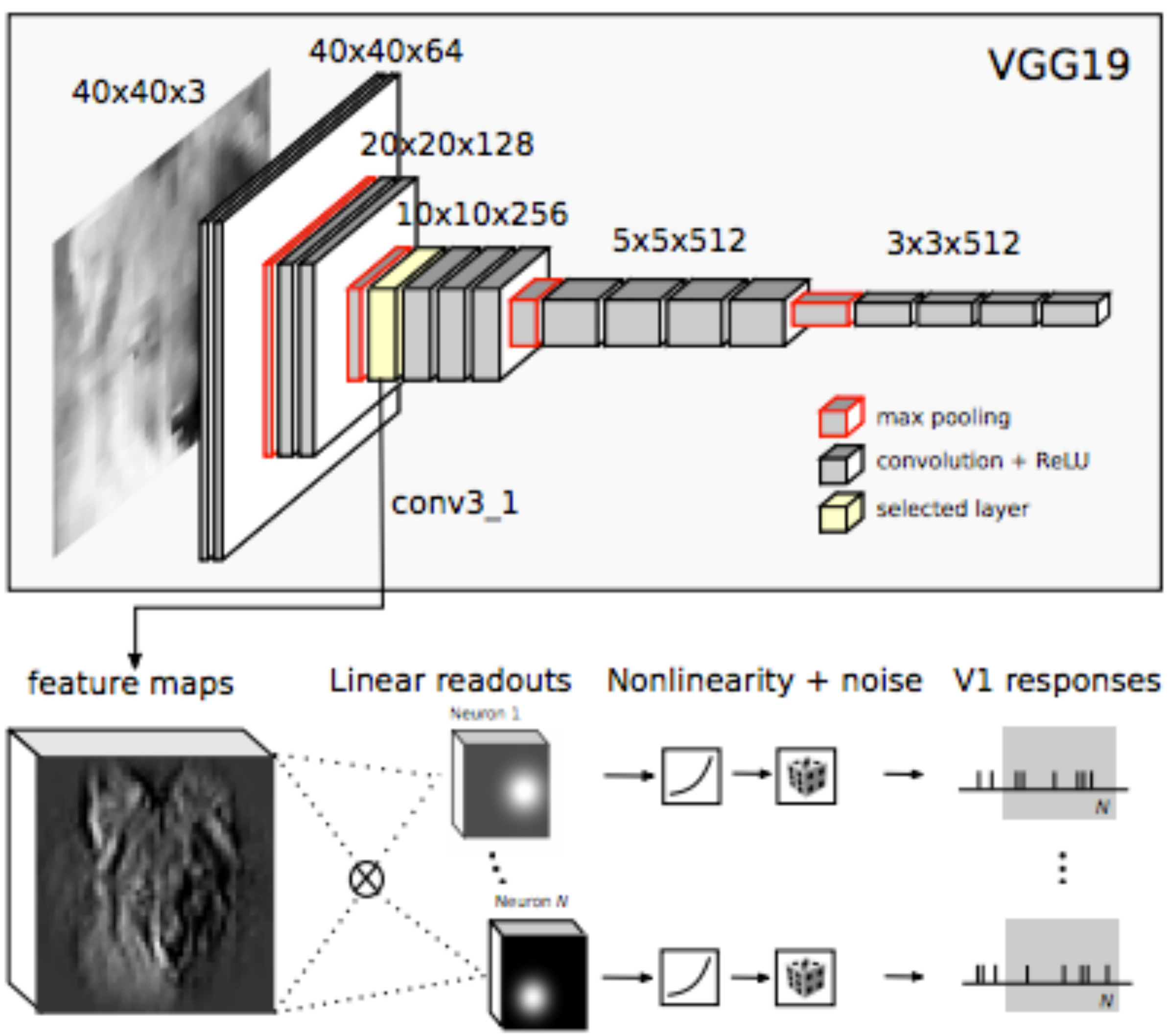
Deep convolutional models improve predictions of macaque V1 responses to natural images

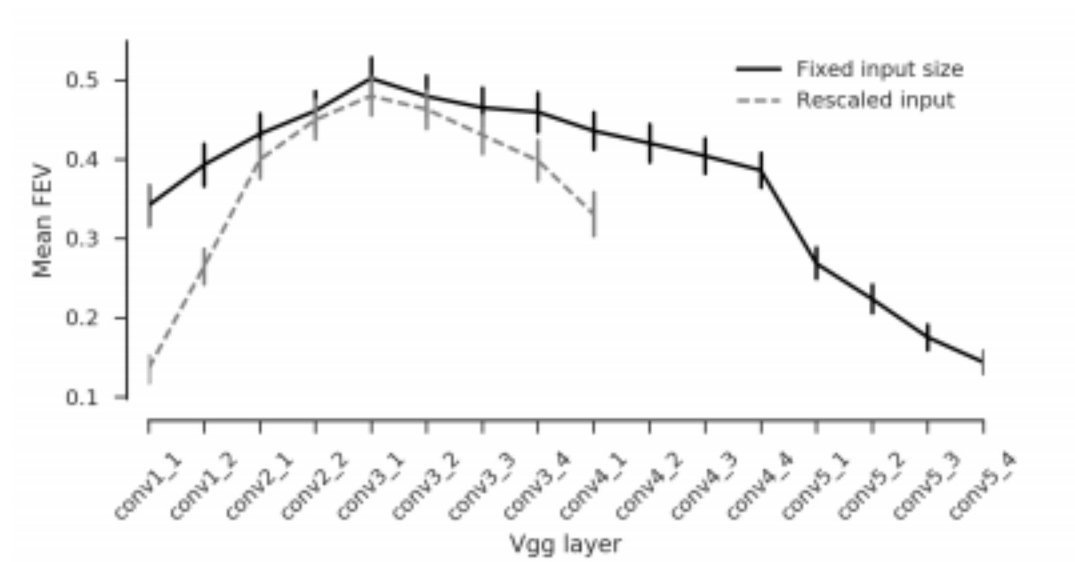
Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, Alexander S Ecker

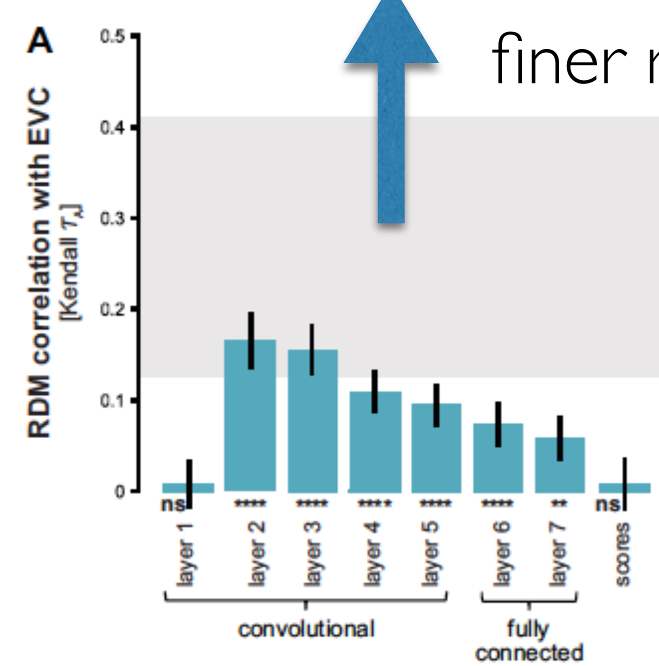
doi: https://doi.org/ 10.1101/201764

50% explained variance vs

- 17% for Linear-Nonlinear-Poisson (with gabor filters)
- 39% for Berkeley Wavelet Transform







finer resolution

Layer-area correspondence

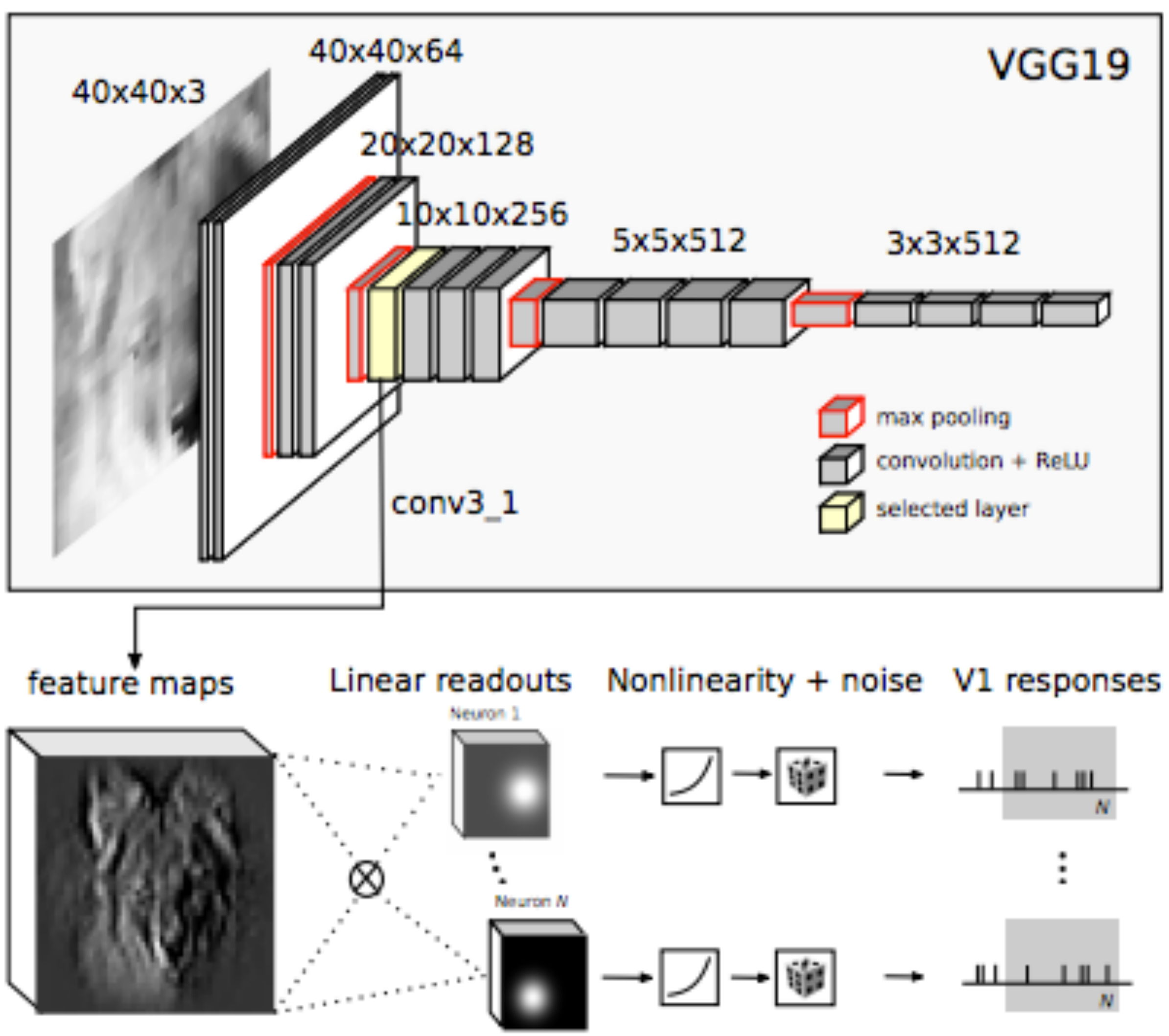
Deep convolutional models improve predictions of macaque V1 responses to natural images

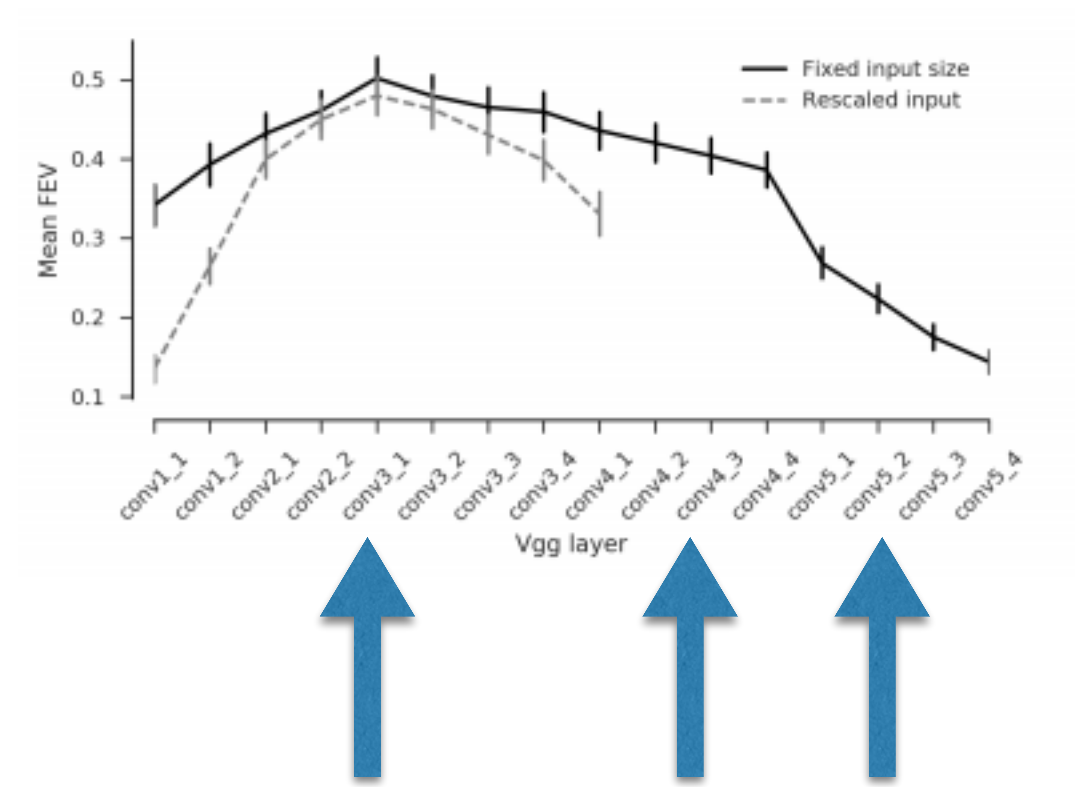
Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, Alexander S Ecker

doi: https://doi.org/ 10.1101/201764

50% explained variance vs

- 17% for Linear-Nonlinear-Poisson (with gabor filters)
- 39% for Berkeley Wavelet Transform





Peak V1

Peak V4 (unpublished)

Peak IT (unpublished)

Layer-area correspondence

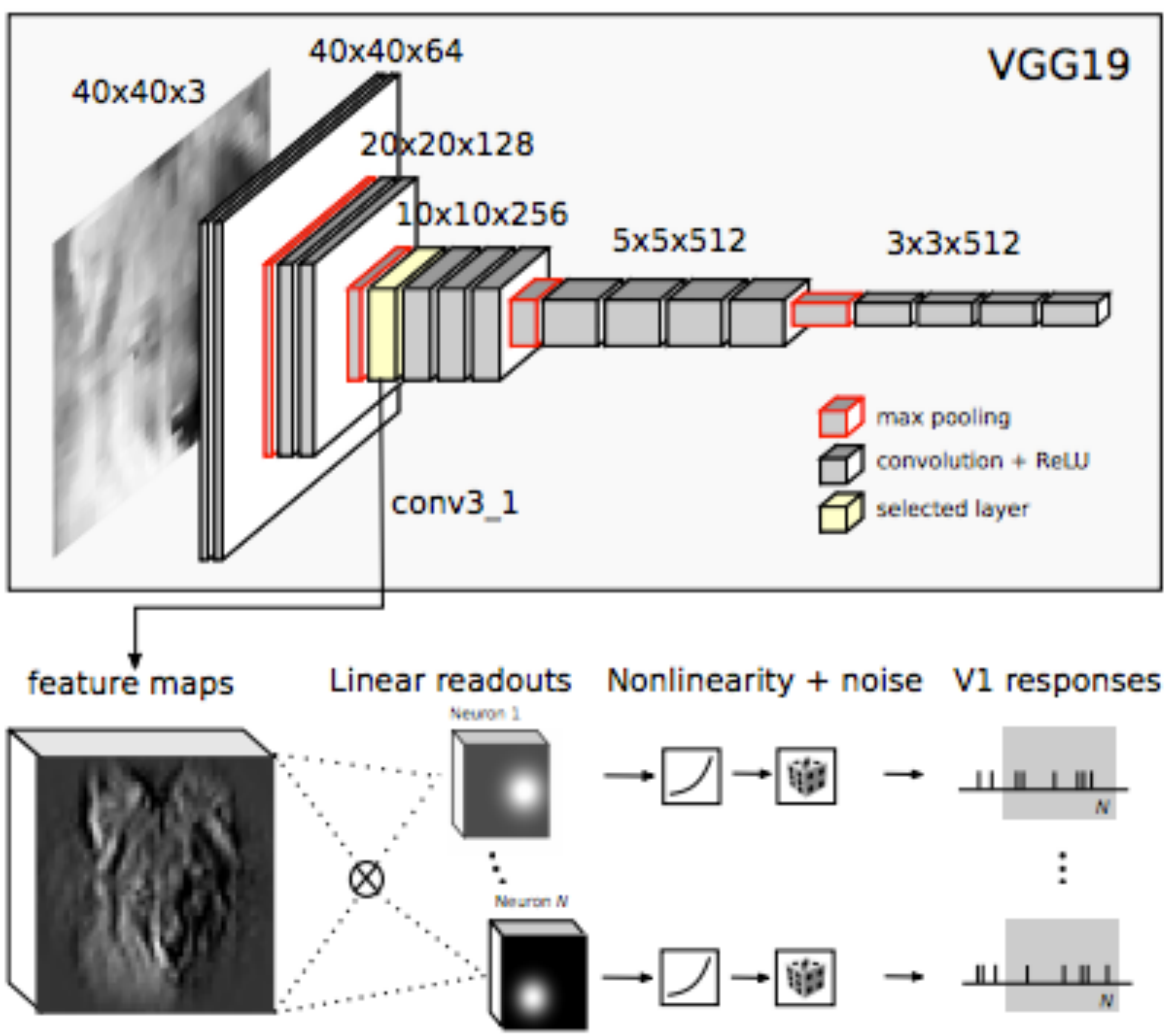
Deep convolutional models improve predictions of macaque V1 responses to natural images

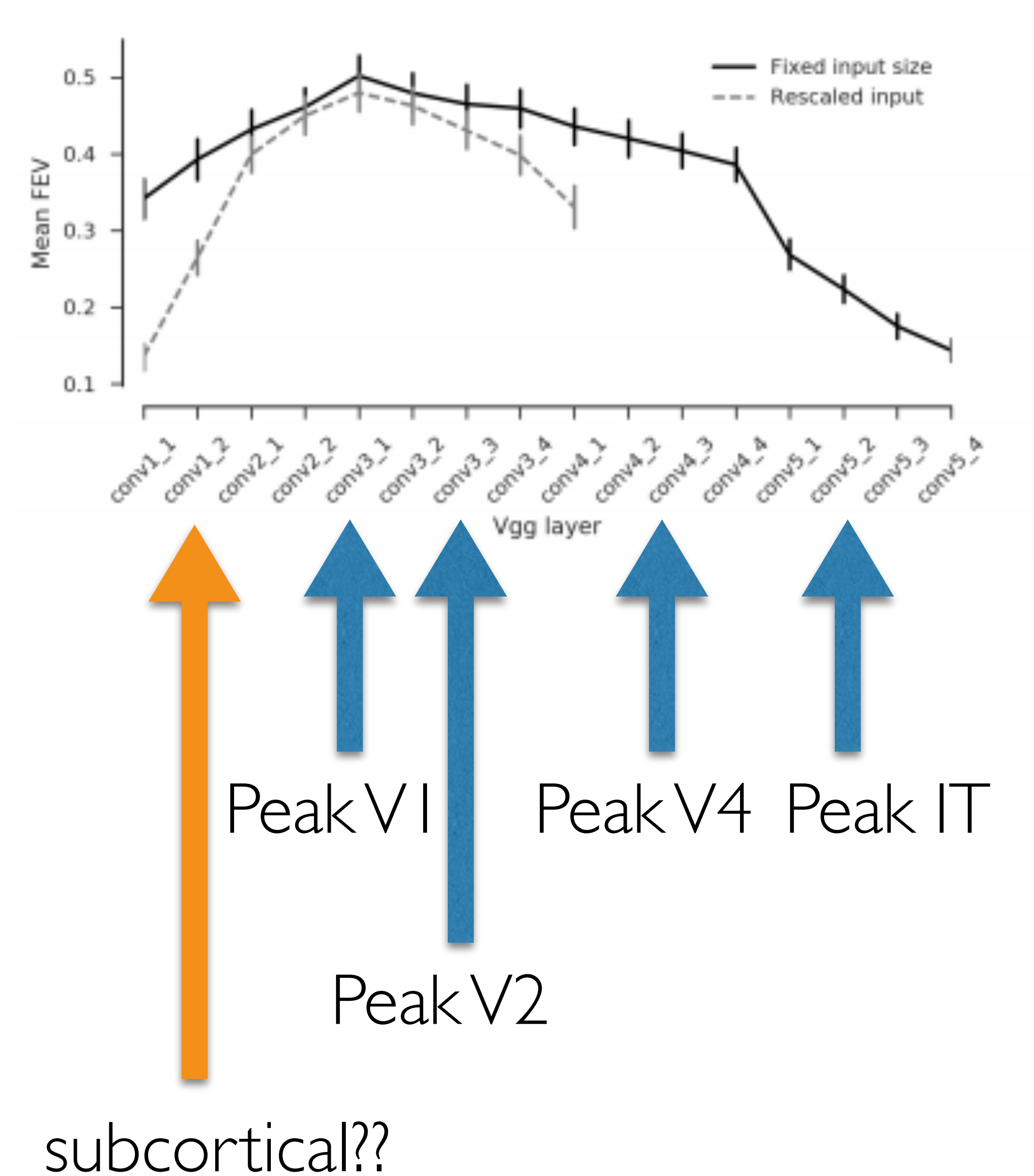
Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, Alexander S Ecker

doi: https://doi.org/ 10.1101/201764

50% explained variance vs

- 17% for Linear-Nonlinear-Poisson (with gabor filters)
- 39% for Berkeley Wavelet Transform





Layer-area correspondence

Deep Learning Models of the Retinal Response to Natural Scenes

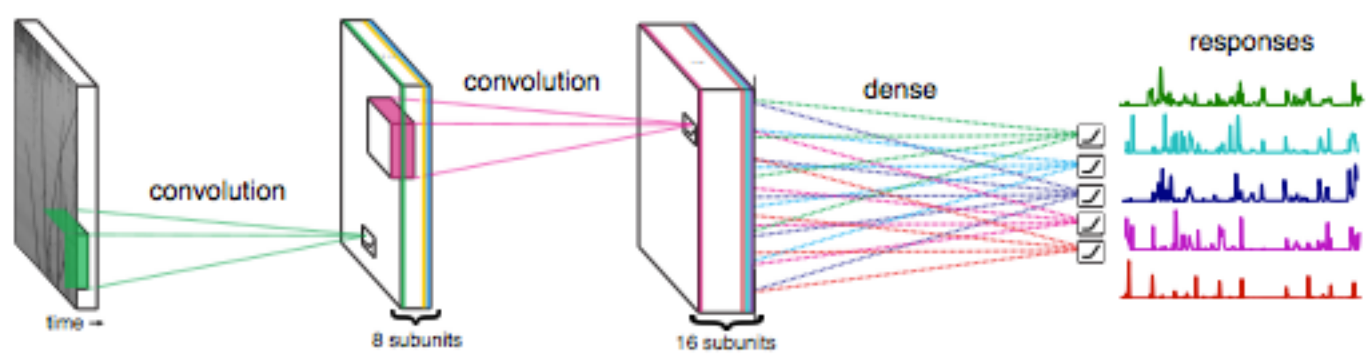
Lane T. McIntosh[*1], Niru Maheswaranathan[*1], Aran Nayebi[1],
Surya Ganguli[2,3], Stephen A. Baccus[3]
[1]Neurosciences PhD Program, [2]Department of Applied Physics, [3]Neurobiology Department
Stanford University
{lmcintosh, nirum, anayebi, sganguli, baccus}@stanford.edu

Layer-area correspondence

Deep Learning Models of the Retinal Response to Natural Scenes

Lane T. McIntosh*[1], Niru Maheswaranathan*[1], Aran Nayebi[1],
Surya Ganguli[2,3], Stephen A. Baccus[3]
[1]Neurosciences PhD Program, [2]Department of Applied Physics, [3]Neurobiology Department
Stanford University
{lmcintosh, nirum, anayebi, sganguli, baccus}@stanford.edu

Three-layer CNN best fits retinal ganglion cell response patterns to natural images.

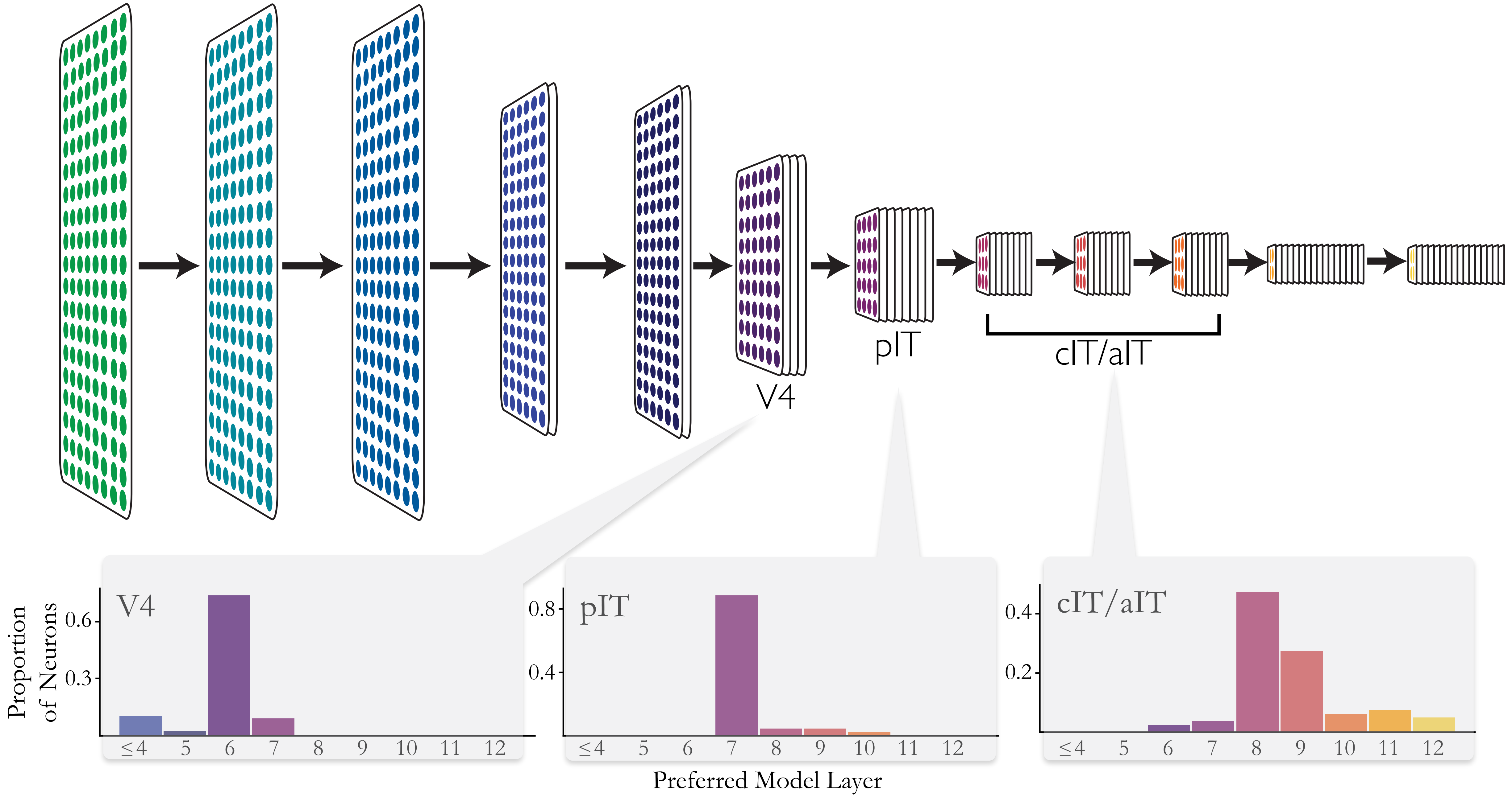


Layer-area correspondence

Better models of the ventral visual stream:

- V4 at 6th convolutional layer
- pIT at 7th convolutional layer
- cIT/aIT at layers 8-10, depending on neurons position on A/P axis

Dan Bear Jonas Kubilius



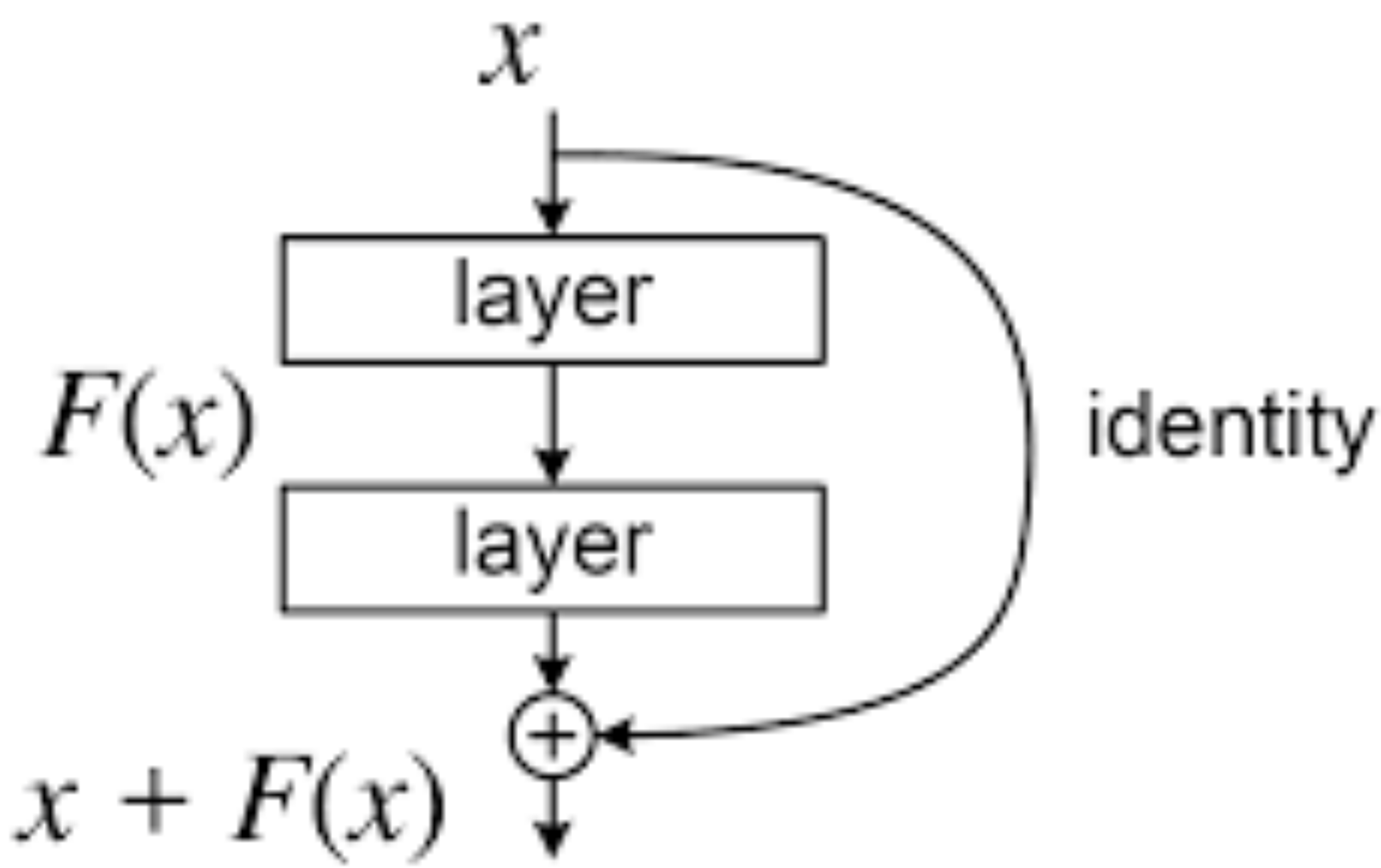
Post-AlexNet Developments

(1) Residual Connections and ResNets

(2) Vision Transformers

Post-AlexNet Developments

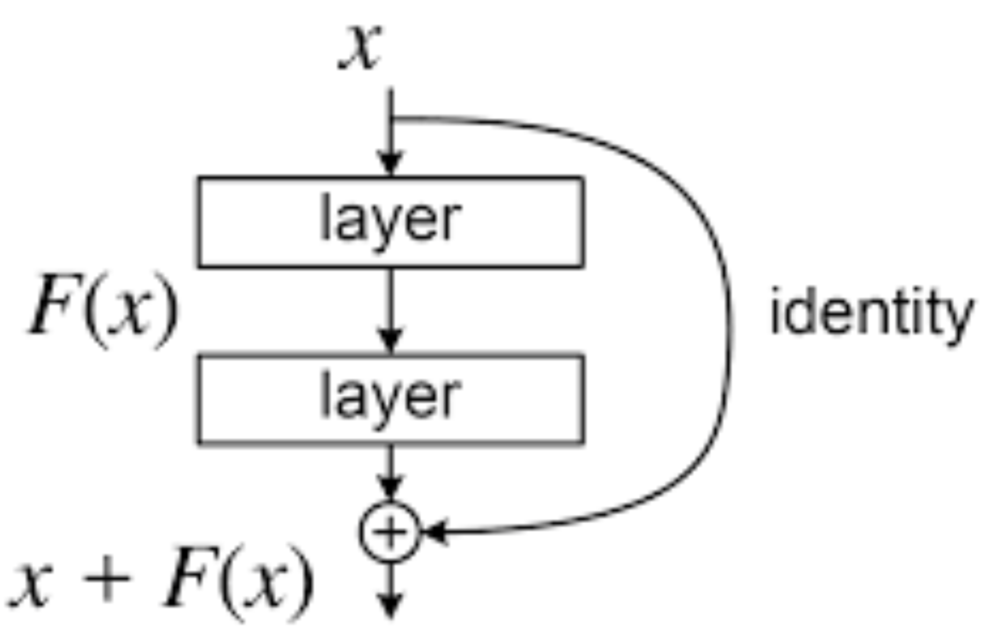
(1) Residual Connections and ResNets

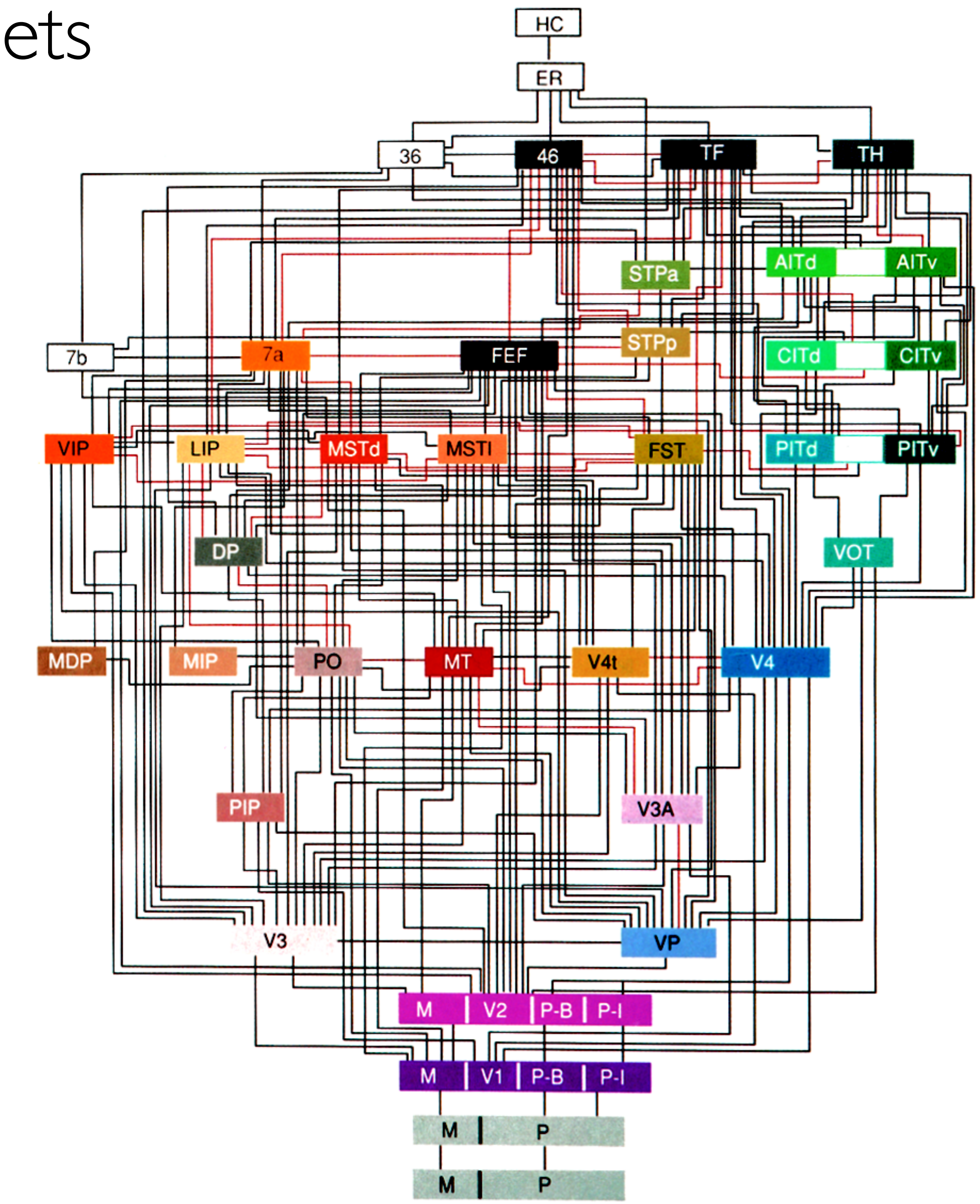


Residual connection stabilizes gradient backflow.

Post-AlexNet Developments

(1) Residual Connections and ResNets



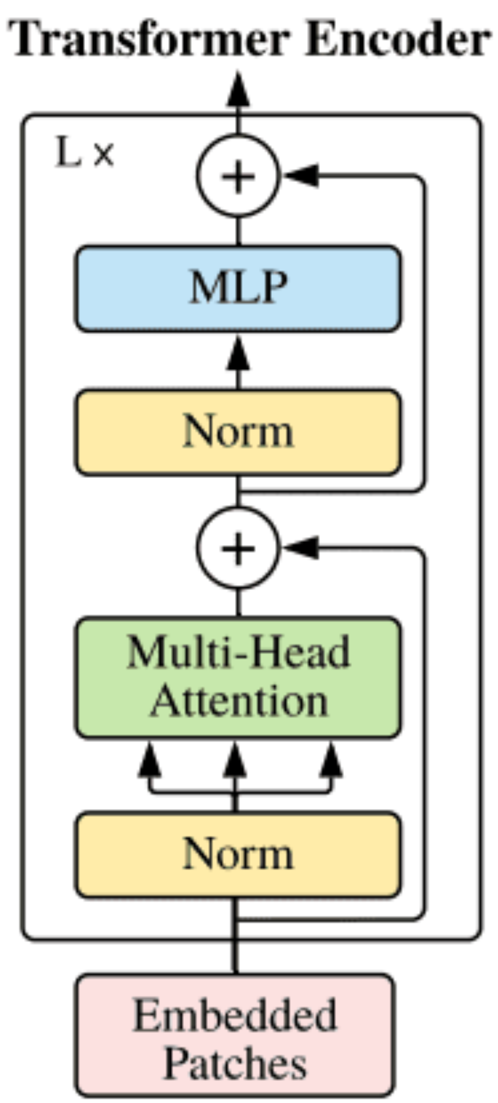


Residual connection stabilizes gradient backflow.

Lots of skip connections present in actual brain.

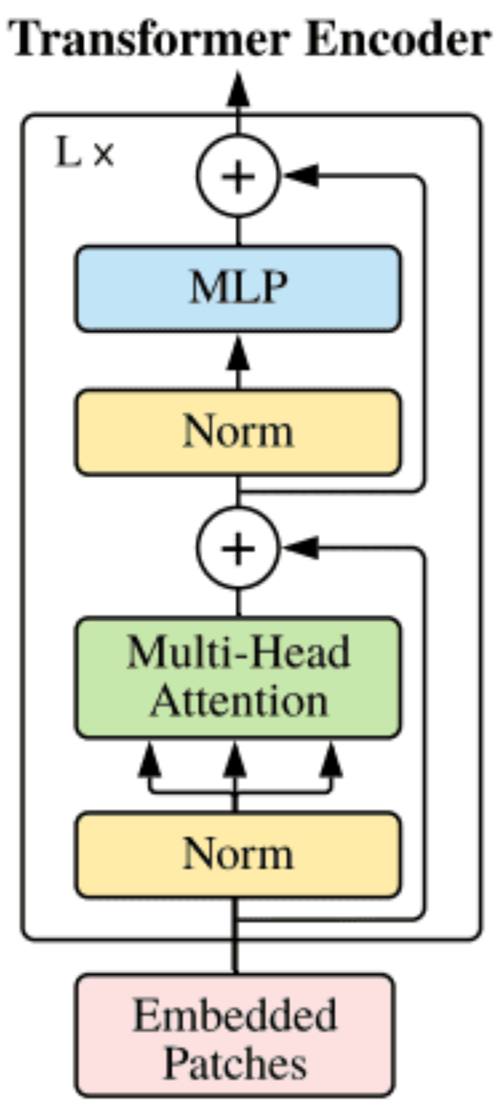
Post-AlexNet Developments

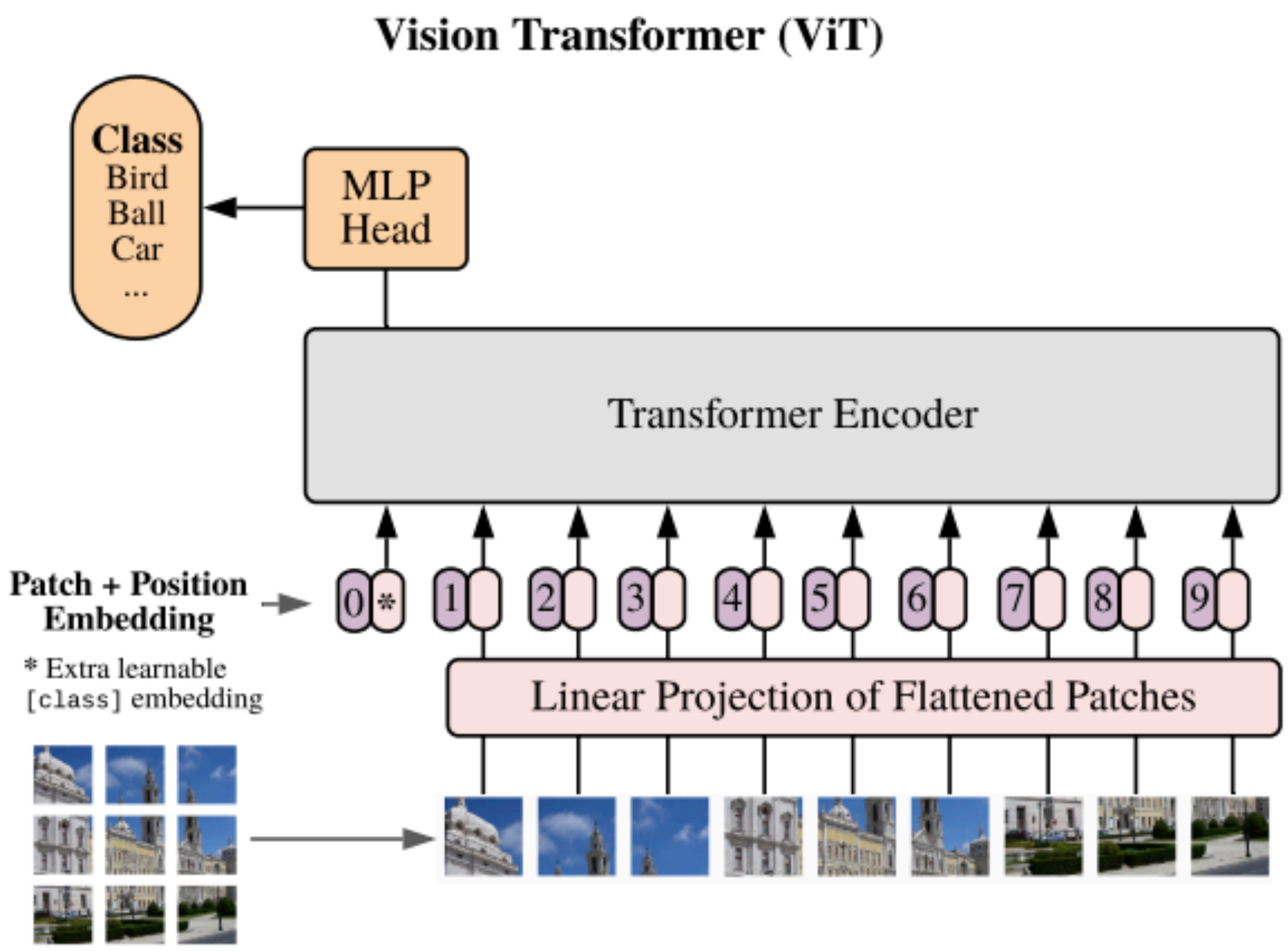
(2) Vision Transformers



Post-AlexNet Developments

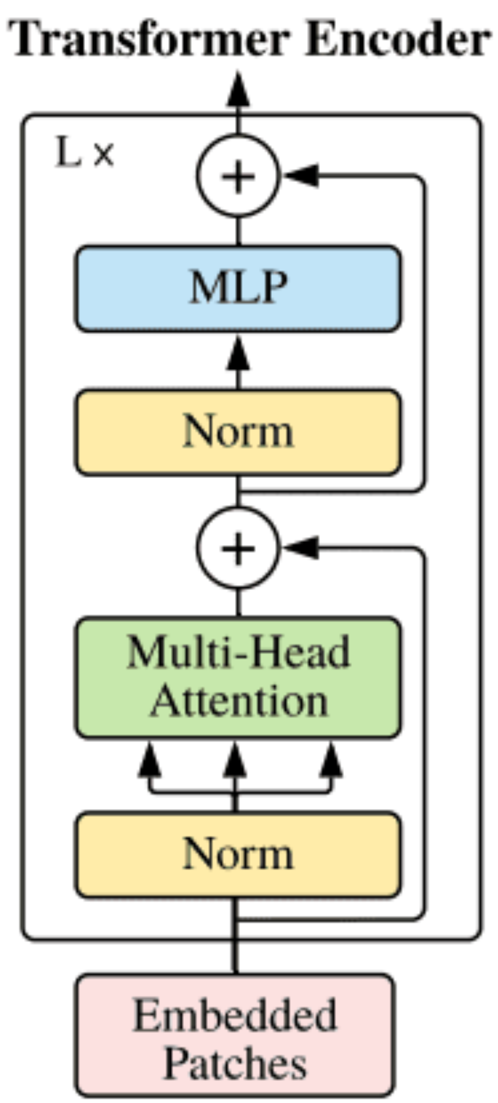
(2) Vision Transformers

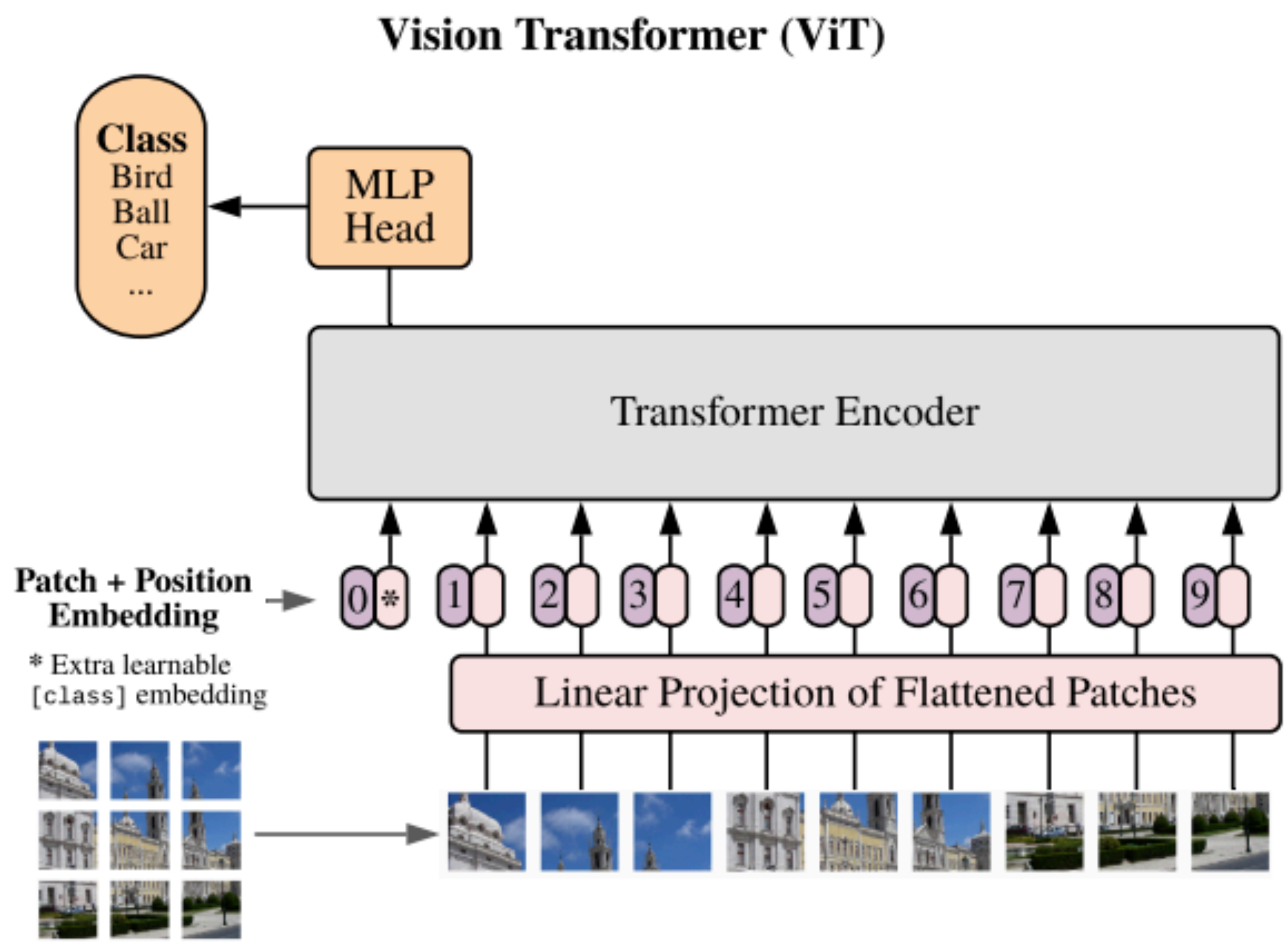




Post-AlexNet Developments

(2) Vision Transformers





NB: still hierarchical, still with residual connections, potential locality from patches . . .

Post-AlexNet Developments

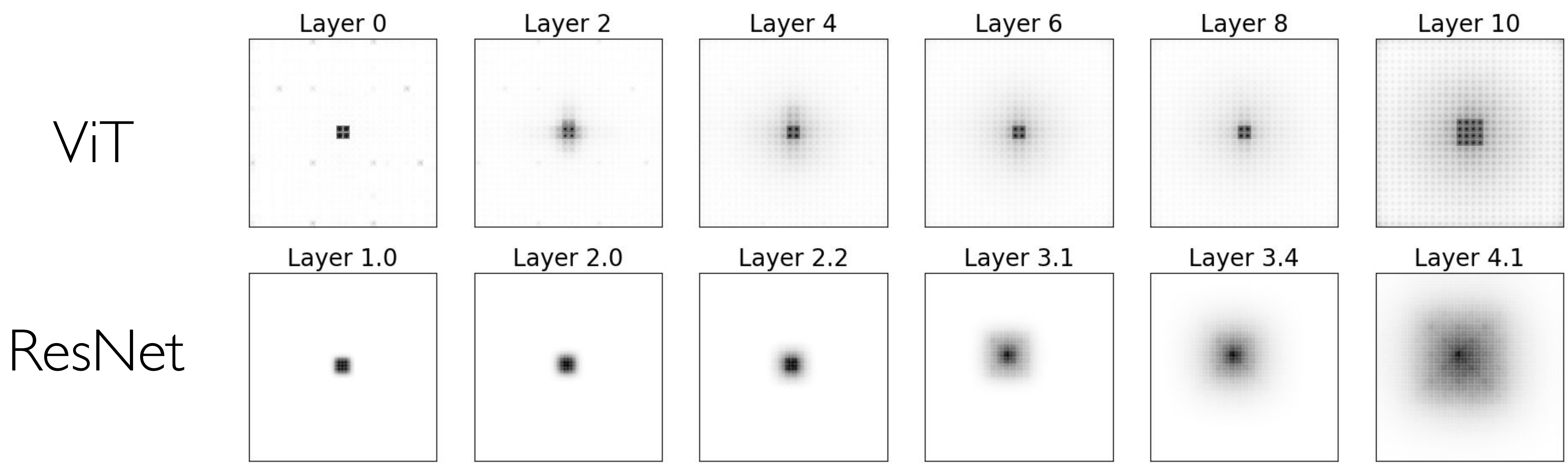
(2) Vision Transformers

Looking at receptive field analysis of ViTs vs ResNet:

Post-AlexNet Developments

(2) Vision Transformers

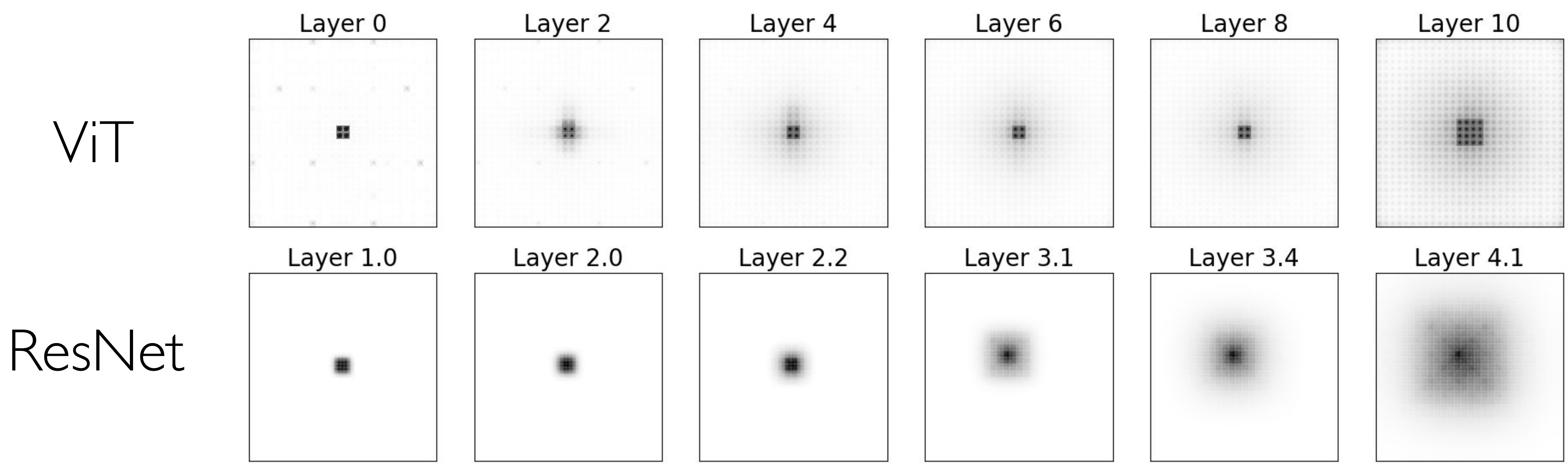
Looking at receptive field analysis of ViTs vs ResNet:



Post-AlexNet Developments

(2) Vision Transformers

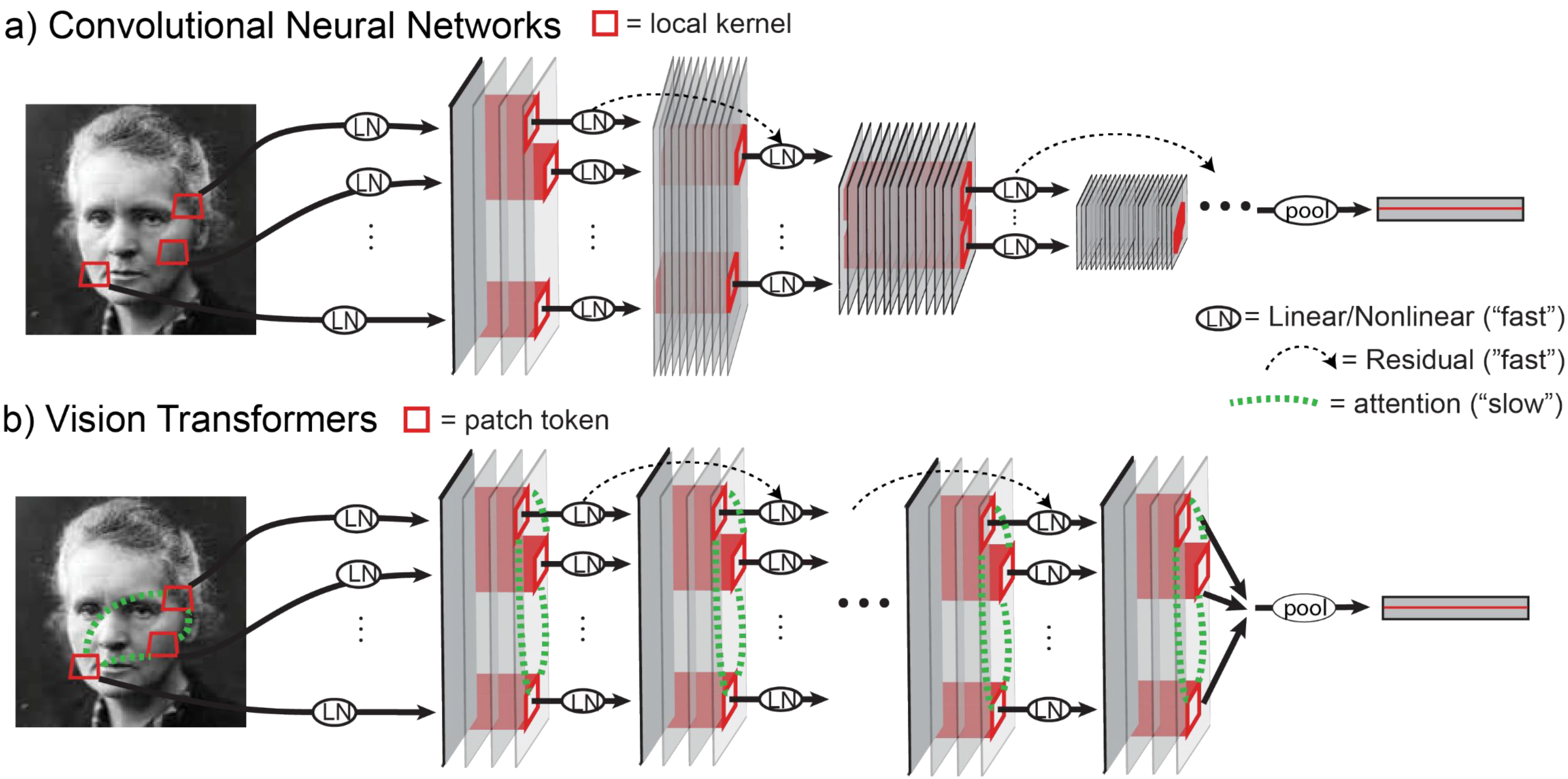
Looking at receptive field analysis of ViTs vs ResNet:



. . . we see learned ViT is <u>mostly local</u>, with increasing receptive field sizes.

Post-AlexNet Developments

(2) Vision Transformers

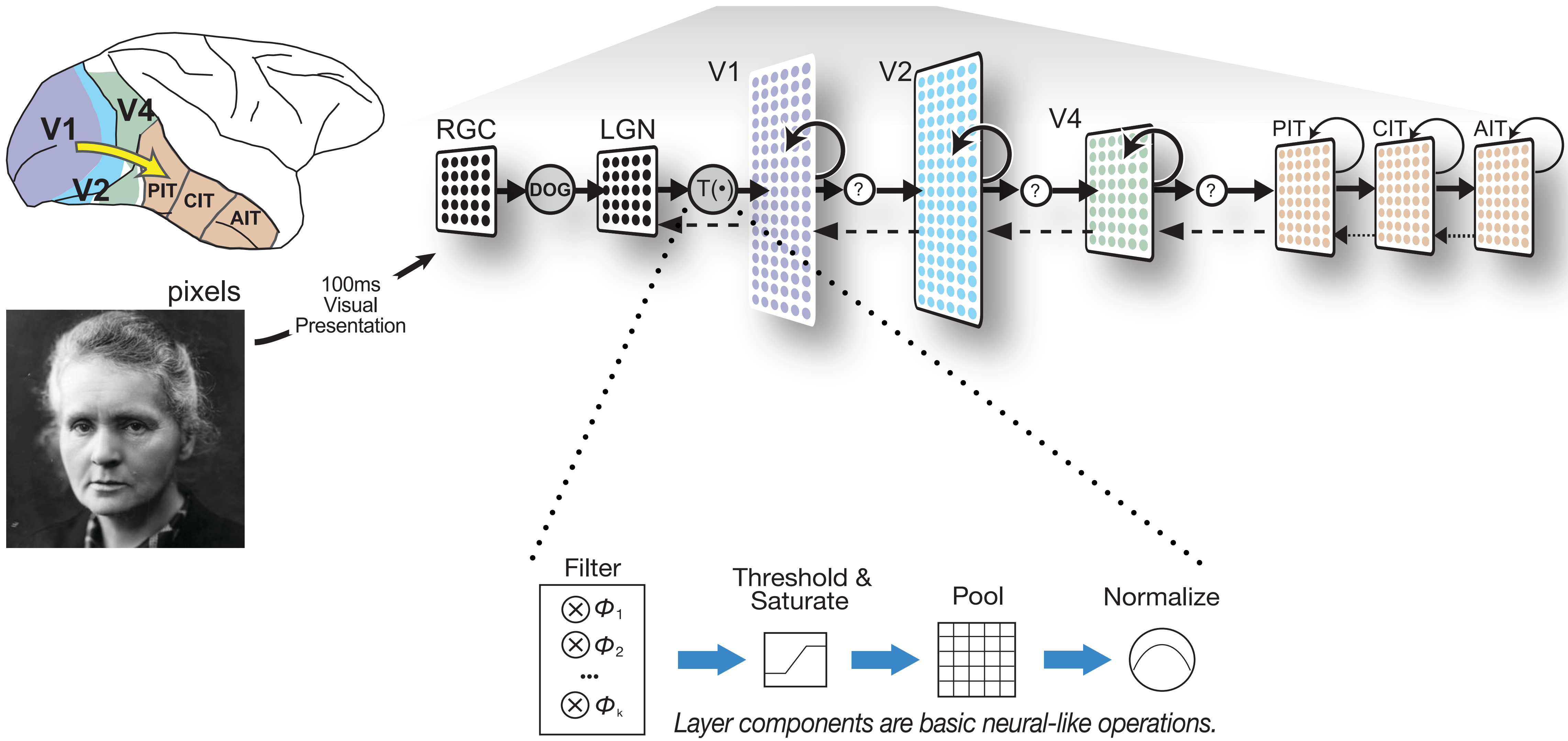


<u>ViT is a bit like a CNN with sparse global connections.</u>

Principles of Visual Architecture

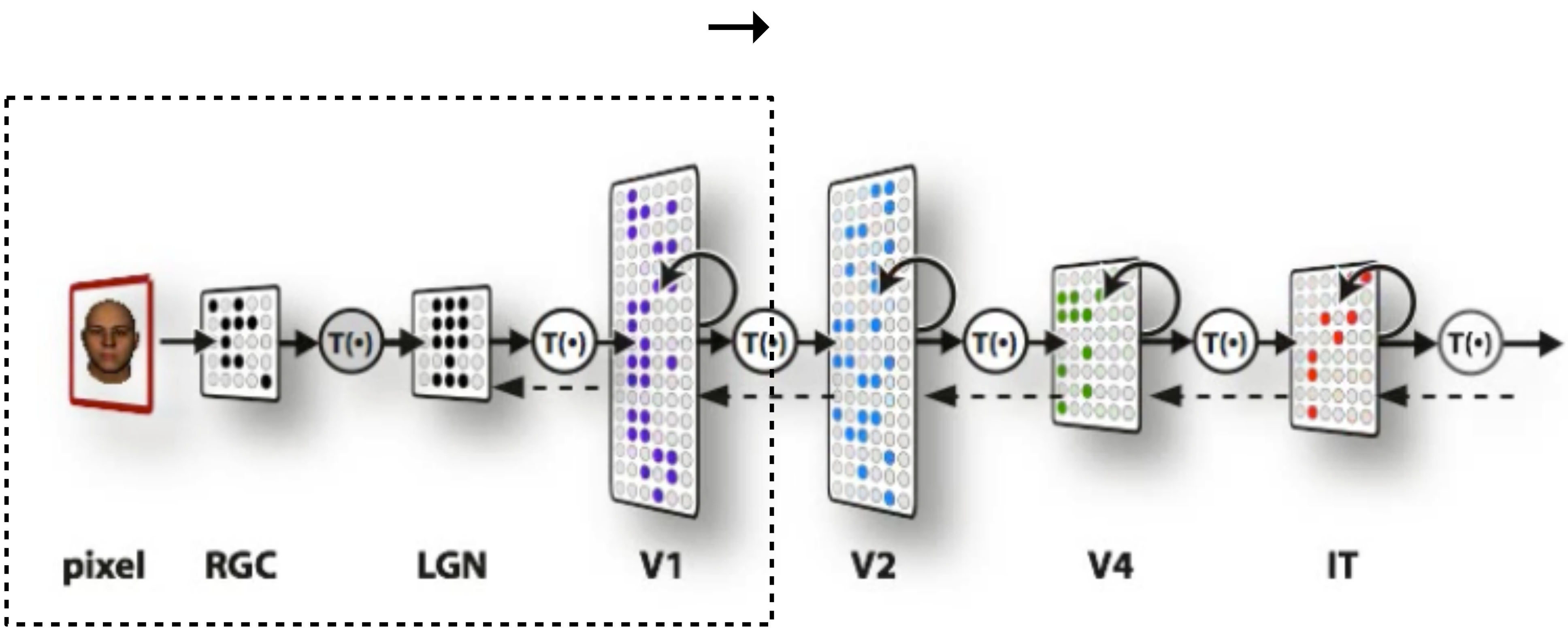
(1) Hierarchical (2) Mostly local (3) Rectification-like nonlinearity

(4) Some residual connections (5) Normalization



Behavioral "Top-Down" constraints

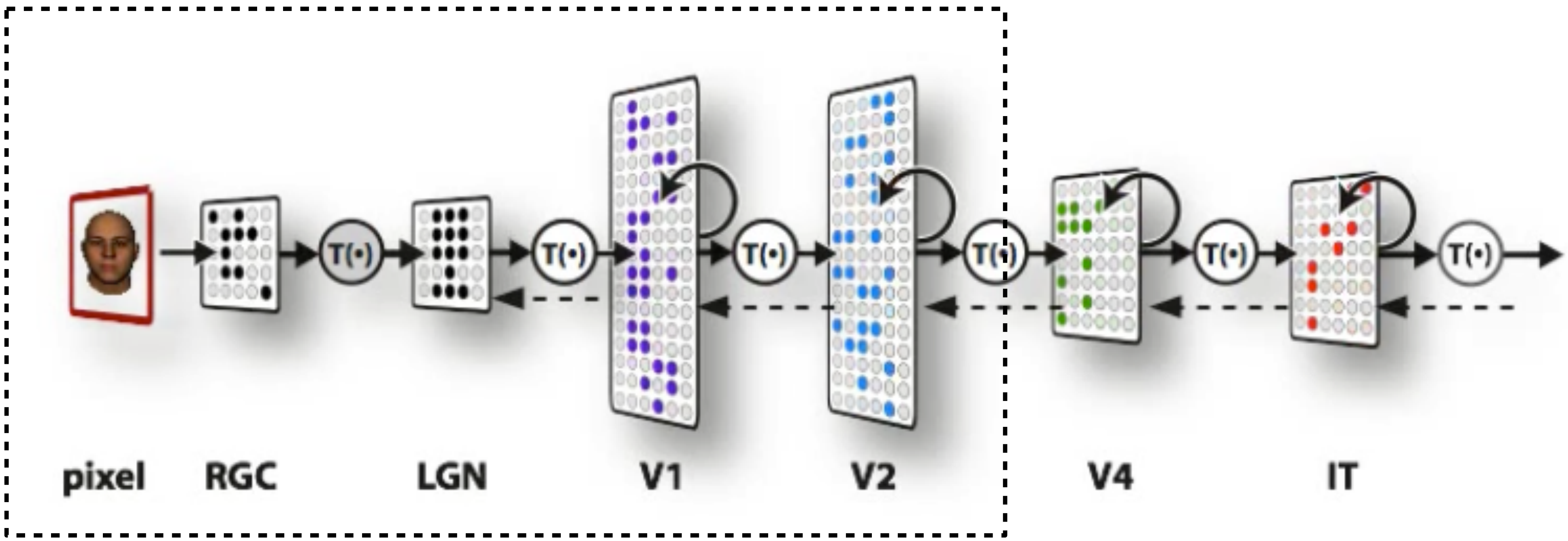
Complement standard "from below" approach …



Behavioral "Top-Down" constraints

Complement standard "from below" approach …

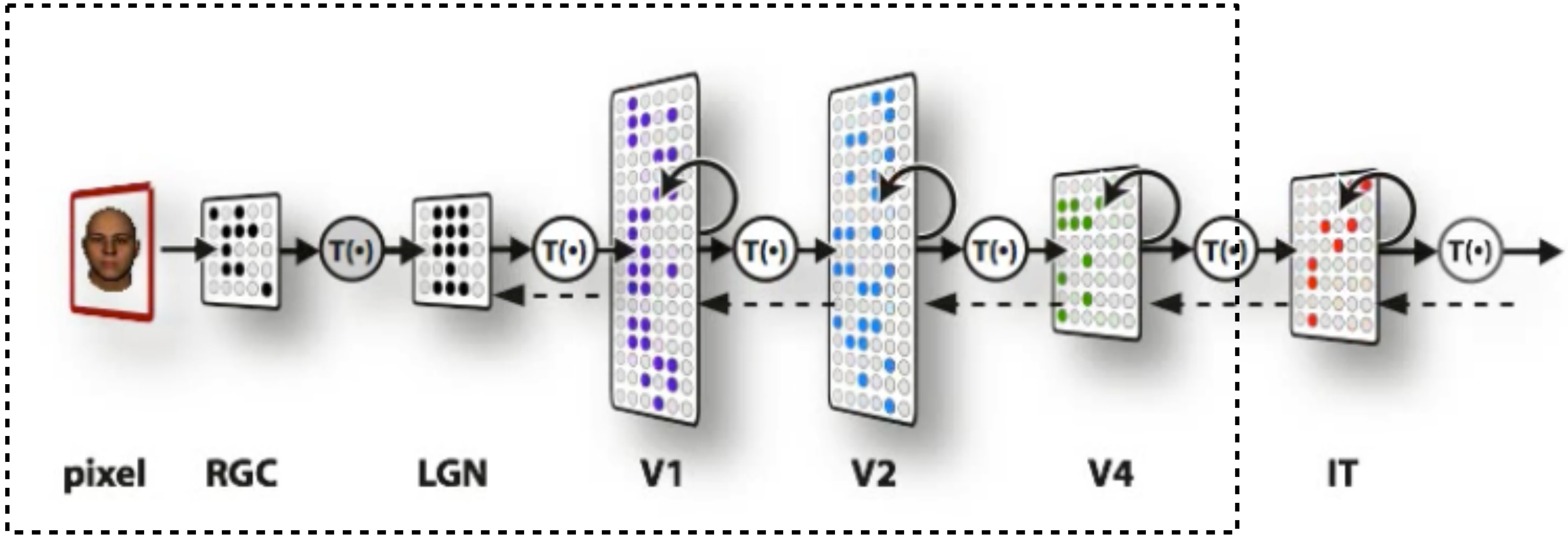
→



Behavioral "Top-Down" constraints

Complement standard "from below" approach …

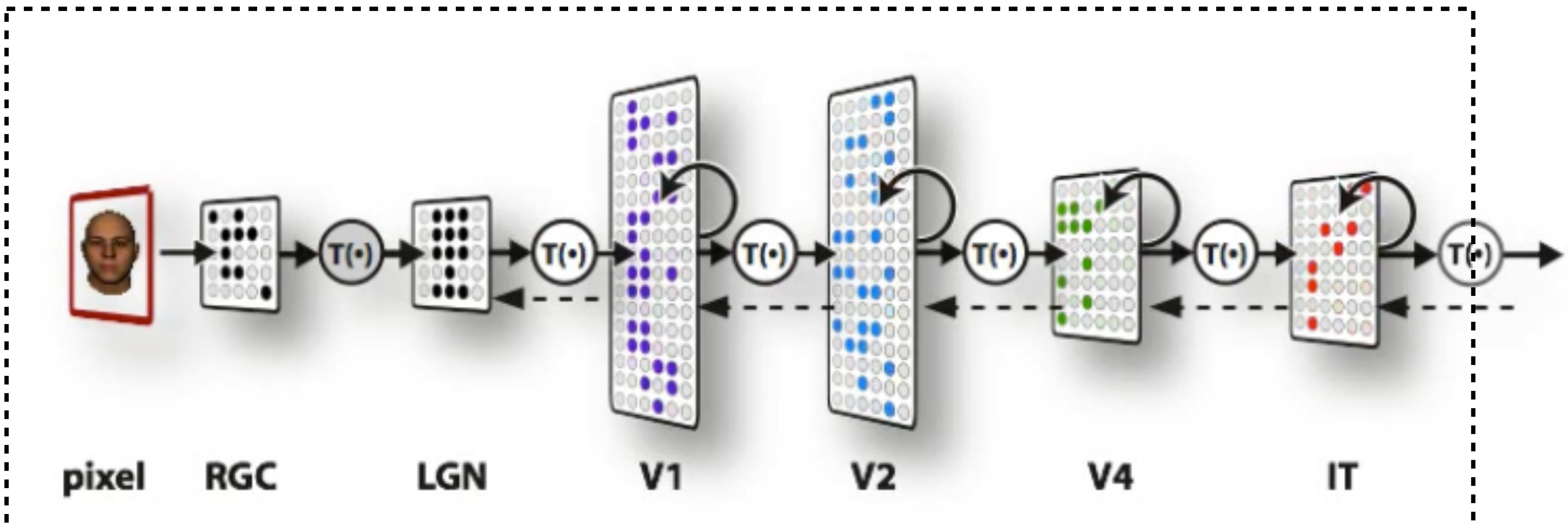
→



Behavioral "Top-Down" constraints

Complement standard "from below" approach …

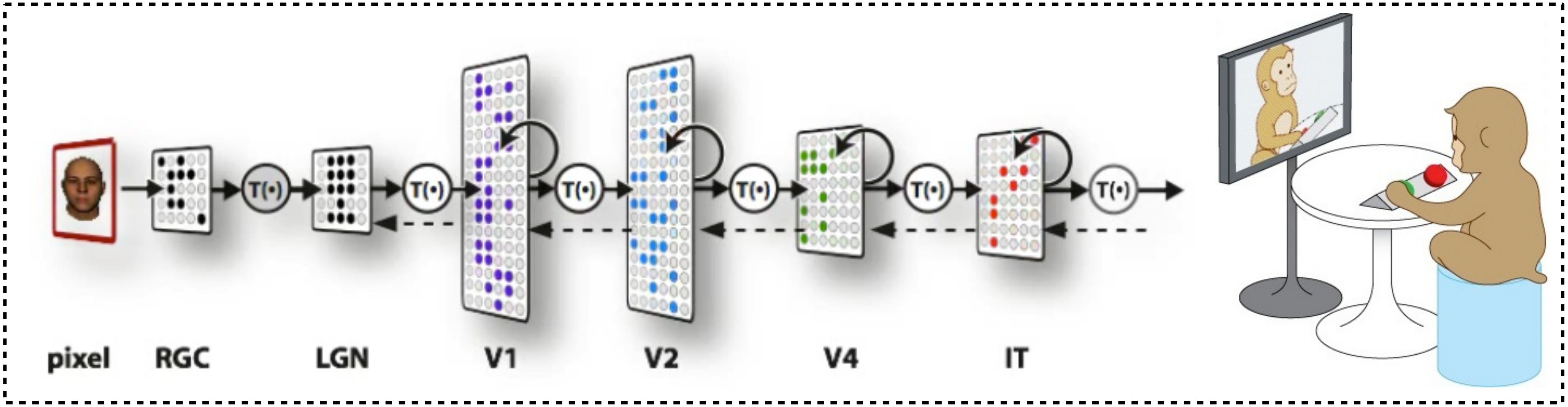
→



Behavioral "Top-Down" constraints

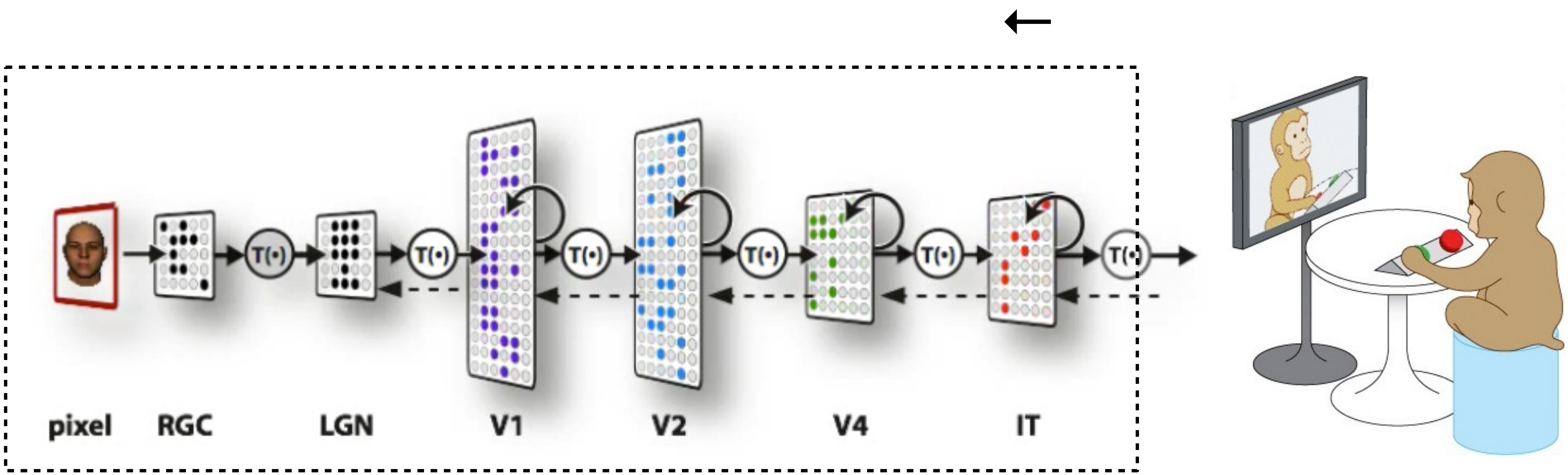
Complement standard "from below" approach … with behavioral constraints

←



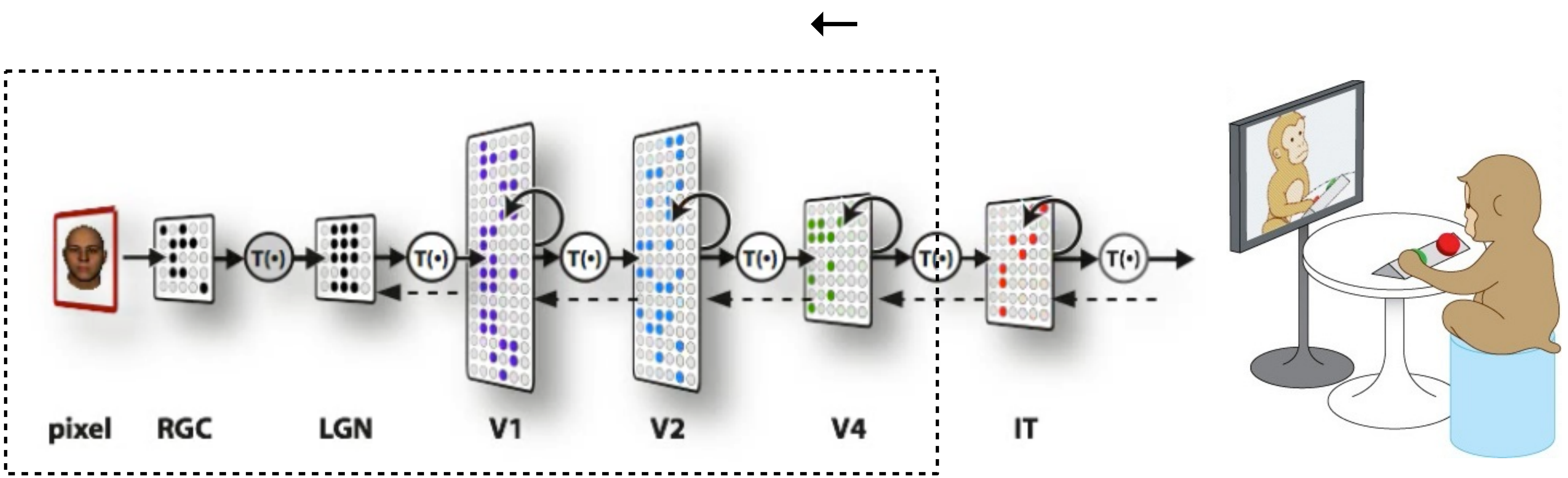
Behavioral "Top-Down" constraints

Complement standard "from below" approach … with behavioral constraints



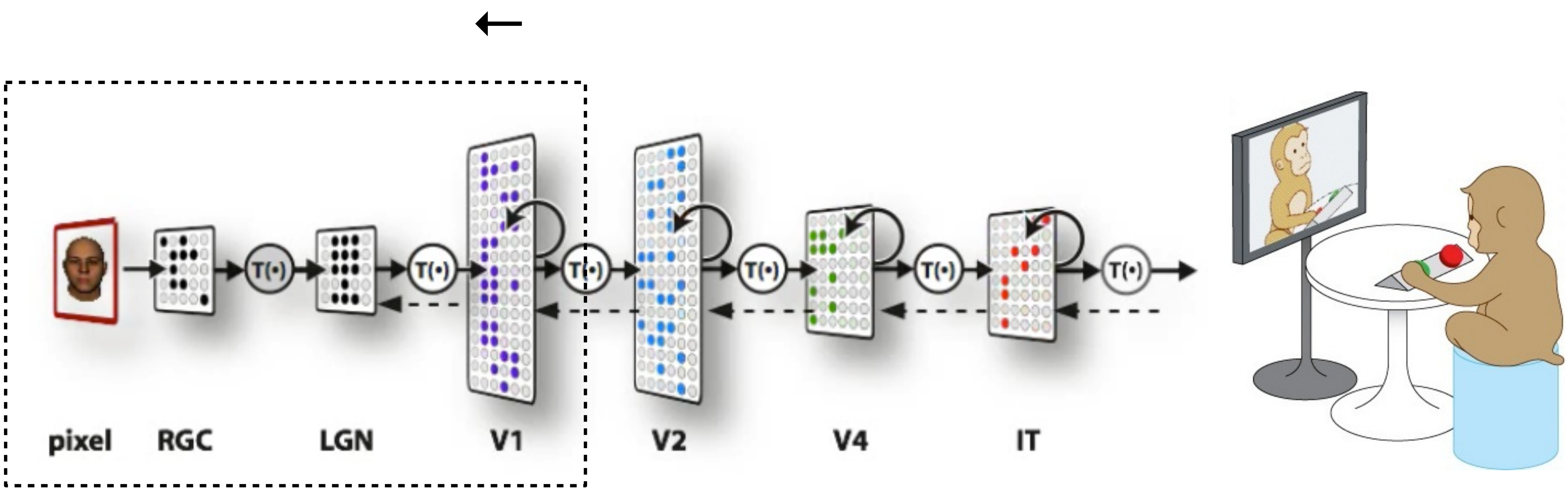
Behavioral "Top-Down" constraints

Complement standard "from below" approach … with behavioral constraints



Behavioral "Top-Down" constraints

Complement standard "from below" approach … with behavioral constraints



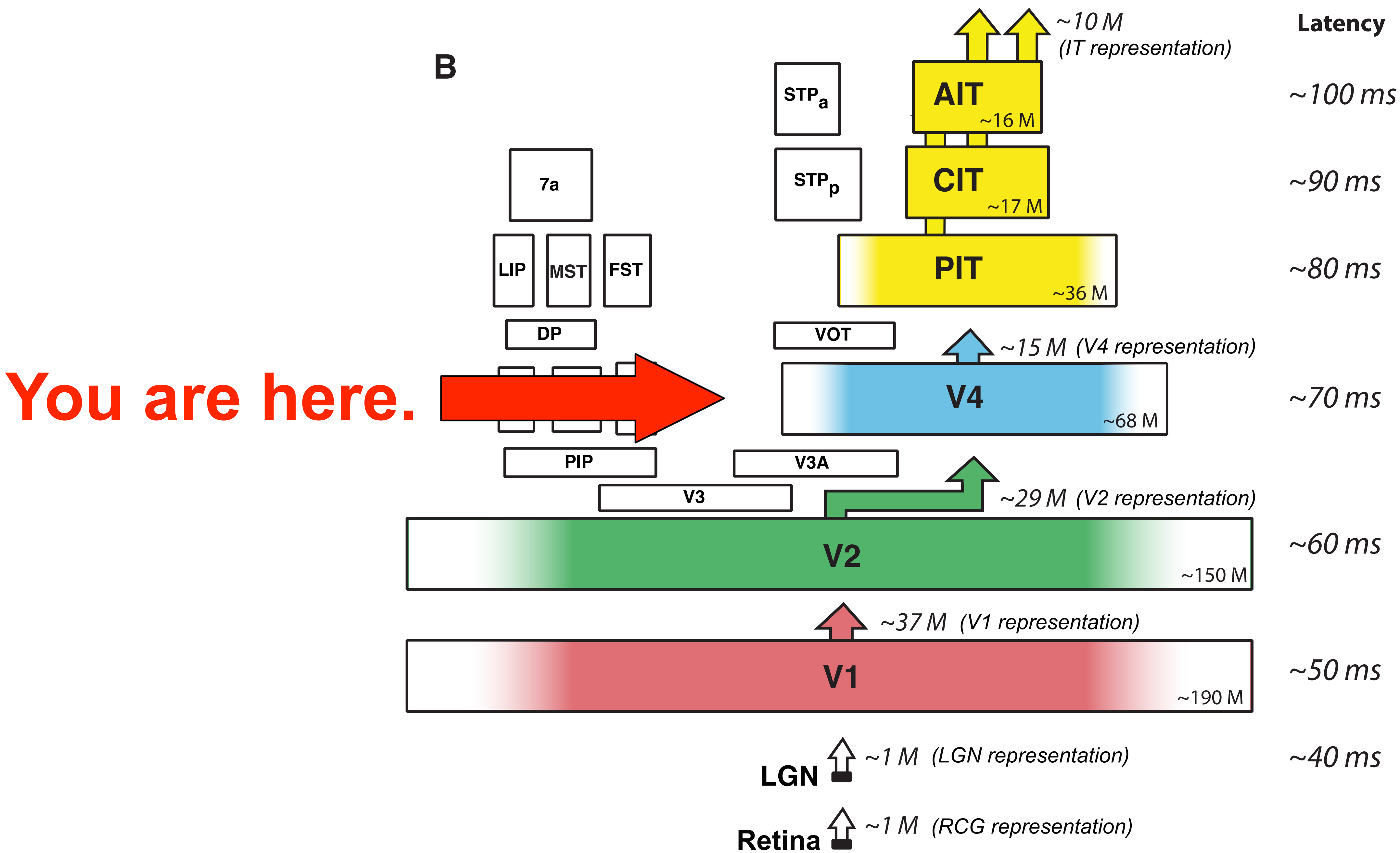
RESEARCH ARTICLE

NEUROSCIENCE

Neural population control via deep image synthesis

Pouya Bashivan*, Kohitij Kar*, James J. DiCarlo†

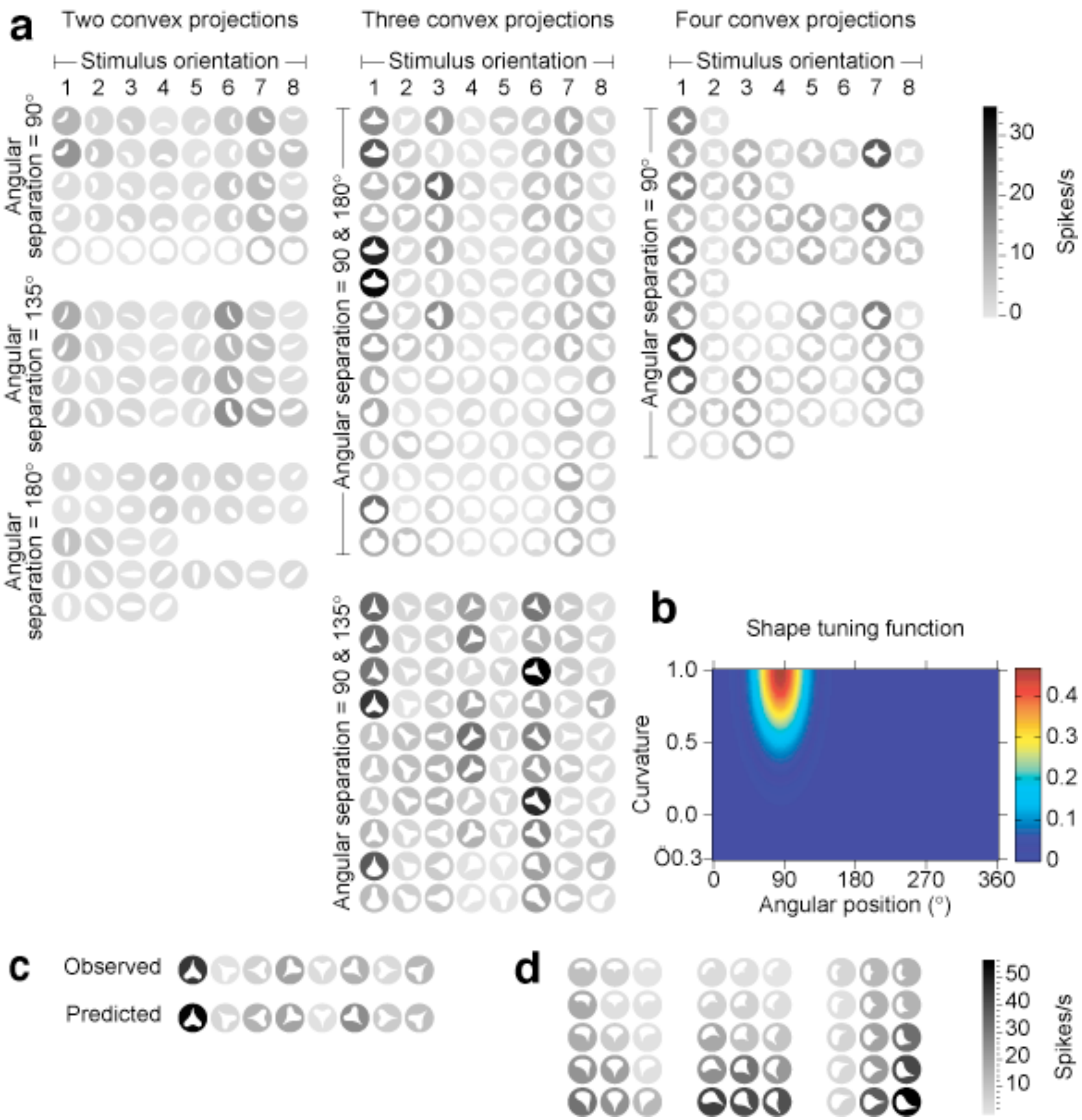
Particular deep artificial neural networks (ANNs) are today's most accurate models of the primate brain's ventral visual stream. Using an ANN-driven image synthesis method, we found that luminous power patterns (i.e., images) can be applied to primate retinae to predictably push the spiking activity of targeted V4 neural sites beyond naturally occurring levels. This method, although not yet perfect, achieves unprecedented independent control of the activity state of entire populations of V4 neural sites, even those with overlapping receptive fields. These results show how the knowledge embedded in today's ANN models might be used to noninvasively set desired internal brain states at neuron-level resolution, and suggest that more accurate ANN models would produce even more accurate control.



Adapted from DiCarlo et al. 2012

Recall ….

Adapted from C.E. Connor



<u>Make a basis for shapes</u>:
each shape = set of curved elements
each element = (ang position, curvature)

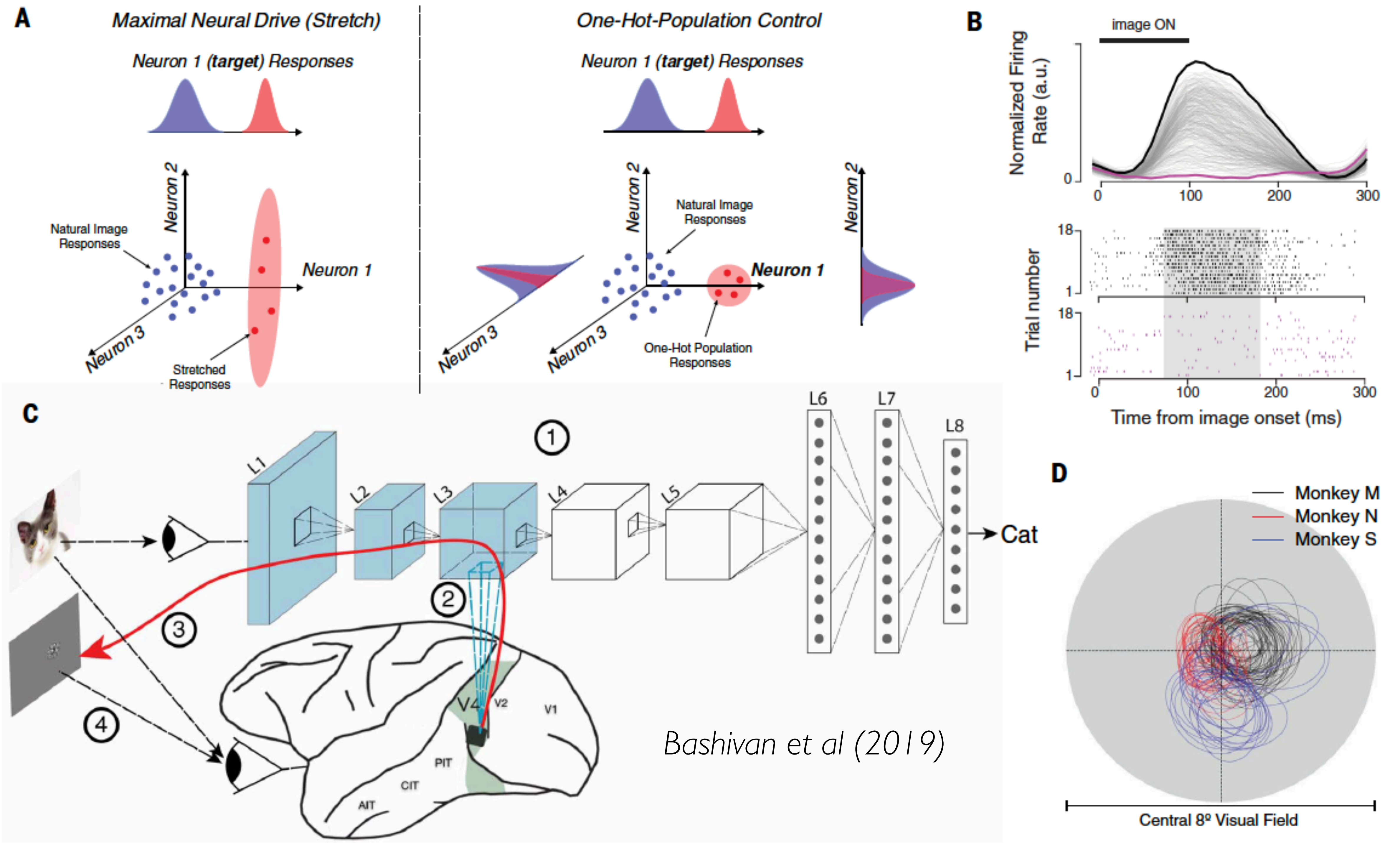
<u>Hypothesis</u>:
V4 neurons are tuned in this basis

<u>Experimental result</u>:
Hypothesis explains ~50% of the explainable response variance for these types of stimuli

Problem:
No predictions for any other images.
i.e.
is not an "image-computable" model

Pasupathy and Connor (V4)
Brincat and Connor (PIT)

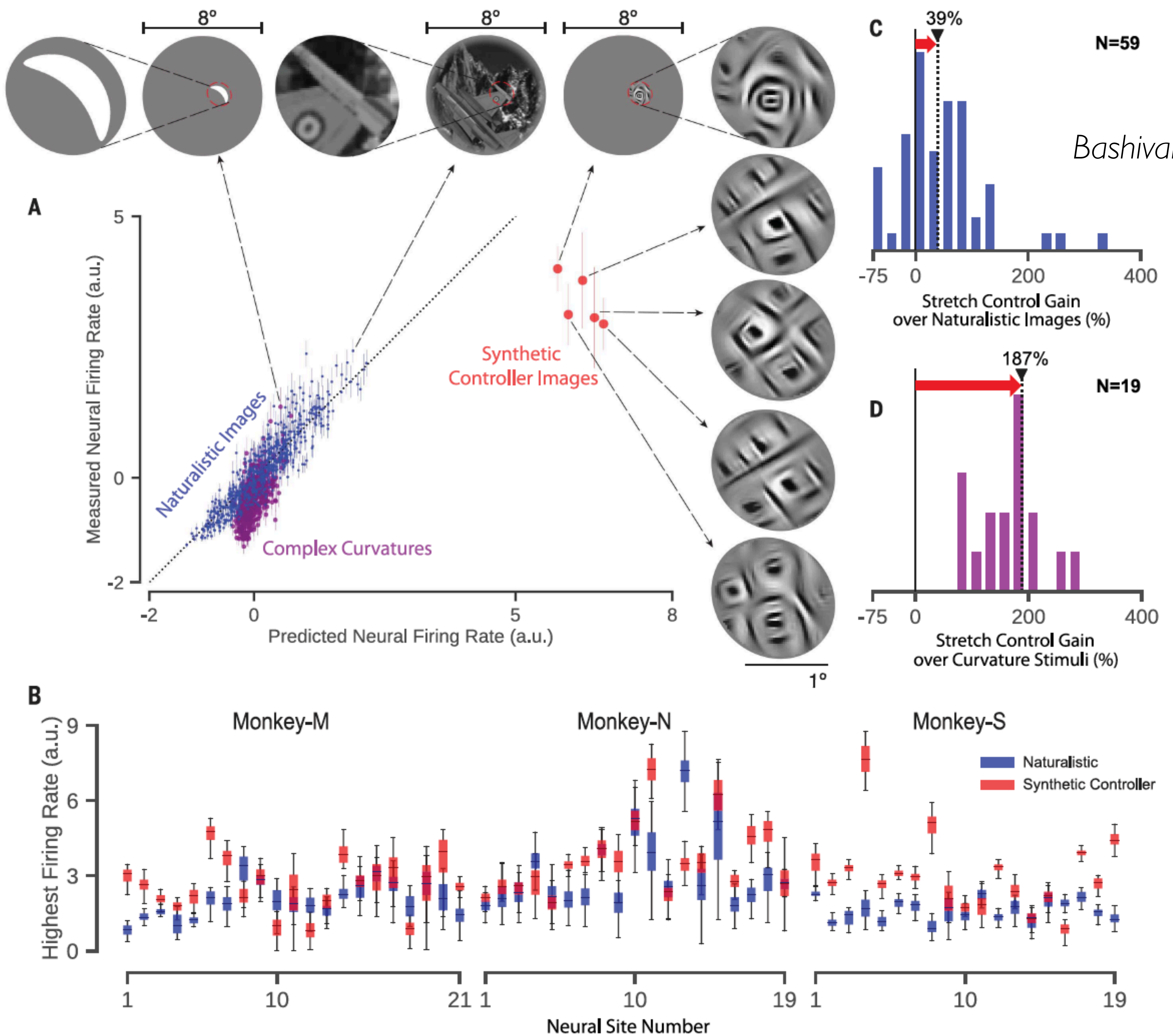
"Further Confirmation"



Bashivan et al (2019)

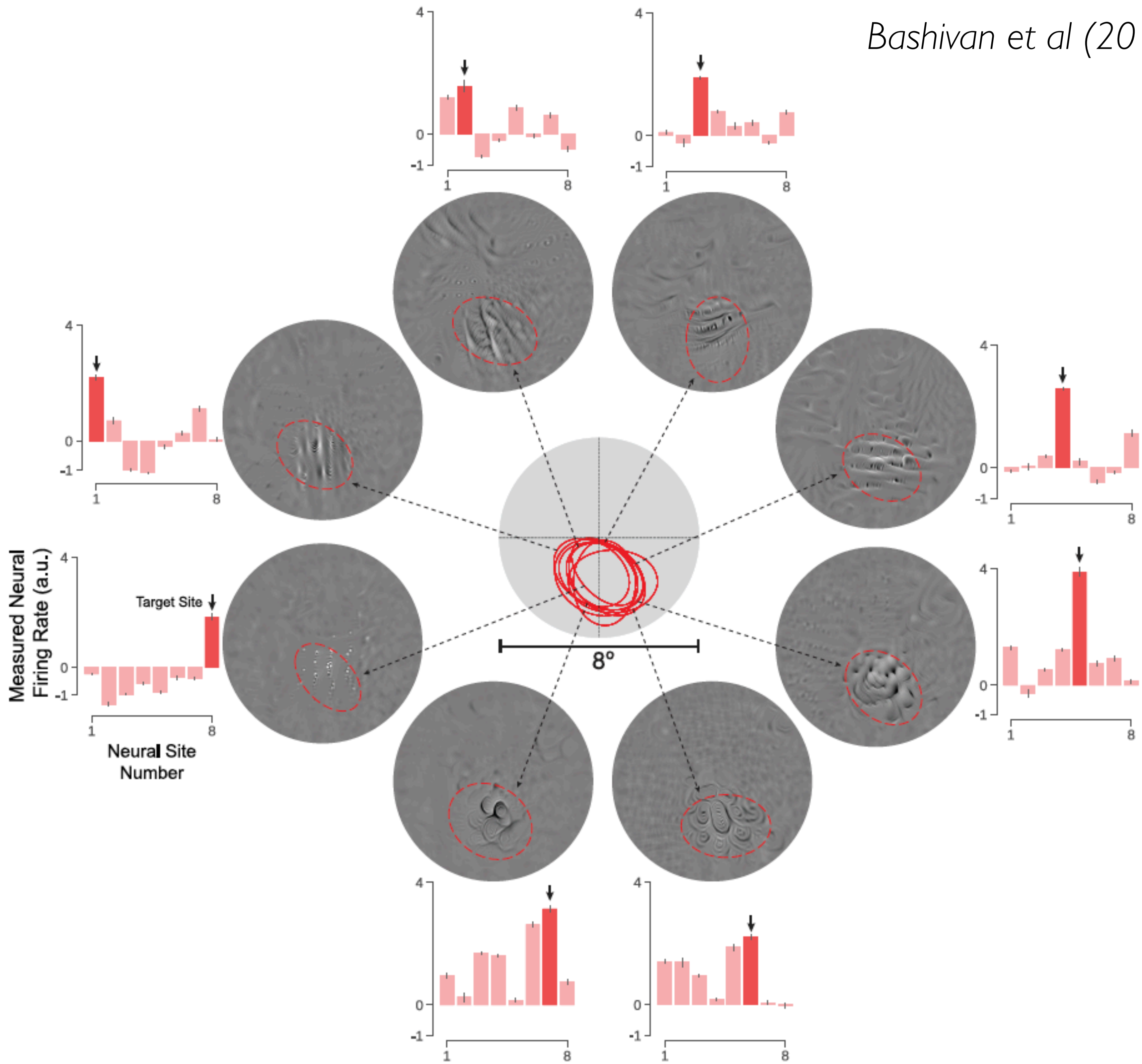
“Further Confirmation”

Bashivan et al (2019)



"Further Confirmation"

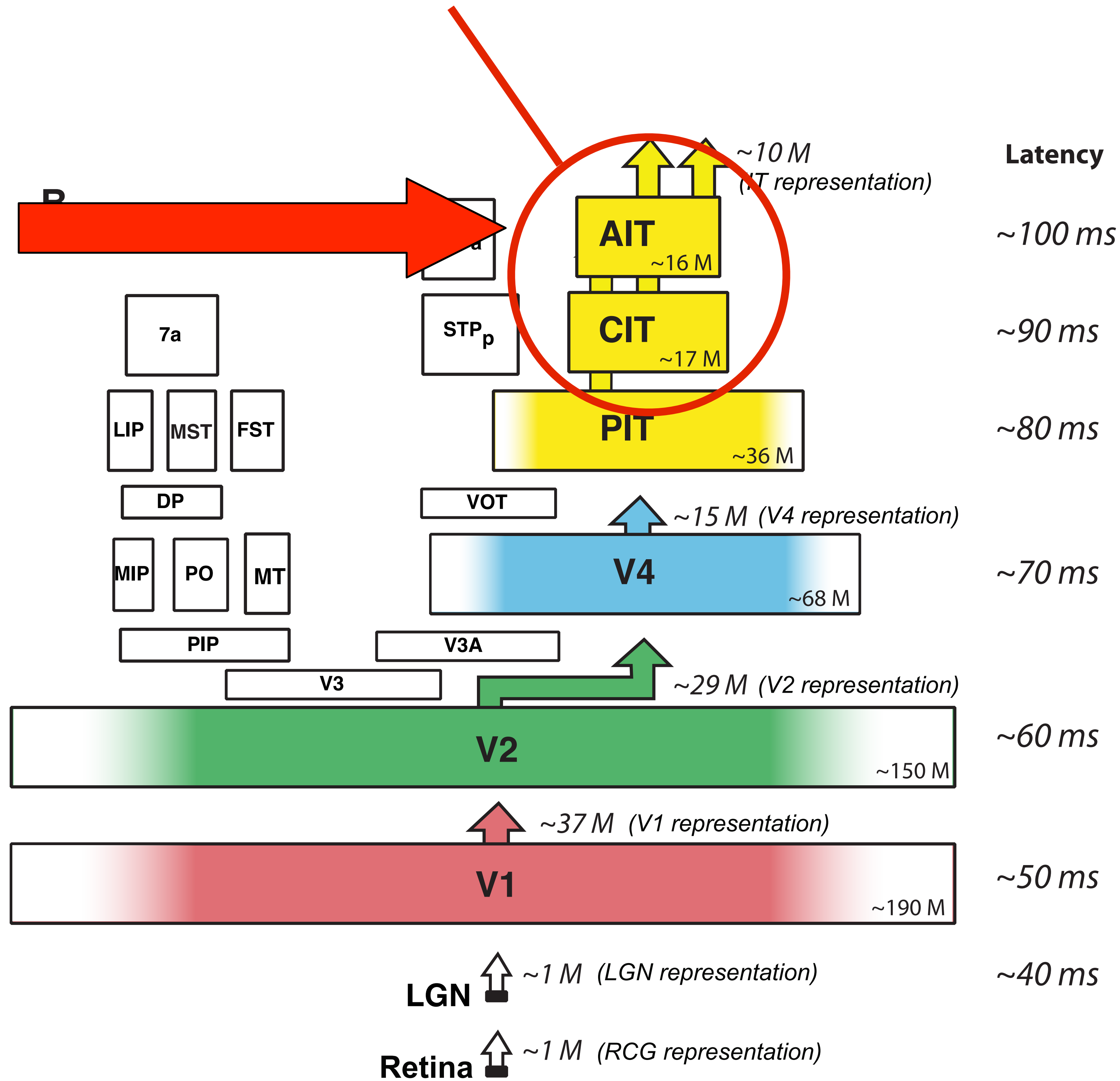
Fig. 4. Example of independent control of each neural site on a subset of V4 neural sites with highly overlapping cRFs. Controller images were synthesized to try to achieve a one-hot population over a population of eight neural sites (in each control test, the target neural site is shown in dark red and designated by an arrow). Despite highly overlapping receptive fields (center), most of the neural sites could be individually controlled to a reasonable degree. Controller images are shown along with the extended cRF (2 SD) of each site (red dashed ovals). Error bars denote 95% confidence interval.



Bashivan et al (2019)

"IT" (Inferior temporal cortex)

You are here.



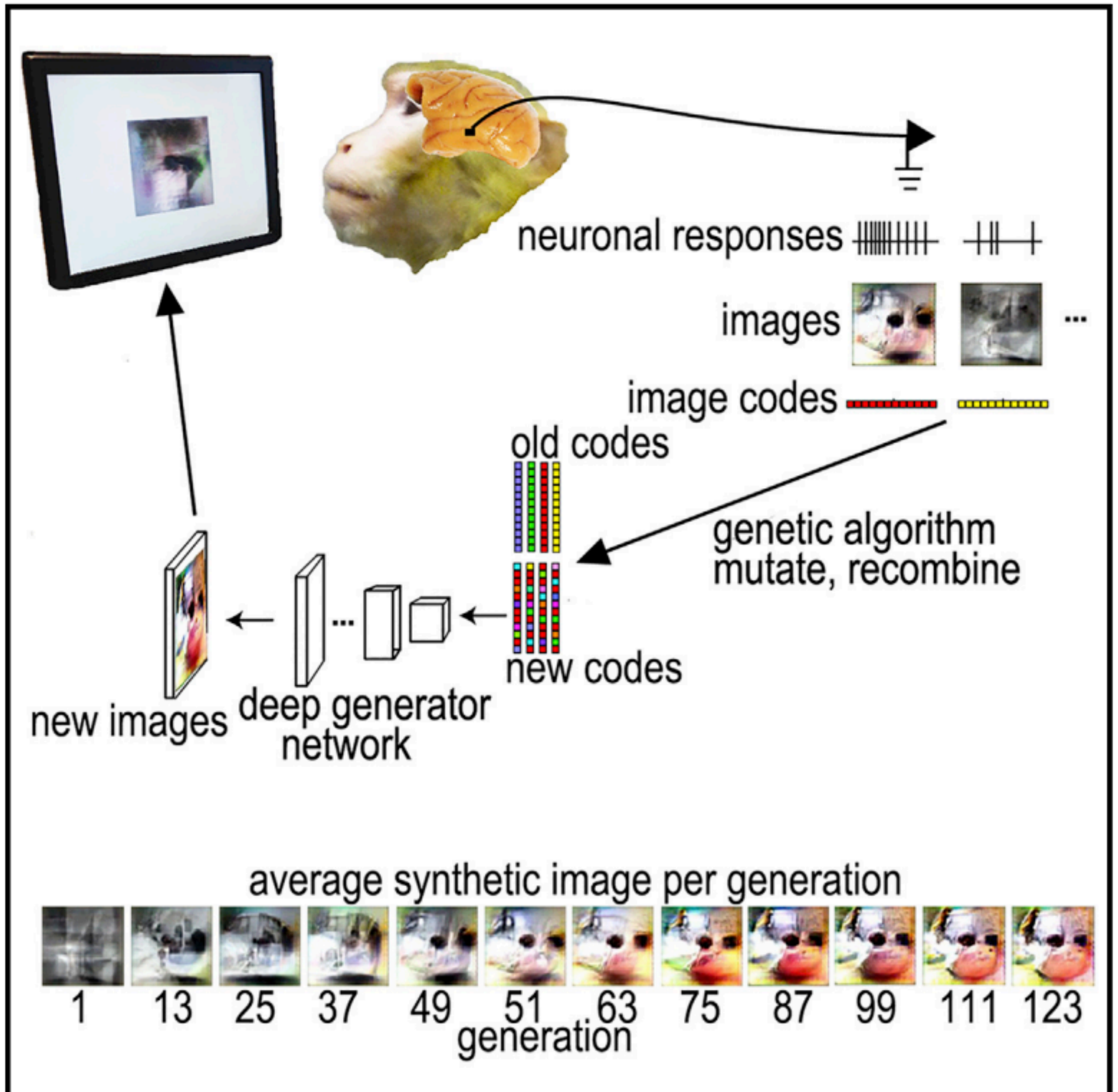
Adapted from DiCarlo et al. 2012

Cell

Article

Evolving Images for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences

Graphical Abstract



Authors

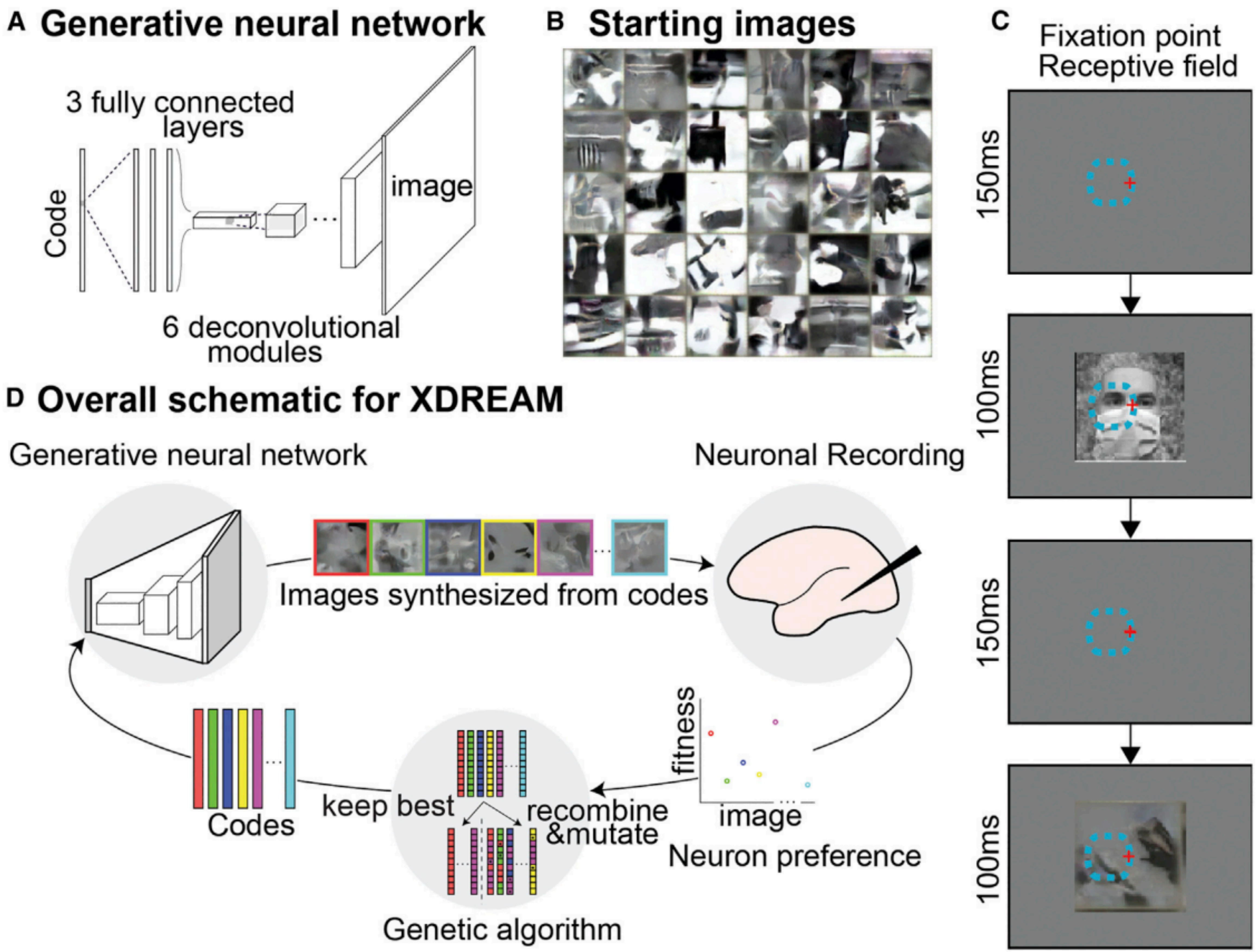
Carlos R. Ponce, Will Xiao,
Peter F. Schade, Till S. Hartmann,
Gabriel Kreiman, Margaret S. Livingstone

Correspondence

crponce@wustl.edu (C.R.P.),
mlivingstone@hms.harvard.edu (M.S.L.)

In Brief

Neurons guided the evolution of their own best stimuli with a generative deep neural network.



Ponce et al (2020)

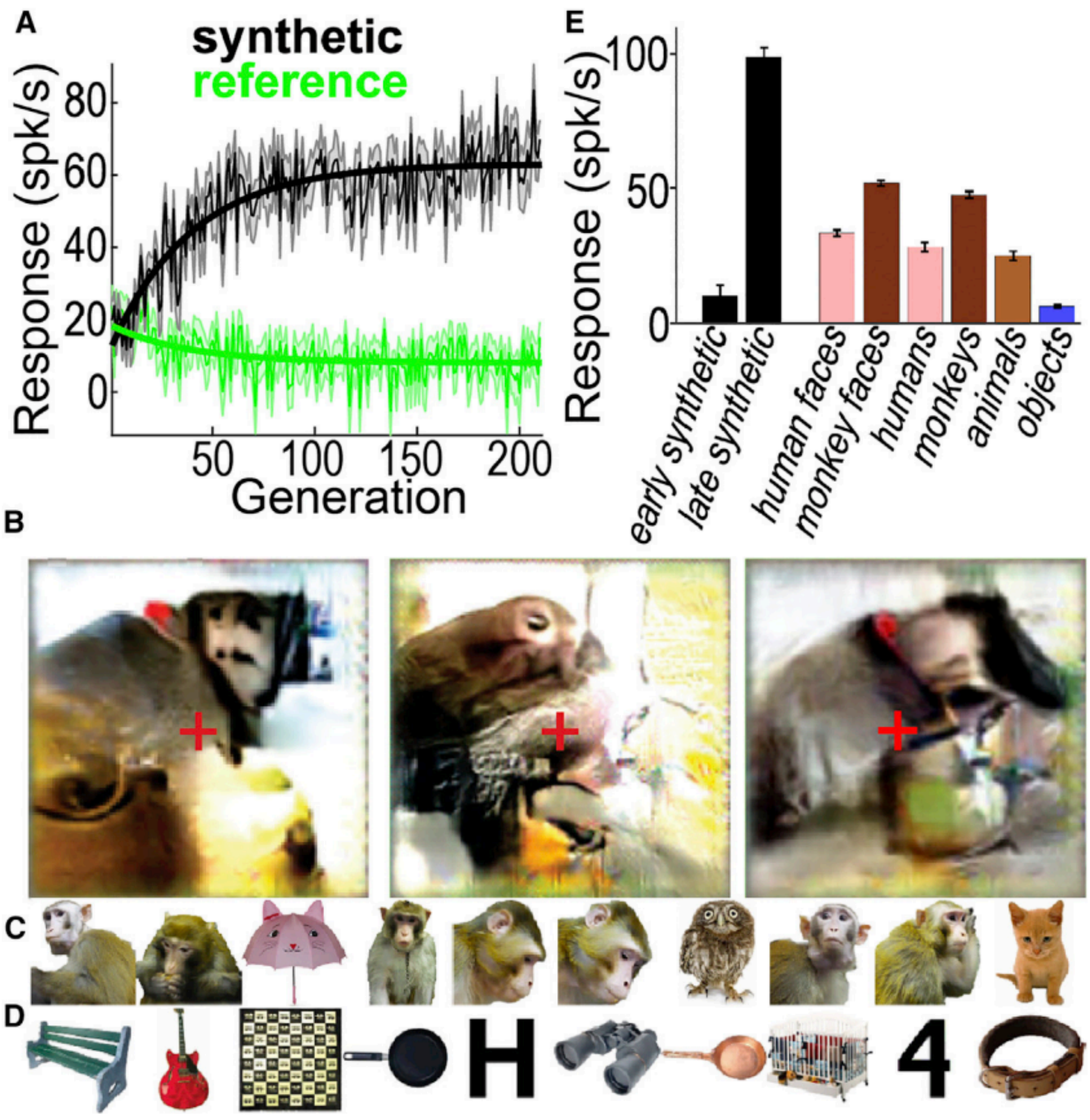


Figure 4. Evolution of Synthetic Images by Maximizing Responses of Single Neuron Ri-10, Same Unit as Figure 3

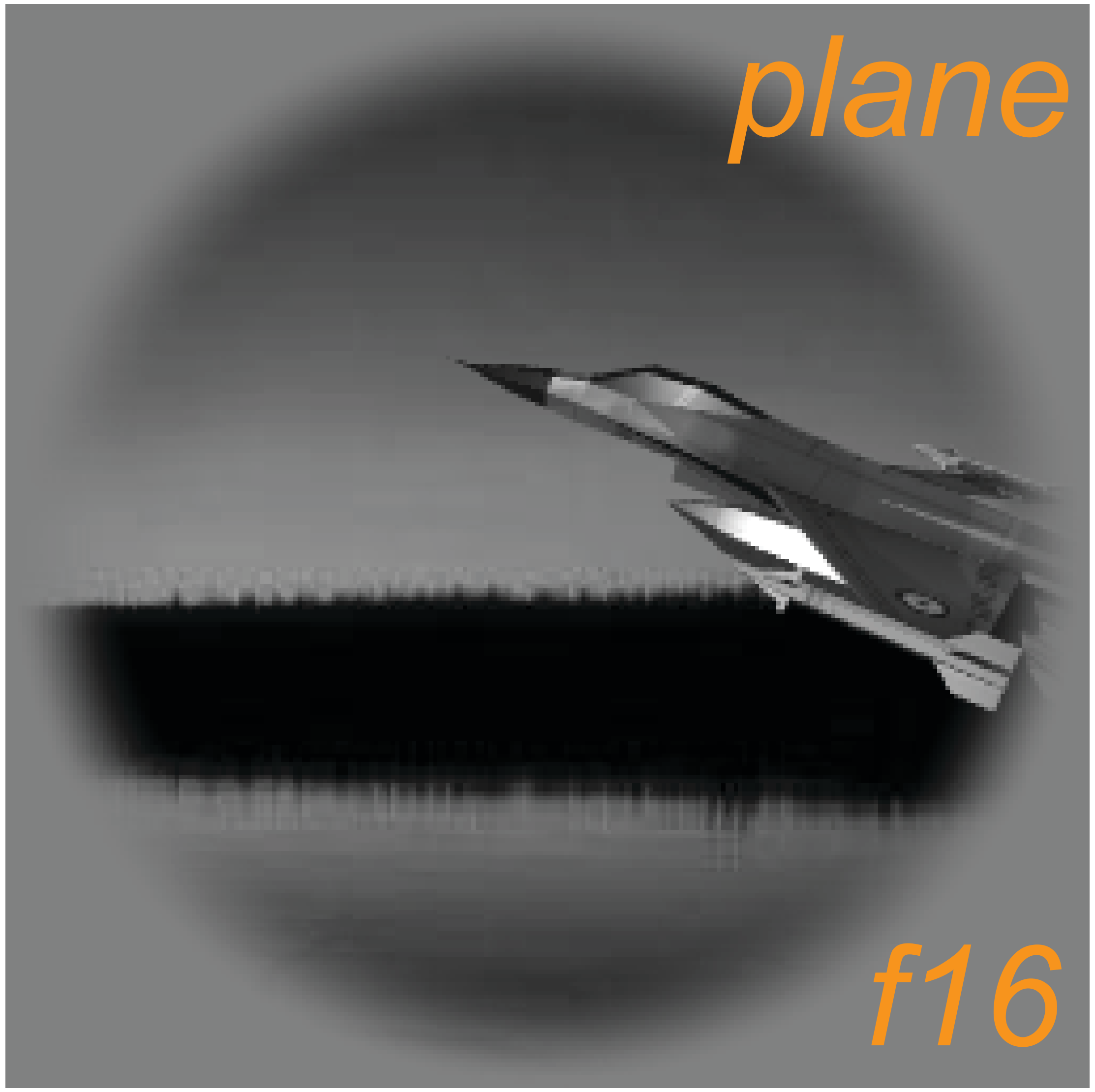
(A) Mean response to synthetic (black) and reference (green) images for every generation (spikes per s ± SEM). Solid straight lines show an exponential fit to the response over the experiment.

(B) Last-generation images evolved during three independent evolution experiments; the leftmost image corresponds to the evolution in (A); the other two evolutions were carried out on the same single unit on different days. Red crosses indicate fixation. The left half of each image corresponds to the contralateral visual field for this recording site. Each image shown here is the average of the top 5 images from the final generation.

(C–E) Selectivity of this neuron to 2,550 natural images. (C) In (C) are the top 10 images from this image set for this neuron. (D) In (D) are the worst 10 images from this image set for this neuron. The entire rank ordered natural image set is shown in Figure S2. (E) In (E) is the selectivity of this neuron to different image categories (mean ± SEM). The entire image set comprised 2,550 natural images plus selected synthetic images. Early synthetic is defined as the best image from each of the first 10 generations and late from the last 10. Each image response is the average over 10–12 repeated presentations. See Figure S3 for additional independent evolutions from this site.

Ponce et al (2020)

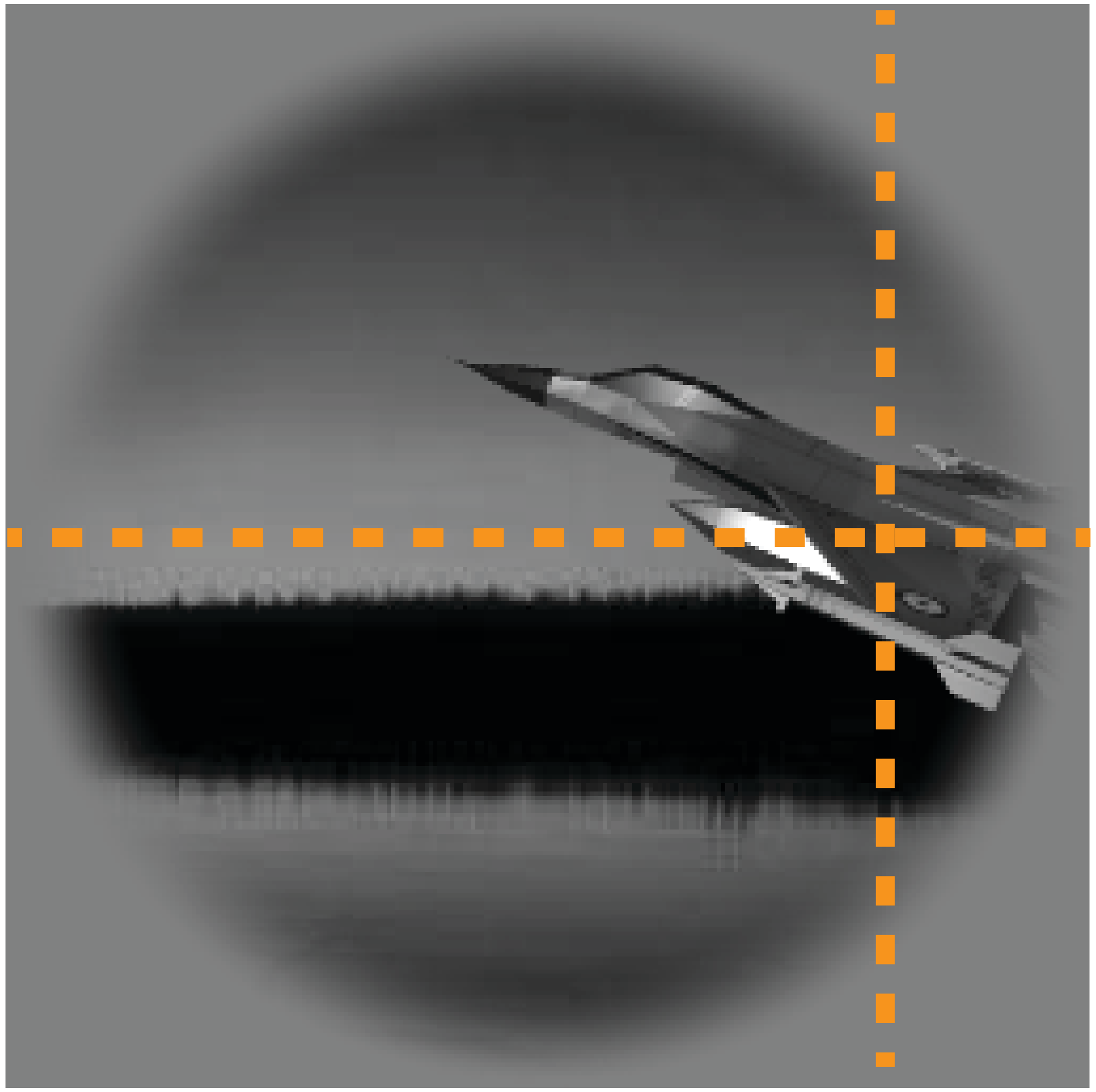
Beyond categorization



Category

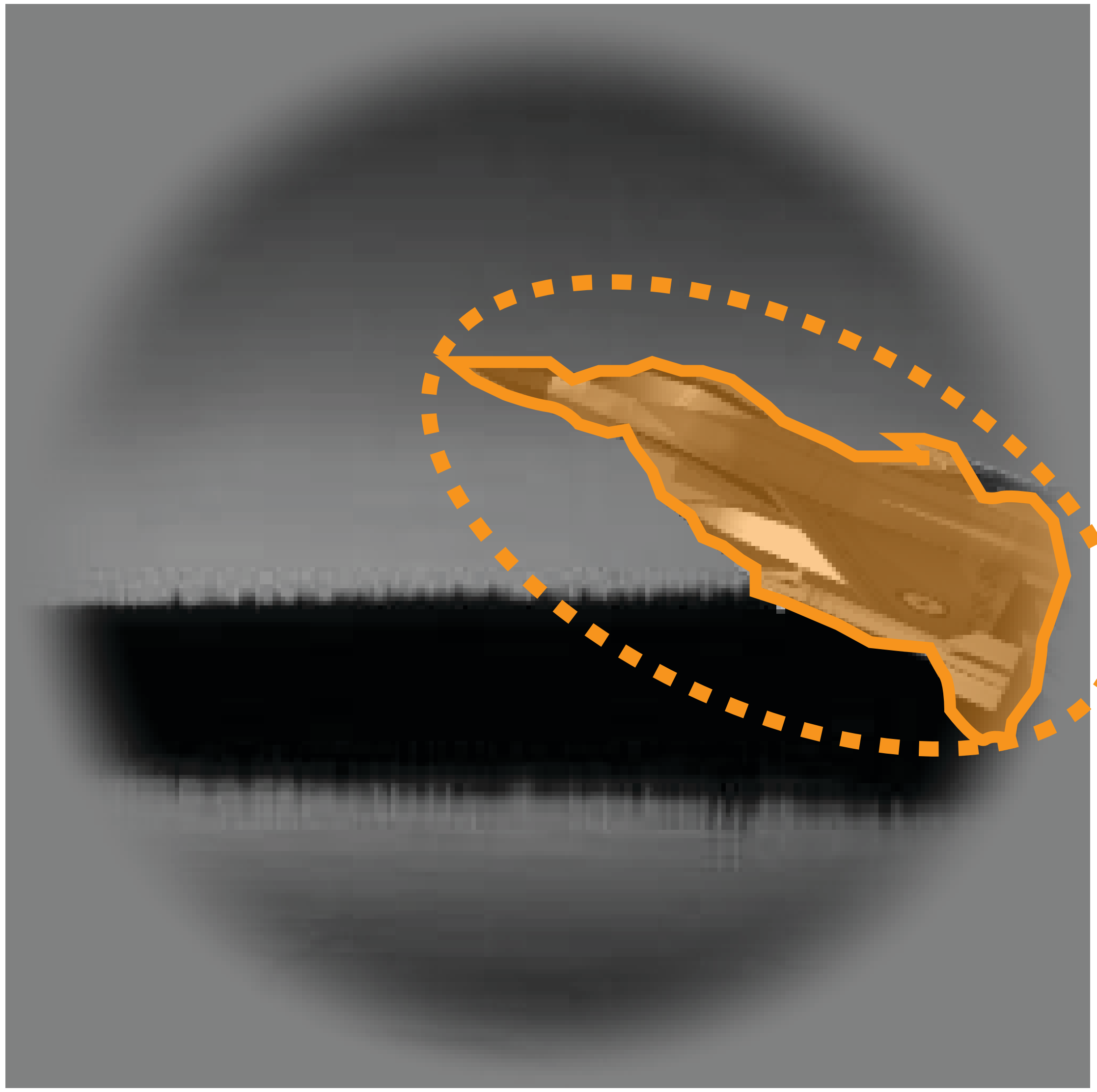
Identity

Beyond categorization

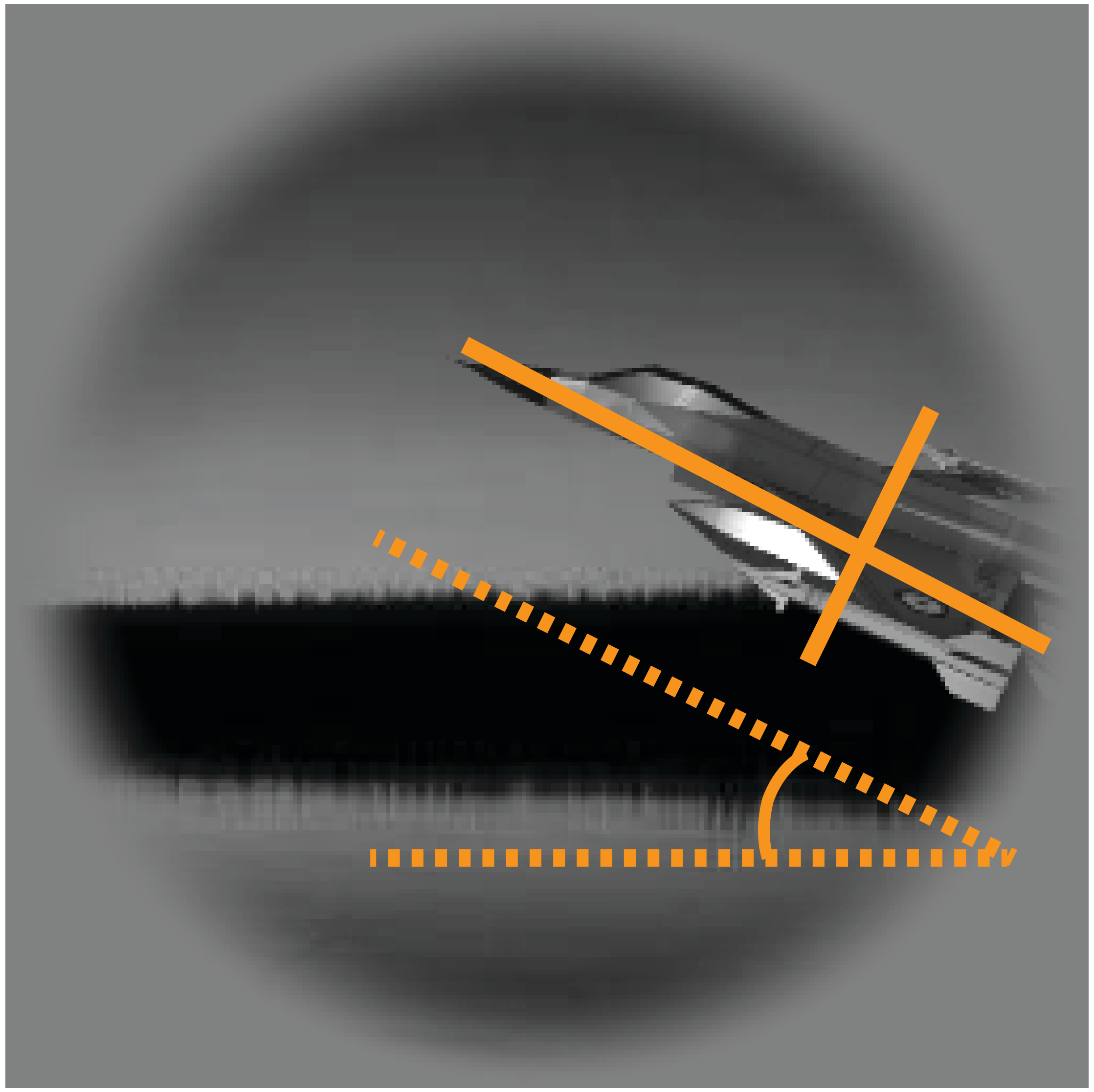


Position

Beyond categorization



Beyond categorization



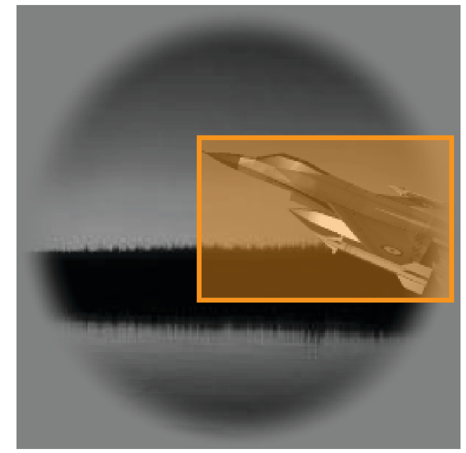
Aspect Ratio and Angle

Beyond categorization

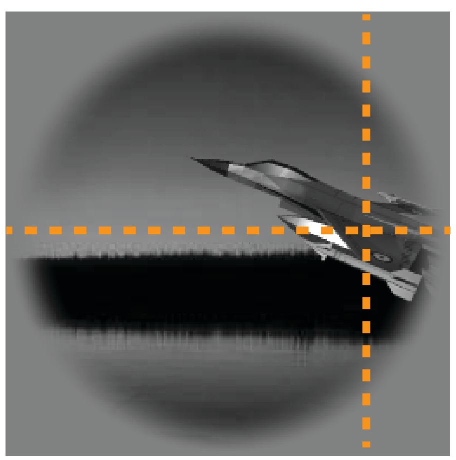
We can quickly assess the scene as a whole.

Category

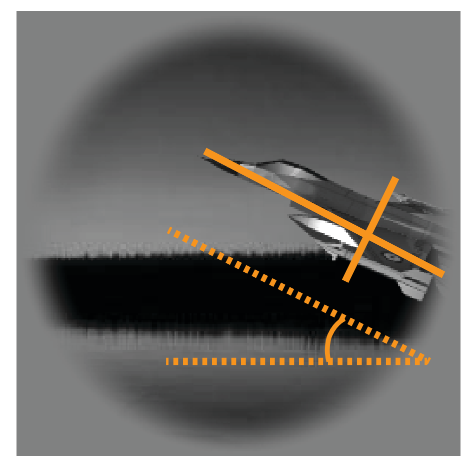
Identity



Bounding Box



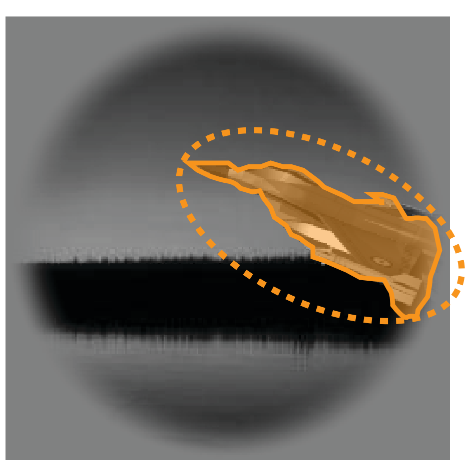
X and Y Axis Position



Aspect Ratio

Major Axis Length

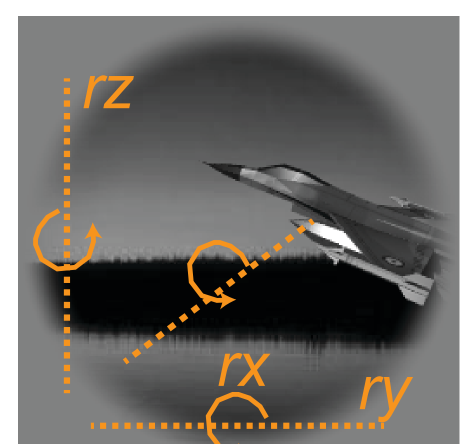
Major Axis Angle



Perimeter

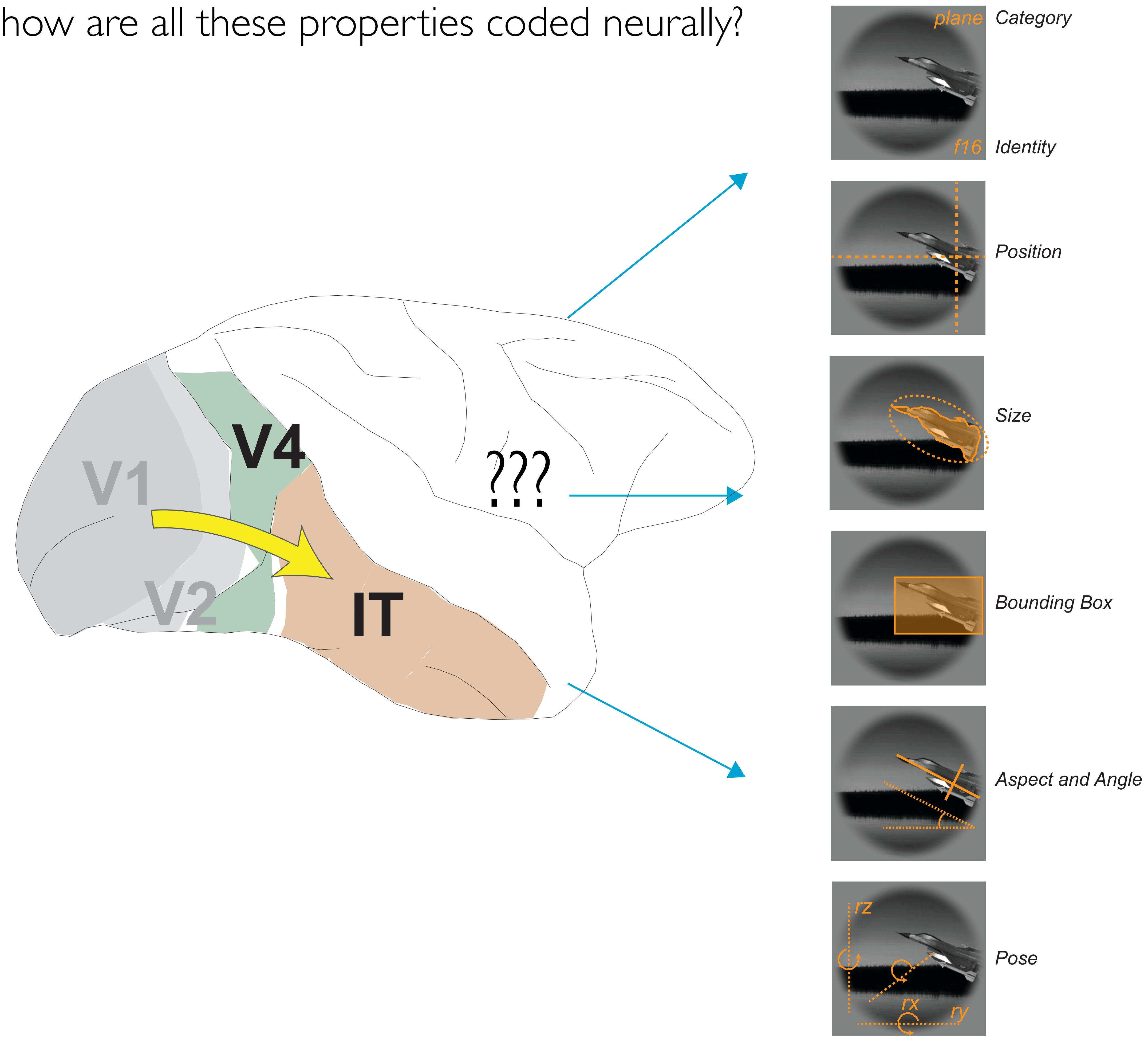
2-D Retinal Area

3-D Object Scale



Pose in each axis

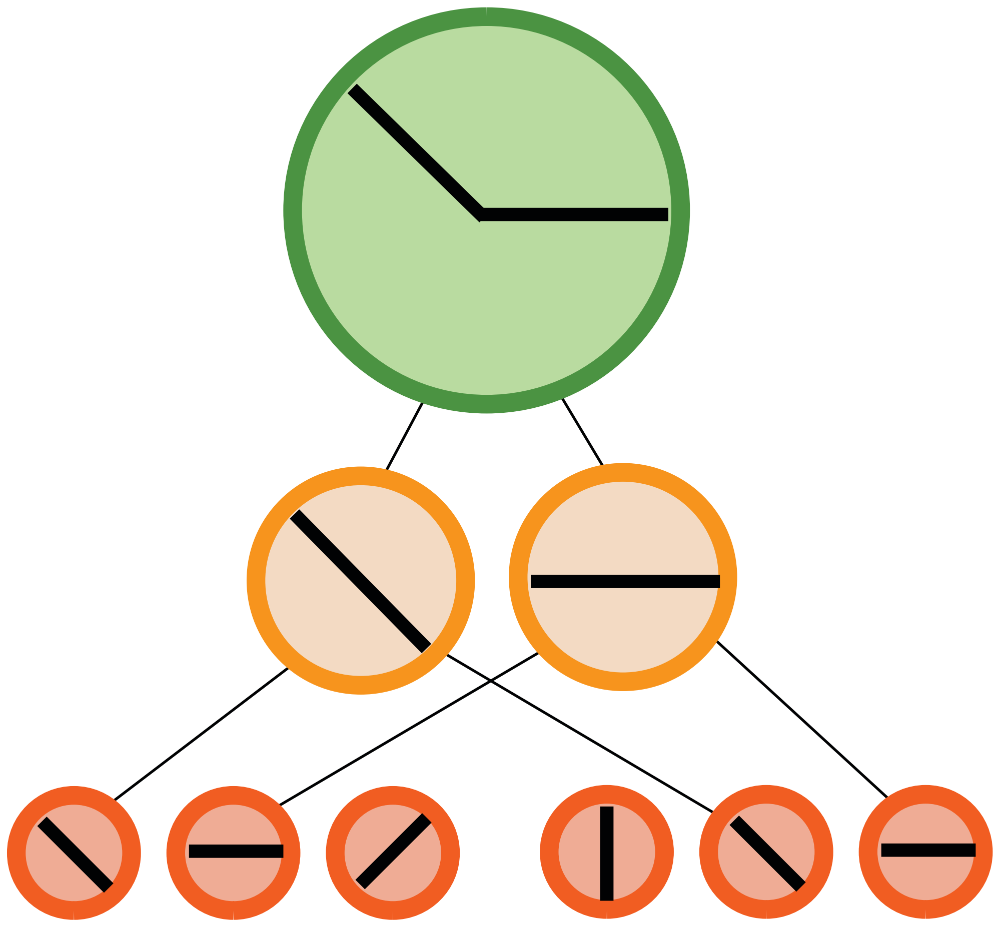
Where and how are all these properties coded neurally?



Beyond categorization

"Standard word model" predicts: **not at the top of the ventral stream**.

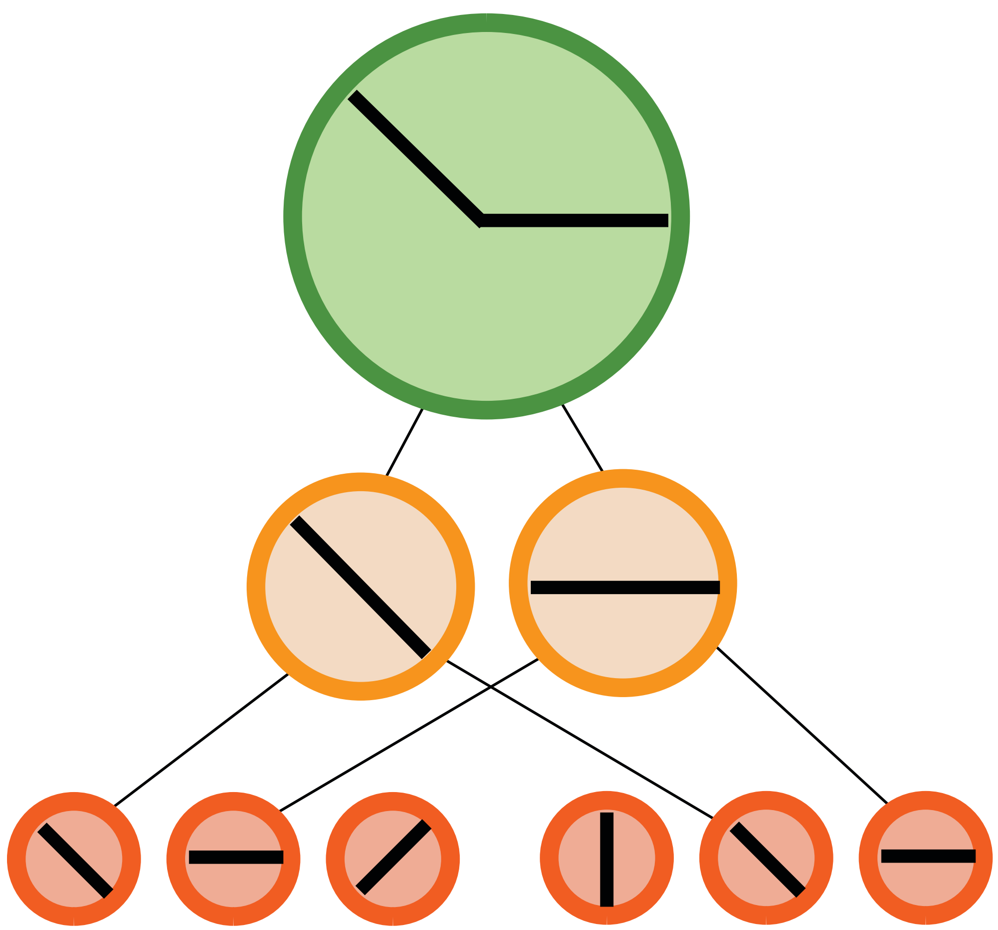
Aggregation over identity-preserving transformations, e.g. translation.



Beyond categorization

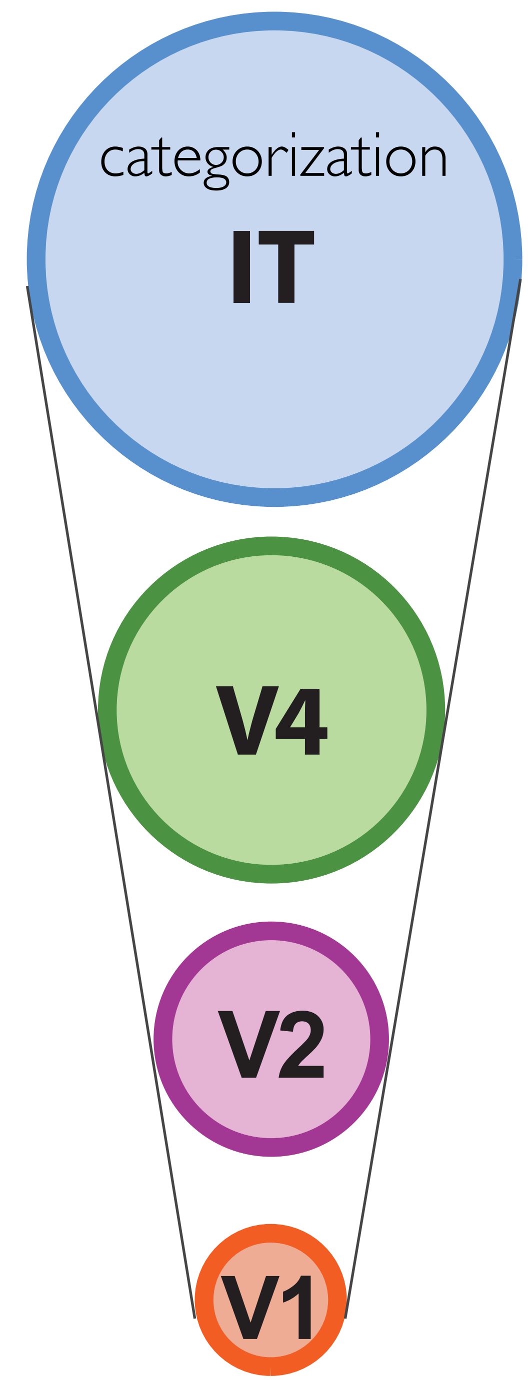
"Standard word model" predicts: **not at the top of the ventral stream**.

Aggregation over identity-preserving transformations, e.g. translation.



Receptive Field Size ↑

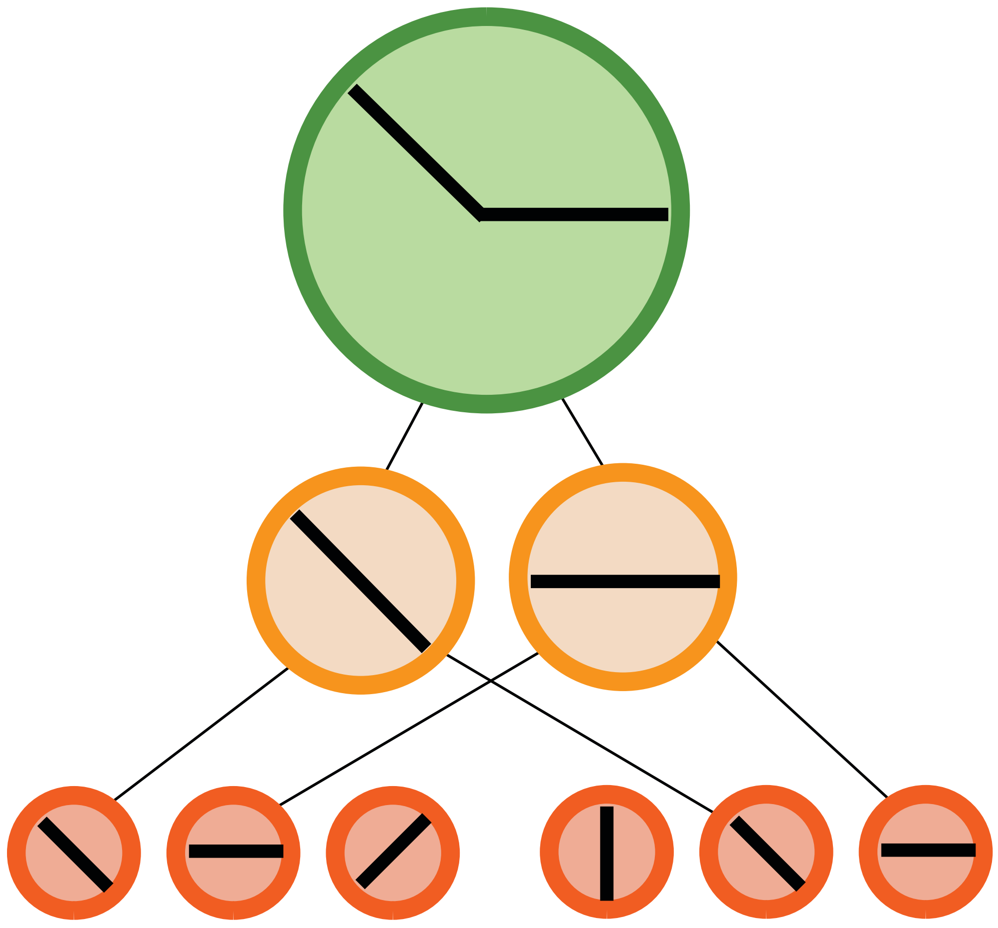
Category Invariance ↑



Beyond categorization

"Standard word model" predicts: **not at the top of the ventral stream**.

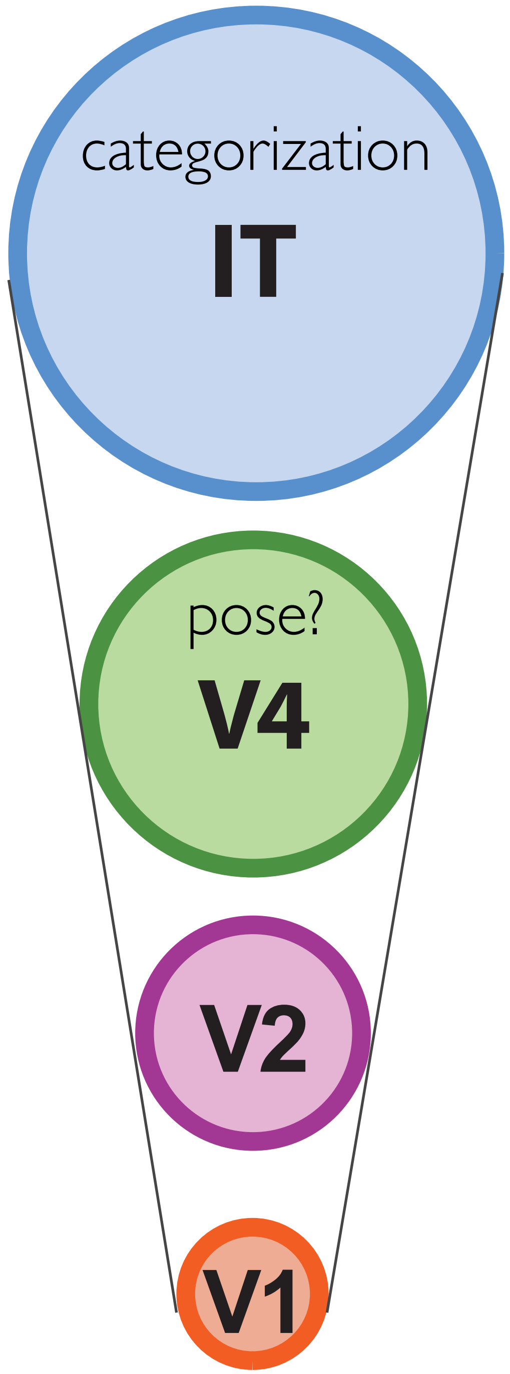
Aggregation over identity-preserving transformations, e.g. translation.



Receptive Field Size ↑

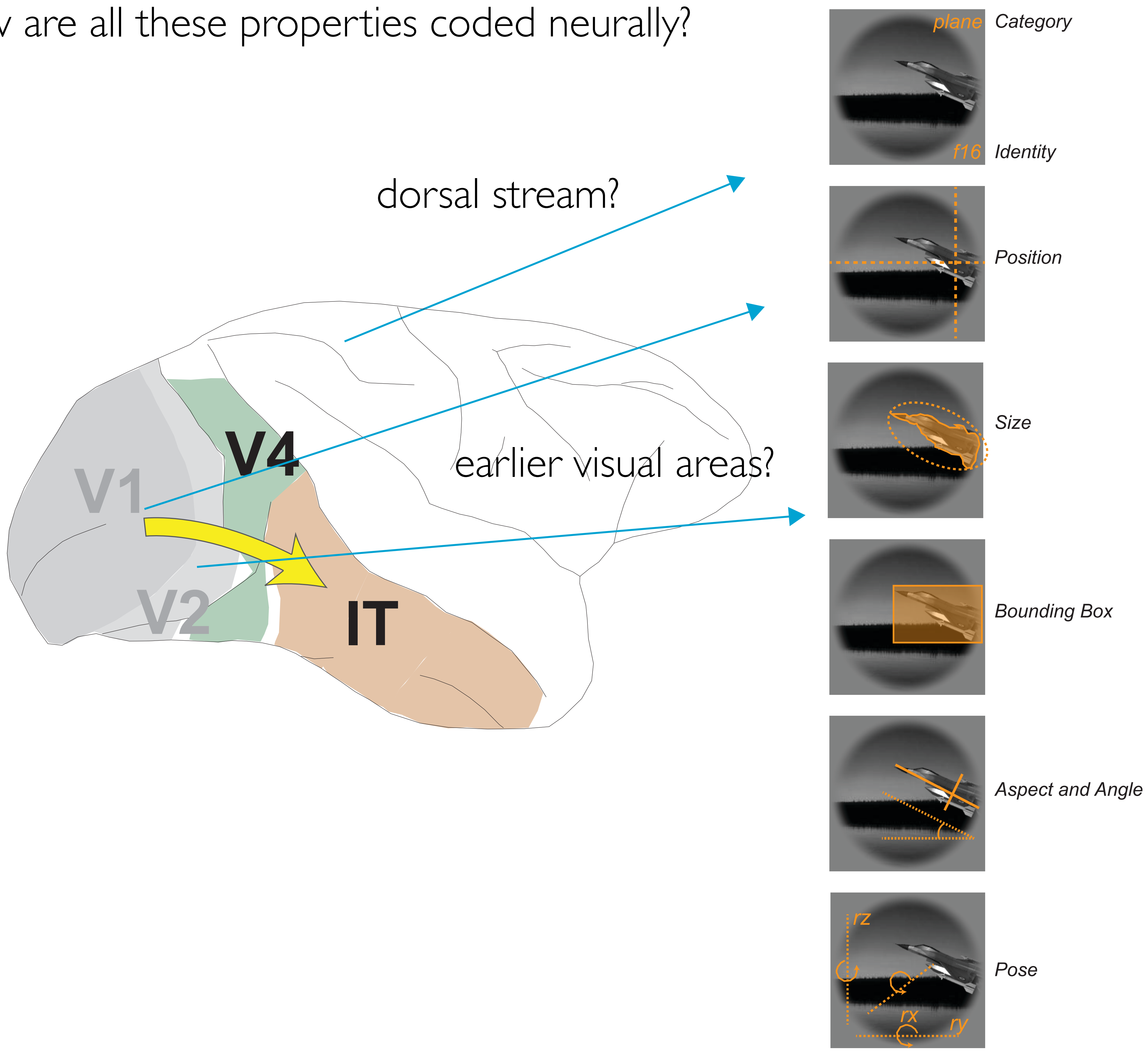
Category Invariance ↑

(e.g.) Position Sensitivity ↓

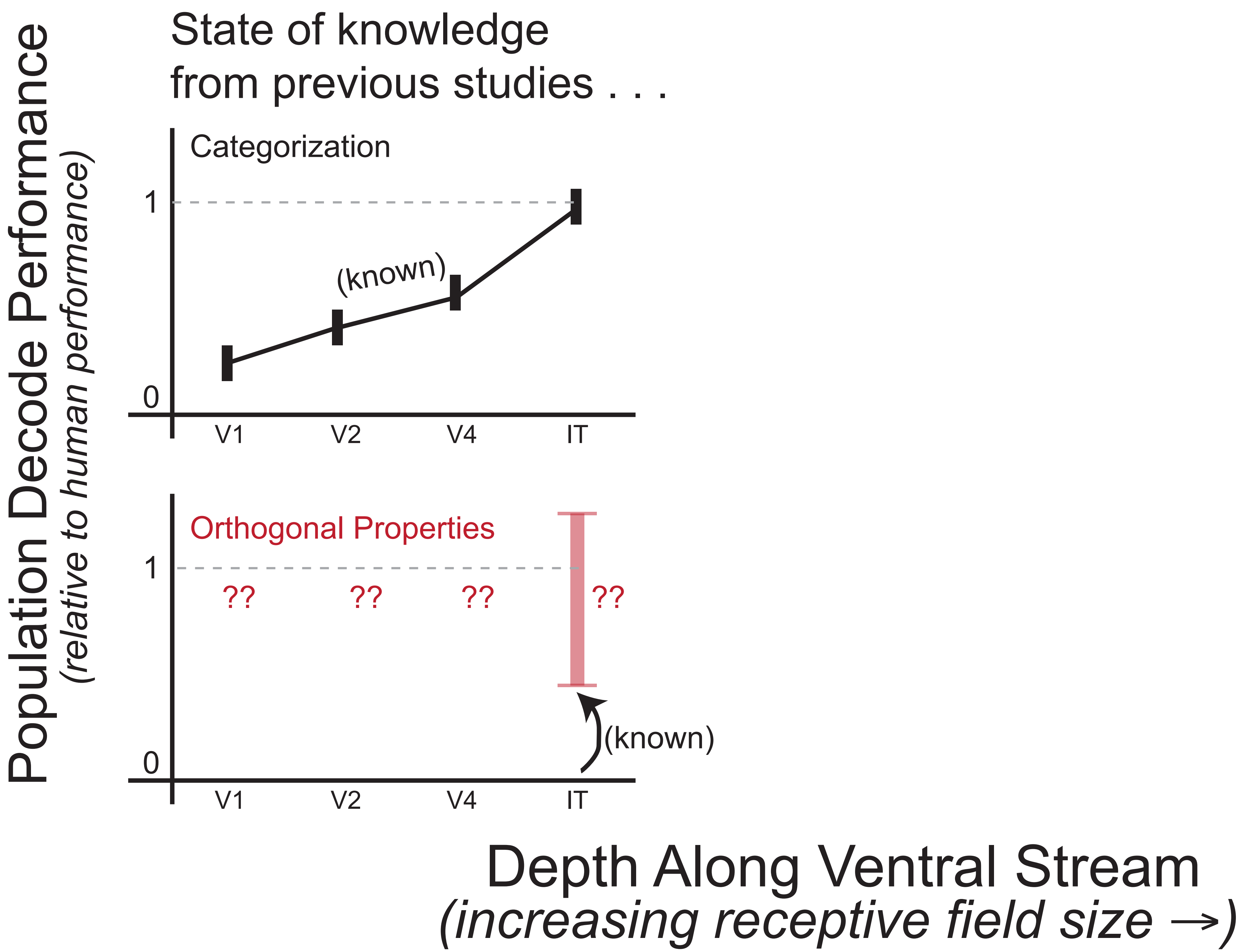


position / size estimation

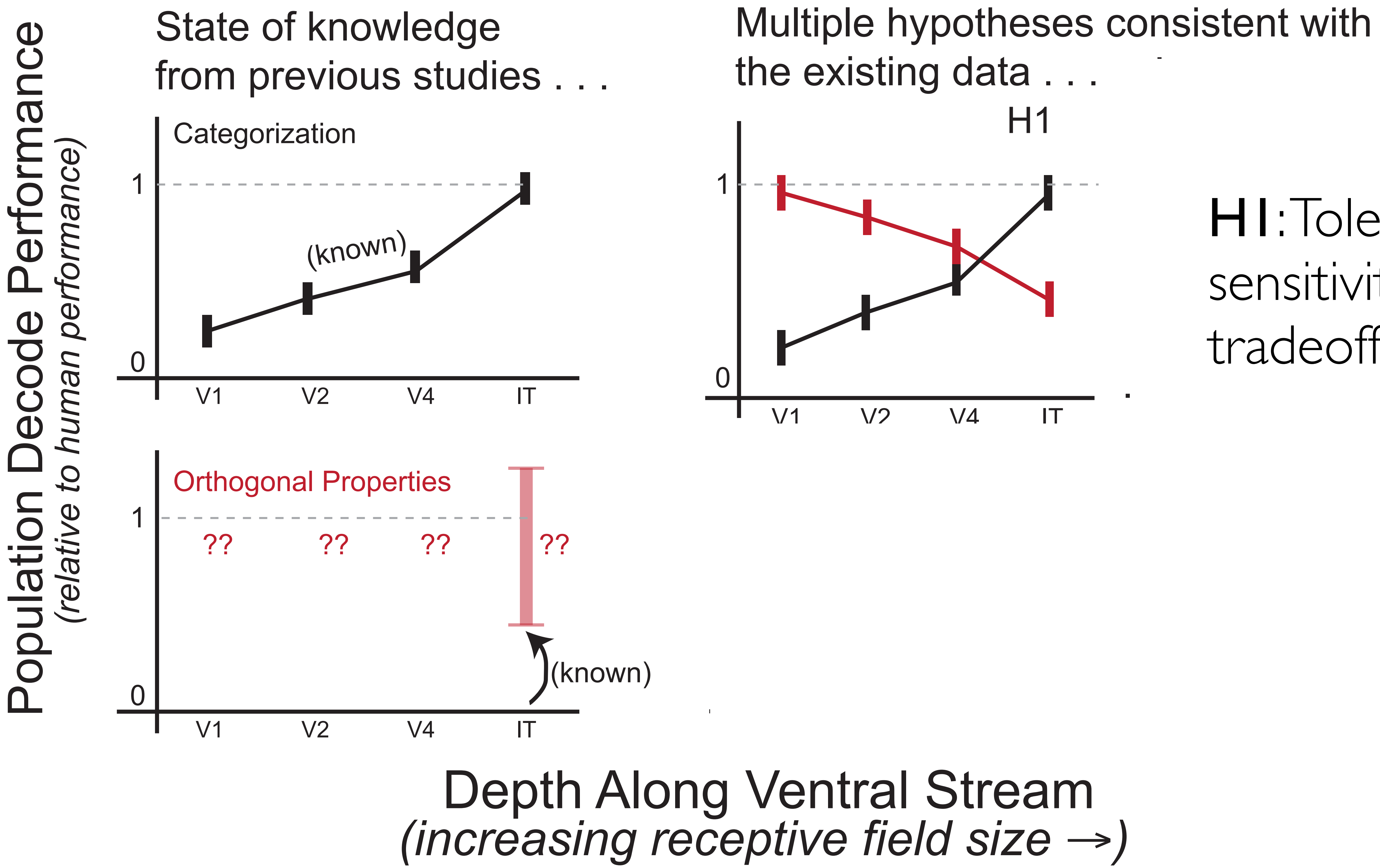
Where and how are all these properties coded neurally?



Somewhat newish ideas about IT?

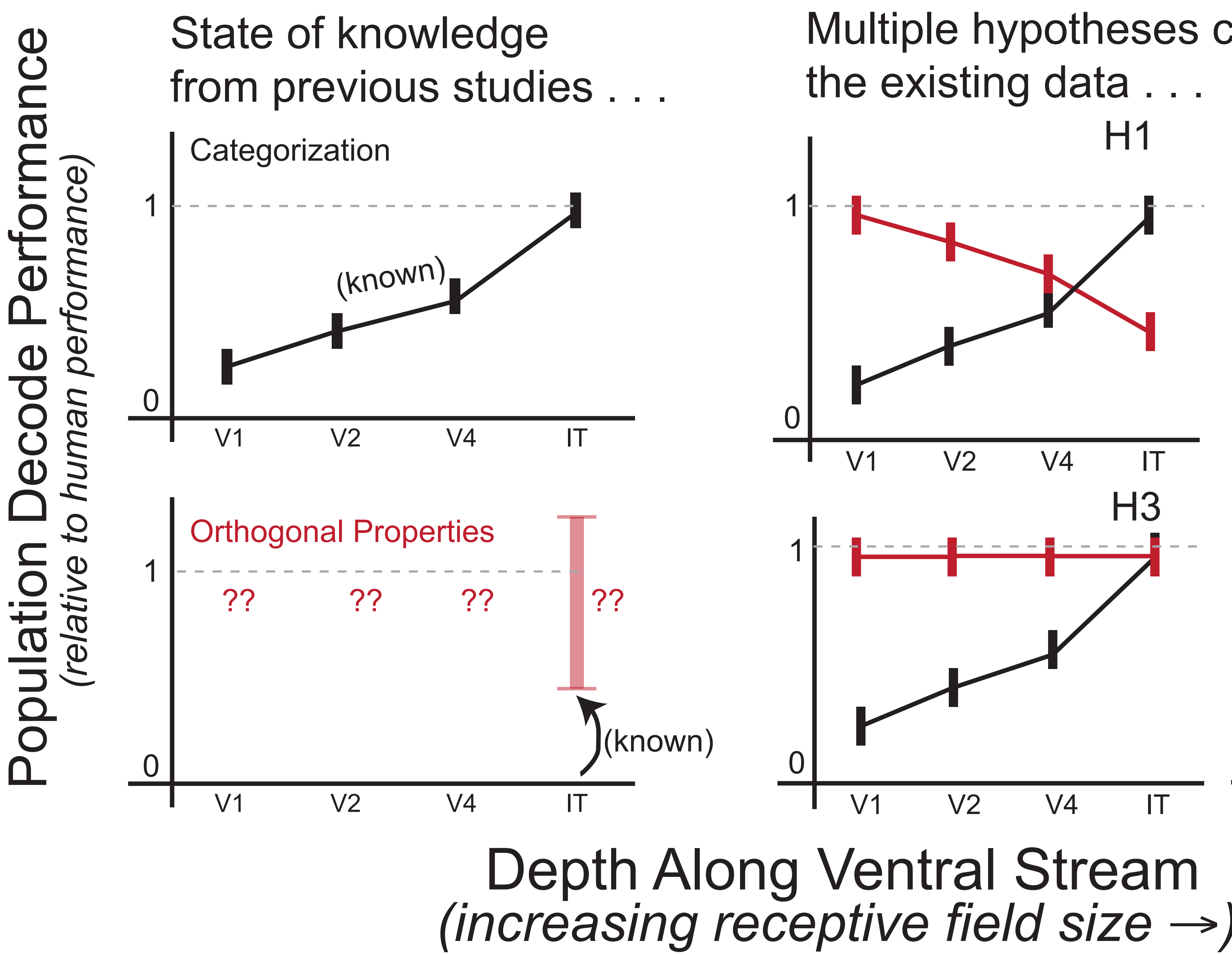


Somewhat newish ideas about IT?



H1: Tolerance / sensitivity tradeoff?

Somewhat newish ideas about IT?

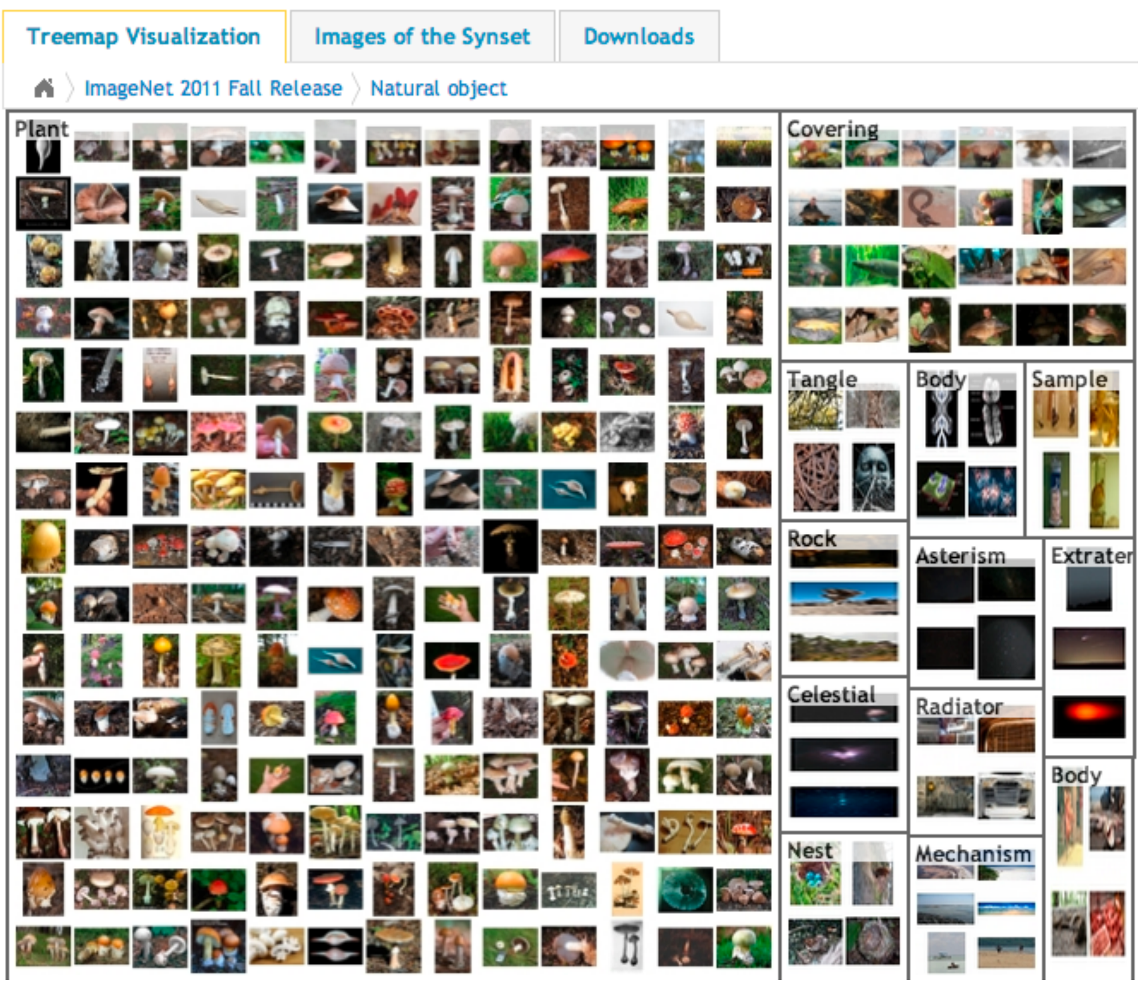


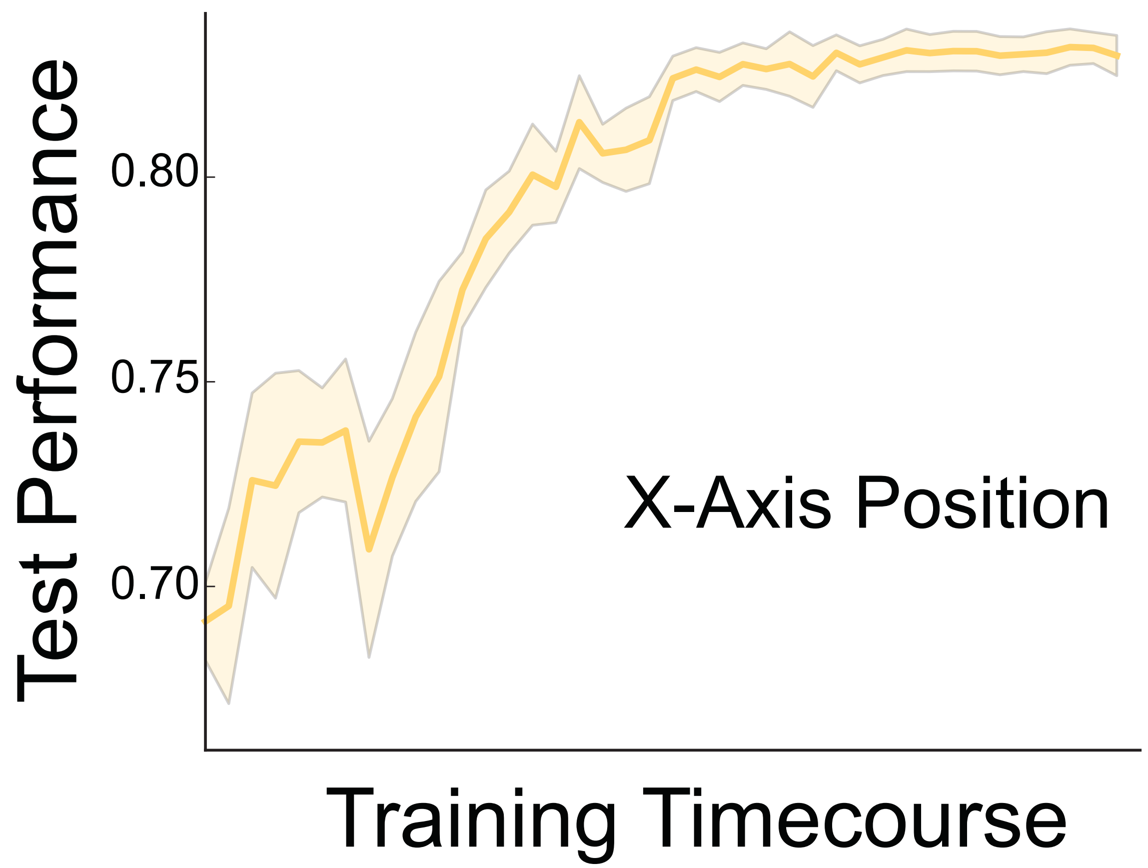
H3: Information preservation?

Beyond categorization

Unexpected observation:

Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*



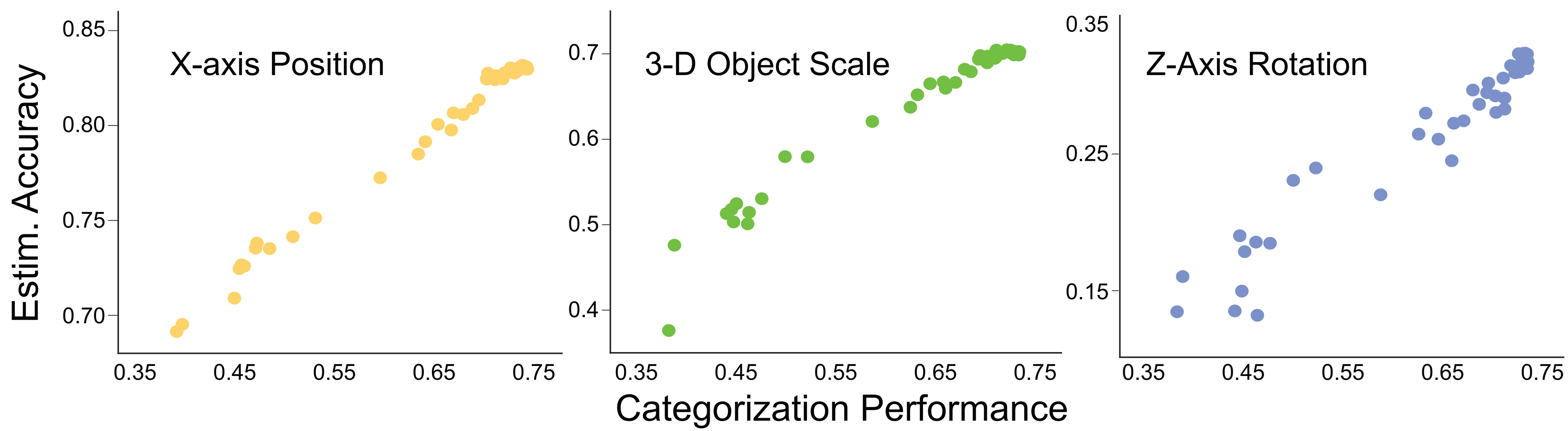


Training on categorization task → *Increased* performance on position estimation task.

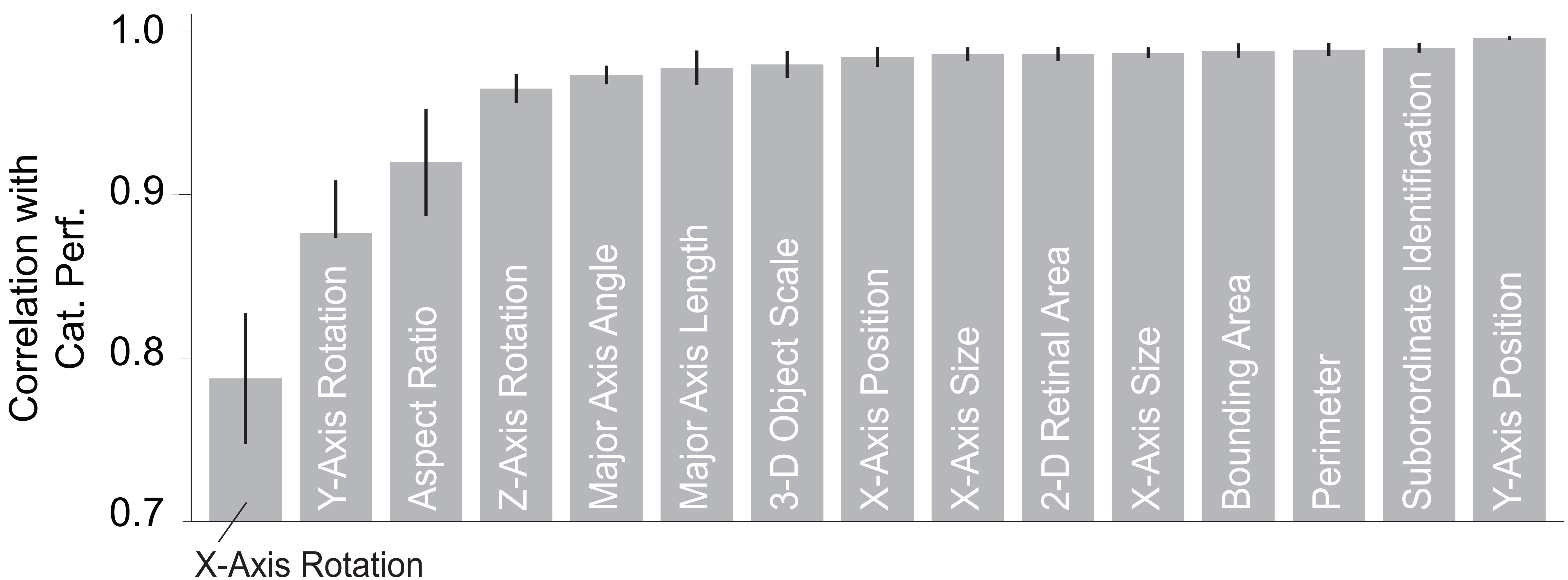
even though the goal was to become INVARIANT to position

Beyond categorization

Category optimization → improved performance on non-categorical tasks.

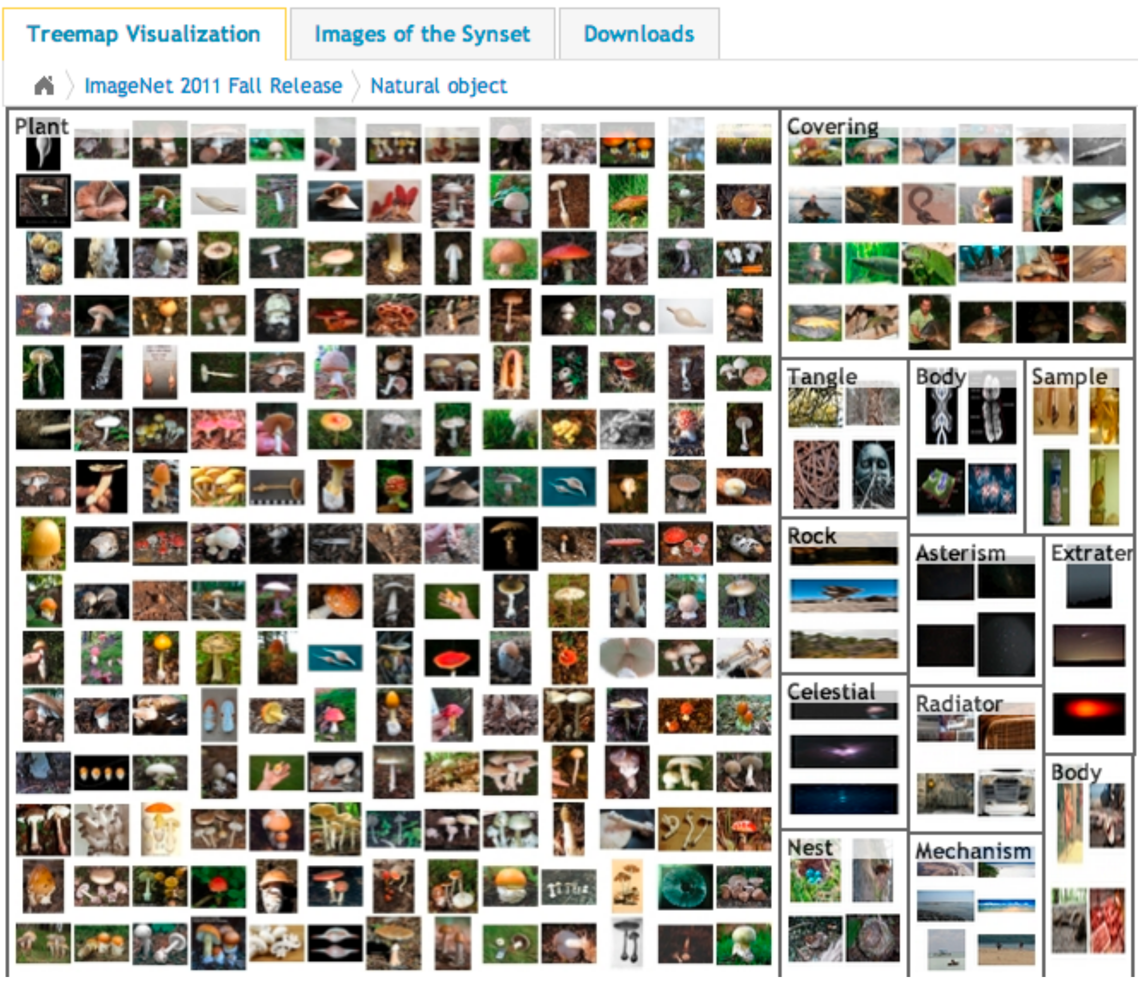


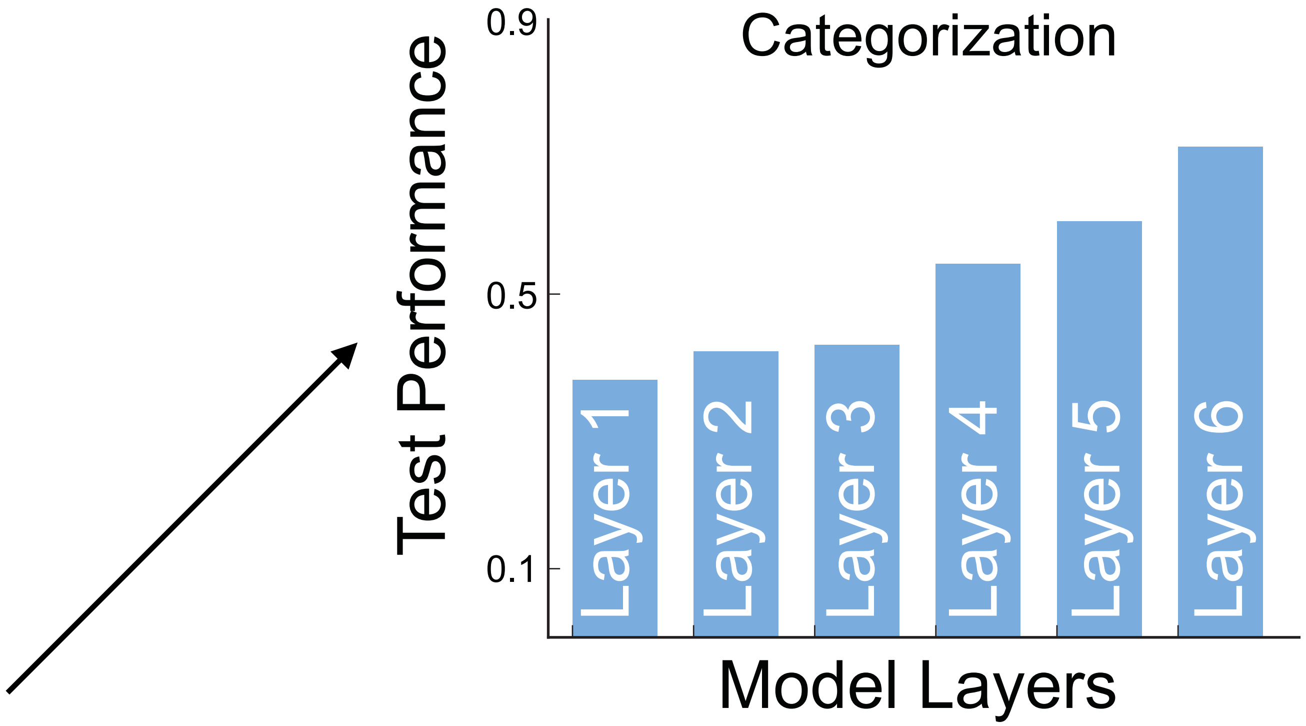
Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*



Beyond categorization

Unexpected observation #2:

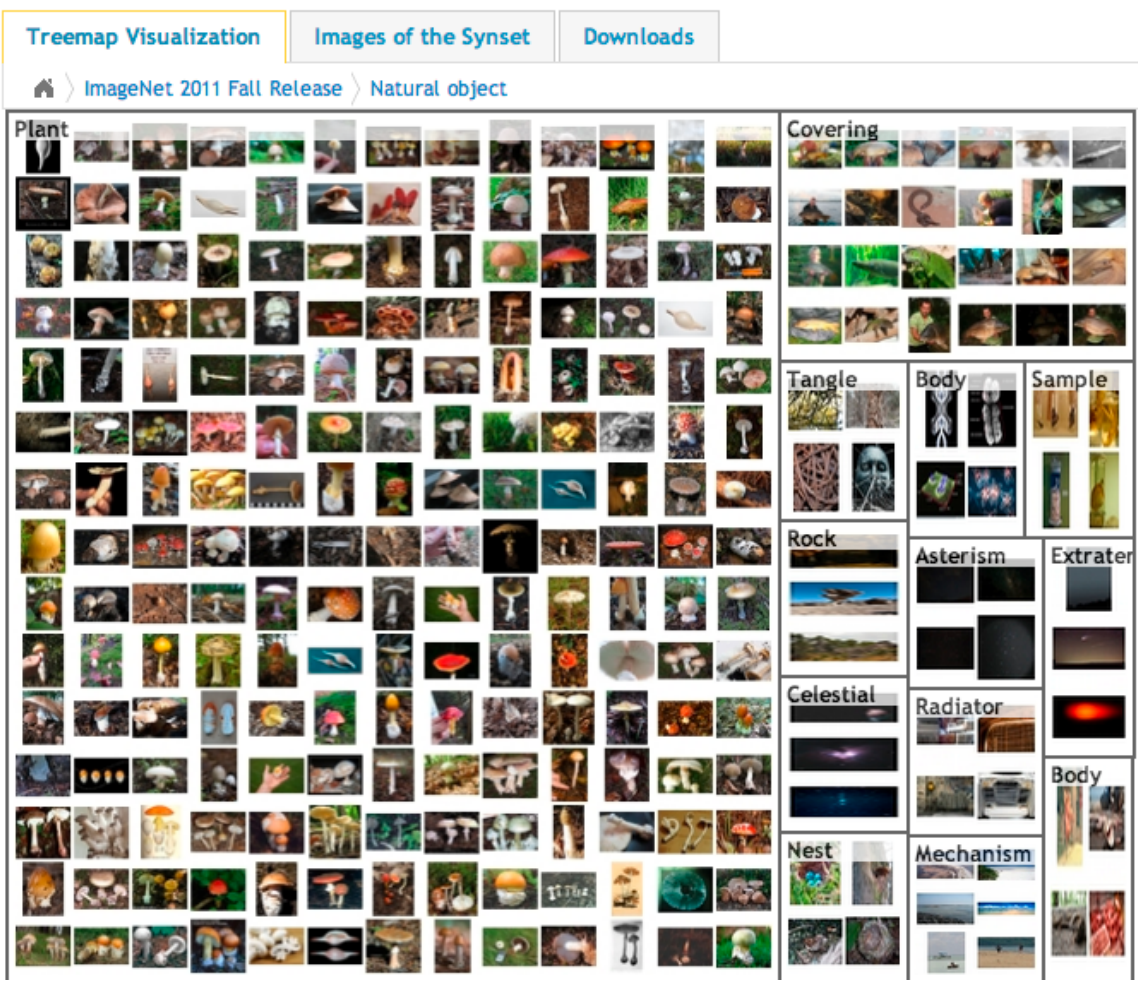


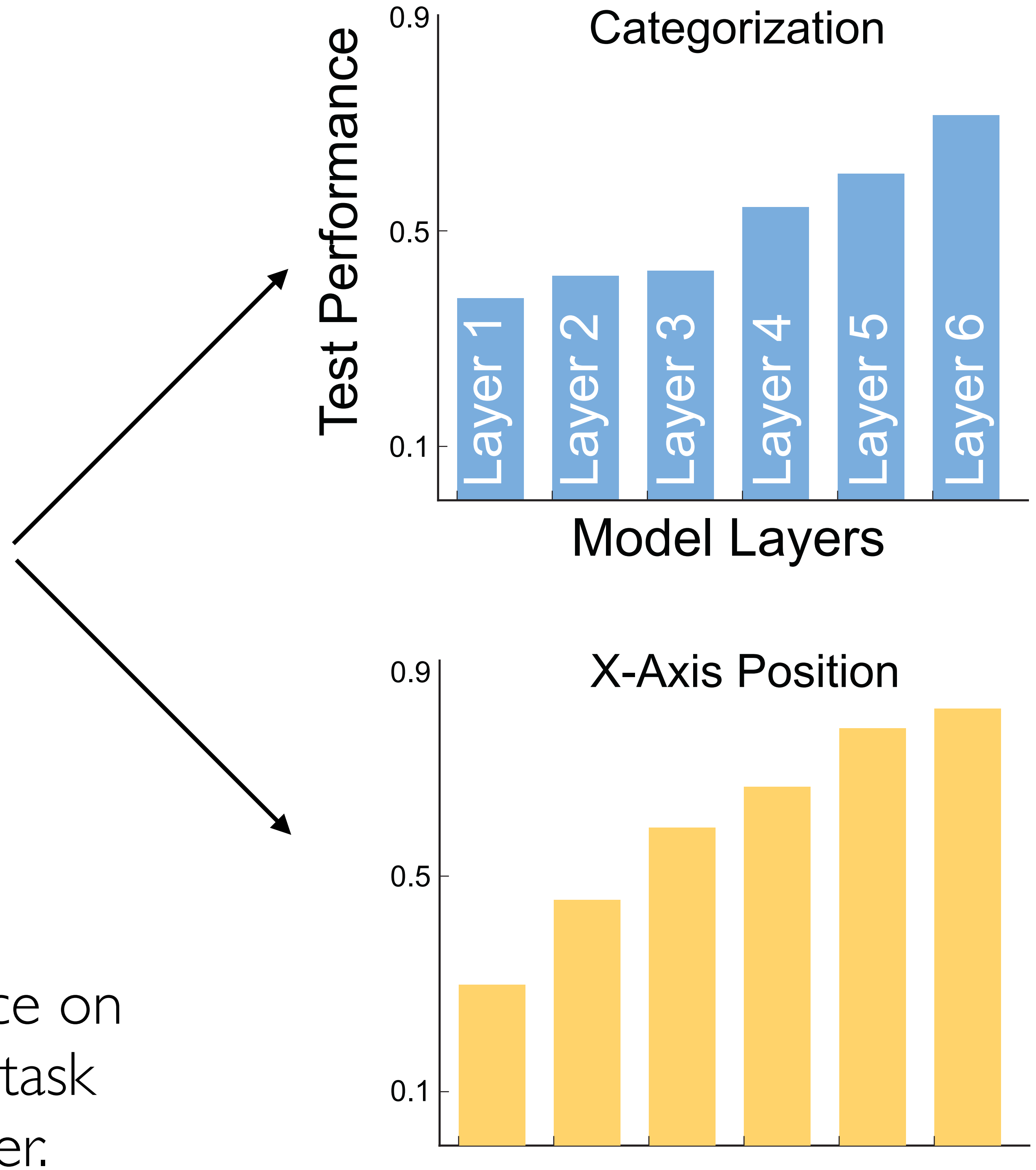


Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

Beyond categorization

Unexpected observation #2:



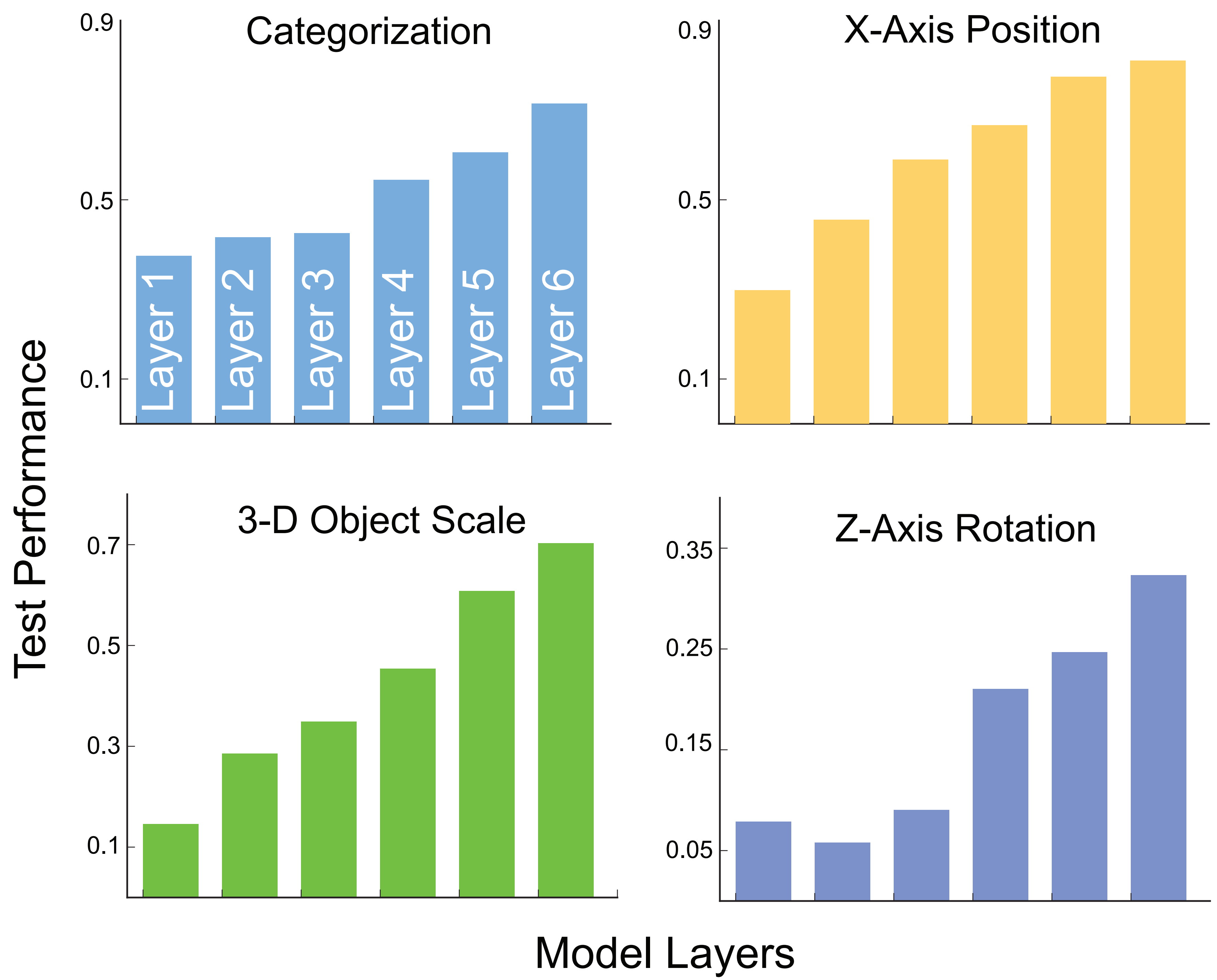


Increased performance on position estimation task at each model layer.

Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

Beyond categorization

For all tasks of visual interest we could measure in our test dataset:

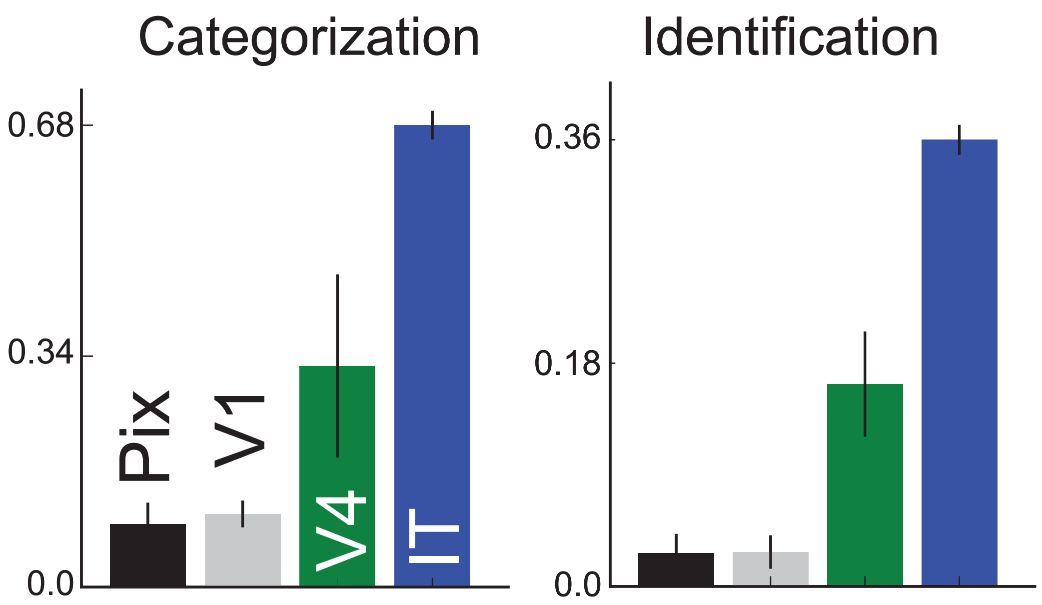


Performance on non-categorical tasks increases at each layer.

Beyond categorization

What do the data say?

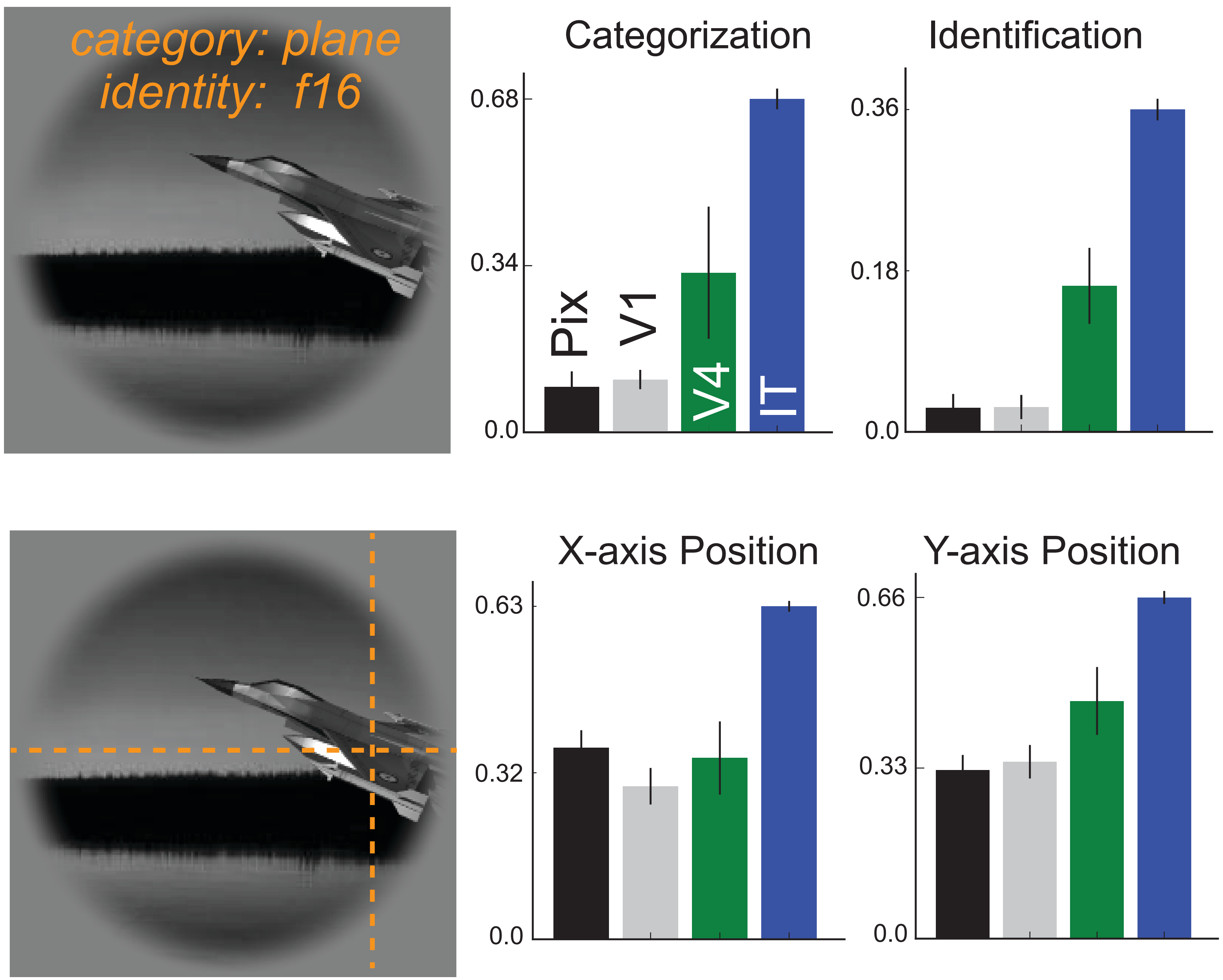
Population Decoding



Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

IT cortex
V4 cortex
V1-like model
pixel control

Population Decoding



Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

IT cortex

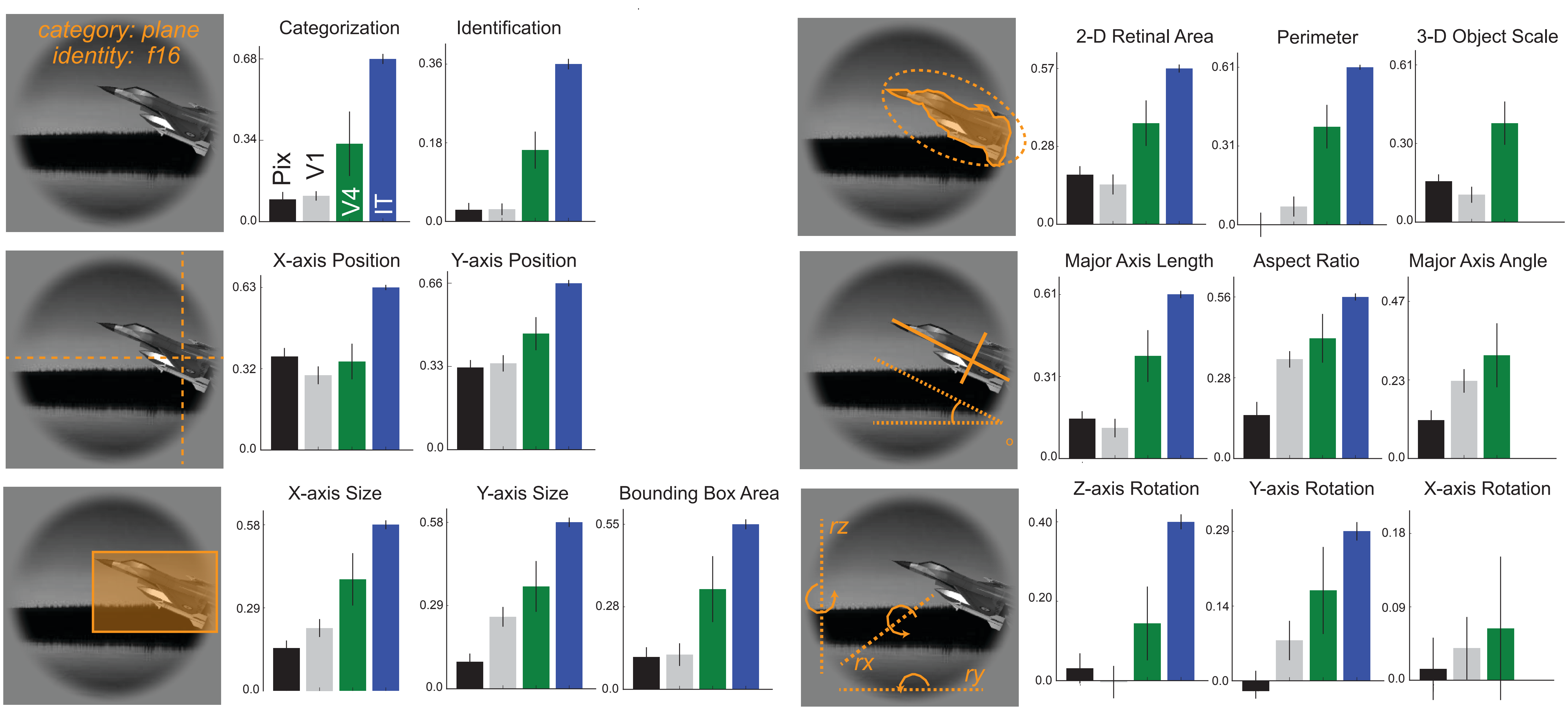
V4 cortex

V1-like model

pixel control

Population Decoding

IT > V4, V1 for all tasks **V4 > V1** for most tasks



Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

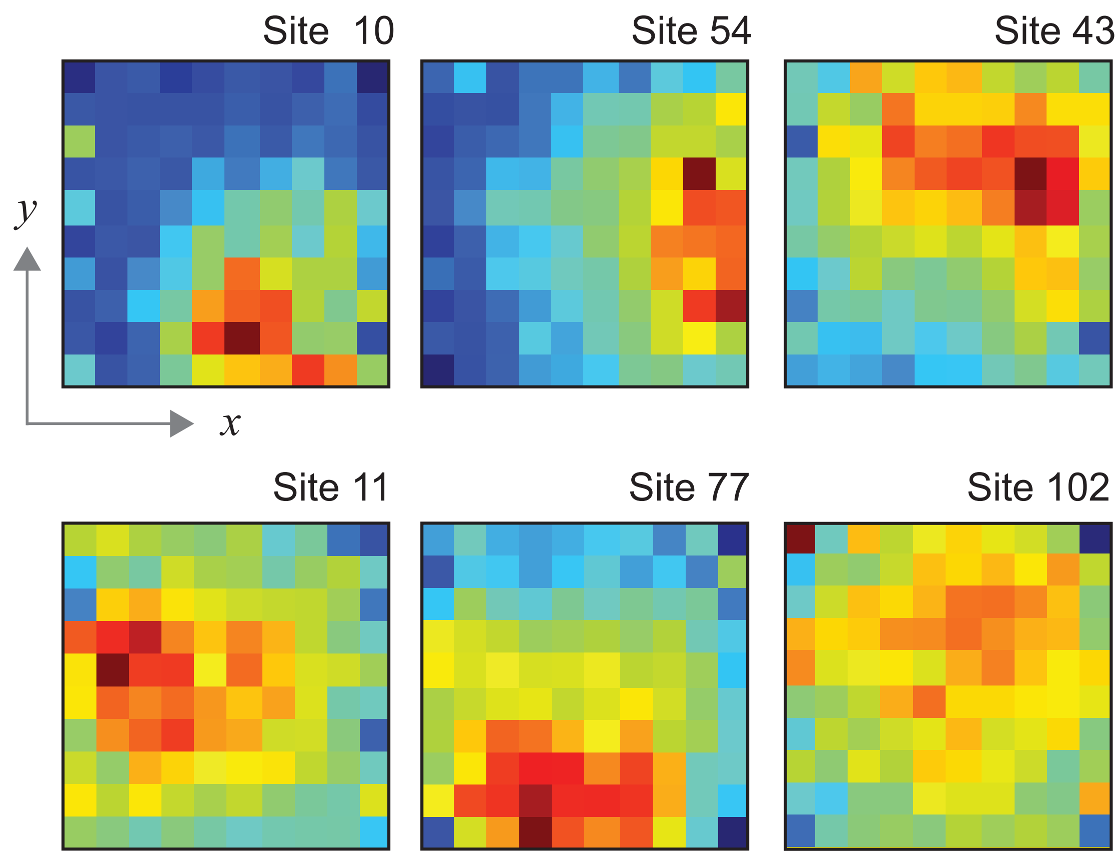
IT cortex

V4 cortex

V1-like model

pixel control

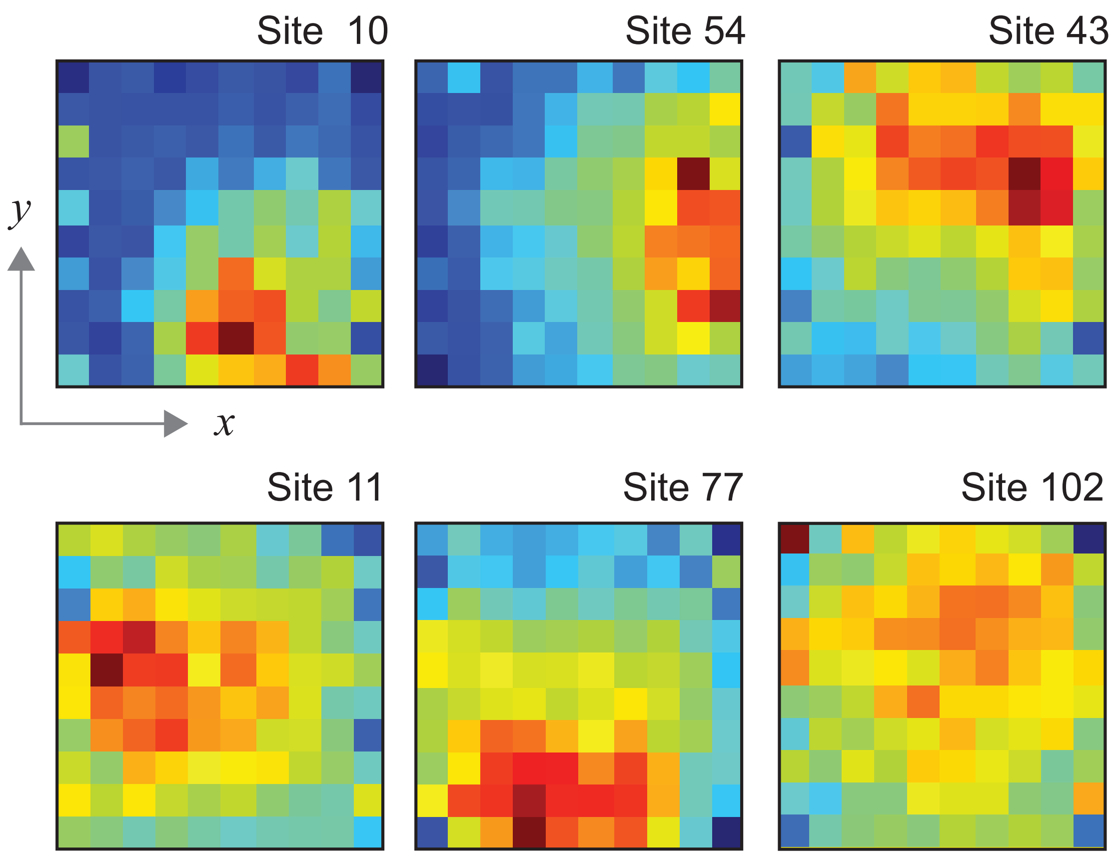
Single Site Responses



Best single position-encoding sites.

heat map value at x, y =
response averaged over all images where object center is in position x, y

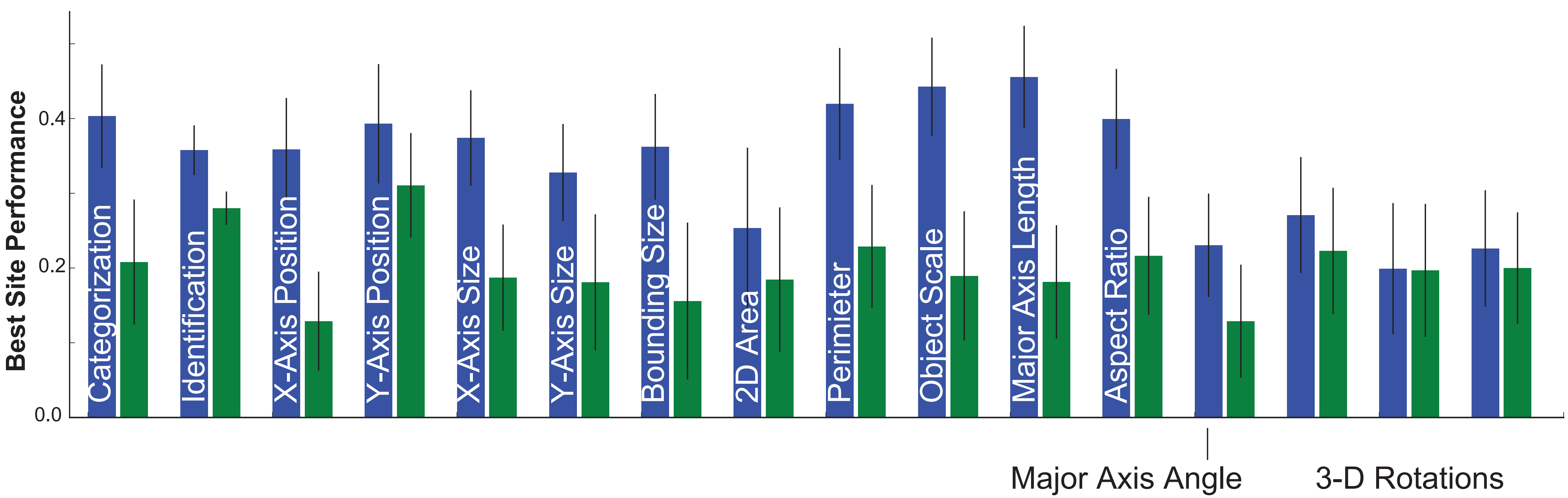
Single Site Responses



Best single position-encoding sites.

heat map value at x, y =
response averaged over all images where object center is in position x, y

Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*



IT cortex   V4 cortex

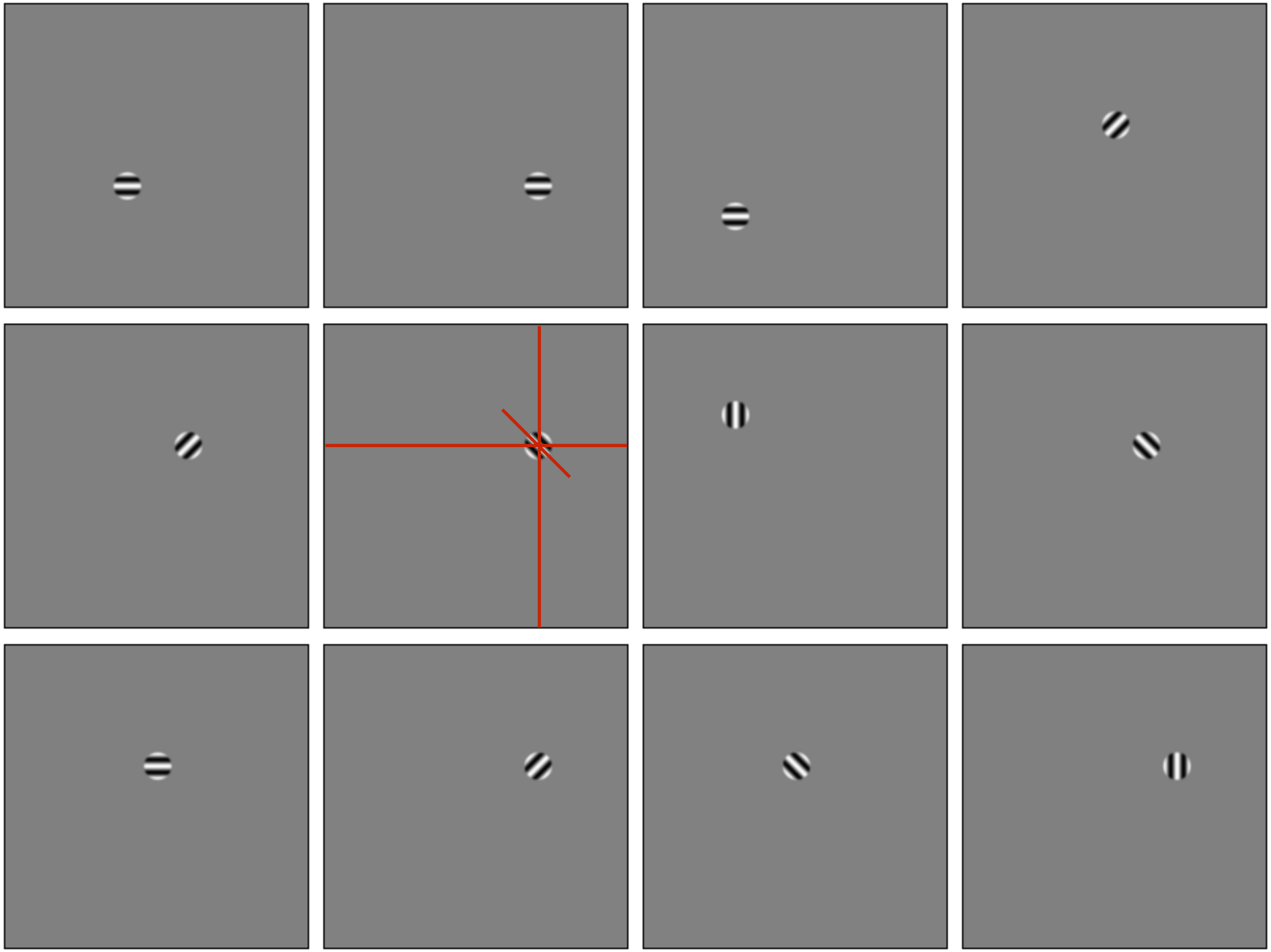
Population Decoding

"Standard" receptive field-mapping stimuli w/ position and orientation variation:

X-position

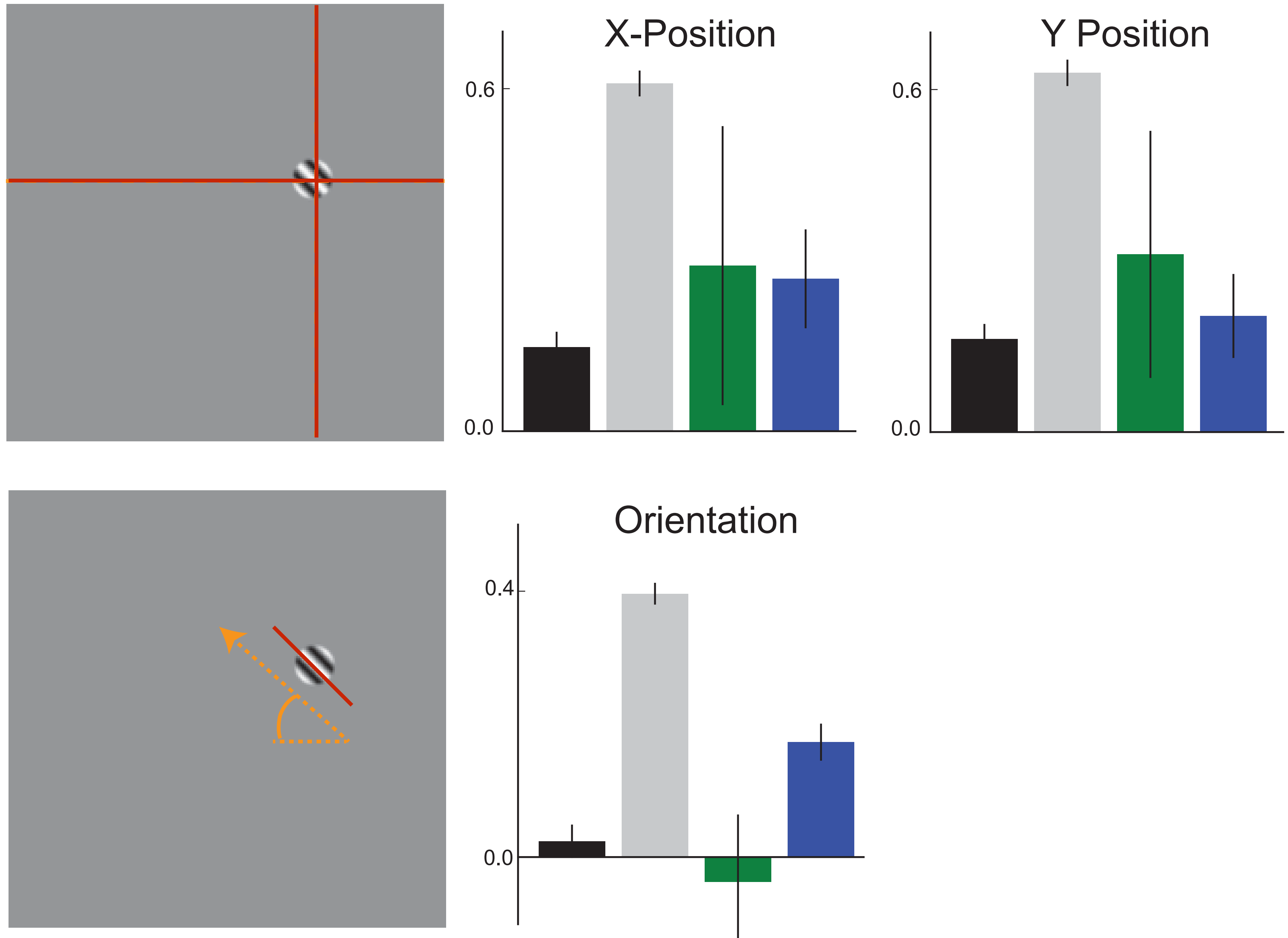
Y-position

Orientation



Population Decoding

V1 > V4, IT for "standard" tasks



Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

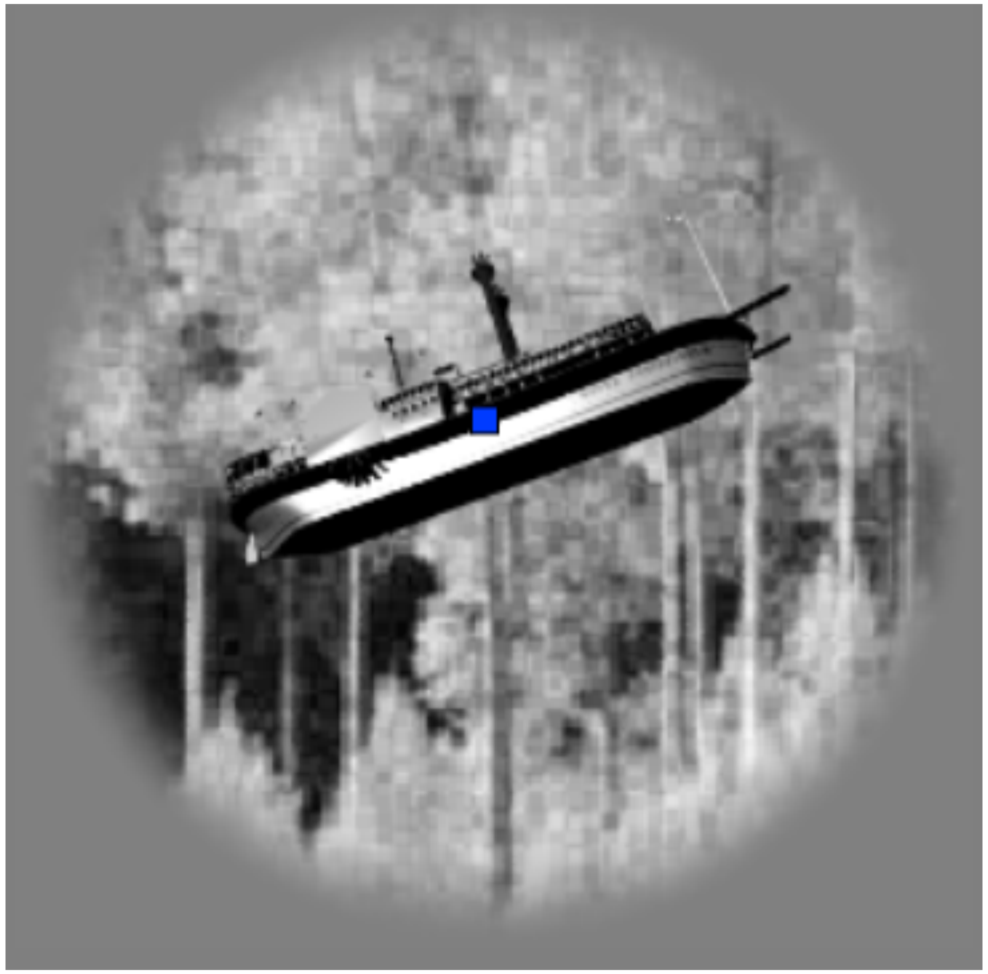
IT cortex

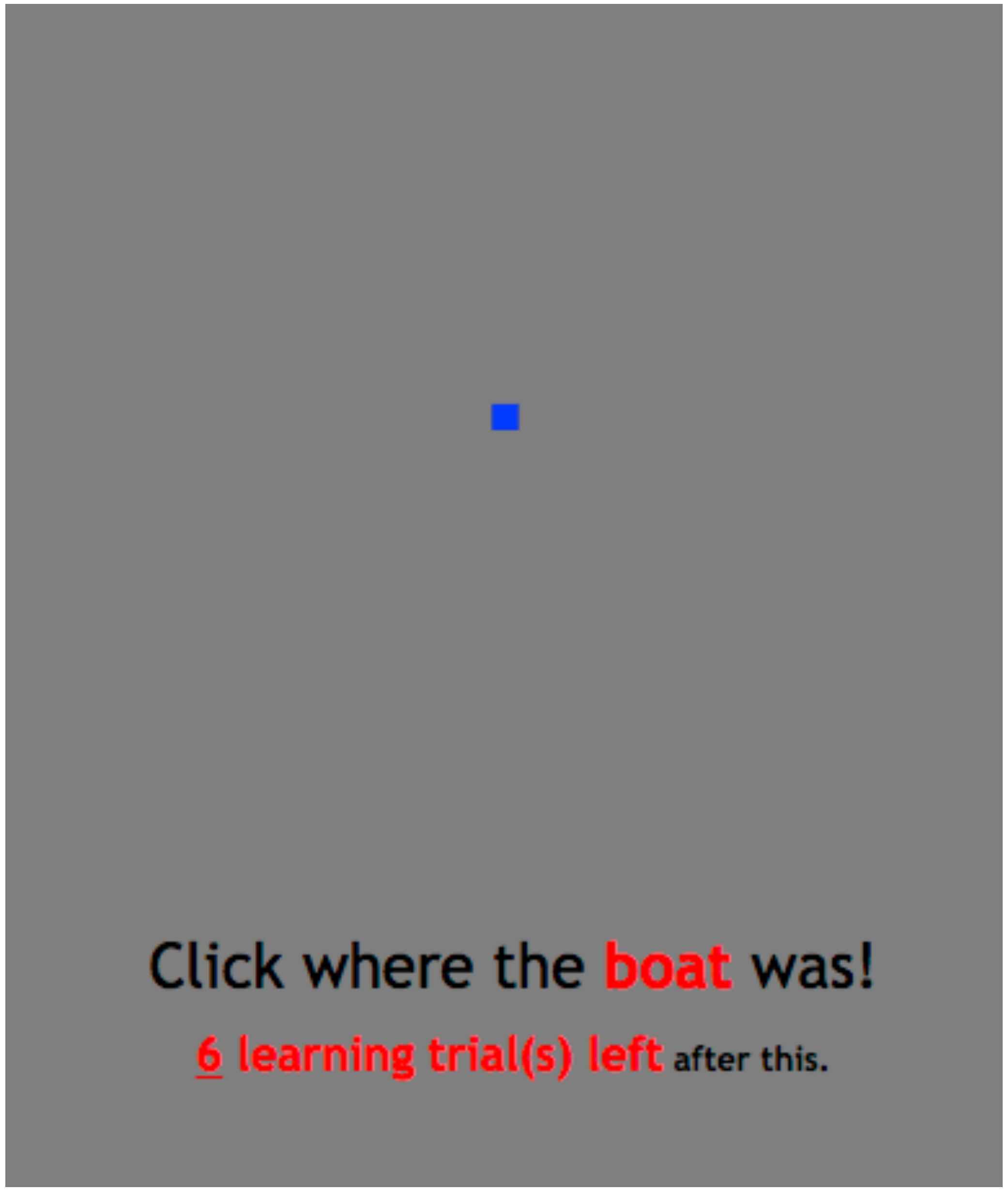
V4 cortex

V1-like model

pixel control

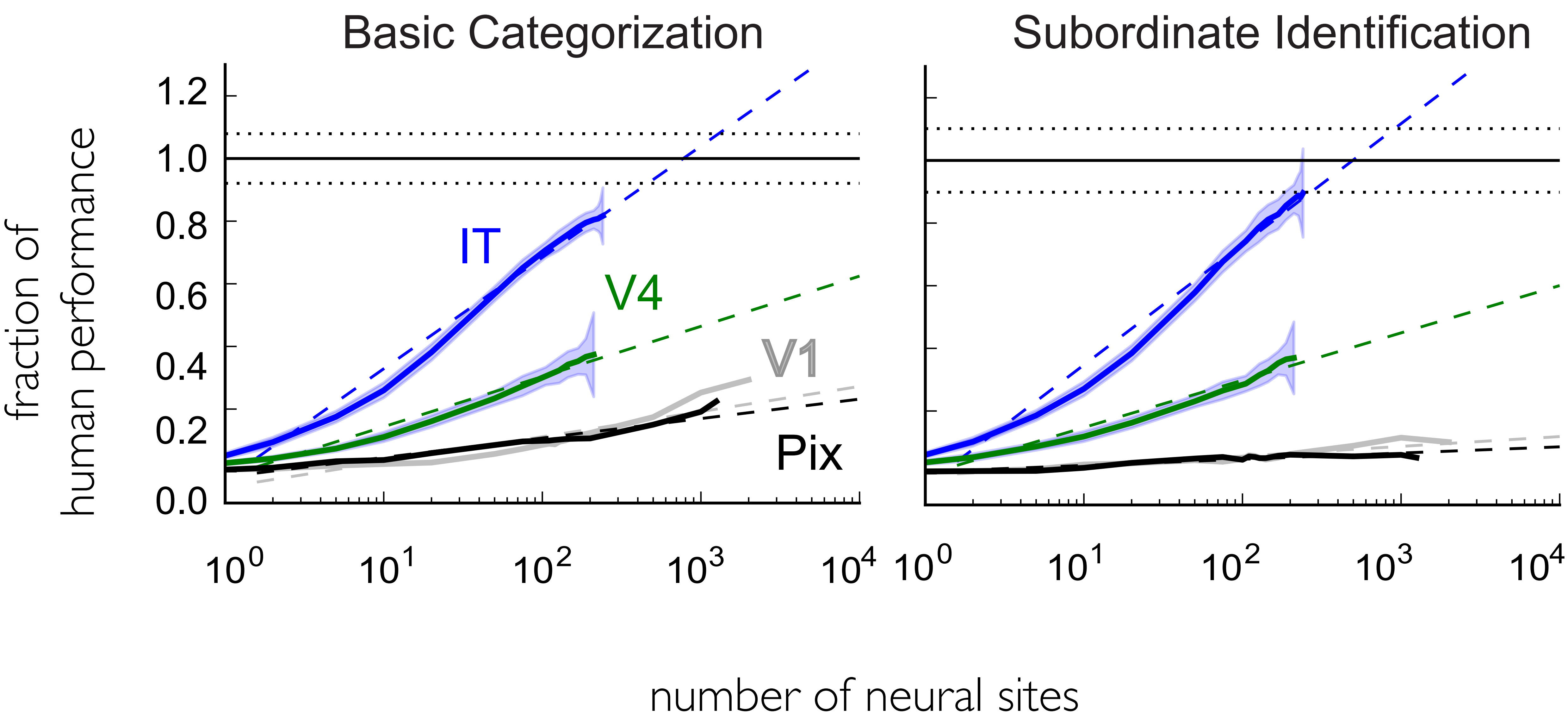
Human Psychophysical Measurements





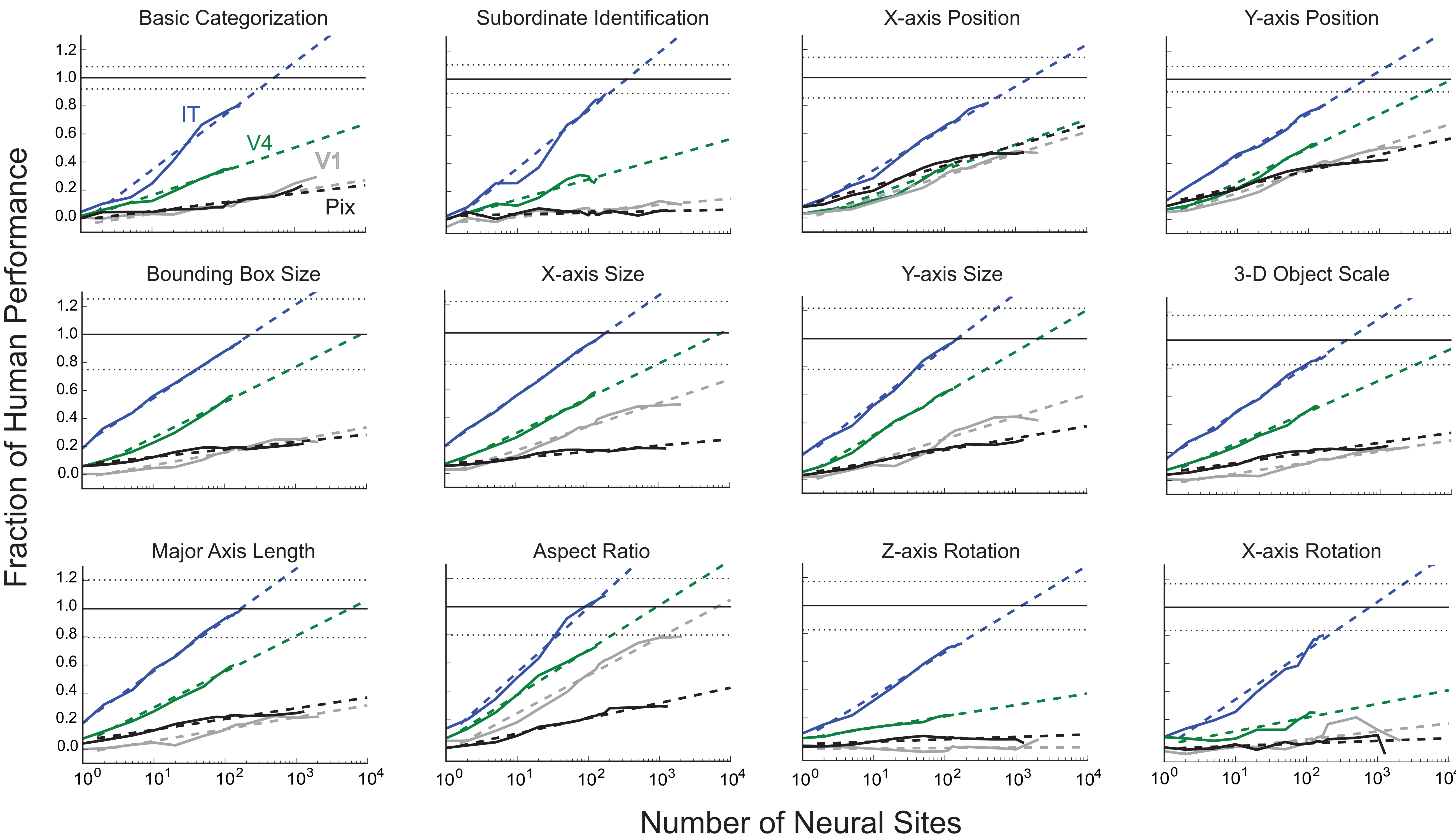
Monkey Neurons vs Humans

performance ~ k * log(N)



Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

Monkey Neurons vs Humans



Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

Monkey Neurons vs Humans

	IT	V4	V1	Pix
Basic Categorization	773 ± 185	2.2×10^6	—	—
Subordinate Identification	496 ± 93	4.4×10^6	—	—
X-axis Position	1414 ± 403	5.2×10^5	3.0×10^7	—
Y-axis Position	918 ± 309	2.5×10^4	8.7×10^6	—
Bounding Box Size	322 ± 90	1.7×10^4	—	—
X-axis Size	256 ± 87	9.8×10^3	3.4×10^7	—
Y-axis Size	237 ± 87	3.8×10^3	9.5×10^6	—
3-D Object Scale	401 ± 90	3.2×10^4	—	—
Major Axis Length	201 ± 70	1.1×10^4	—	—
Aspect Ratio	163 ± 61	951 ± 59	6.5×10^3	—
Major Axis Angle	804 ± 136	3.2×10^6	—	—
Z-axis Rotation	1932 ± 1061	—	—	—
Y-axis Rotation	369 ± 115	2.8×10^5	—	—
X-axis Rotation	1570 ± 530	—	—	—

— = more than 10 billion sites required

Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

Mean over tasks, human-parity for IT is at ~**700** multi-unit trial-averaged sites.

Monkey Neurons vs Humans

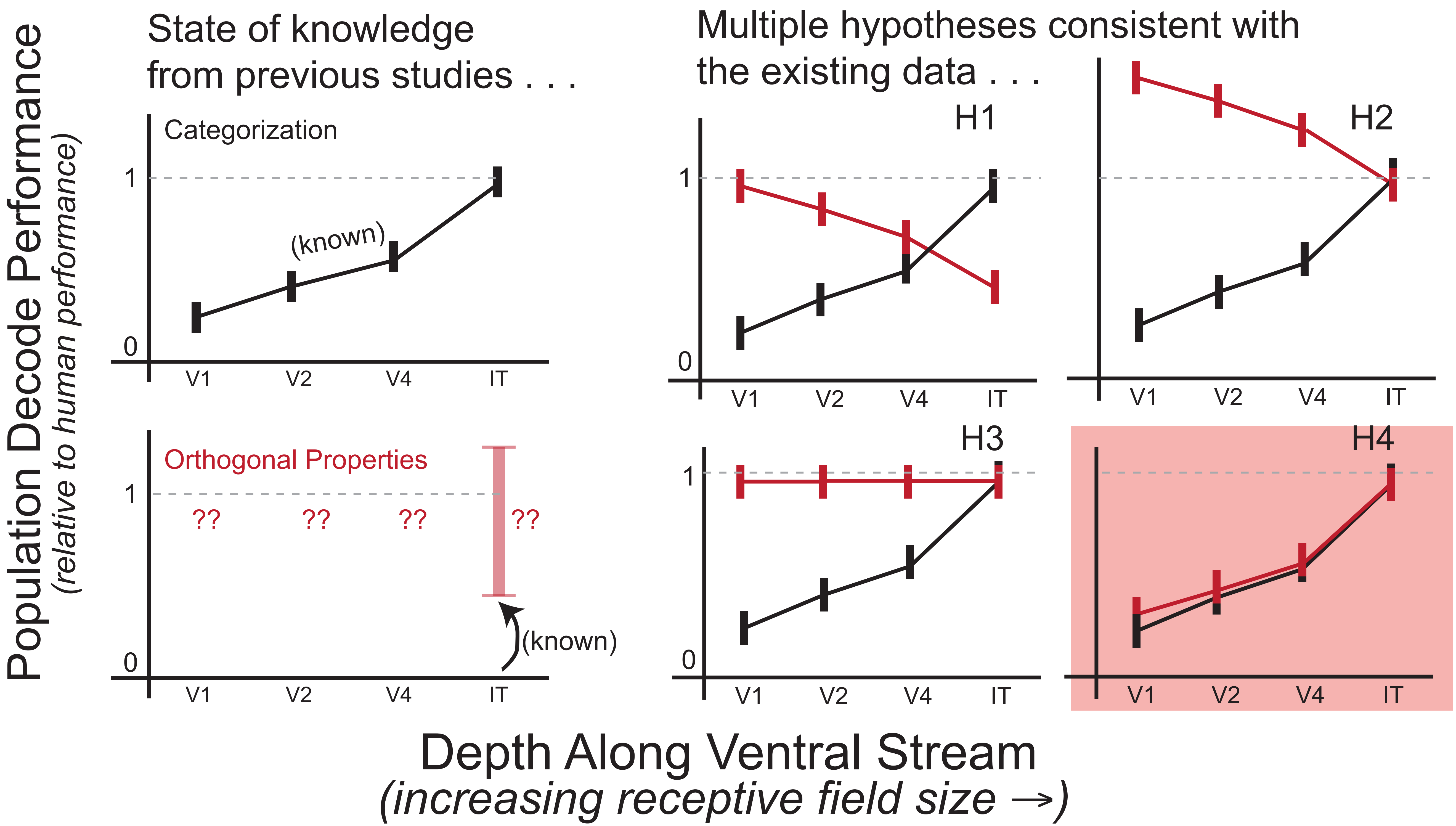
	IT	V4	V1	Pix
Basic Categorization	773 ± 185	2.2×10^6	—	—
Subordinate Identification	496 ± 93	4.4×10^6	—	—
X-axis Position	1414 ± 403	5.2×10^5	3.0×10^7	—
Y-axis Position	918 ± 309	2.5×10^4	8.7×10^6	—
Bounding Box Size	322 ± 90	1.7×10^4	—	—
X-axis Size	256 ± 87	9.8×10^3	3.4×10^7	—
Y-axis Size	237 ± 87	3.8×10^3	9.5×10^6	—
3-D Object Scale	401 ± 90	3.2×10^4	—	—
Major Axis Length	201 ± 70	1.1×10^4	—	—
Aspect Ratio	163 ± 61	951 ± 59	6.5×10^3	—
Major Axis Angle	804 ± 136	3.2×10^6	—	—
Z-axis Rotation	1932 ± 1061	—	—	—
Y-axis Rotation	369 ± 115	2.8×10^5	—	—
X-axis Rotation	1570 ± 530	—	—	—

— = more than 10 billion sites required

Hong, Yamins*, Majaj & DiCarlo.* ***Nat. Neuro.*** *(2016)*

Mean over tasks, human-parity for IT is at ~**350000** single-unit single-trial neurons.

Somewhat newish ideas about IT?



H4: Simultaneous build-up of encoding

Somewhat newish ideas about IT?

1. IT is *NOT* invariant. Strict generalization of simple-to-complex cells: **no**.

2. "Lower-level" properties are not that low-level — at least, with complex objects and backgrounds.

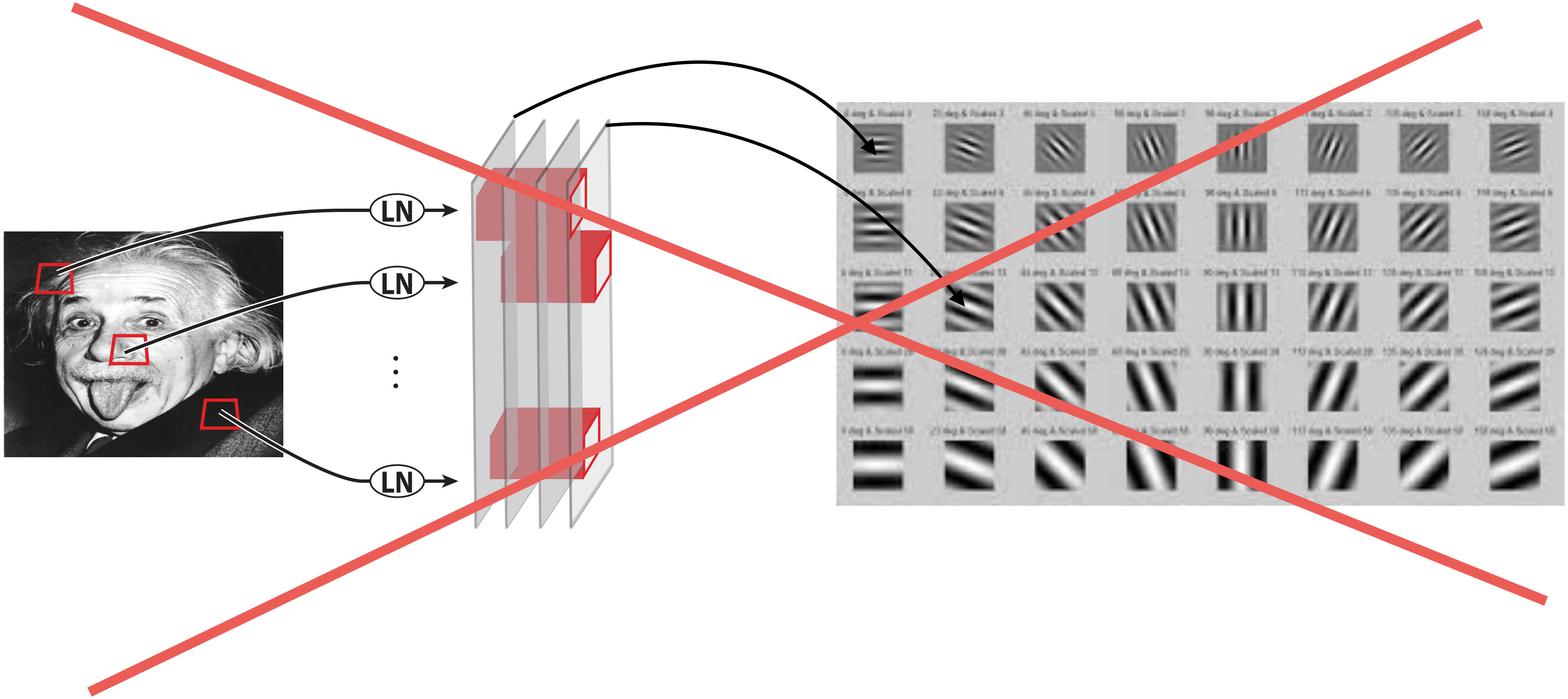
3. Categorization and non-categorical properties "go together" — *not* just that "not all (e.g.) position information is lost" (MacEvoy 2013, DiCarlo 2003)

Provides support to a hypothesis for what IT does:

"Inverting the generative model of the scene"

But what type of understanding is this?

But what type of understanding is this?

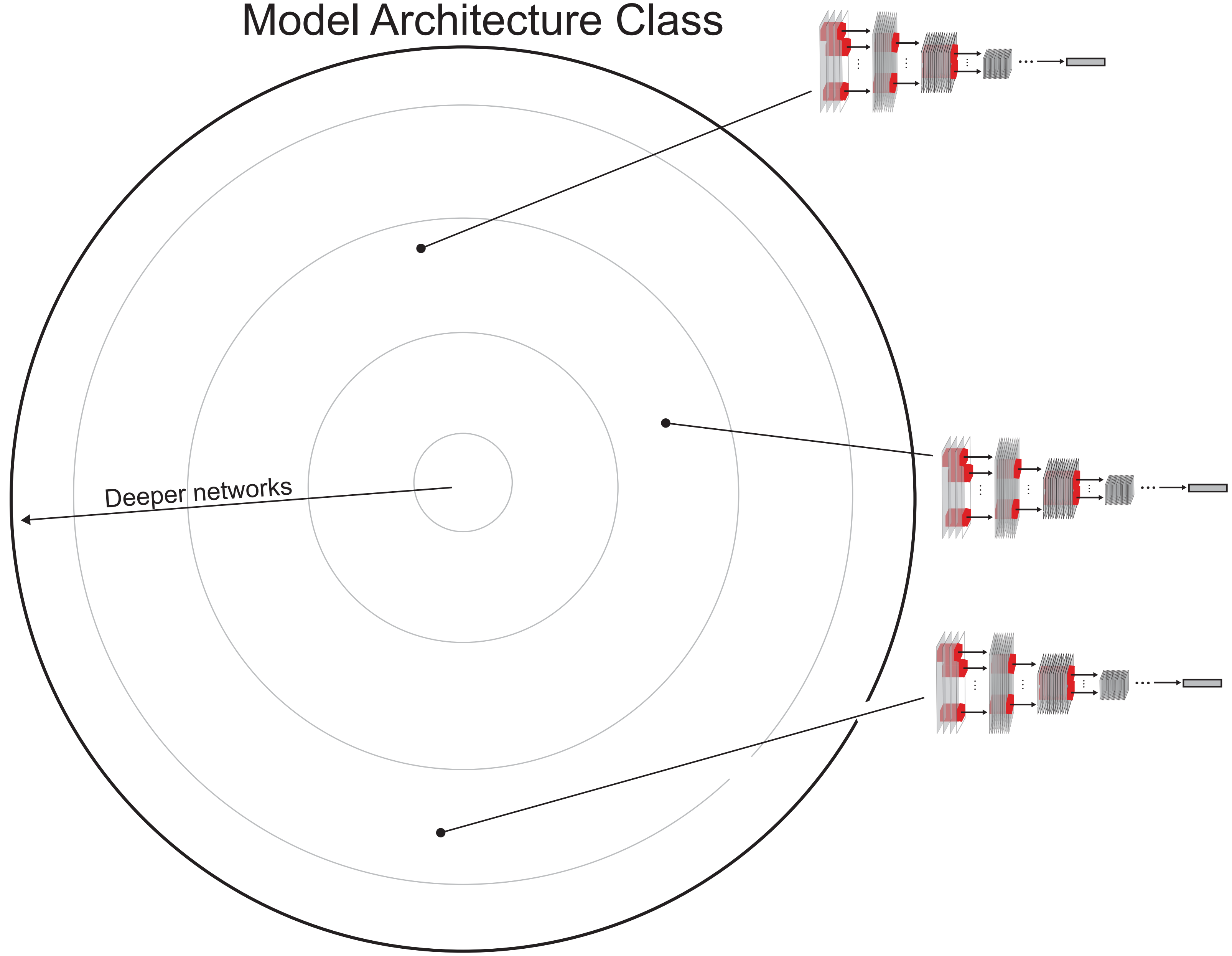


not saying this type of understanding is impossible …

Principle of "Goal-Driven Modeling"

> Formulate comprehensive model class (**CNNs**)

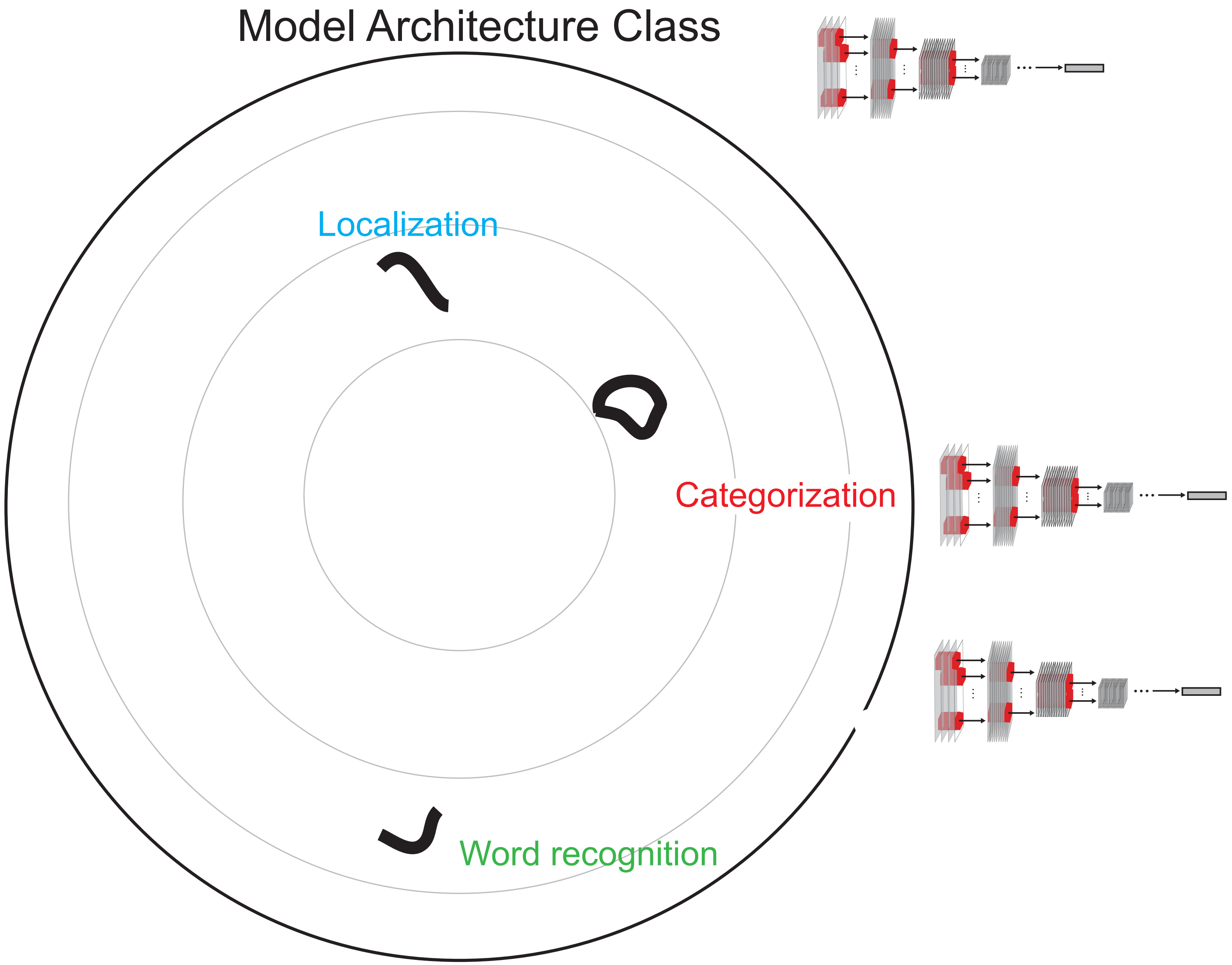
Model Architecture Class



Yamins & DiCarlo.
***Nat. Neuro.** (2016)*

> Formulate comprehensive model class (**CNNs**)

> Choose challenging, ethologically-valid tasks (**categorization**)

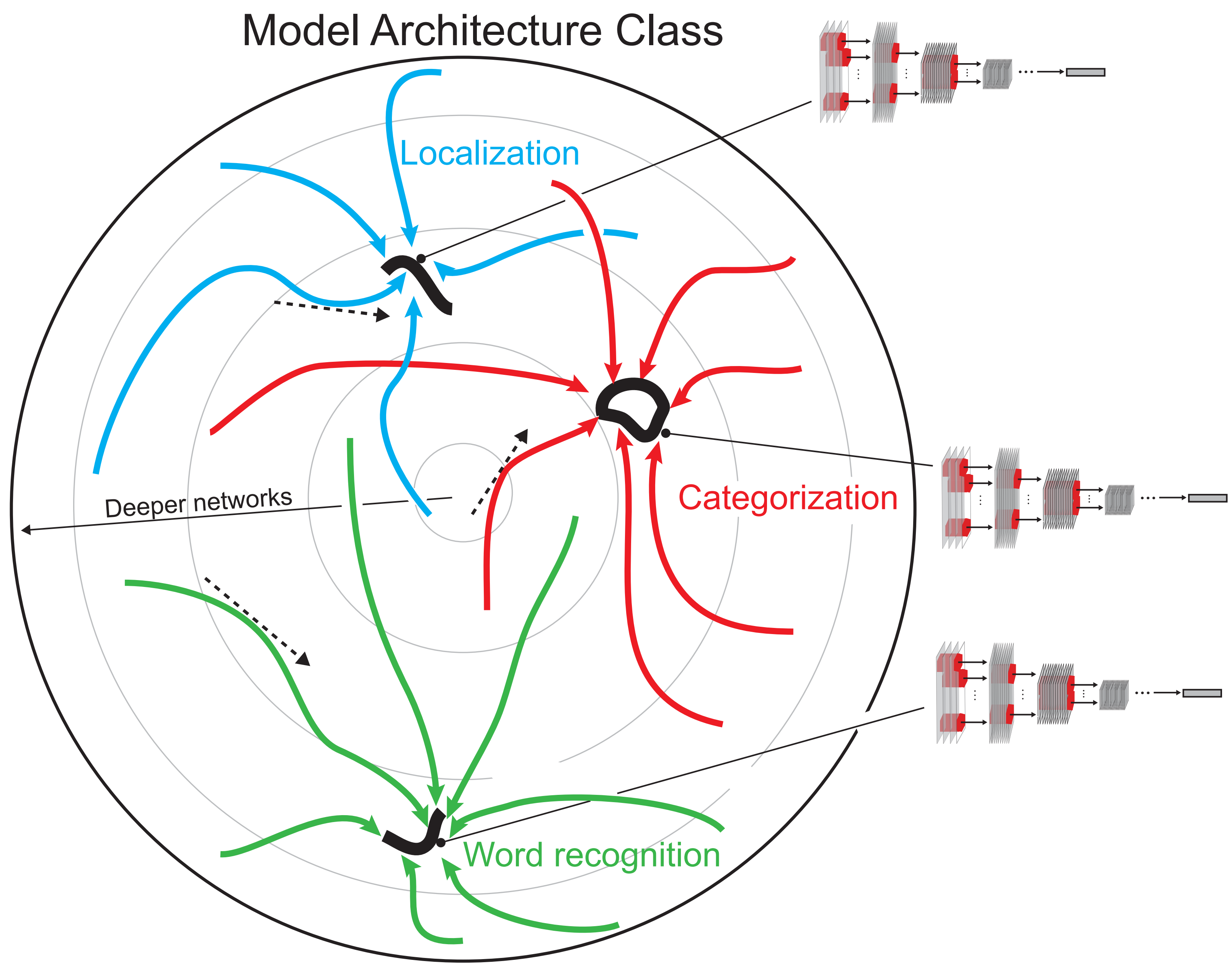


Yamins & DiCarlo.
***Nat. Neuro.** (2016)*

> Formulate comprehensive model class (**CNNs**)

> Choose challenging, ethologically-valid tasks (**categorization**)

> Implement generic learning rules (**gradient descent**)



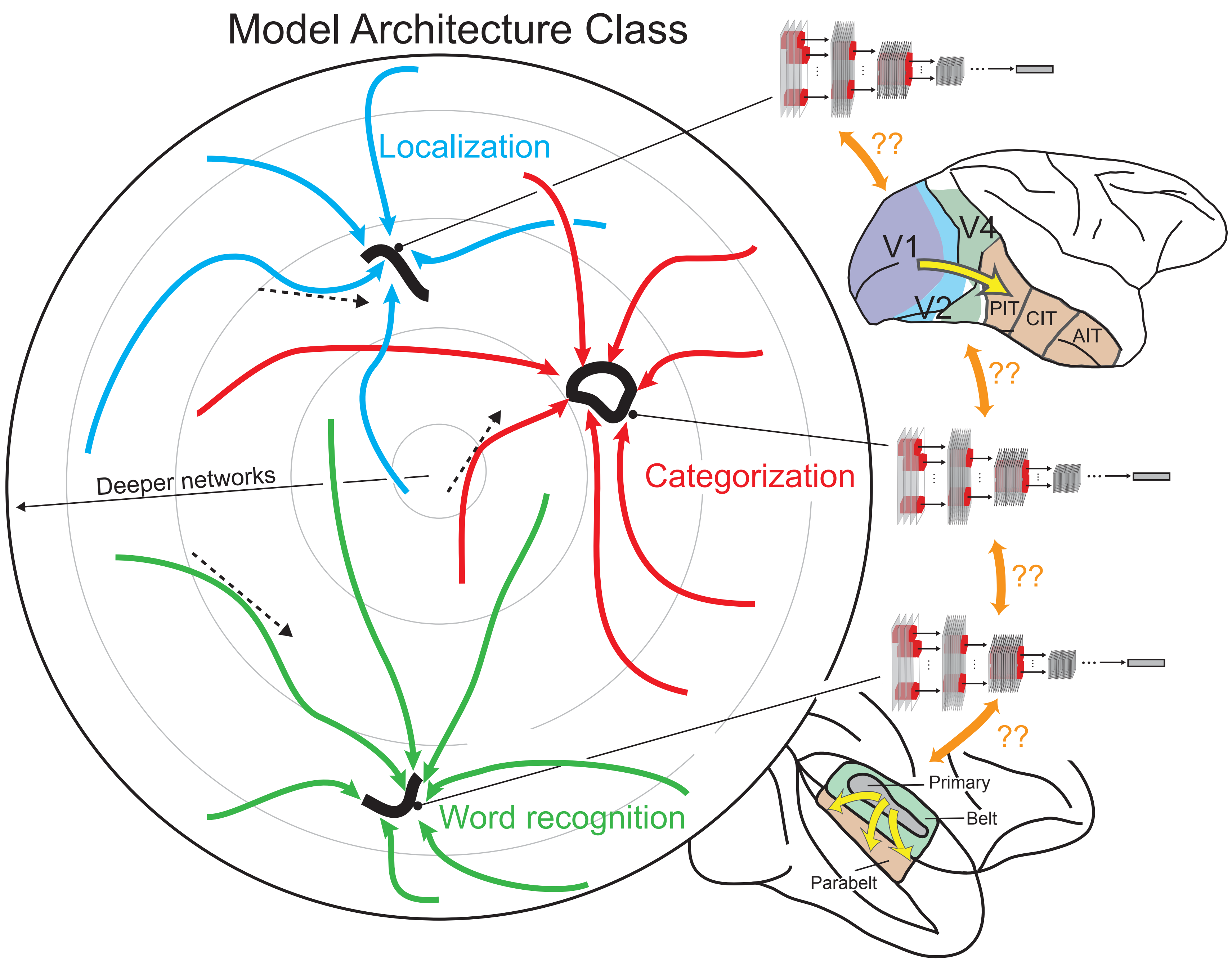
Yamins & DiCarlo.
Nat. Neuro. *(2016)*

> Formulate comprehensive model class (**CNNs**)

> Choose challenging, ethologically-valid tasks (**categorization**)

> Implement generic learning rules (**gradient descent**)

> Map to brain data. (**ventral stream**)



Yamins & DiCarlo.
Nat. Neuro. *(2016)*

Four Principles of Goal-Driven Modeling

1.

A = *architecture class*

2.

T = *task/objective*

3.

D = *dataset*

4.

L = *learning rule*

Four Principles of Goal-Driven Modeling

1.

A = *architecture class*

2.

T = *task/objective*

3.

D = *dataset*

4.

L = *learning rule*

Best proxies thus far for ventral stream:

A = *ConvNets of reasonable depth*

T = *multi-way object categorization*

D = *ImageNet images*

L = *evolutionary architecture search + filter learning through gradient descent*

Four Principles of Goal-Driven Modeling

1.

A = *architecture class* ***= circuit neuro-anatomy***

2.

T = *task/objective* ***= ecological niche***

3.

D = *dataset* ***= environment***

4.

L = *learning rule* ***= natural selection + synaptic plasticity***

Best proxies thus far for ventral stream:

A = *ConvNets of reasonable depth*

T = *multi-way object categorization*

D = *ImageNet images*

L = *evolutionary architecture search + filter learning through gradient descent*

Four Principles of Goal-Driven Modeling

1.

A = *architecture class* ***= circuit neuro-anatomy***

solving

2.

T = *task/objective* ***= ecological niche***

situated in

3.

D = *dataset* ***= environment***

updating according to

4.

L = *learning rule* ***= natural selection + synaptic plasticity***

Best proxies thus far for ventral stream:

A = *ConvNets of reasonable depth*

T = *multi-way object categorization*

D = *ImageNet images*

L = *evolutionary architecture search + filter learning through gradient descent*

"Nothing in biology makes sense except in light of evolution"

Theo Dobzhansky

"Nothing in biology makes sense except in light of evolution"

Theo Dobzhansky

"Nothing in neuroscience makes sense except in light of behavior"

Gordon Shepherd

"Nothing in biology makes sense except in light of evolution"

Theo Dobzhansky

"Nothing in neuroscience makes sense except in light of behavior"

Gordon Shepherd

computational

Nothing in neuroscience makes sense except in light of ***optimization****.*

Stanford Neuro AI Lab

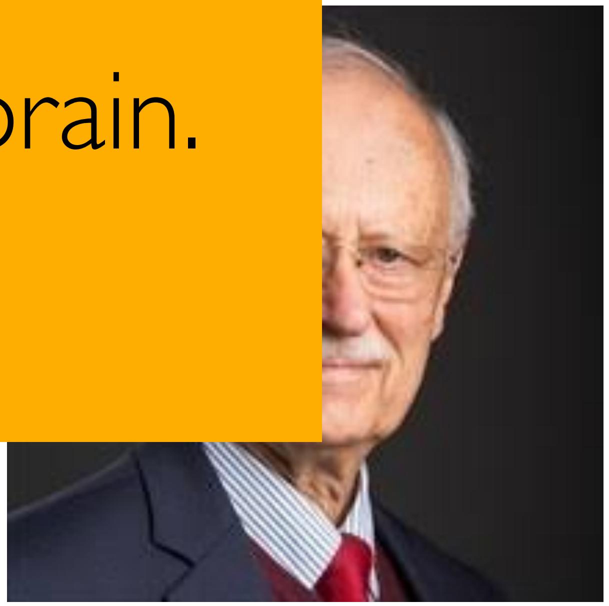
"Nothing in biology makes sense except in light of evolution"

Dobzhansky

"Nothing

Restated:

Behavior is highly constraining of the brain.



Gordon Shepherd

computational

Nothing in neuroscience makes sense except in light of ***optimization****.*

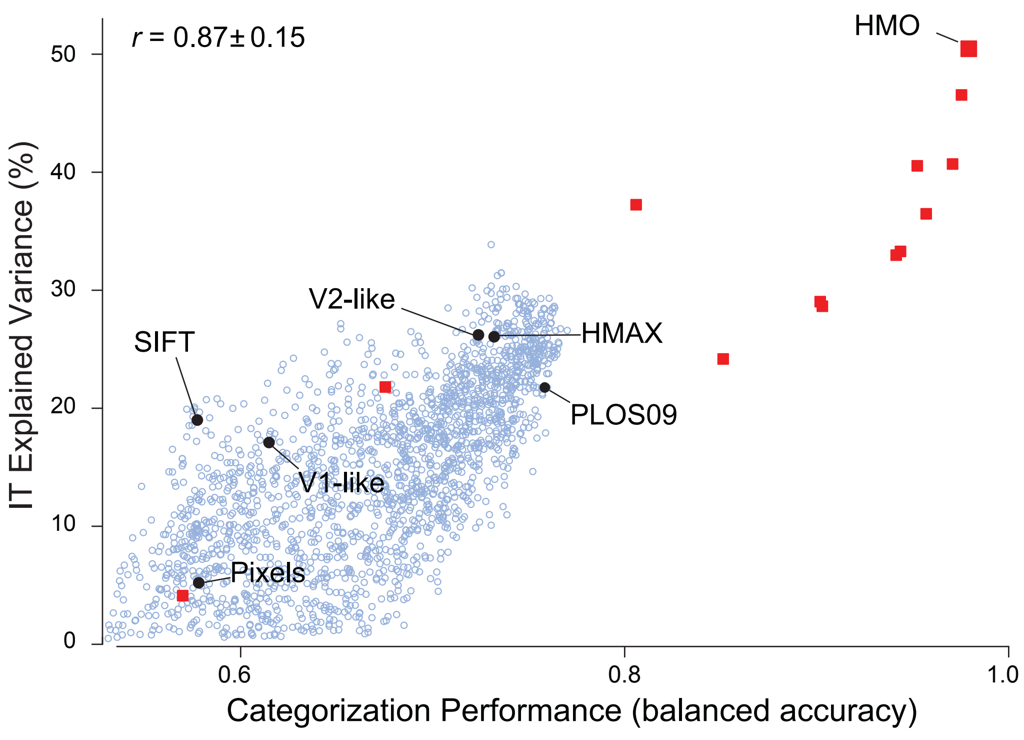
Stanford
Neuro AI Lab

~~Principle~~ of "Goal-Driven Modeling"

Heuristic of "Goal-Driven Modeling"

~~Principle~~ of "Goal-Driven Modeling"

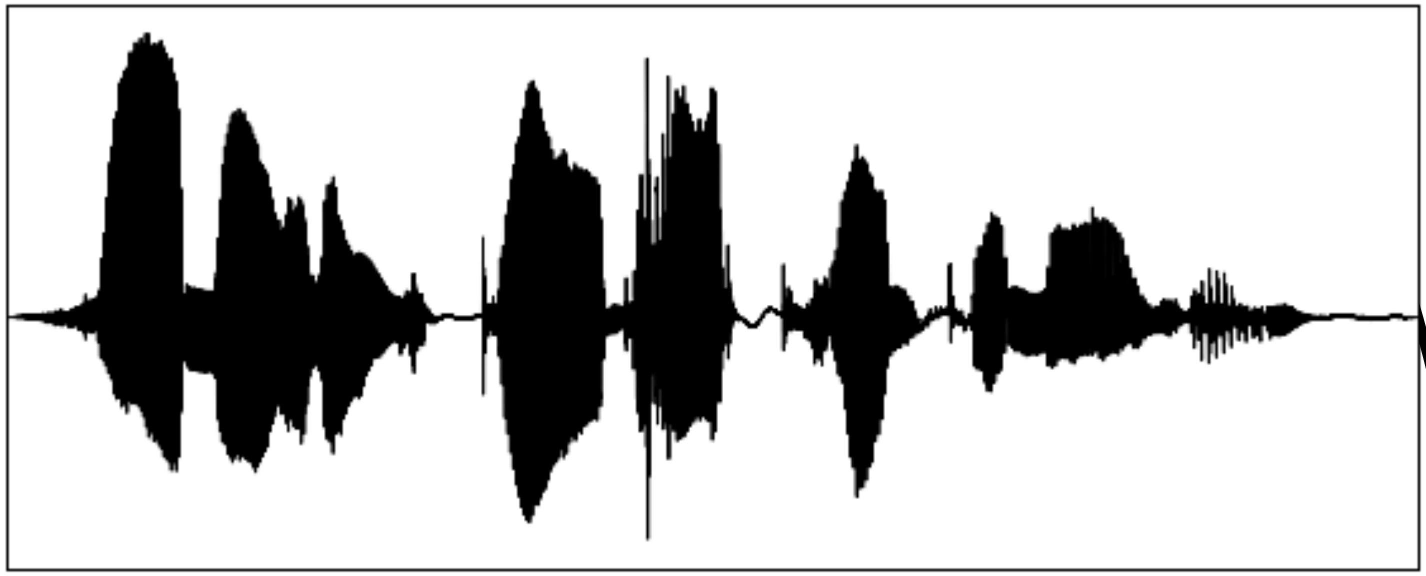
Heuristic of "Goal-Driven Modeling"



res-net?

… after all at some point, for any given task, you'll probably "go over the hump" … perhaps when you exceed human performance or overfit on that task

Can we go beyond vision?



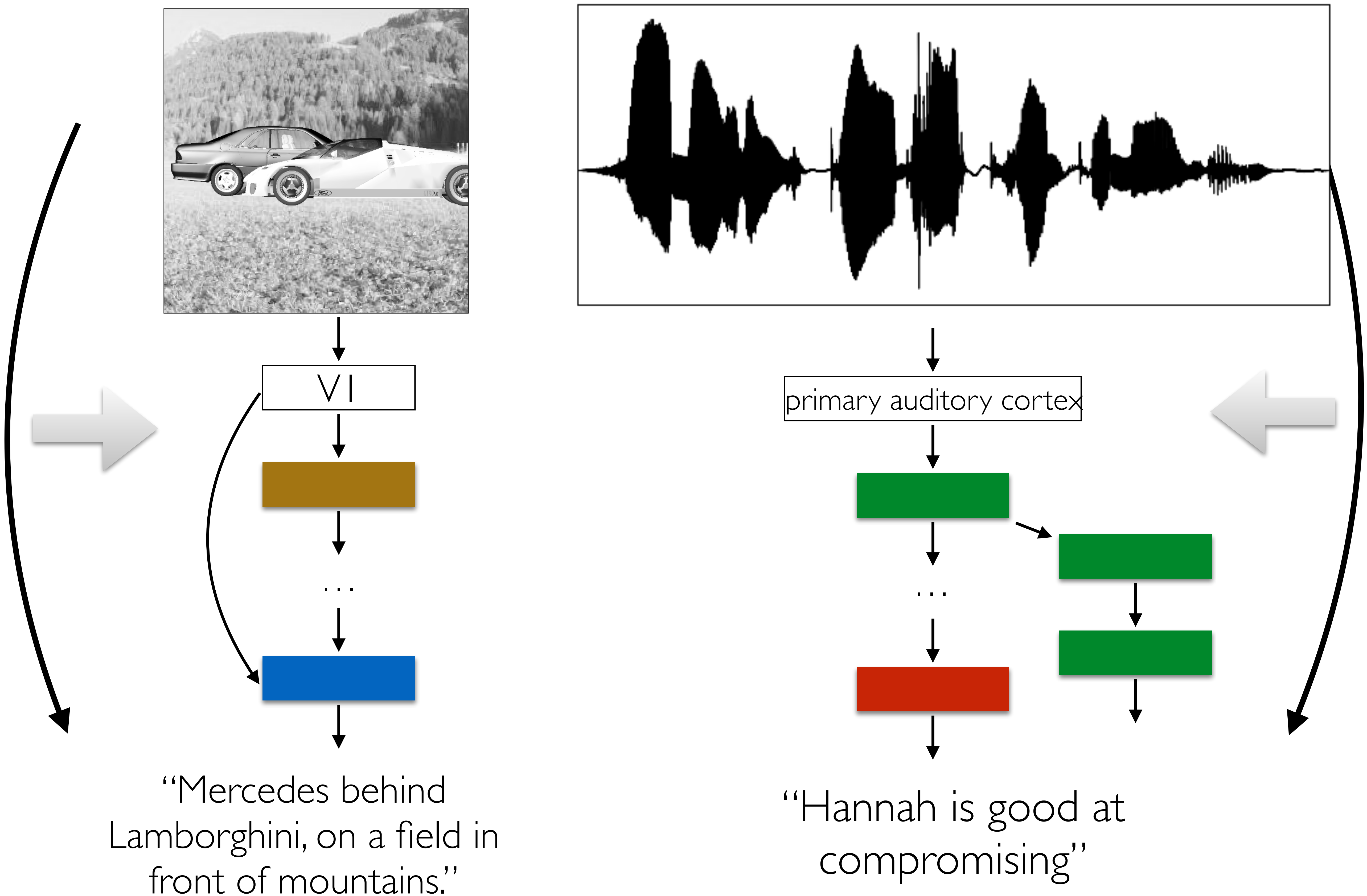
visual cortex

auditory cortex

"Mercedes behind Lamborghini, on a field in front of mountains."

"Hannah is good at compromising"

Can we go beyond vision?

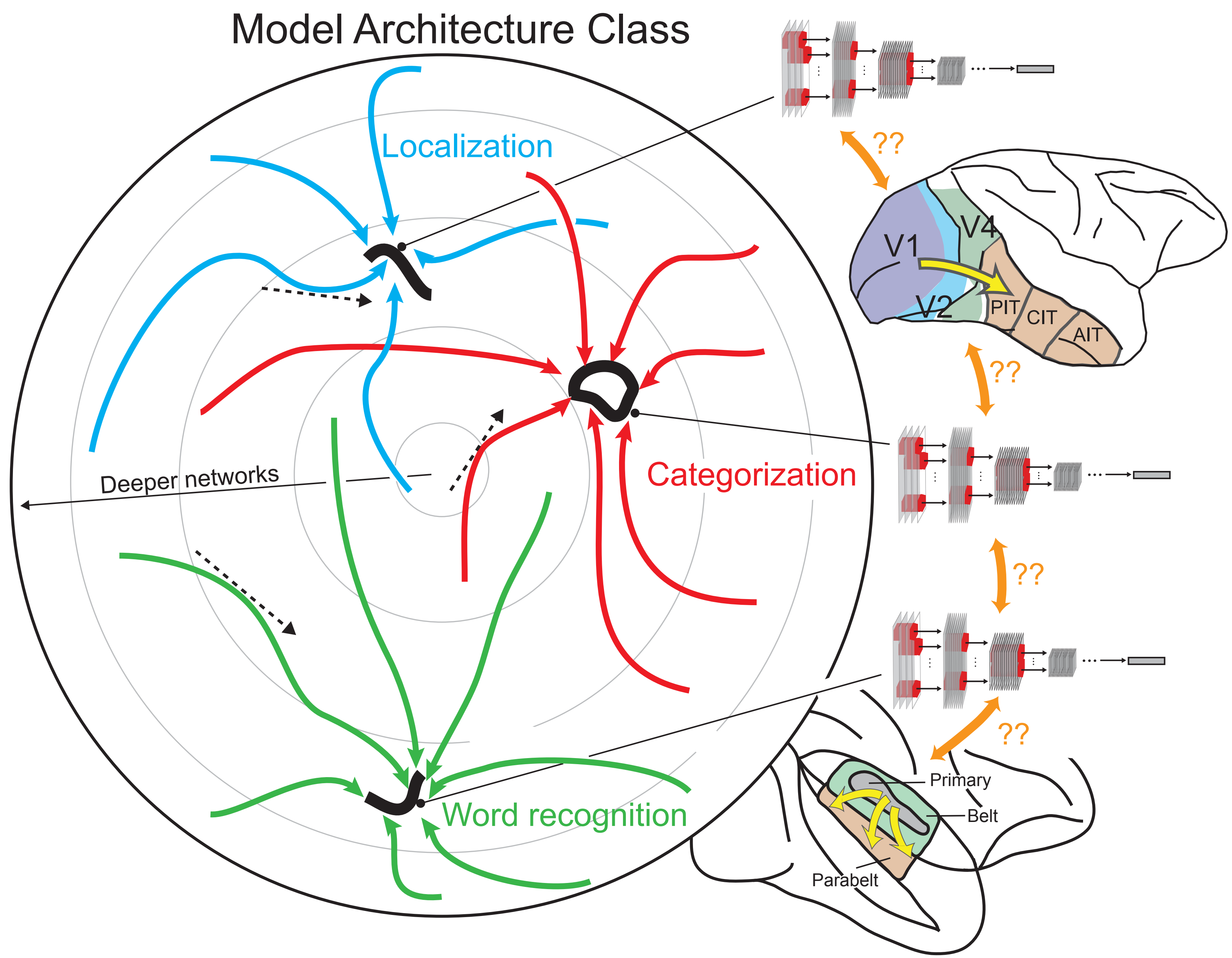


> Formulate comprehensive model class (**RNNs**)

> Choose challenging, ethologically-valid tasks (**task switching/ memory**)

> Implement generic learning rules (**??**)

> Map to brain data. (**Parietal cortex, PFC**)



Yamins & DiCarlo.
Nat. Neuro. *(2016)*

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

1. ✗ ***bad***

***A** = architecture class*

e.g. **CNNs**

2.

***T** = task/objective*

e.g. **Object Categorization**

3.

***D** = dataset*

e.g. **ImageNet**

4.

***L** = learning rule*

e.g. **Arch. Srch.** + **Grad. Desc.**

PROBLEM

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

1. ✗ ***bad***

***A** = architecture class*

e.g. **CNNs**

2.

***T** = task/objective*

e.g. **Object Categorization**

3.

***D** = dataset*

e.g. **ImageNet**

4.

***L** = learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

PROBLEM

RECURRENCE and FEEDBACK!!?

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

1. ❌ *bad*

A = *architecture class*

e.g. **CNNs**

2. ❌ *bad*

T = *task/objective*

e.g. **Object Categorization**

3.

D = *dataset*

e.g. **ImageNet**

4.

L = *learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

PROBLEM

RECURRENCE and FEEDBACK!!?

TOO MUCH LABELLED DATA REQUIRED!!?

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

PROBLEM

1. ✗ ***bad***

A = *architecture class*

e.g. **CNNs**

RECURRENCE and FEEDBACK!!?

2. ✗ ***bad***

T = *task/objective*

e.g. **Object Categorization**

TOO MUCH LABELLED DATA REQUIRED!!?

3. ✗ ***bad***

D = *dataset*

e.g. **ImageNet**

REAL NOISY VIDEO DATASTREAMS vs STEREOTYPED CLEAN STILL IMAGES

4.

L = *learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

PROBLEM

1. ✗**bad**

A = *architecture class*

e.g. **CNNs**

RECURRENCE and FEEDBACK!!?

2. ✗**bad**

T = *task/objective*

e.g. **Object Categorization**

TOO MUCH LABELLED DATA REQUIRED!!?

3. ✗**bad**

D = *dataset*

e.g. **ImageNet**

REAL NOISY VIDEO DATASTREAMS vs STEREOTYPED CLEAN STILL IMAGES

4. ✗**bad**

L = *learning rule*

e.g. **Arch. Srch.** + **Grad. Desc.**

BACKPROP AND ITS DISCONTENTS

So far, we've done the basic idea

Date	Session
01/06	Introduction to NeuroAI
01/08	Visual Systems Neuroscience Background
01/13	DNN Models of the Visual System I
01/15	DNN Models of the Visual System II
01/20	[NO CLASS-MLK DAY]
01/22	Recurrent Models in Vision and Beyond
01/27	**Guest Lecture — Meenakshi Khosla (USCD):** *Mapping Neural Networks to the Brain*
01/28	
01/29	Unsupervised Learning and the Brain
02/03	**Guest Lecture — Arash Afraz (NIH):** *Model-Driven Brain Perturbation*
02/05	Auditory and Somatosensory Models
02/10	**Guest Lecture — Rhodri Cusack (Trinity):** *Models of Development and Learning*
02/11	
02/12	**Guest Lecture — Josh McDermott (MIT):** *Leveraging Models of Auditory Cortex*
02/17	[NO CLASS-PRESIDENT'S DAY]
02/19	Learning Rules in the Brain
02/24	Models of the Motor System
02/25	
02/26	**Guest Lecture — Scott Linderman (Stanford):** *Dynamical Systems Models in Neuroscience*
03/03	**Guest Lecture — Greta Tuckute (MIT):** *The Human Language Network & LLMs*
03/05	The Hippocampus: Memory and Spatial Navigation
03/10	Topographic Models: A Unified Theory of the Brain
03/12	**Guest Lecture — Robert Hawkins (Stanford):** *Cognitive Modeling*

Basic idea (01/08 – 01/15)

Next we'll fix some of the problems . . .

Date	Session
01/06	Introduction to NeuroAI
01/08	Visual Systems Neuroscience Background
01/13	DNN Models of the Visual System I
01/15	DNN Models of the Visual System II
01/20	[NO CLASS-MLK DAY]
01/22	Recurrent Models in Vision and Beyond
01/27	**Guest Lecture — Meenakshi Khosla (USCD):** *Mapping Neural Networks to the Brain*
01/28	
01/29	Unsupervised Learning and the Brain
02/03	**Guest Lecture — Arash Afraz (NIH):** *Model-Driven Brain Perturbation*
02/05	Auditory and Somatosensory Models
02/10	**Guest Lecture — Rhodri Cusack (Trinity):** *Models of Development and Learning*
02/11	
02/12	**Guest Lecture — Josh McDermott (MIT):** *Leveraging Models of Auditory Cortex*
02/17	[NO CLASS-PRESIDENT'S DAY]
02/19	Learning Rules in the Brain
02/24	Models of the Motor System
02/25	
02/26	**Guest Lecture — Scott Linderman (Stanford):** *Dynamical Systems Models in Neuroscience*
03/03	**Guest Lecture — Greta Tuckute (MIT):** *The Human Language Network & LLMs*
03/05	The Hippocampus: Memory and Spatial Navigation
03/10	Topographic Models: A Unified Theory of the Brain
03/12	**Guest Lecture — Robert Hawkins (Stanford):** *Cognitive Modeling*

Basic idea (01/08–01/15)

Fixing problems (01/22–02/03)

. . . and then go beyond vision.

Date	Session
01/06	Introduction to NeuroAI
01/08	Visual Systems Neuroscience Background
01/13	DNN Models of the Visual System I
01/15	DNN Models of the Visual System II
01/20	[NO CLASS-MLK DAY]
01/22	Recurrent Models in Vision and Beyond
01/27	**Guest Lecture — Meenakshi Khosla (USCD):** *Mapping Neural Networks to the Brain*
01/28	
01/29	Unsupervised Learning and the Brain
02/03	**Guest Lecture — Arash Afraz (NIH):** *Model-Driven Brain Perturbation*
02/05	Auditory and Somatosensory Models
02/10	**Guest Lecture — Rhodri Cusack (Trinity):** *Models of Development and Learning*
02/11	
02/12	**Guest Lecture — Josh McDermott (MIT):** *Leveraging Models of Auditory Cortex*
02/17	[NO CLASS-PRESIDENT'S DAY]
02/19	Learning Rules in the Brain
02/24	Models of the Motor System
02/25	
02/26	**Guest Lecture — Scott Linderman (Stanford):** *Dynamical Systems Models in Neuroscience*
03/03	**Guest Lecture — Greta Tuckute (MIT):** *The Human Language Network & LLMs*
03/05	The Hippocampus: Memory and Spatial Navigation
03/10	Topographic Models: A Unified Theory of the Brain
03/12	**Guest Lecture — Robert Hawkins (Stanford):** *Cognitive Modeling*

Basic idea (01/08–01/15)

Fixing problems (01/22–02/03)

Beyond Vision (02/05–03/12)