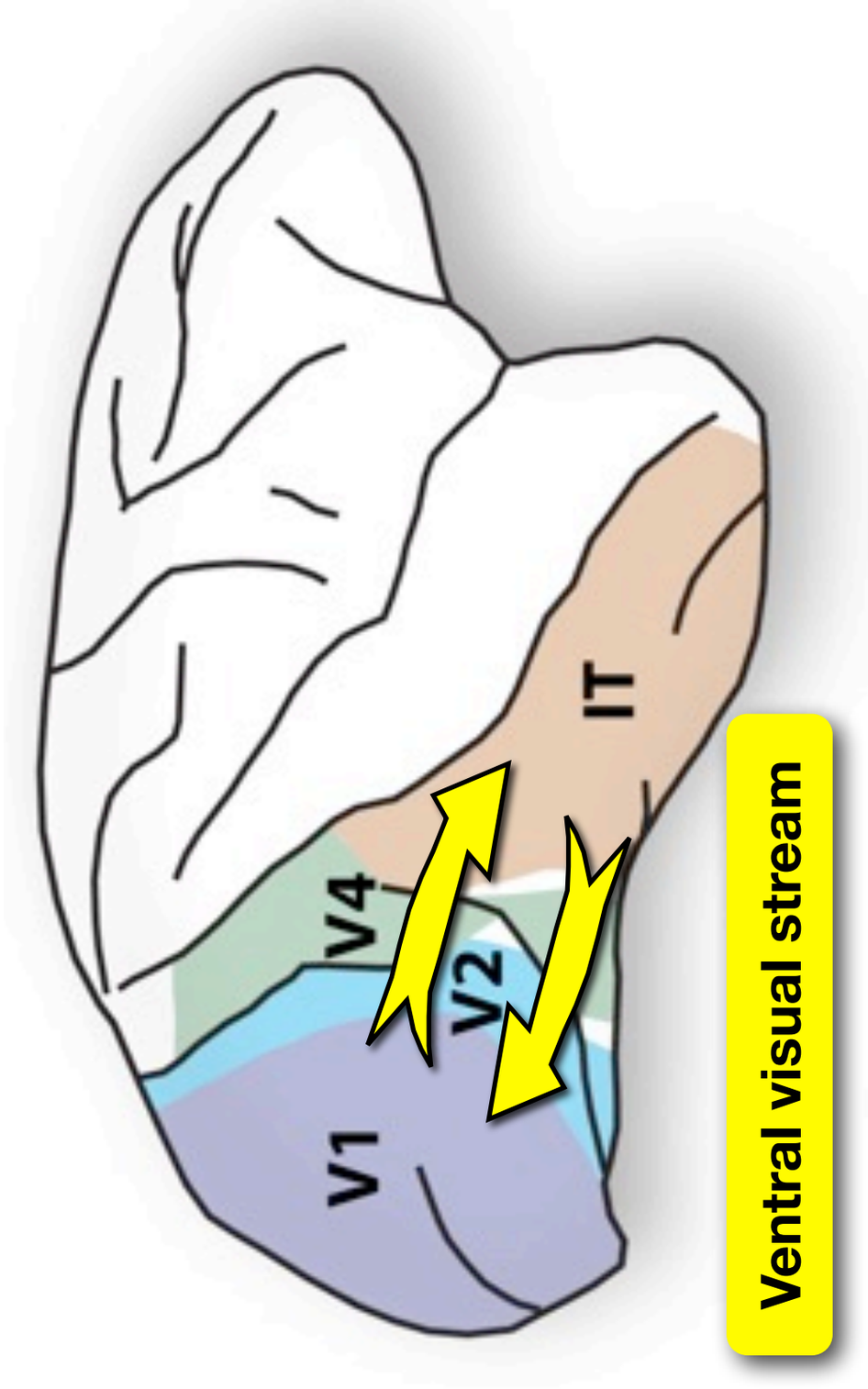
# Deep learning in the visual cortex

Thomas Serre
Brown University

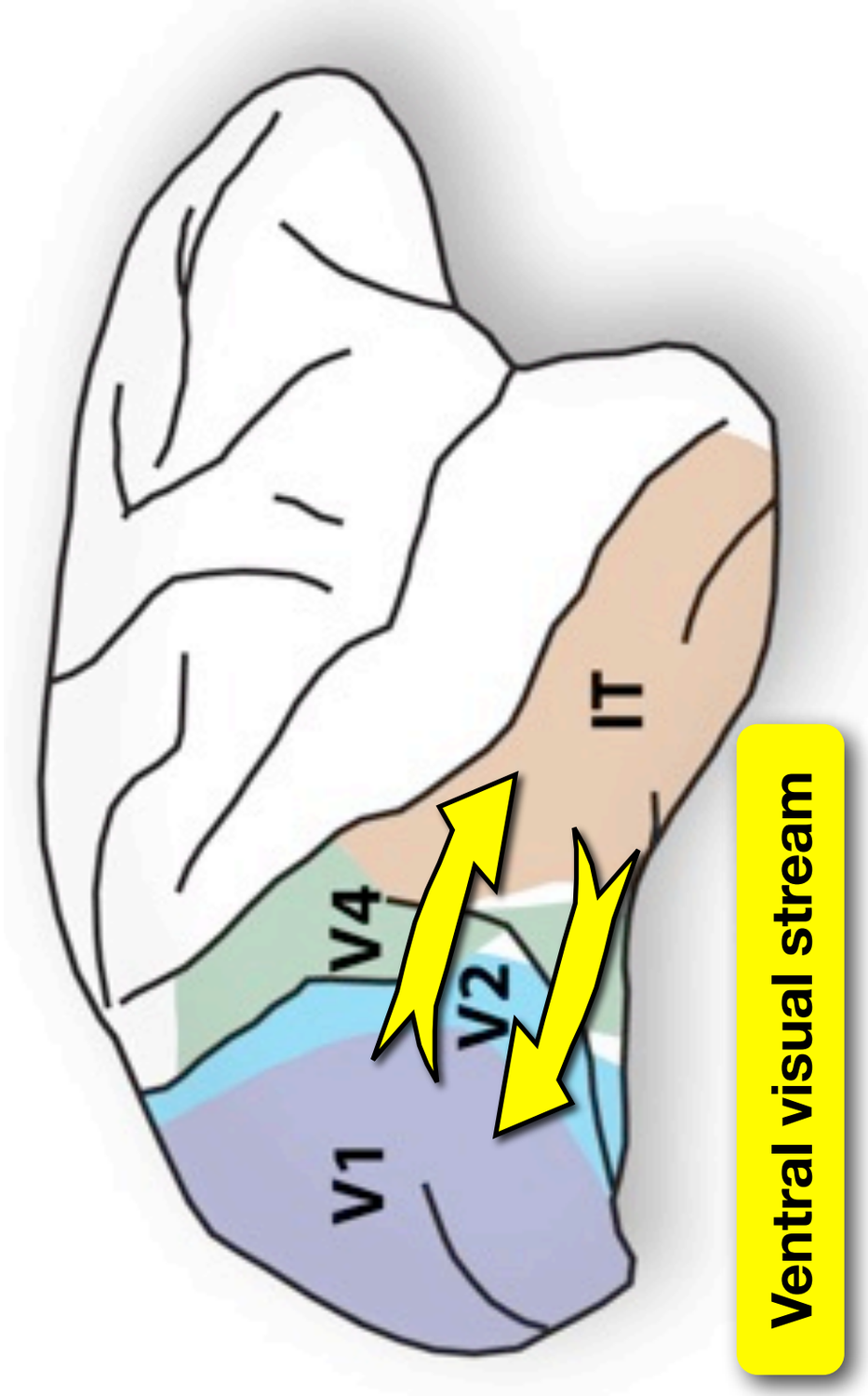
I. Fundamentals of primate vision

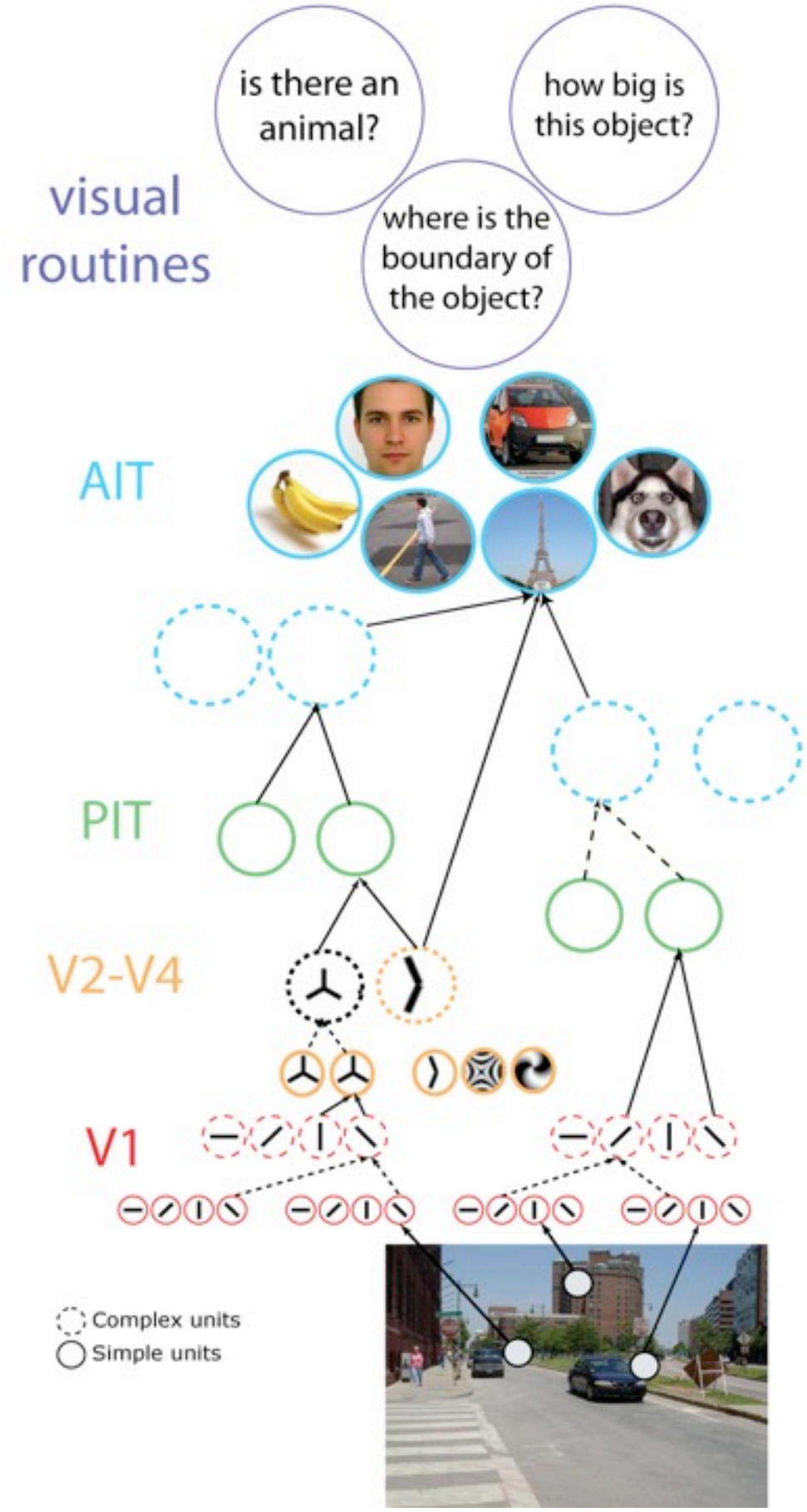
II. Computational mechanisms of rapid recognition and feedforward processing

III. Beyond feedforward processing: Attentional mechanisms and cortical feedback

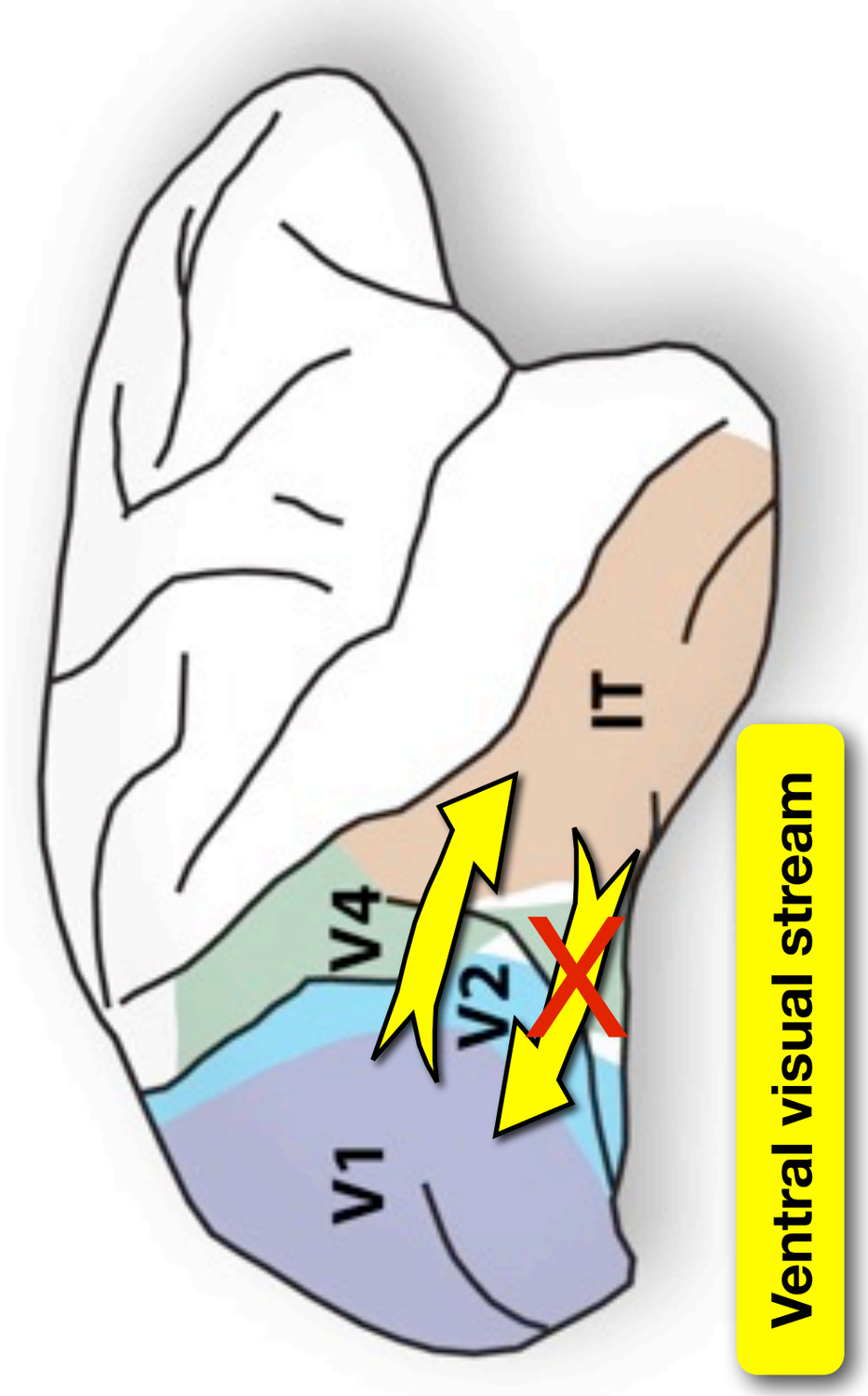


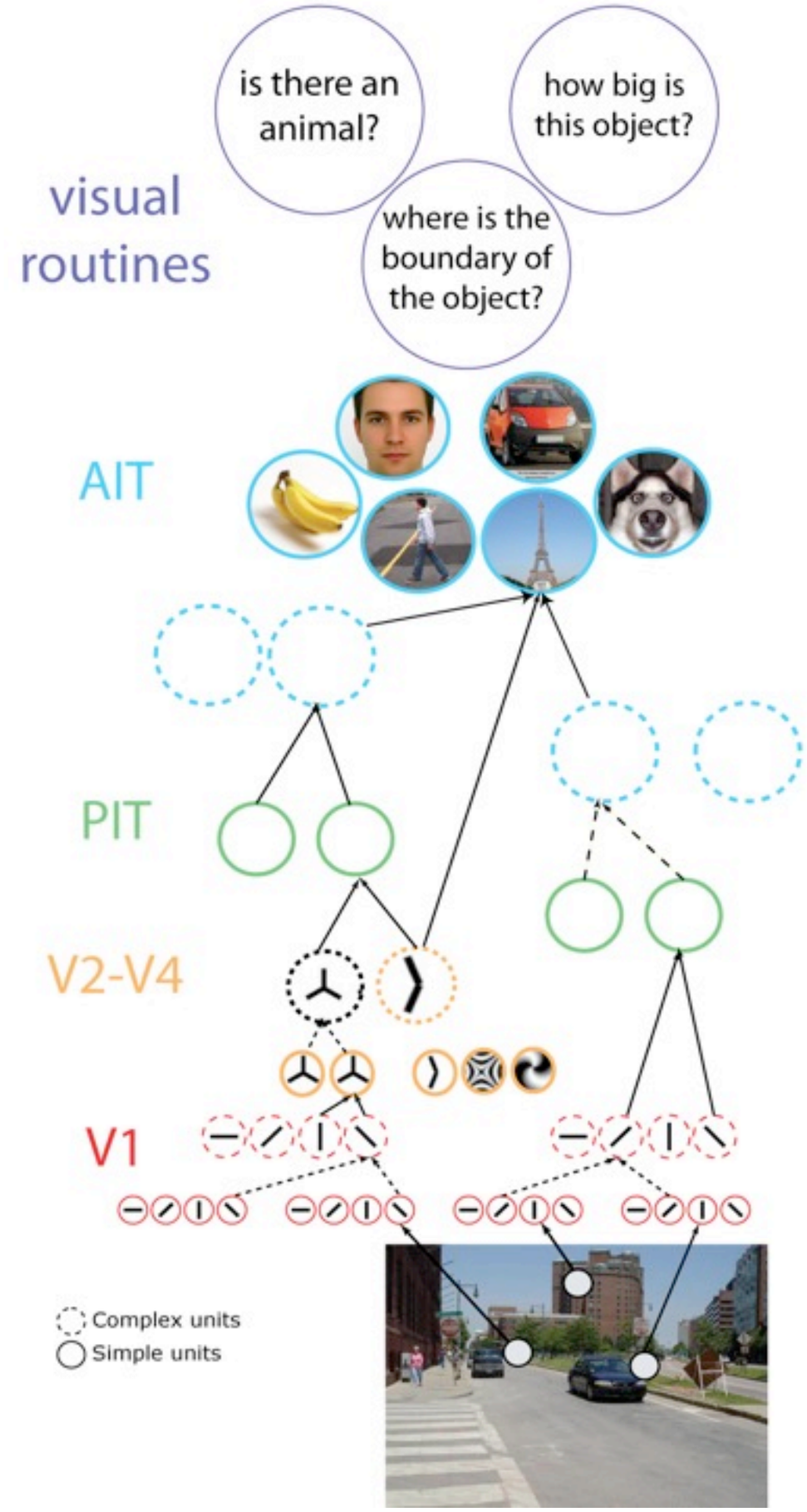

# Feedforward hierarchical model of object recognition




# Feedforward hierarchical model of object recognition



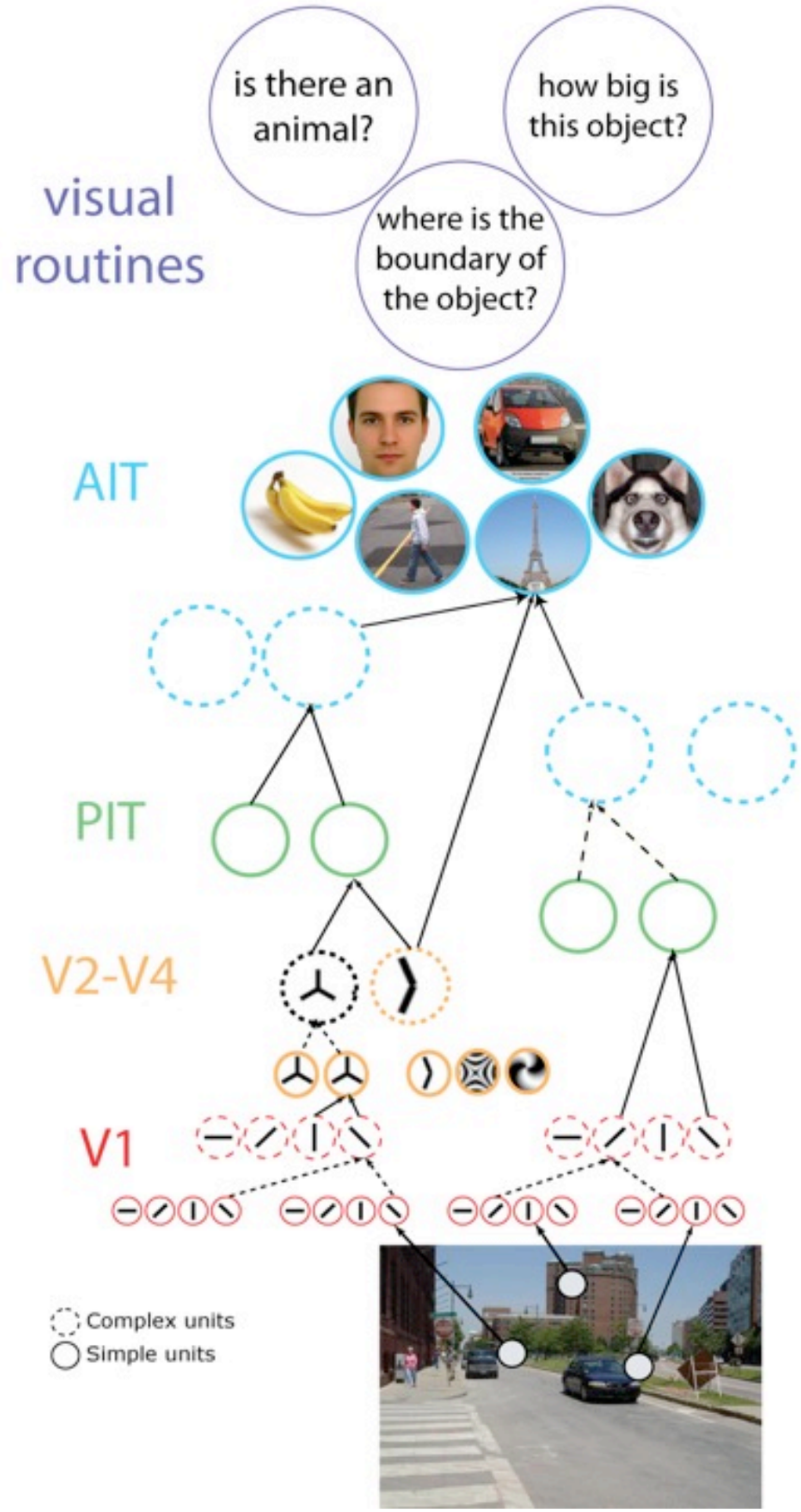




# Feedforward hierarchical model of object recognition

- Qualitative neurobiological models
  (Hubel & Wiesel '59; Perrett & Oram '93)

- Biologically-inspired computer vision systems
  (Fukushima '80; Mel '97; LeCun et al '98; Lowe, '00; Thorpe '02; Ullman et al '02; Wersing and Koerner '03; Mutch & Lowe '06)

- Quantitative neurobiological models
  (Wallis & Rolls '97; Riesenhuber & Poggio '99; Amit & Mascaro '03; Serre et al 2005; Deco & Rolls '06; Masquelier & Thorpe '07)

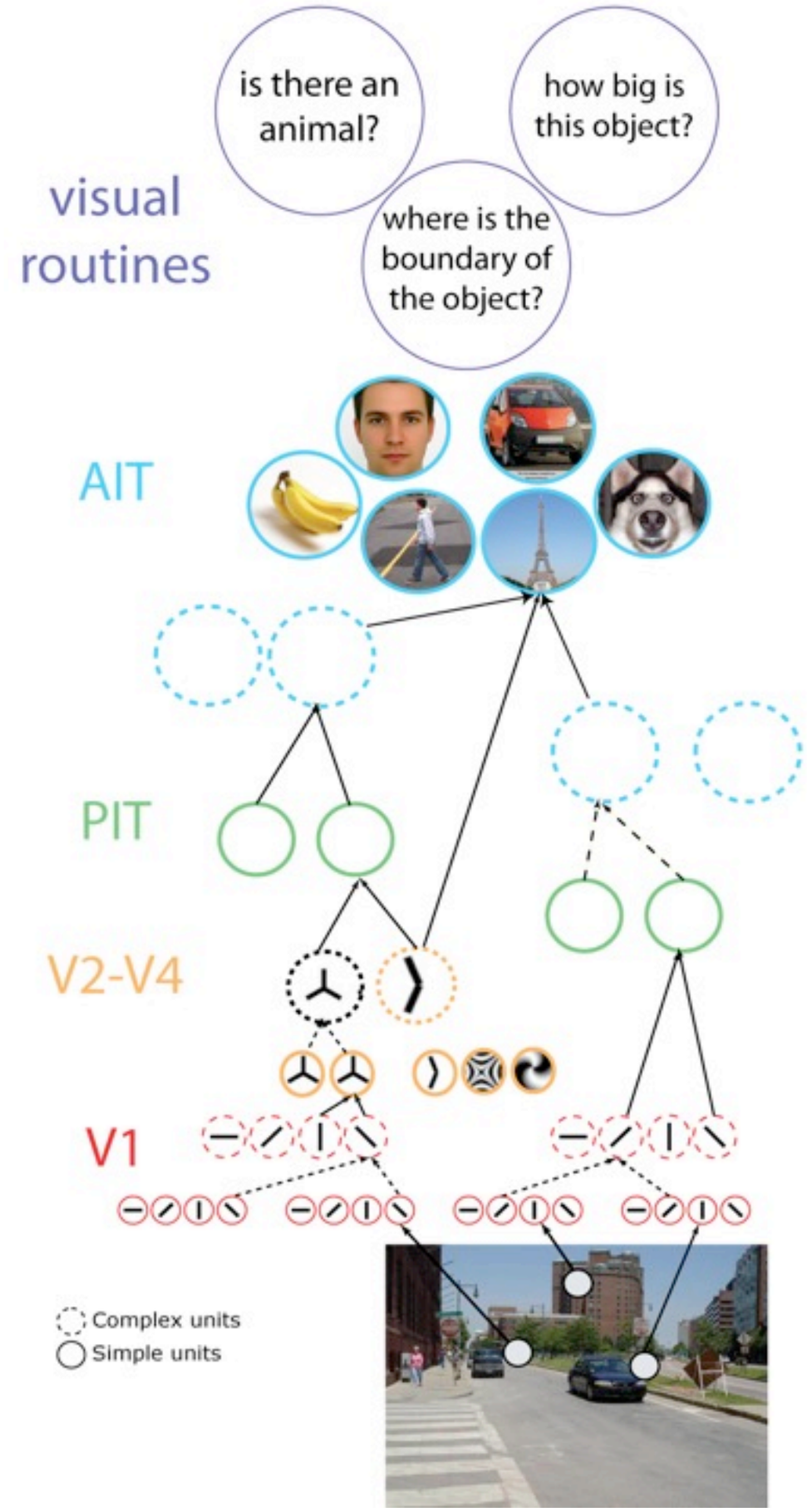



Feedforward hierarchical model of object recognition

- System-level feedforward computational model, large-scale (100M units), spans several areas of the visual cortex

- Some similarities with state-of-the-art computer vision systems (e.g., convolutional and deep belief nets)

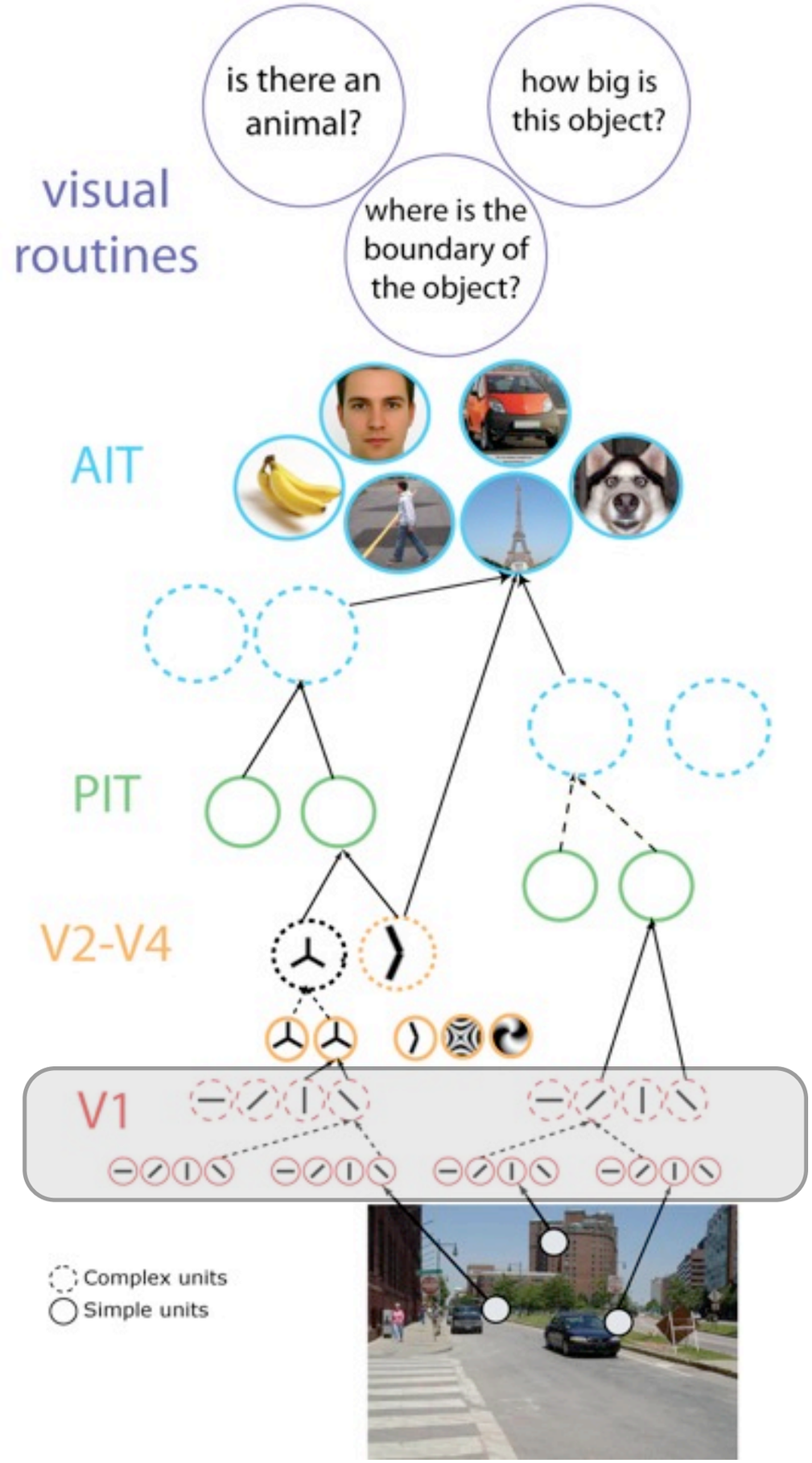
- But **constrained by anatomy and physiology** and shown to be consistent with experimental data across areas of visual cortex




# Feedforward hierarchical model of object recognition

# Cortical mechanisms of invariant recognition

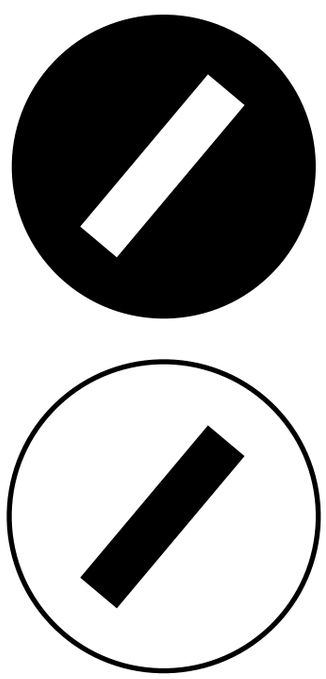
- Processing based on selective pooling mechanisms (to gradually increase selectivity and invariance of units along hierarchy)

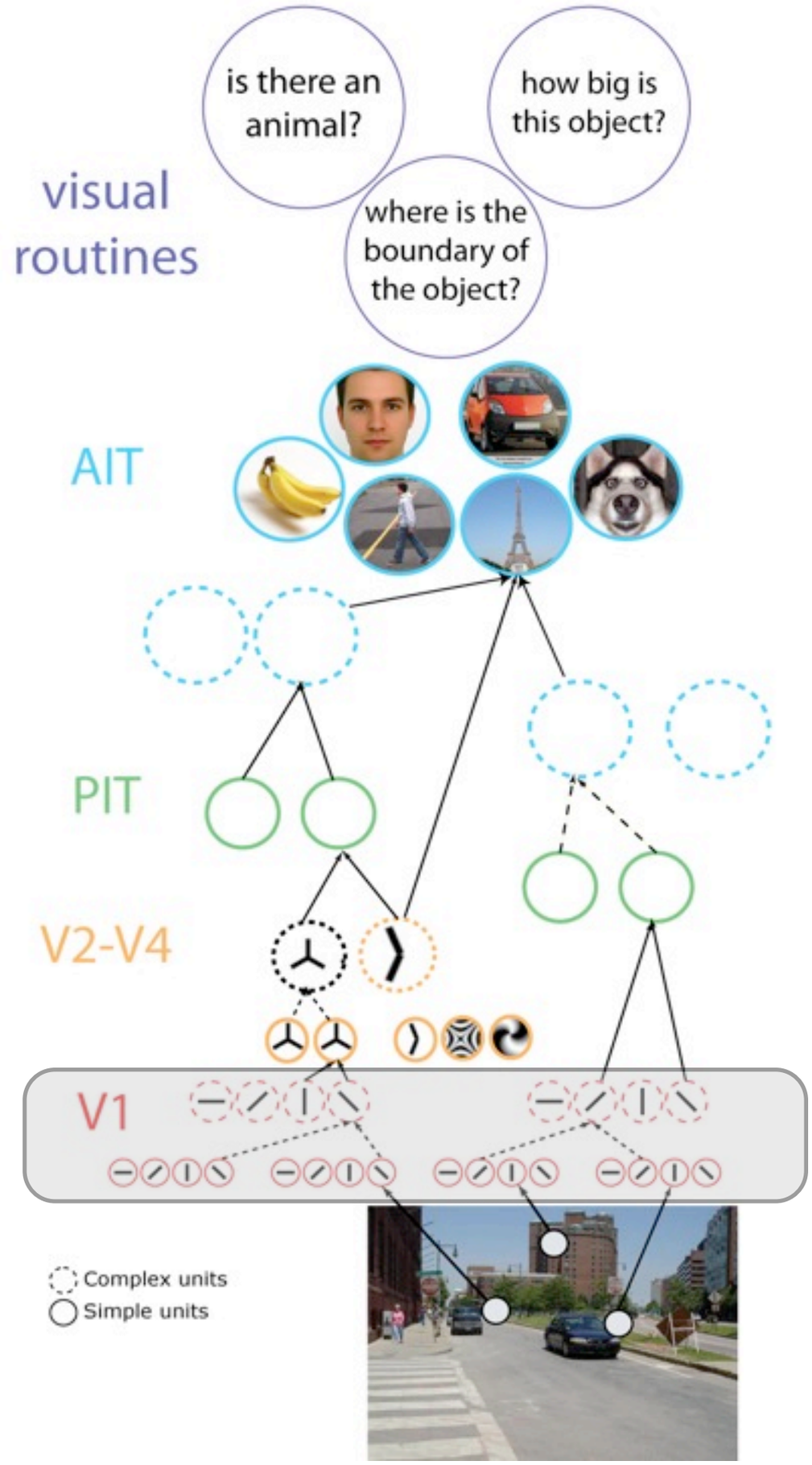


# Cortical mechanisms of invariant recognition

- Processing based on selective pooling mechanisms (to gradually increase selectivity and invariance of units along hierarchy)

simple cells

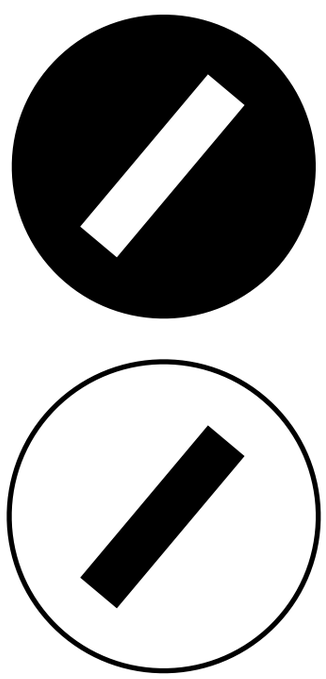




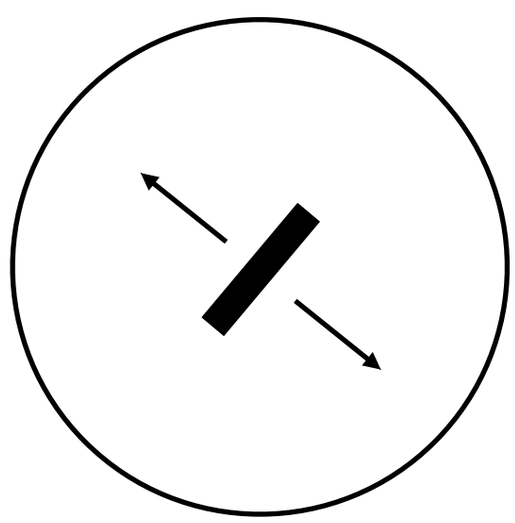
# Cortical mechanisms of invariant recognition

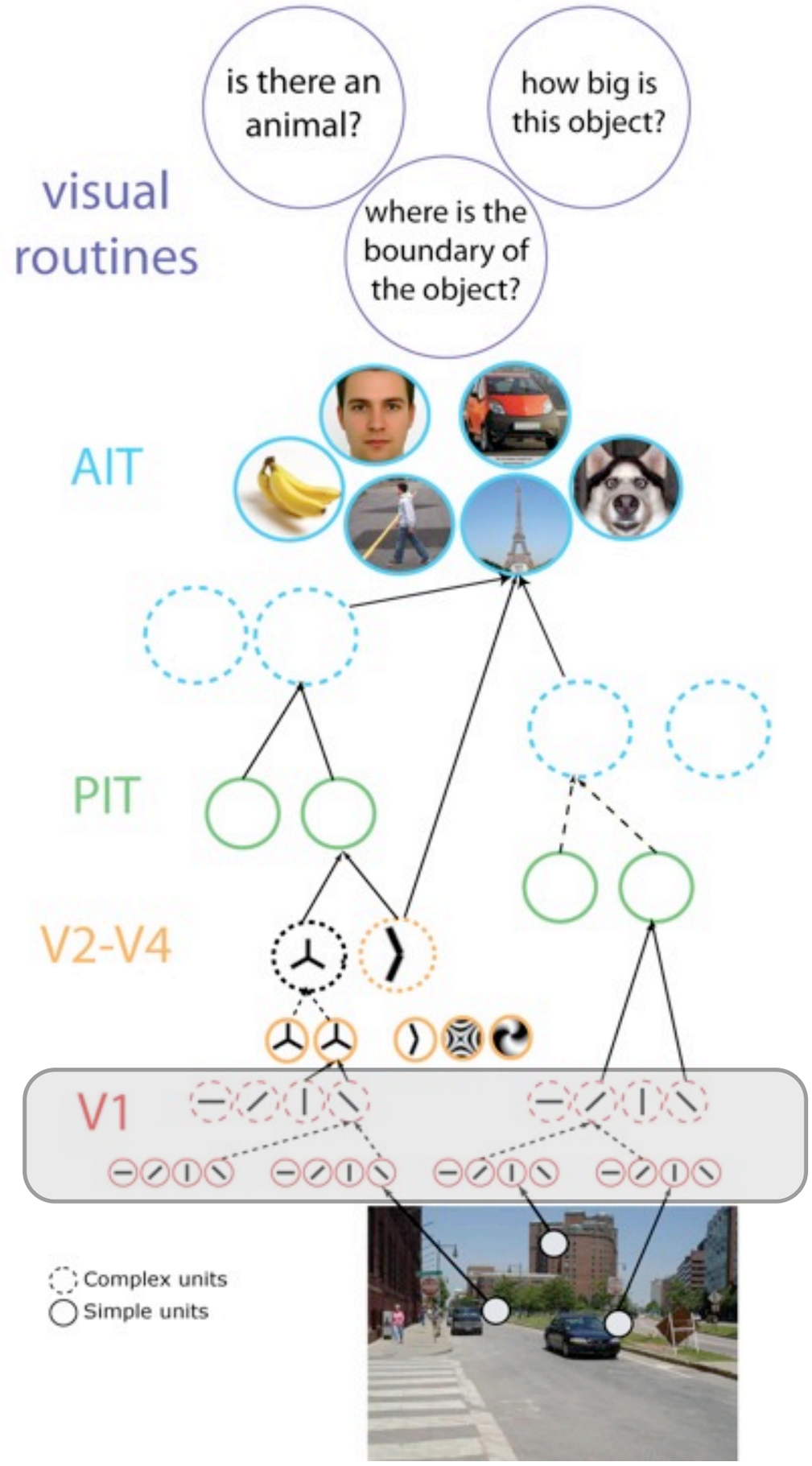
- Processing based on selective pooling mechanisms (to gradually increase selectivity and invariance of units along hierarchy)

simple cells



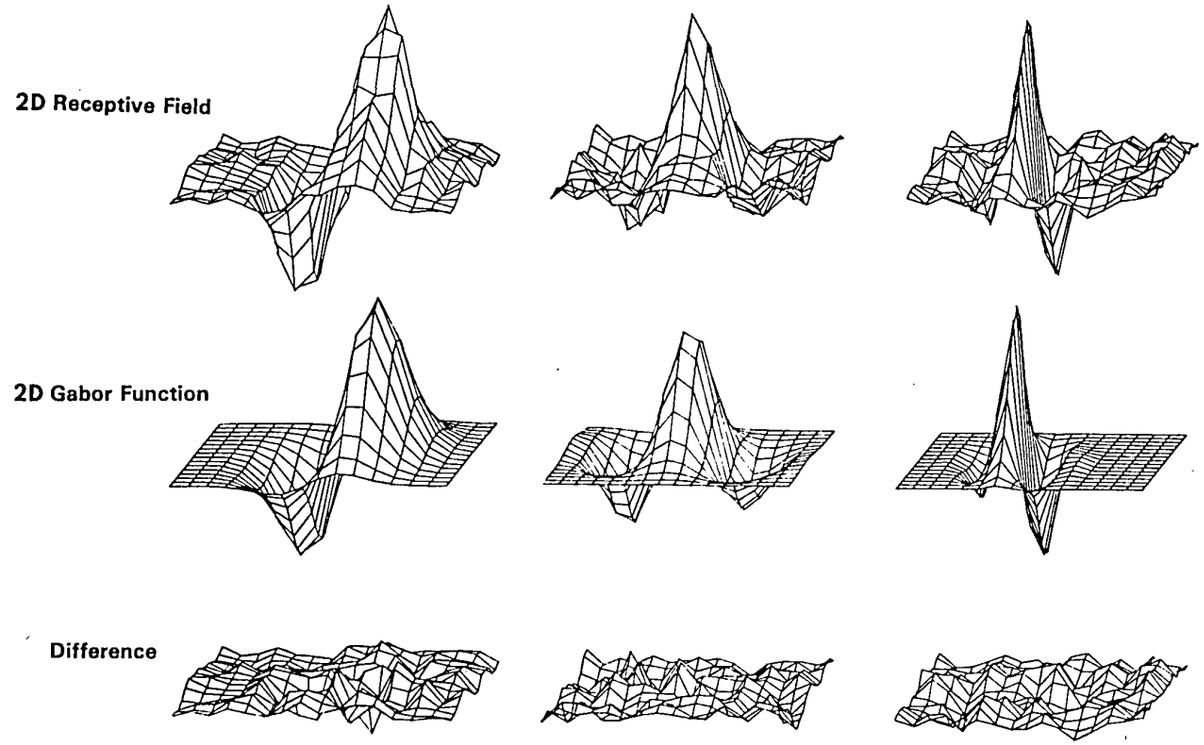
complex cells





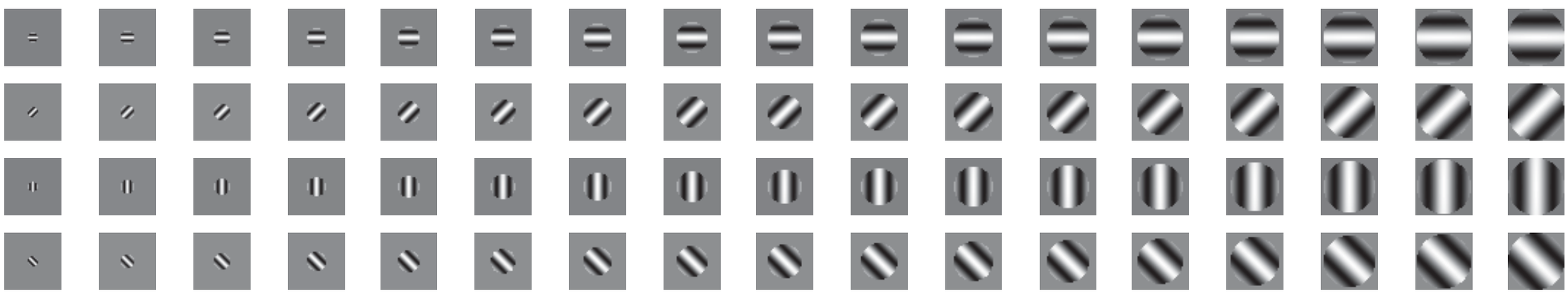
# Cortical mechanisms of invariant recognition

## S1 simple units




Jones & Palmer '87

**Gabor filters at multiple phases (one phase shown), orientations and spatial frequencies/scales (parameters derived from available experimental data)**



# Cortical mechanisms of invariant recognition

## S1 simple units

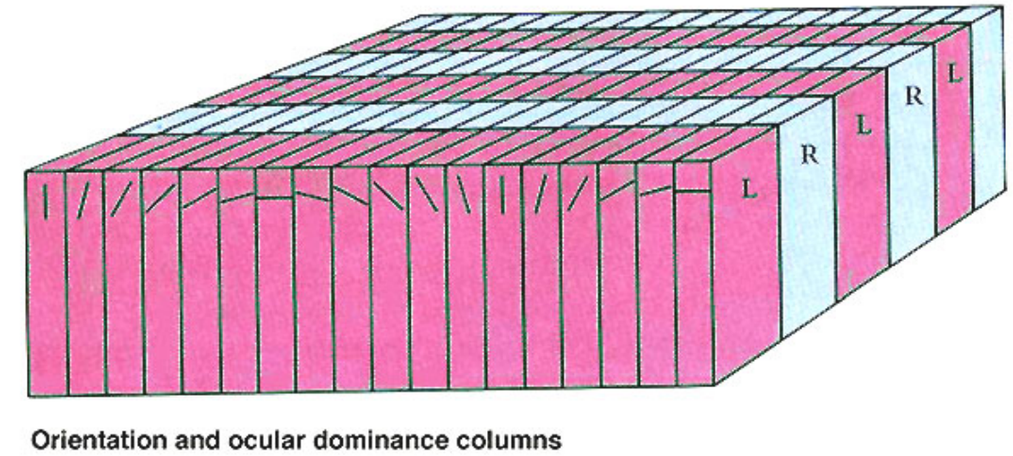
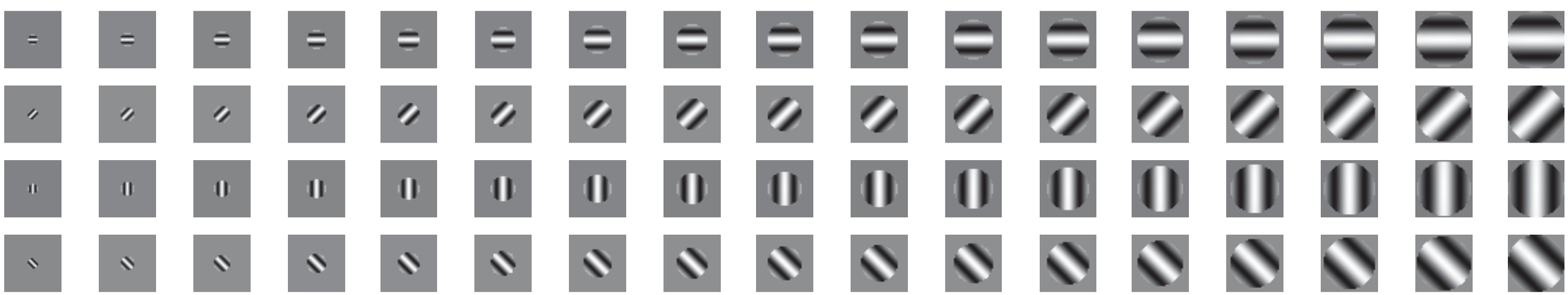



*Figure 23. The ice-cube model of the cortex. It illustrates how the cortex is divided, at the same time, into two kinds of slabs, one set of ocular dominance (left and right) and one set for orientation. The model should not be taken literally: Neither set is as regular as this, and the orientation slabs especially are far from parallel or straight.*

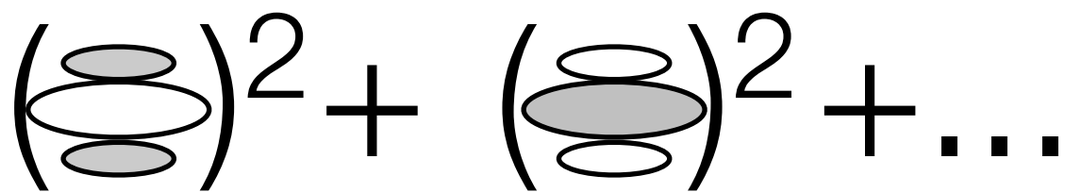
**Gabor filters at multiple phases (one phase shown), orientations and spatial frequencies/scales (parameters derived from available experimental data)**



# Cortical mechanisms of invariant recognition

## C1 complex units

**(1)** half-rectification and summing over phases *at each location* for tolerance to contrast reversal



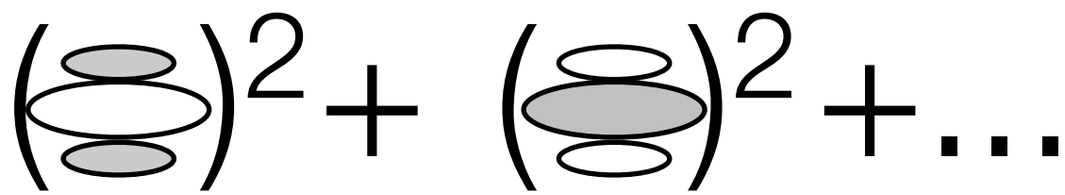


Heeger & Carandini ’94; Lampl et al ’01; Touryan et al ’02; Rust et al ’05; Finn & Ferster ’07

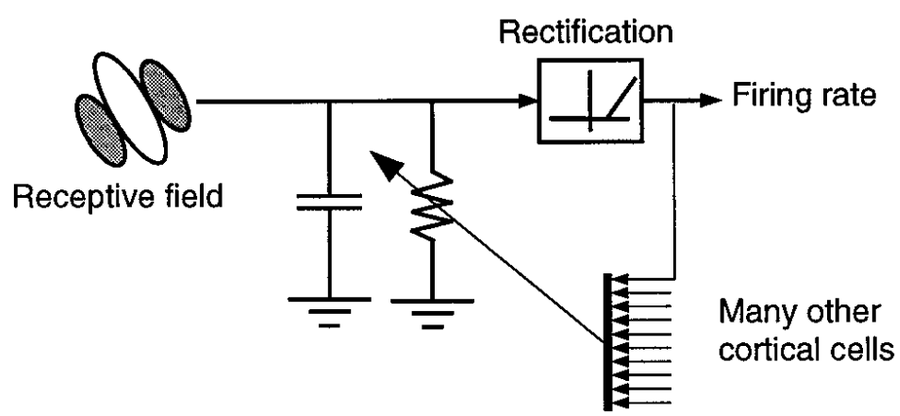
# Cortical mechanisms of invariant recognition

## C1 complex units

**(1)** half-rectification and summing over phases *at each location* for tolerance to contrast reversal



**(2)** gain control / divisive normalization



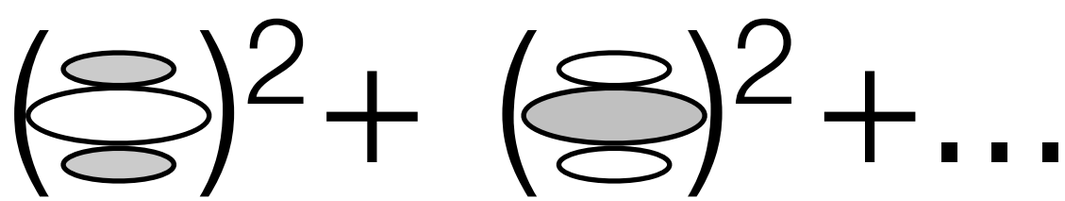



Heeger & Carandini ’94; Lampl et al ’01; Touryan et al ’02; Rust et al ’05; Finn & Ferster ’07

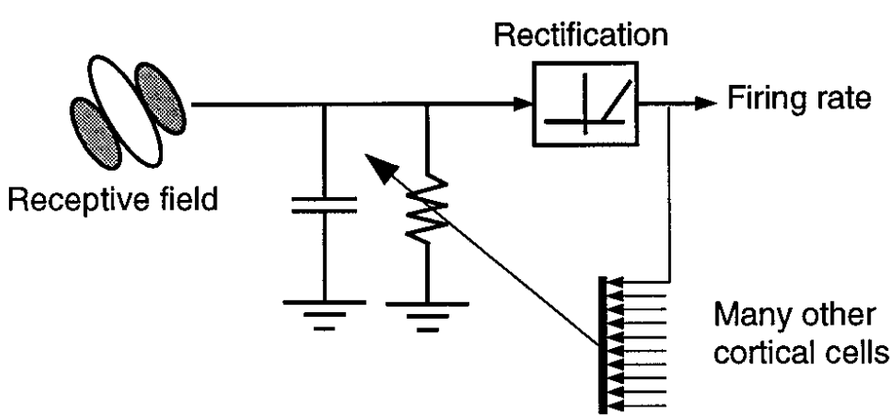
# Cortical mechanisms of invariant recognition

## C1 complex units

**(1)** half-rectification and summing over phases *at each location* for tolerance to contrast reversal

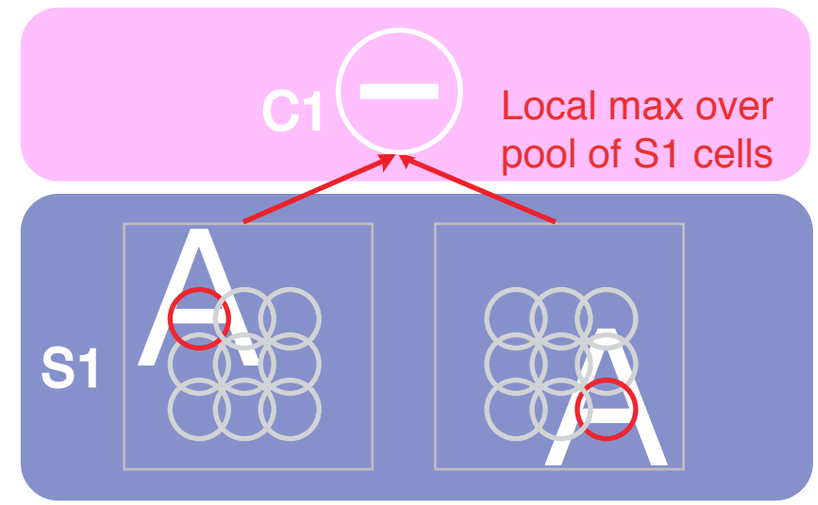


**(2)** gain control / divisive normalization

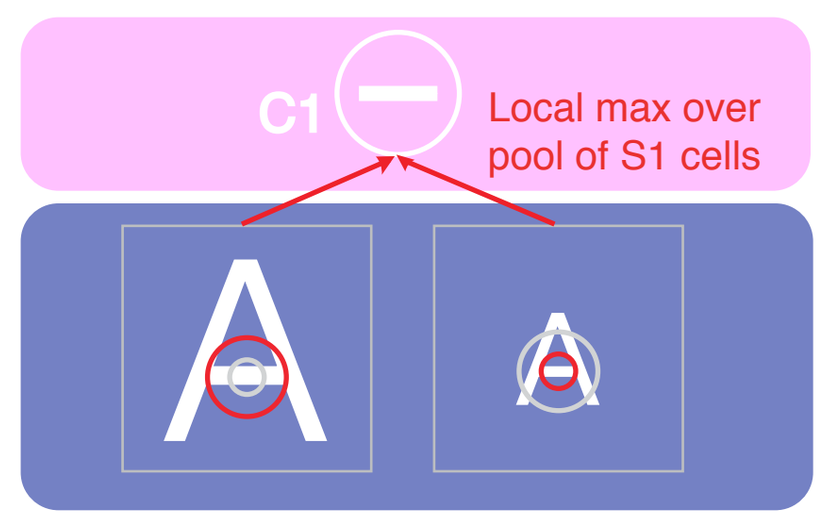



**(3)** selective max-like pooling over nearby positions and scales for tolerance to 2D transformations

Increase in tolerance to position



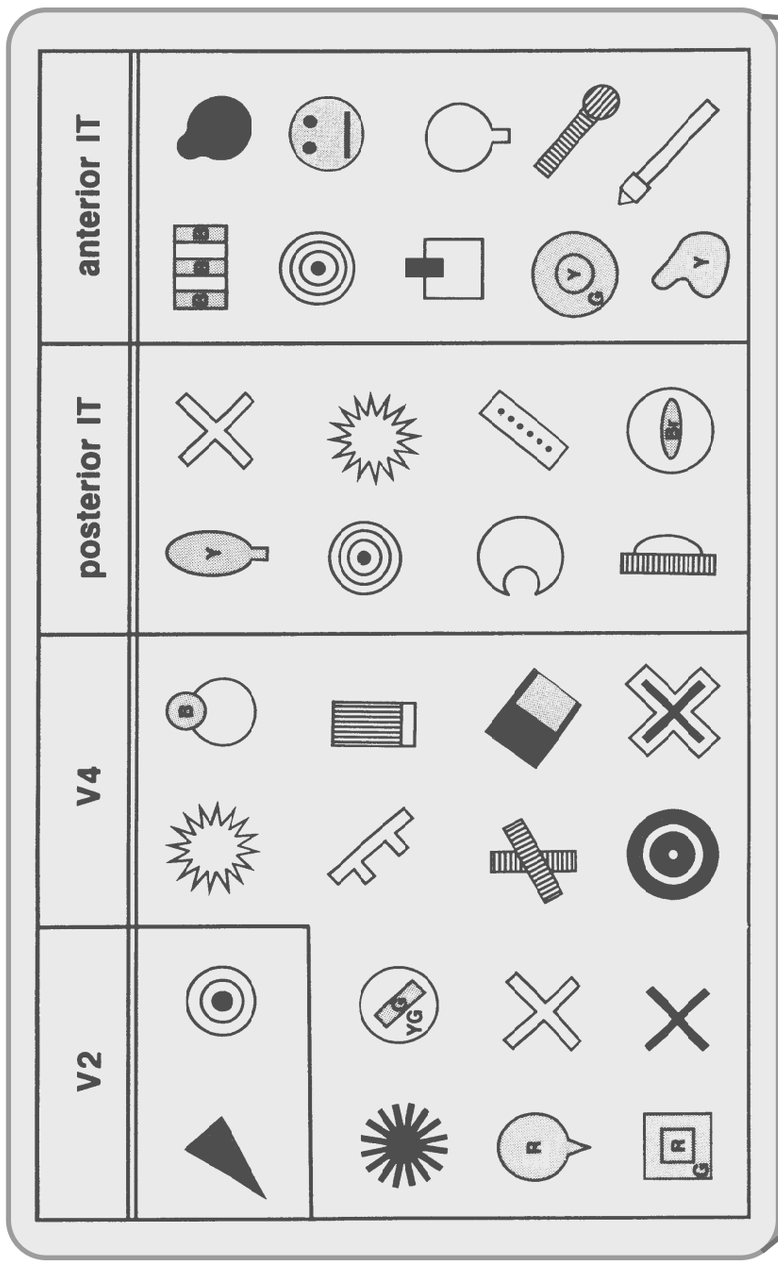

Increase in tolerance to scale

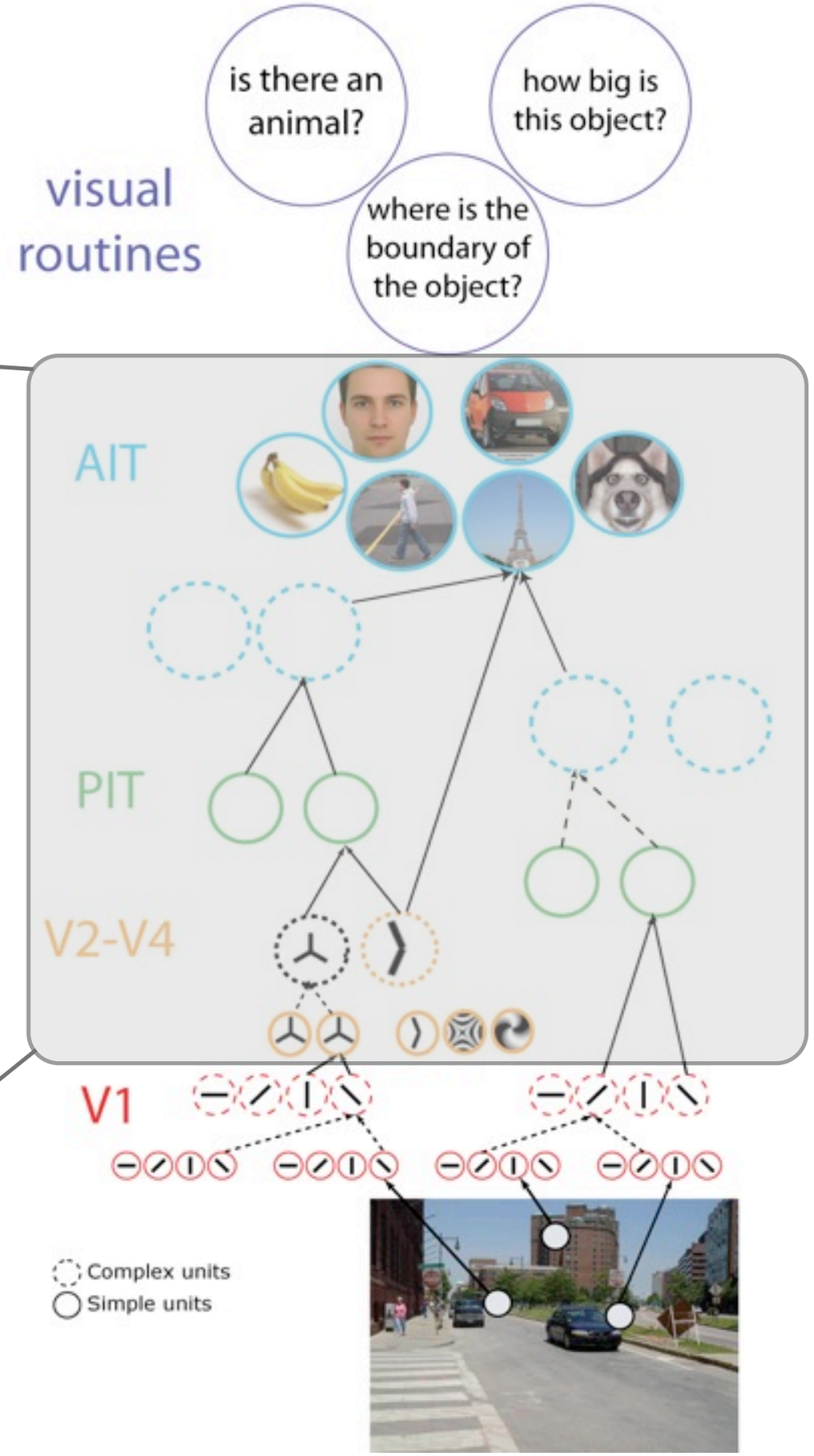





Heeger & Carandini '94; Lampl et al '01; Touryan et al '02; Rust et al '05; Finn & Ferster '07

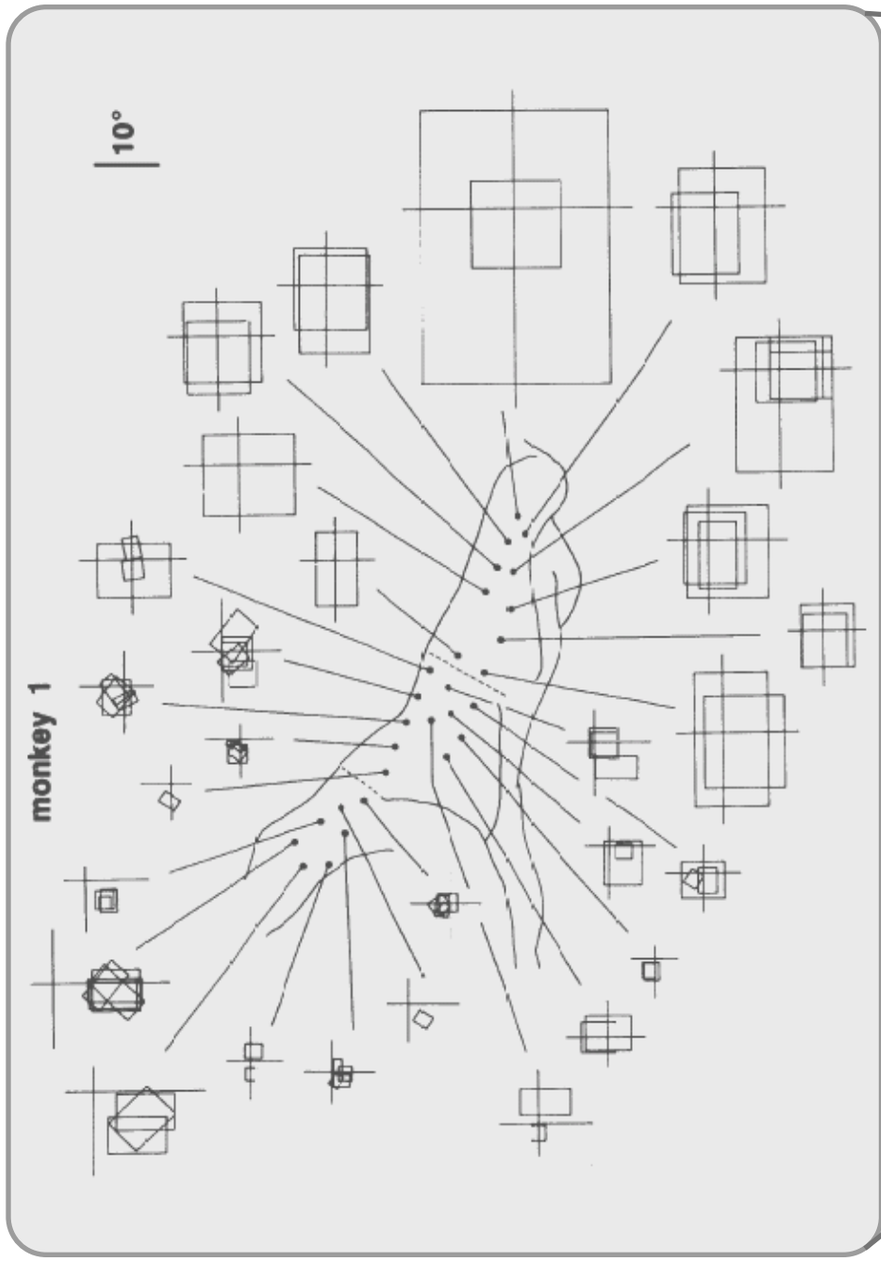
# Cortical mechanisms of invariant recognition

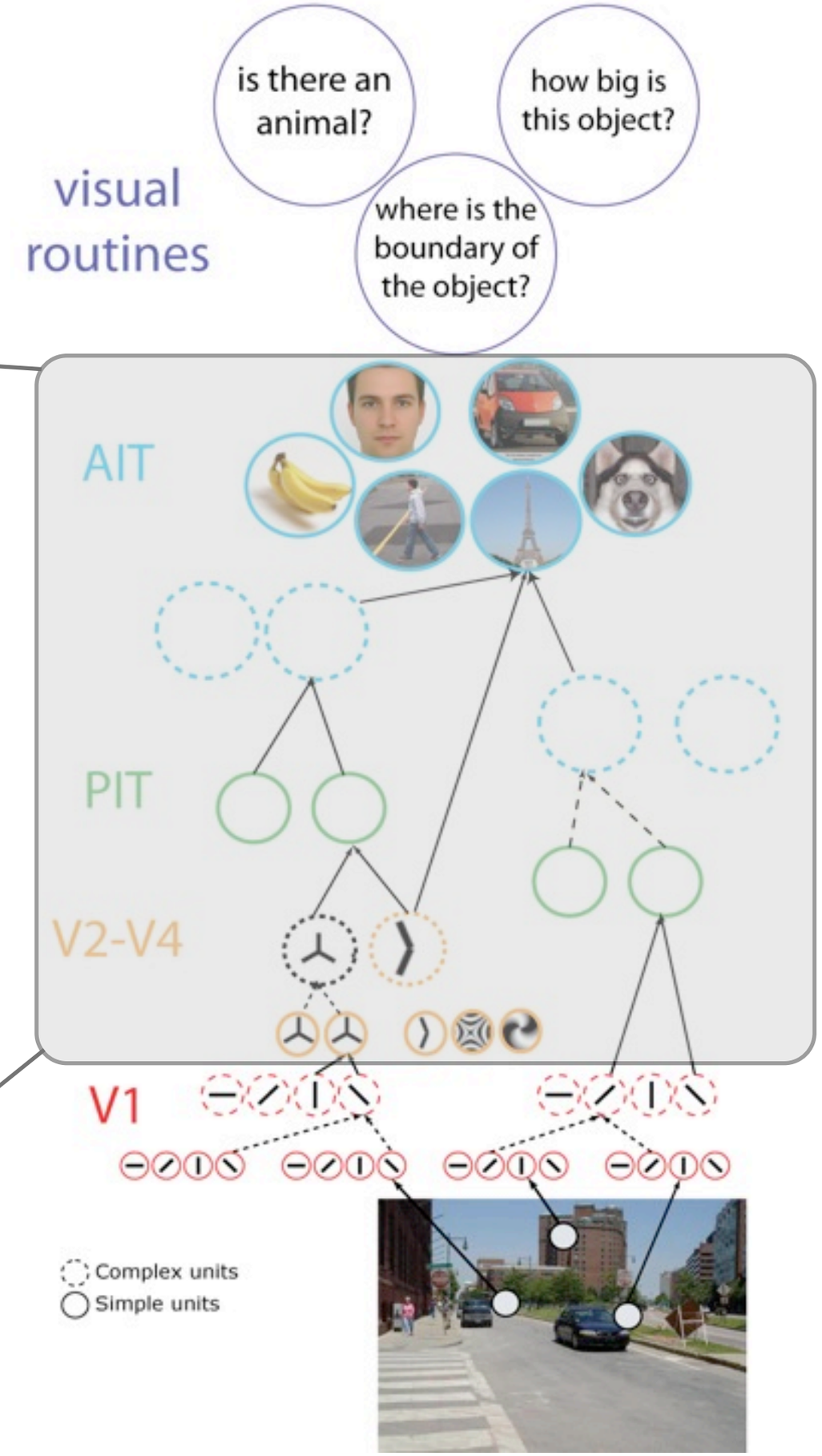



Kobatake & Tanaka 1994

# Cortical mechanisms of invariant recognition



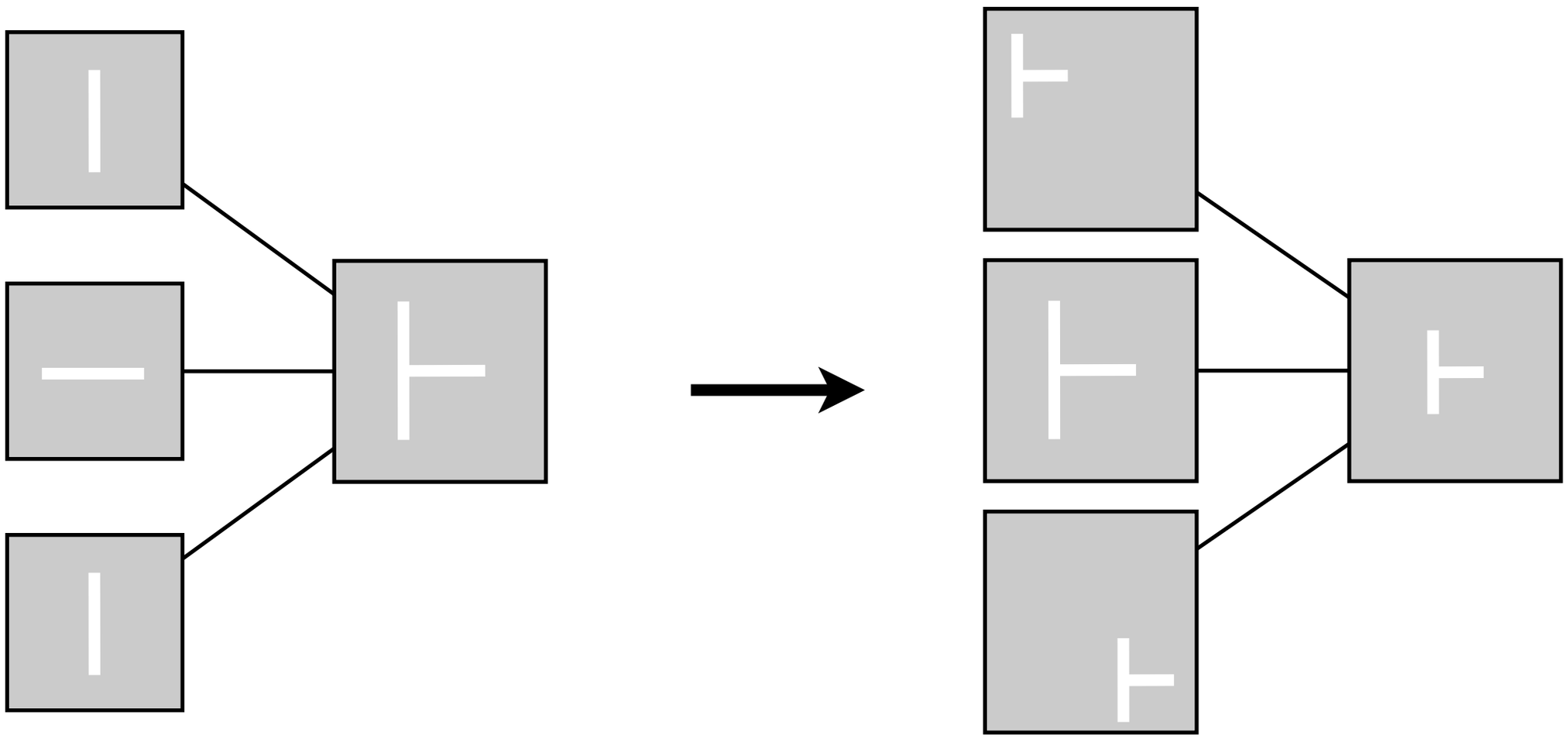

Kobatake & Tanaka 1994

# Simple units

Template matching
Bell-shape like tuning
~ “AND”

# Complex units

Invariance pooling
max-like operation
~”OR”

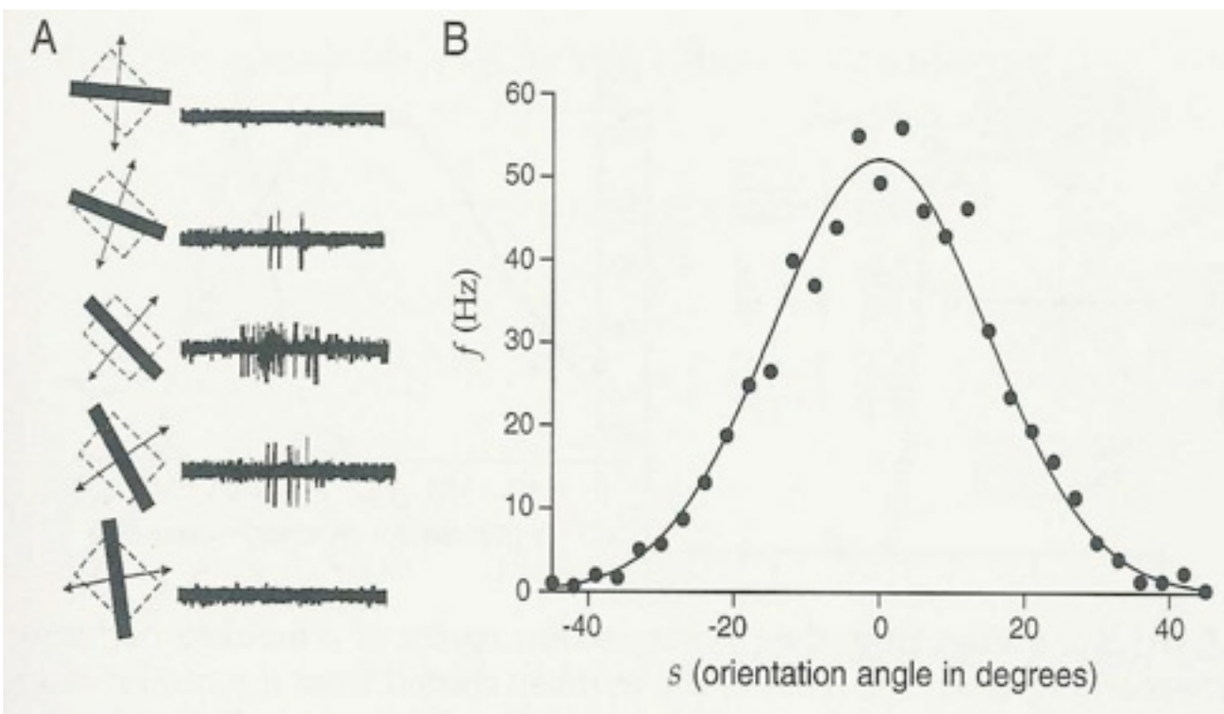


## Selective pooling mechanisms

Riesenhuber & Poggio 1999 (building on Fukushima ‘80 and Hubel & Wiesel ‘62)

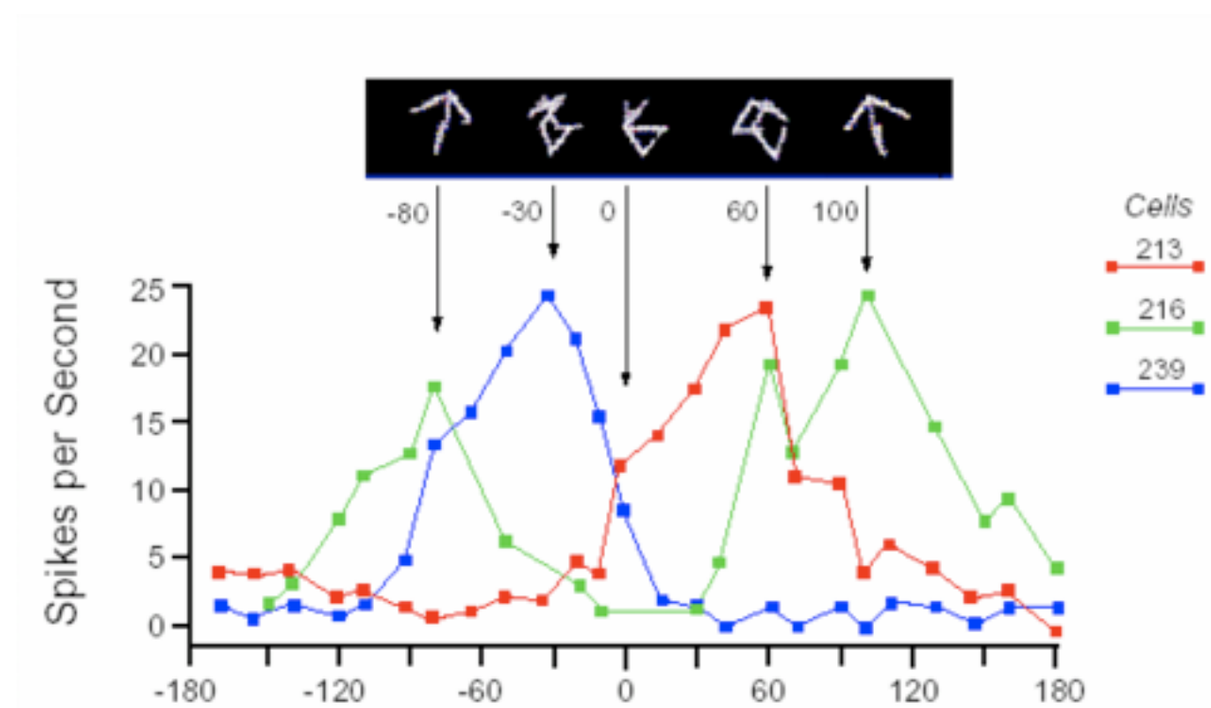
# Tuning in the visual cortex

Dayan & Abbott '01




**Tuning for orientation in V1**

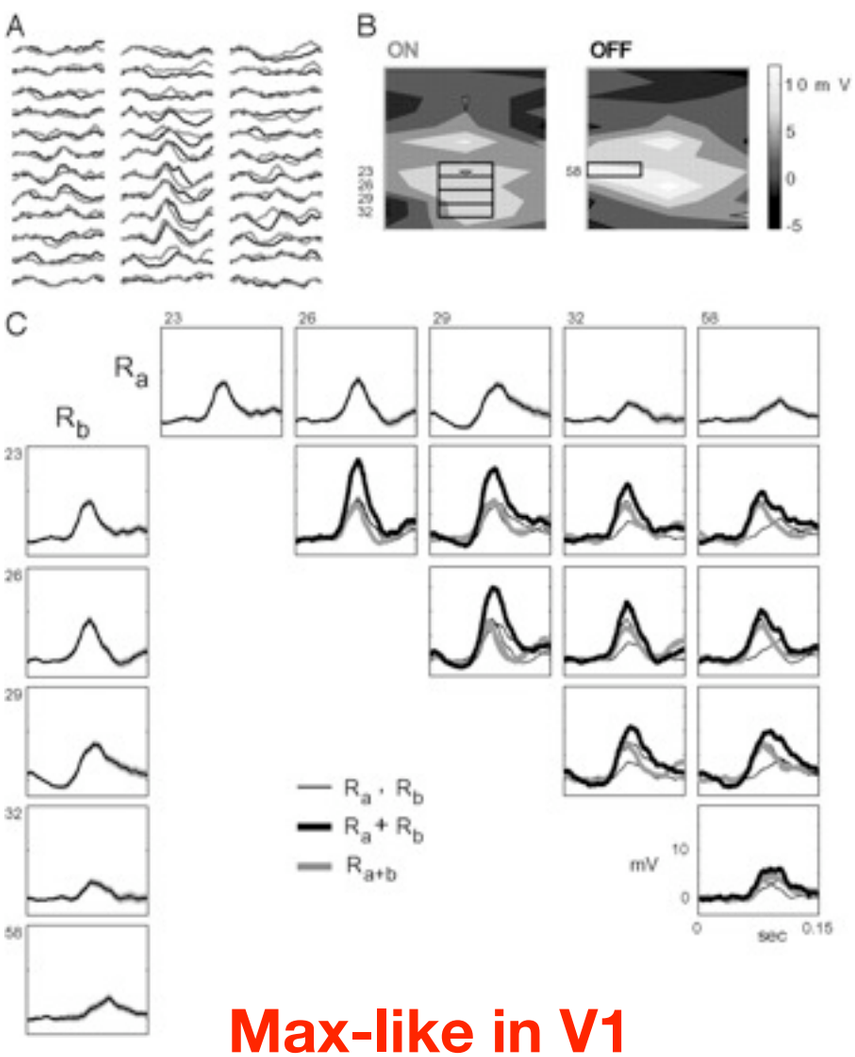
Logothetis Pauls & Poggio '95




**Tuning around 3D views in IT**

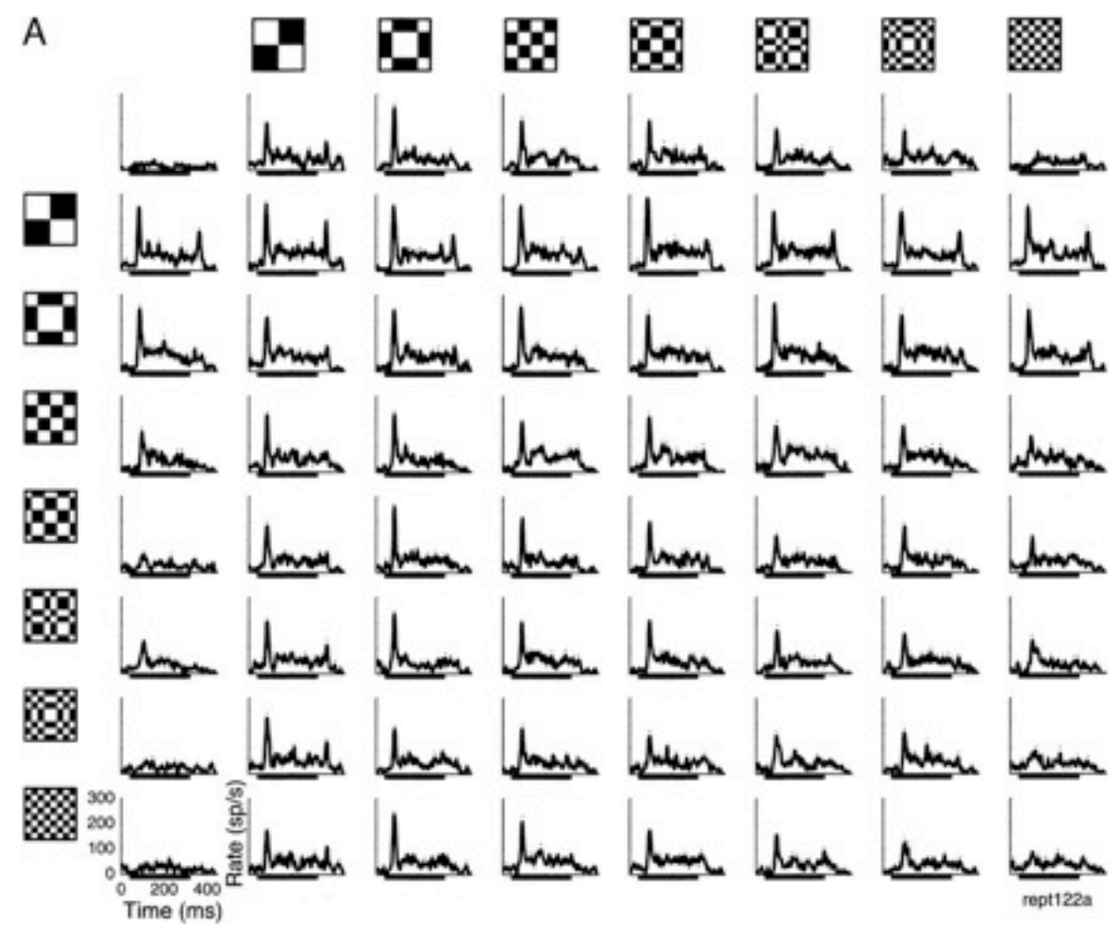
# Max-like computation in the visual cortex

Lampl et al ‘04




**Max-like in V1**

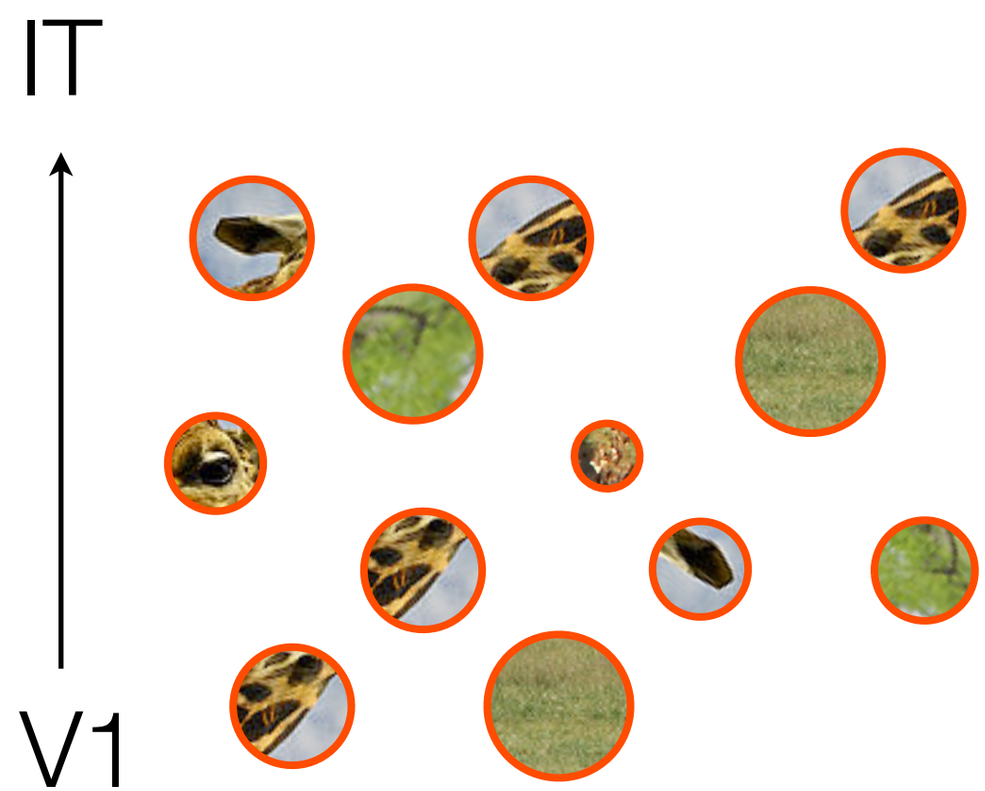
Gawne & Martin ‘02




**Max-like in V4**

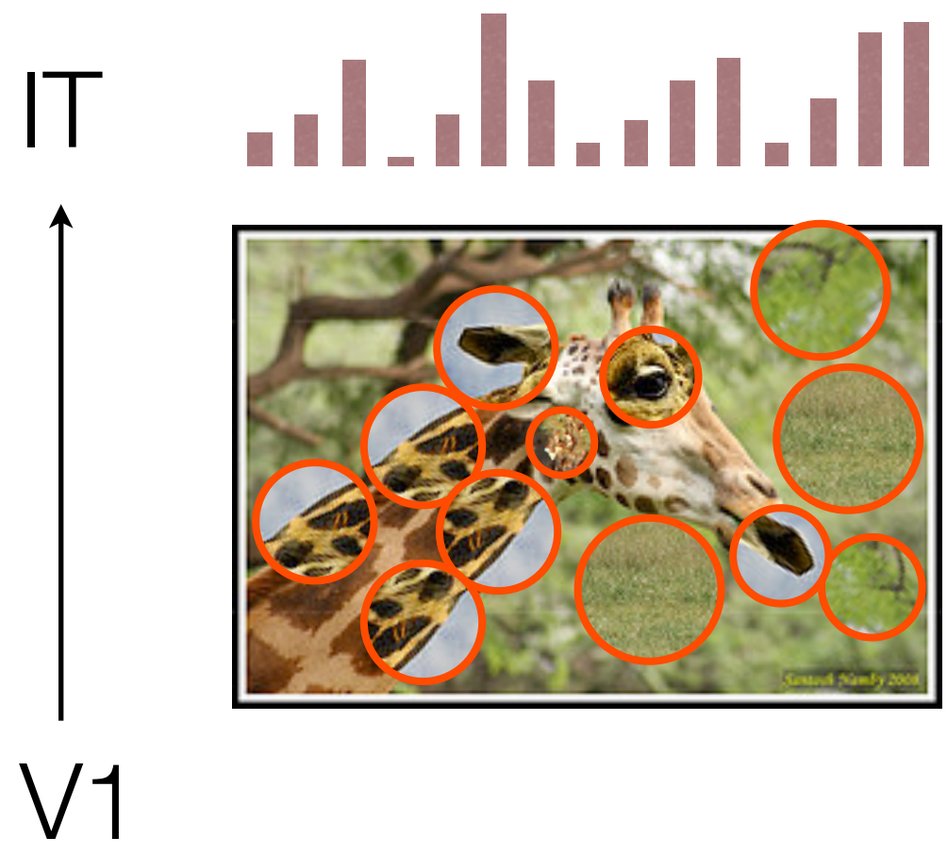
# Mechanisms of invariant recognition

- Unit parameters (i.e., RF sizes, pooling ranges, etc) constrained by available experimental data
- Unsupervised learning of (hierarchy) of frequent image fragments during development (in intermediate stages) shared across categories
- Rapid object recognition based on bottom-up activation of hierarchy of image fragments that allow category info. to be decoded by higher level categorization processes



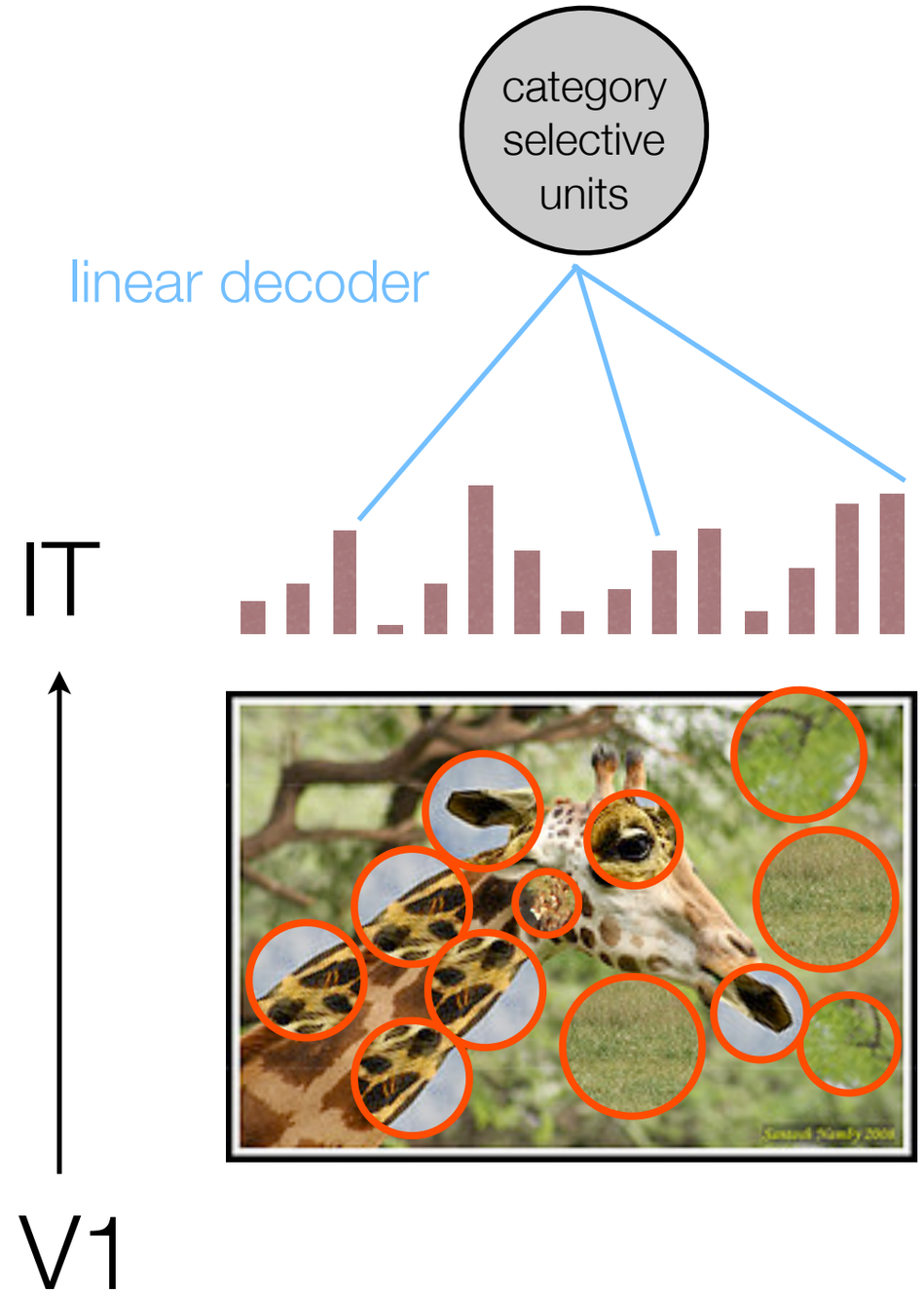
# Mechanisms of invariant recognition

- Unit parameters (i.e., RF sizes, pooling ranges, etc) constrained by available experimental data
- Unsupervised learning of (hierarchy) of frequent image fragments during development (in intermediate stages) shared across categories
- Rapid object recognition based on bottom-up activation of hierarchy of image fragments that allow category info. to be decoded by higher level categorization processes



# Mechanisms of invariant recognition

- Unit parameters (i.e., RF sizes, pooling ranges, etc) constrained by available experimental data
- Unsupervised learning of (hierarchy) of frequent image fragments during development (in intermediate stages) shared across categories
- Rapid object recognition based on bottom-up activation of hierarchy of image fragments that allow category info. to be decoded by higher level categorization processes

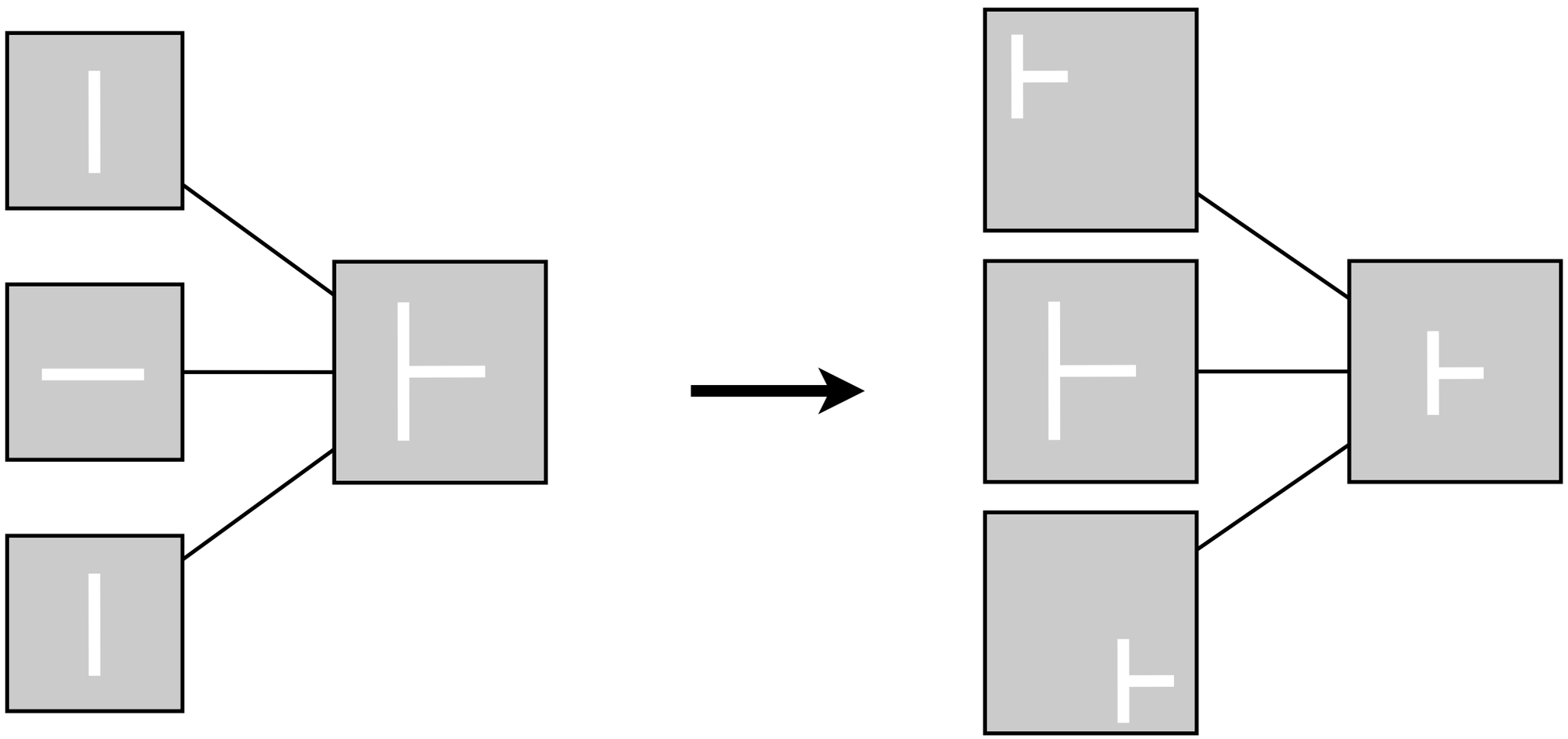


# Simple units

"frequent image features"
i.e., correlation in space

# Complex units

"frequent image transformations"
i.e., correlation in time

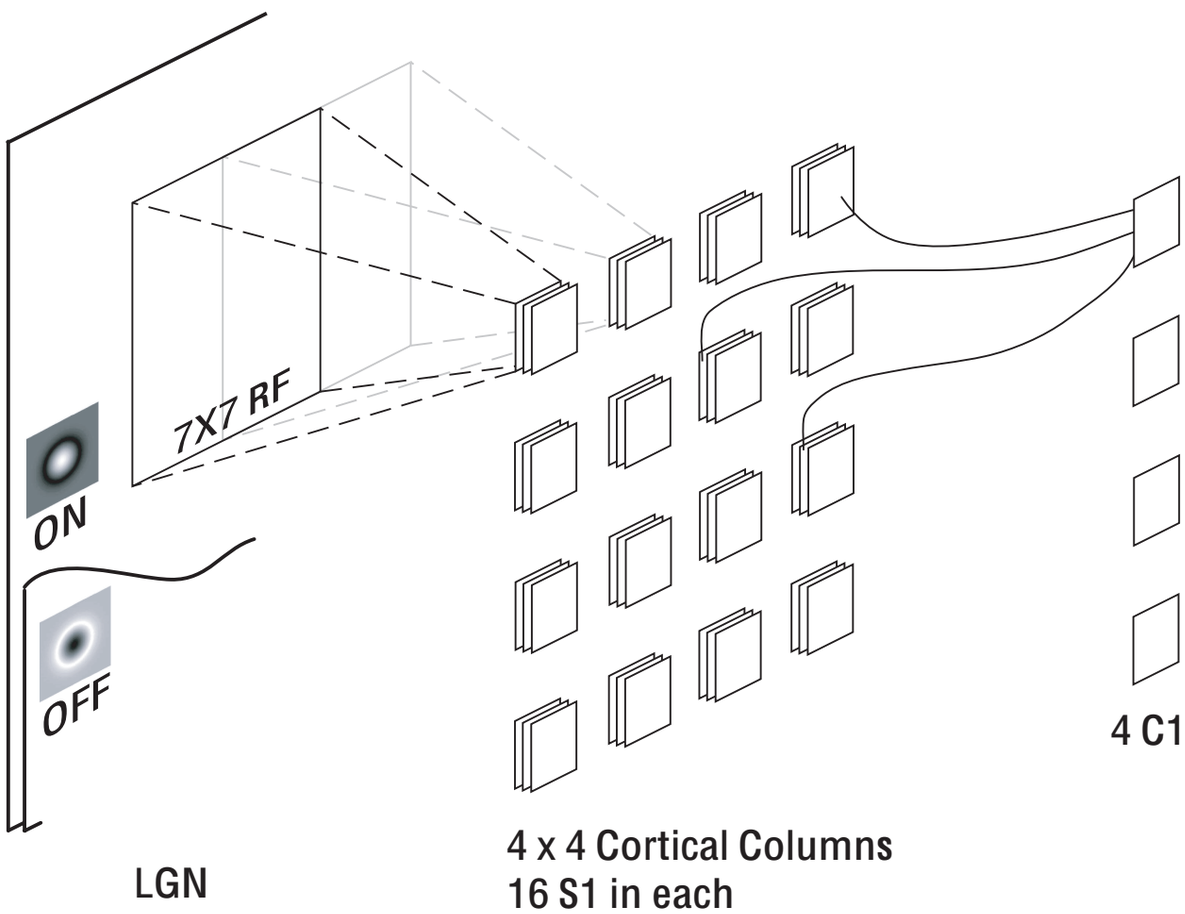


## Selective pooling mechanisms

Masquelier Serre Thorpe & Poggio '07

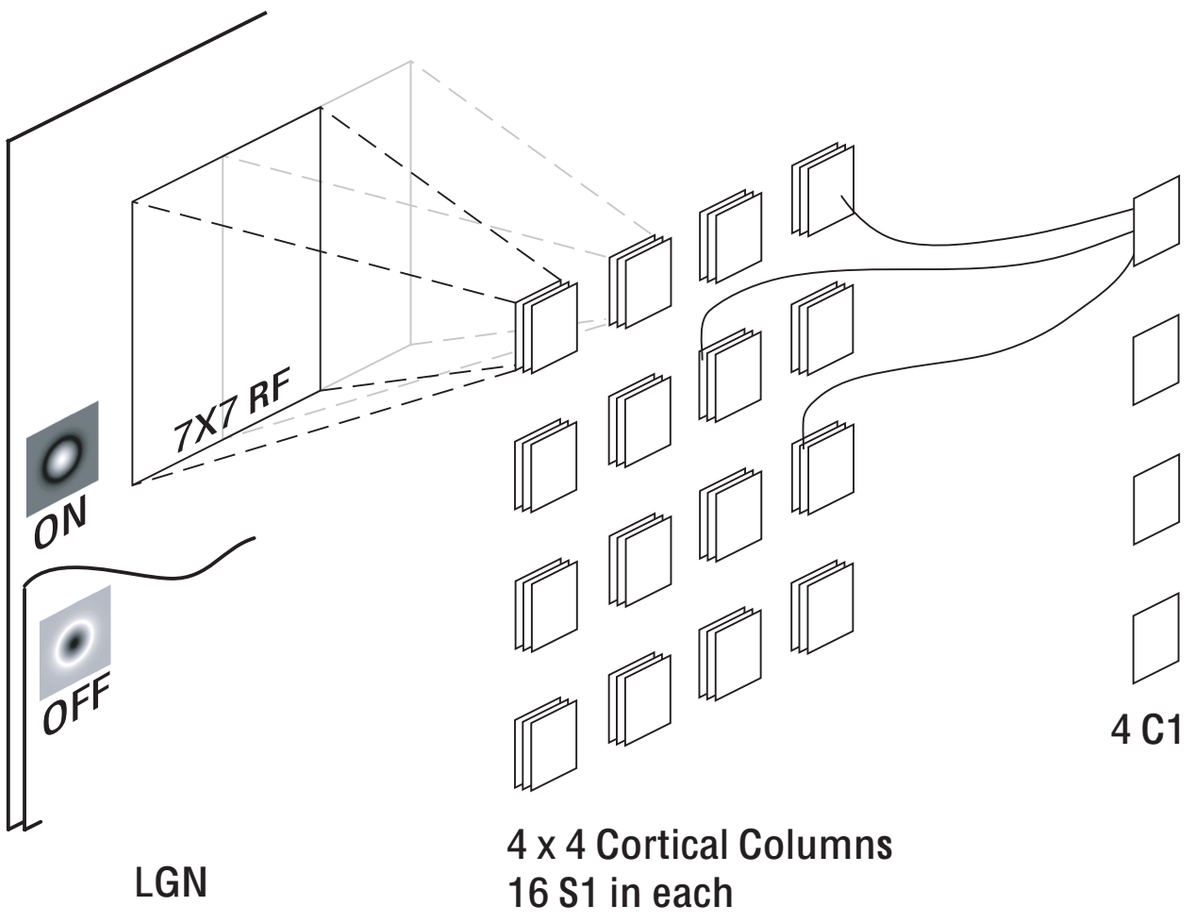
see also Foldiak '91; Hietanen et al '92; Wallis et al '93; Wachsmuth et al '94; Wallis & Rolls '97; Wiskott & Sejnowski '02; Einhauser et al '02; Spratling '05 and many others

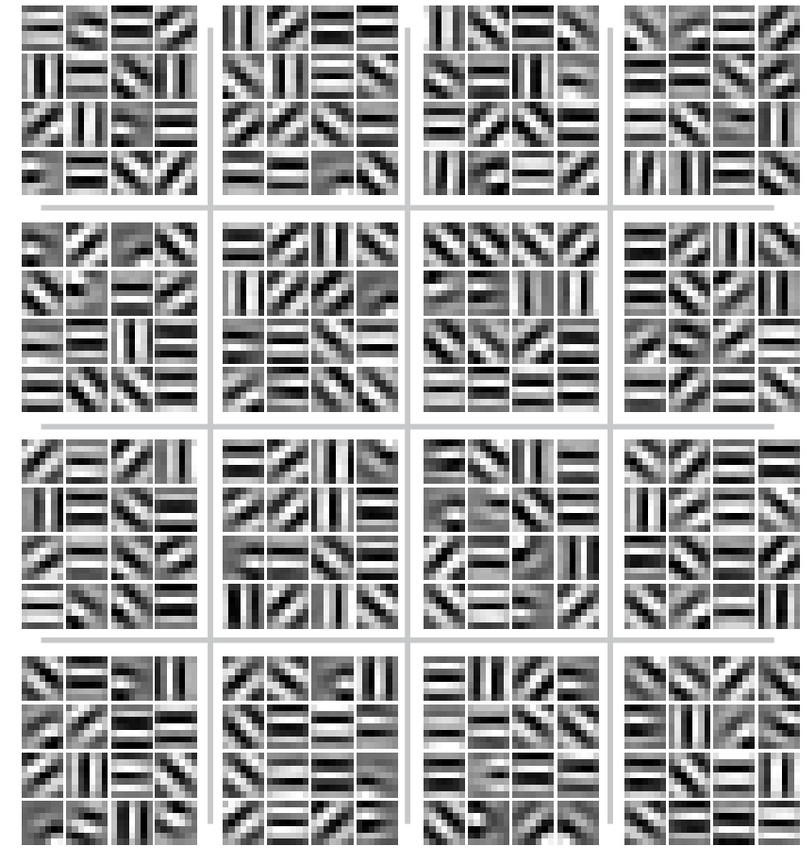
# Learning the invariance from temporal continuity



movie courtesy of Wolfgang Einhauser

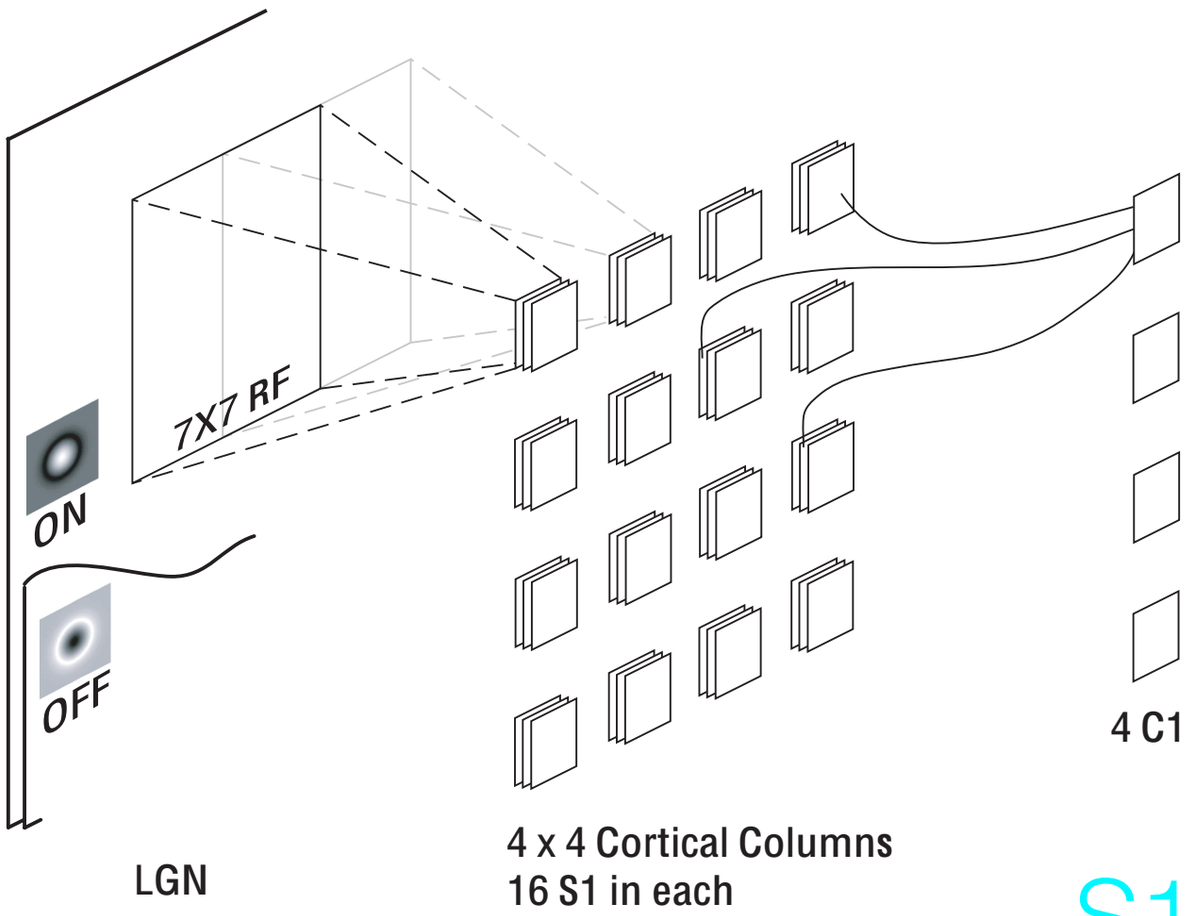
# Learning the invariance from temporal continuity

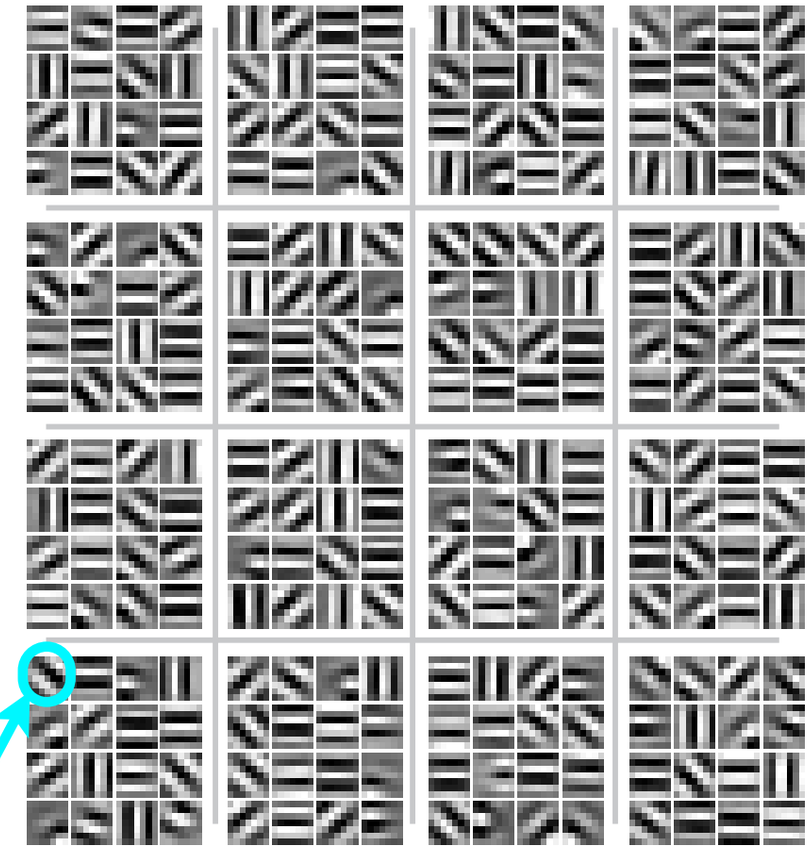


movie courtesy of Wolfgang Einhauser

# Learning the invariance from temporal continuity

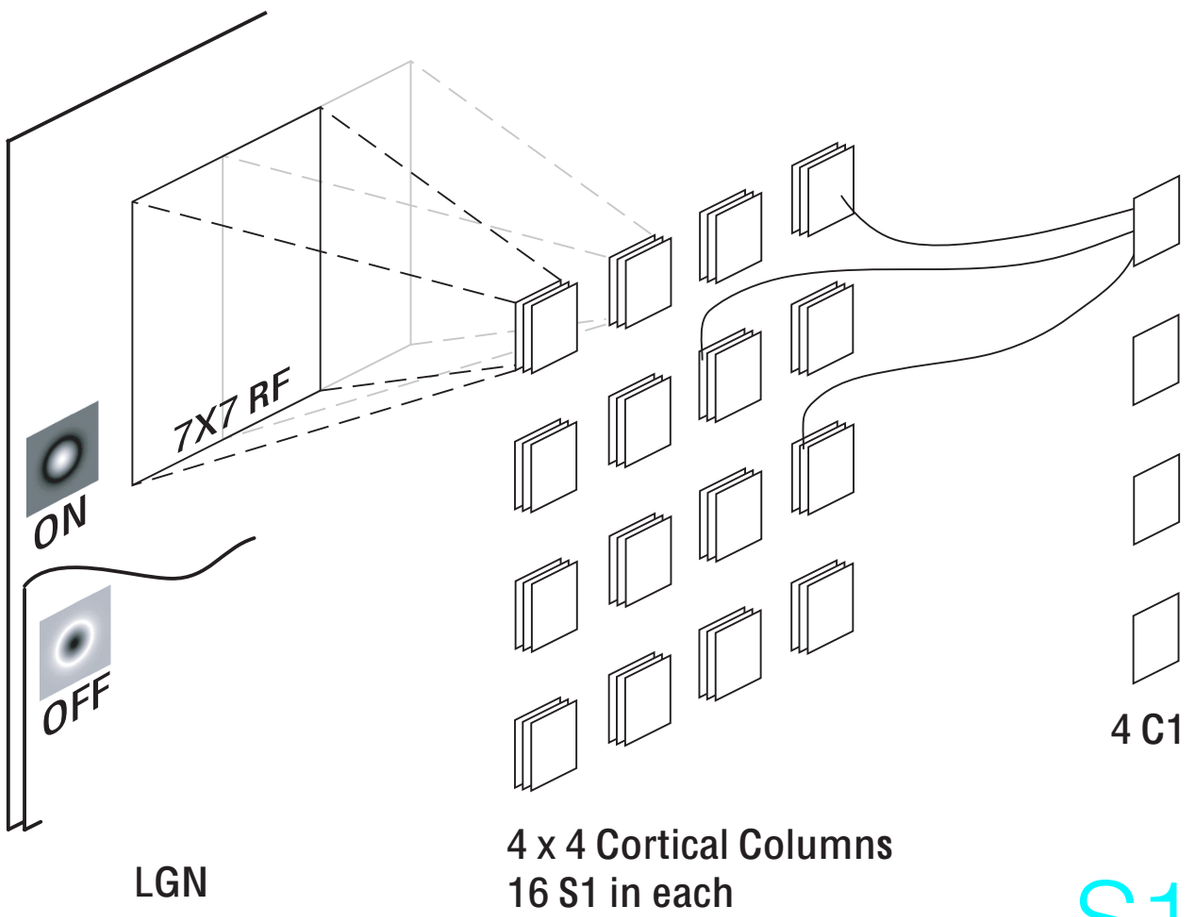


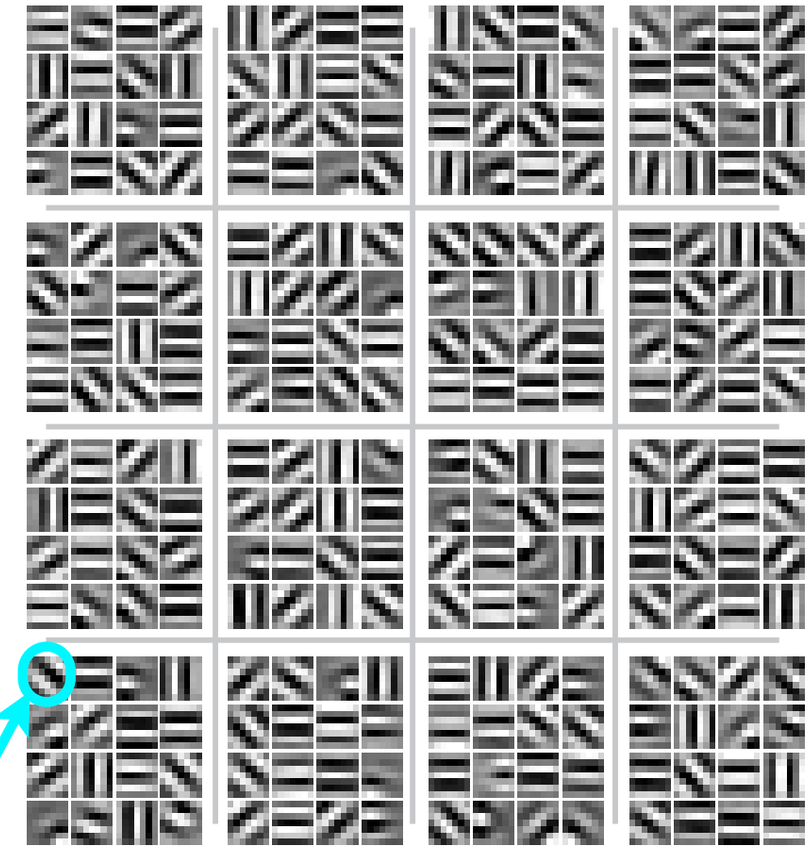



S1 units

movie courtesy of Wolfgang Einhauser

# Learning the invariance from temporal continuity



C1 unit

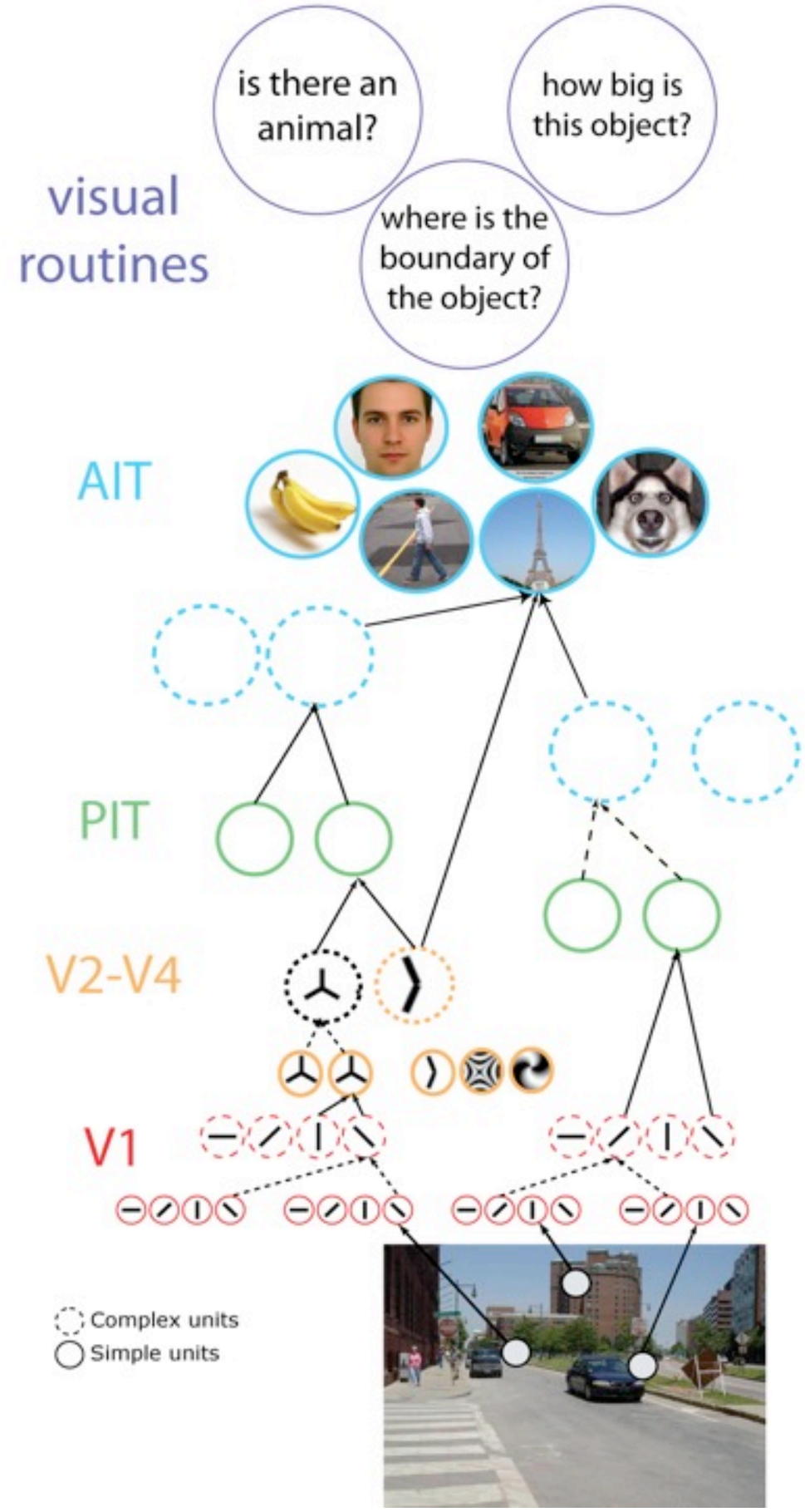
S1 units

movie courtesy of Wolfgang Einhauser

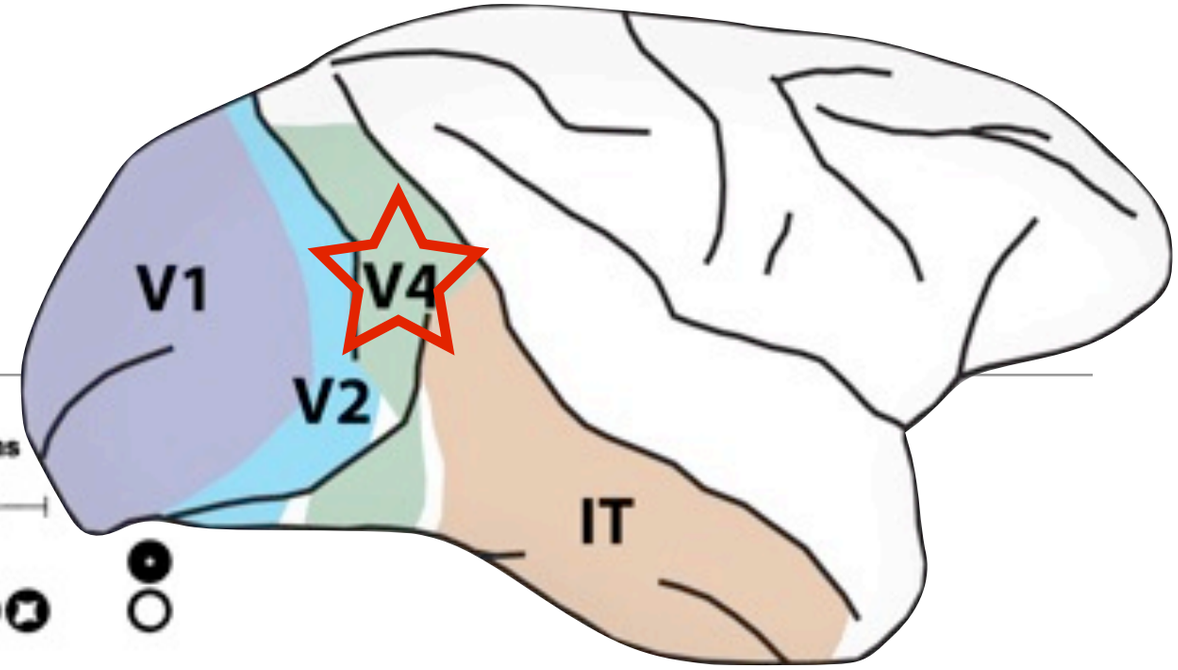
| Area | Type of data | Ref. biol. data | Ref. model data |
|---|---|---|---|
| **Psych.** | Rapid animal categorization | (1) | (1) |
| | Face inversion effect | (2) | (2) |
| **LOC** | Face processing (fMRI) | (3) | (3) |
| **PFC** | Differential role of IT and PFC in categorization | (4) | (5) |
| **IT** | Tuning and invariance properties | (6) | (5) |
| | Read out for object category | (7) | (8,9) |
| | Average effect in IT | (10) | (10) |
| **V4** | MAX operation | (11) | (5) |
| | Tuning for two-bar stimuli | (12) | (8,9) |
| | Two-spot interaction | (13) | (8) |
| | Tuning for boundary conformation | (14) | (8,15) |
| | Tuning for Cartesian and non-Cartesian gratings | (16) | (8) |
| **V1** | Simple and complex cells tuning properties | (17–19) | (8) |
| | MAX operation in subset of complex cells | (20) | (5) |

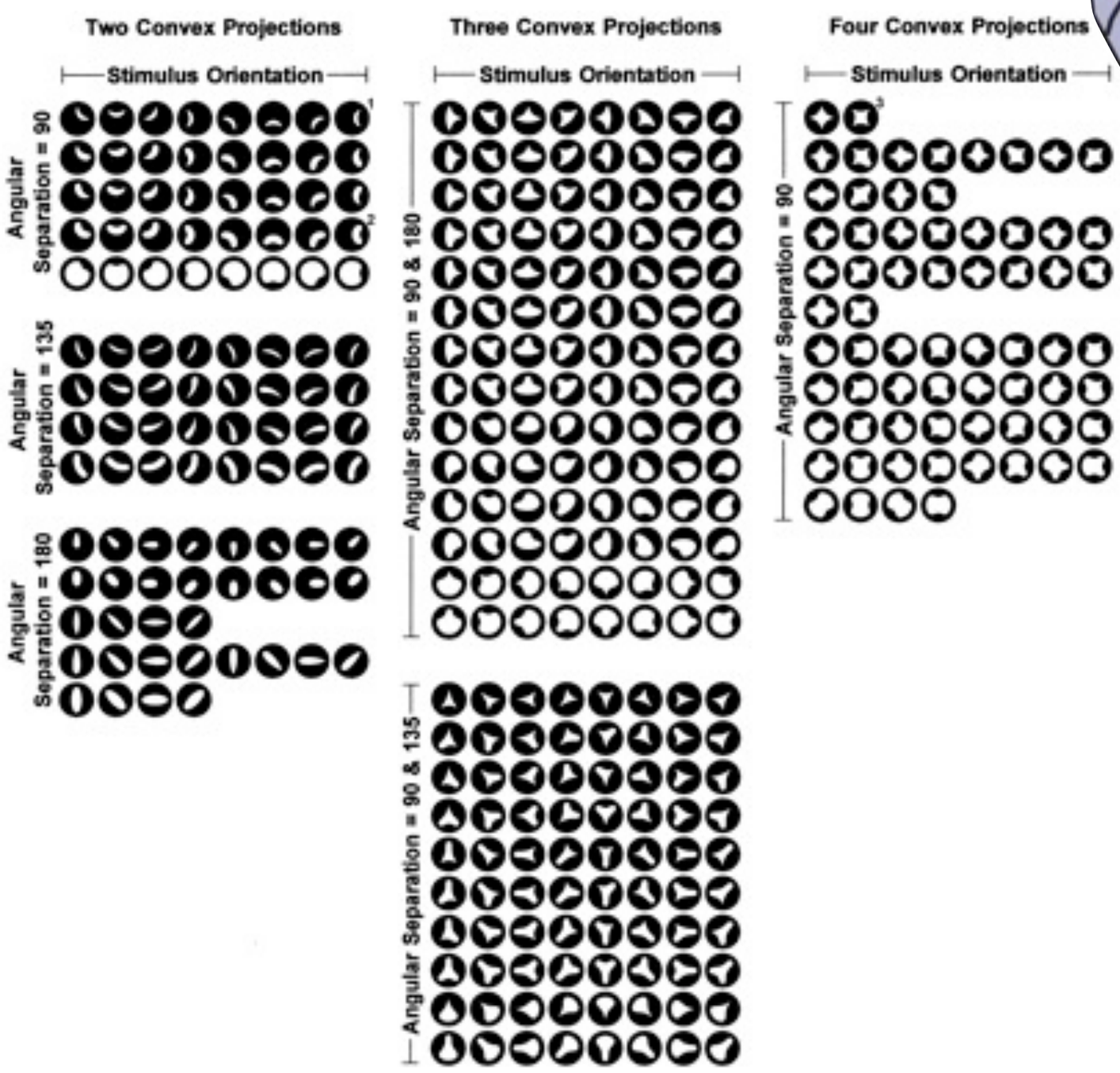

1. Serre, T., Oliva, A., and Poggio, T. *Proc. Natl. Acad. Sci.104*, 6424 (Apr. 2007).
2. Riesenhuber, M. et al. *Proc. Biol. Sci. 271*, S448 (2004).
3. Jiang, X. et al. *Neuron 50*, 159 (2006).
4. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. *Journ. Neurosci. 23*, 5235 (2003).
5. Riesenhuber, M. and Poggio, T. *Nature Neuroscience 2*, 1019 (1999).
6. Logothetis, N.K., Pauls, J., and Poggio, T. *Curr. Biol. 5*, 552 (May 1995).
7. Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, *J.J. Science 310*, 863 (Nov. 2005).
8. Serre, T. et al. *MIT AI Memo 2005-036 / CBCL Memo 259* (2005).
9. Serre, T. et al. *Prog. Brain Res. 165*, 33 (2007).
10. Zoccolan, D., Kouh, M., Poggio, T., and DiCarlo, *J.J. Journ. Neurosci. 27*, 12292 (2007).
11. Gawne, T.J. and Martin, J.M. *Journ. Neurophysiol. 88*, 1128 (2002).
12. Reynolds, J.H., Chelazzi, L., and Desimone, R. *Journ. Neurosci.19*, 1736 (Mar. 1999).
13. Taylor, K., Mandon, S., Freiwald, W.A., and Kreiter, A.K. *Cereb. Cortex 15*, 1424 (2005).
14. Pasupathy, A. and Connor, C. *Journ. Neurophysiol. 82*, 2490 (1999).
15. Cadieu, C. et al. *Journ. Neurophysiol. 98*, 1733 (2007).
16. Gallant, J.L. et al. *Journ. Neurophysiol. 76*, 2718 (1996).
17. Schiller, P.H., Finlay, B.L., and Volman, S.F. *Journ. Neurophysiol. 39*, 1288 (1976).
18. Hubel, D.H. and Wiesel, T.N. *Journ. Physiol. 160*, 106 (1962).
19. De Valois, R.L., Albrecht, D.G., and Thorell, L.G. *Vision Res. 22*, 545 (1982).
20. Lampl, I., Ferster, D., Poggio, T., and Riesenhuber, M. *Journ. Neurophysiol. 92*, 2704 (2004).


# Feedforward hierarchical model of object recognition

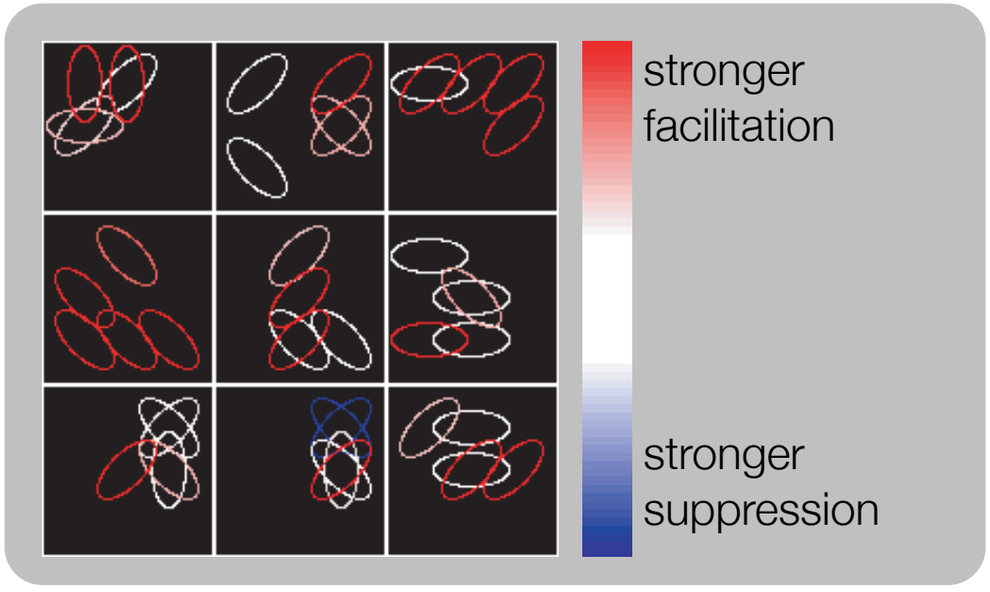


# Comparison w| V4



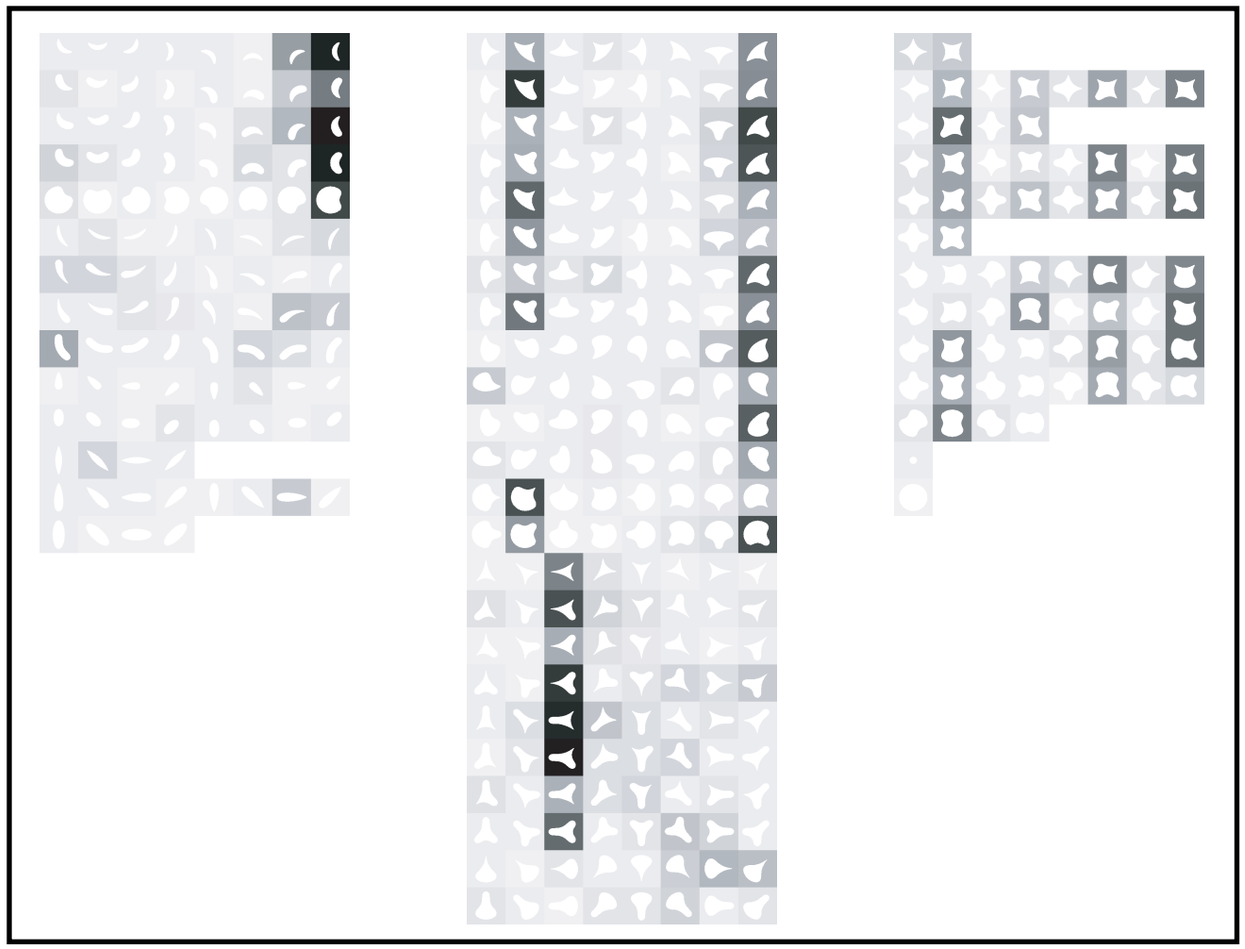

Tuning for curvature and boundary conformations?




Pasupathy & Connor '01

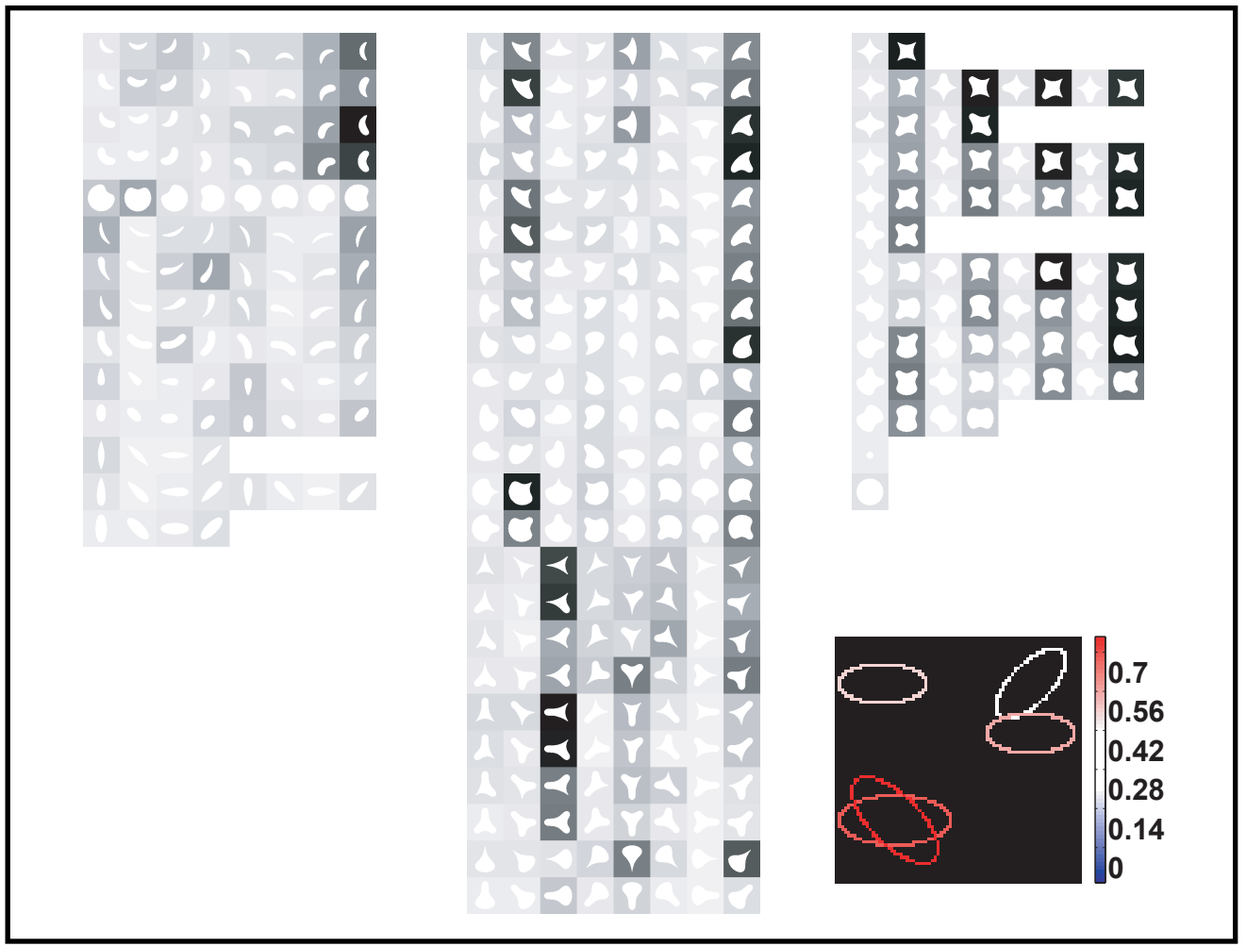
# No parameter fitting!

V4 neuron tuned to boundary conformations



modified from (Pasupathy & Connor 1999)

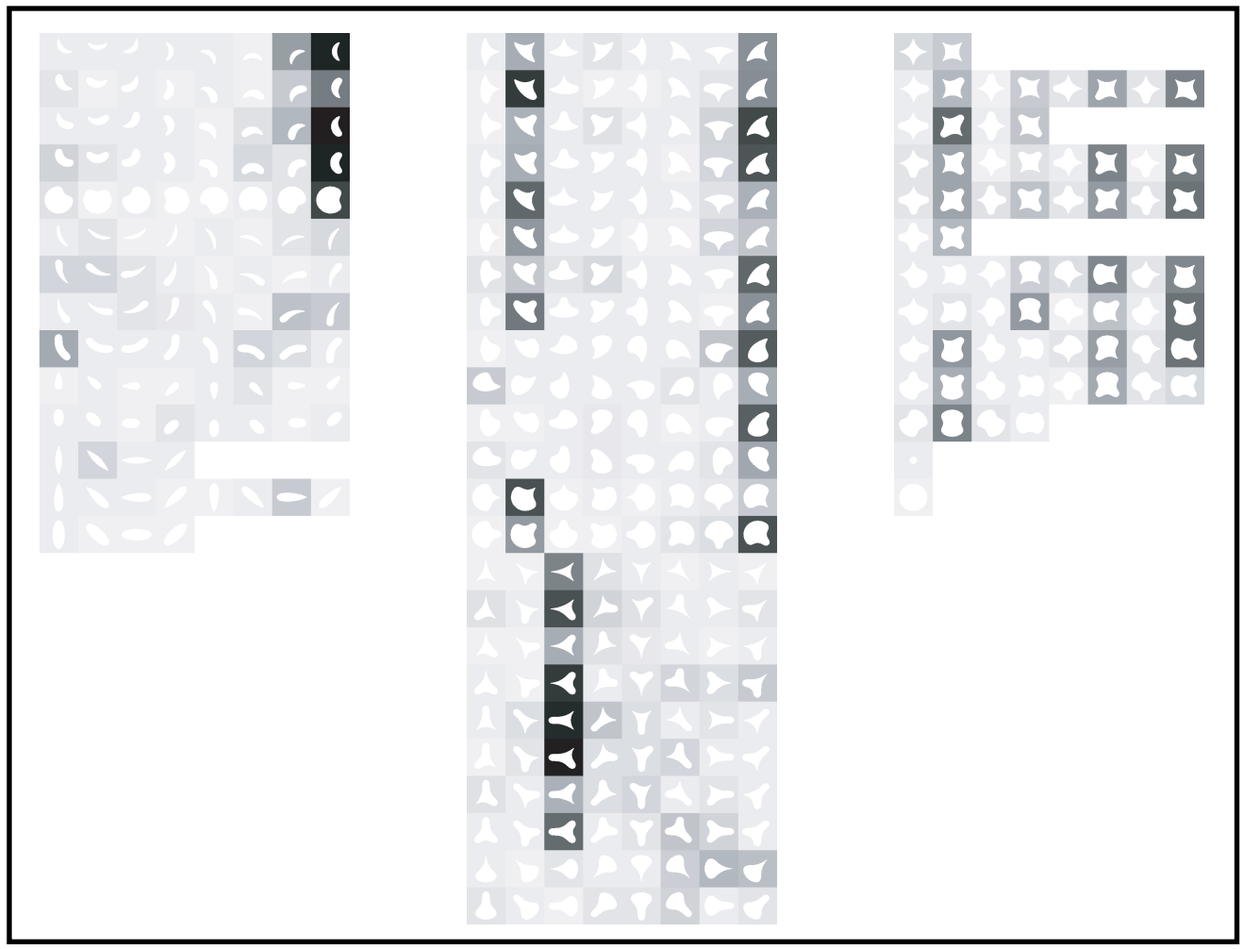
Most similar model C2 unit




Serre Kouh Cadieu Knoblich Kreiman & Poggio ’05

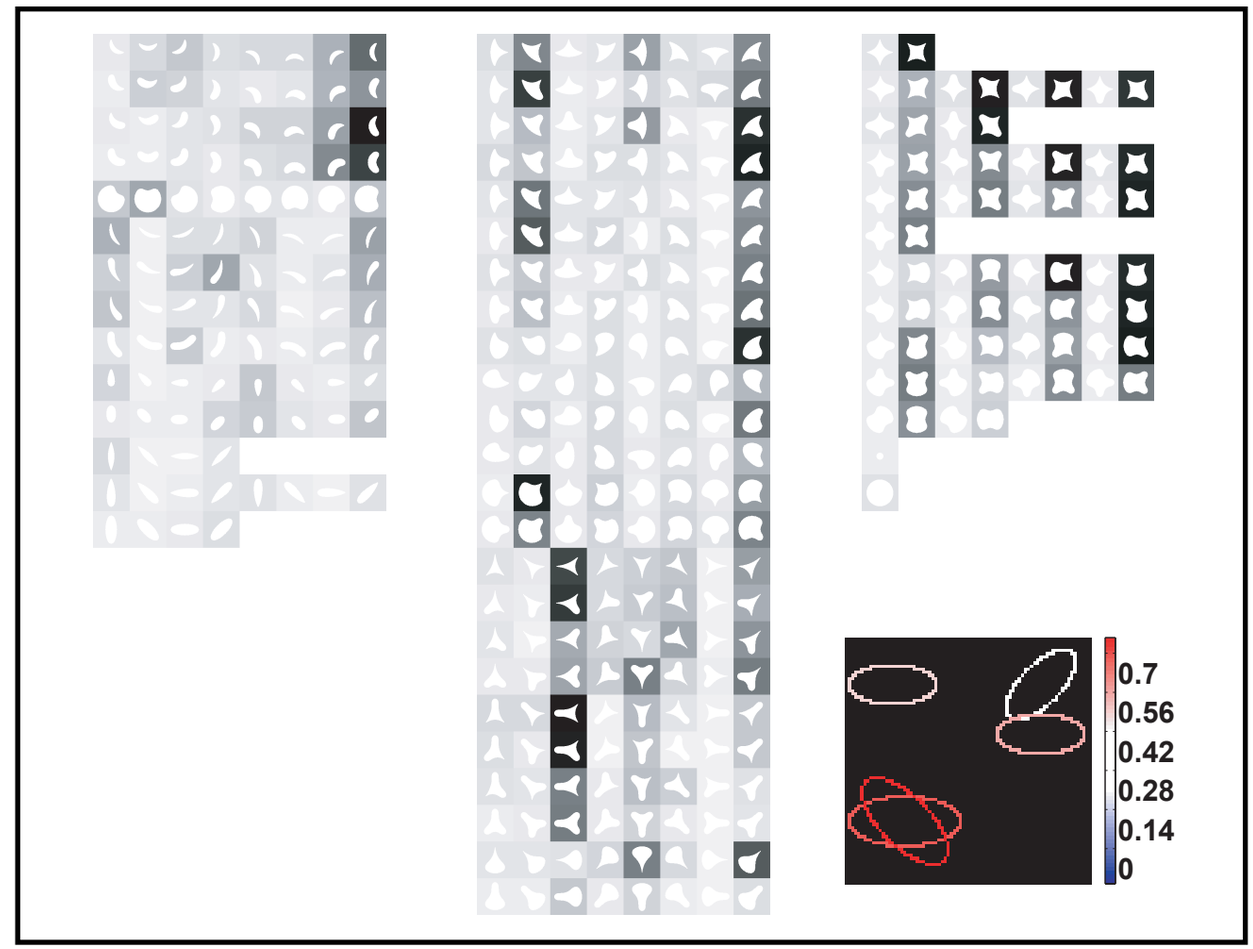
# No parameter fitting!

V4 neuron tuned to boundary conformations



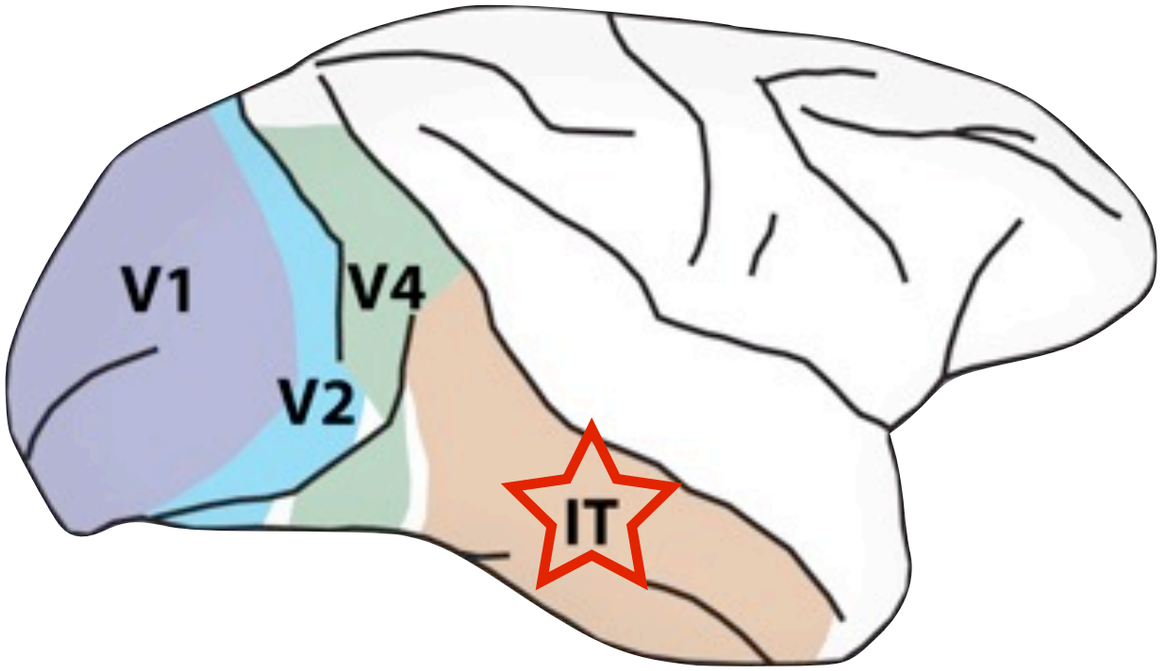
modified from (Pasupathy & Connor 1999)

Most similar model C2 unit



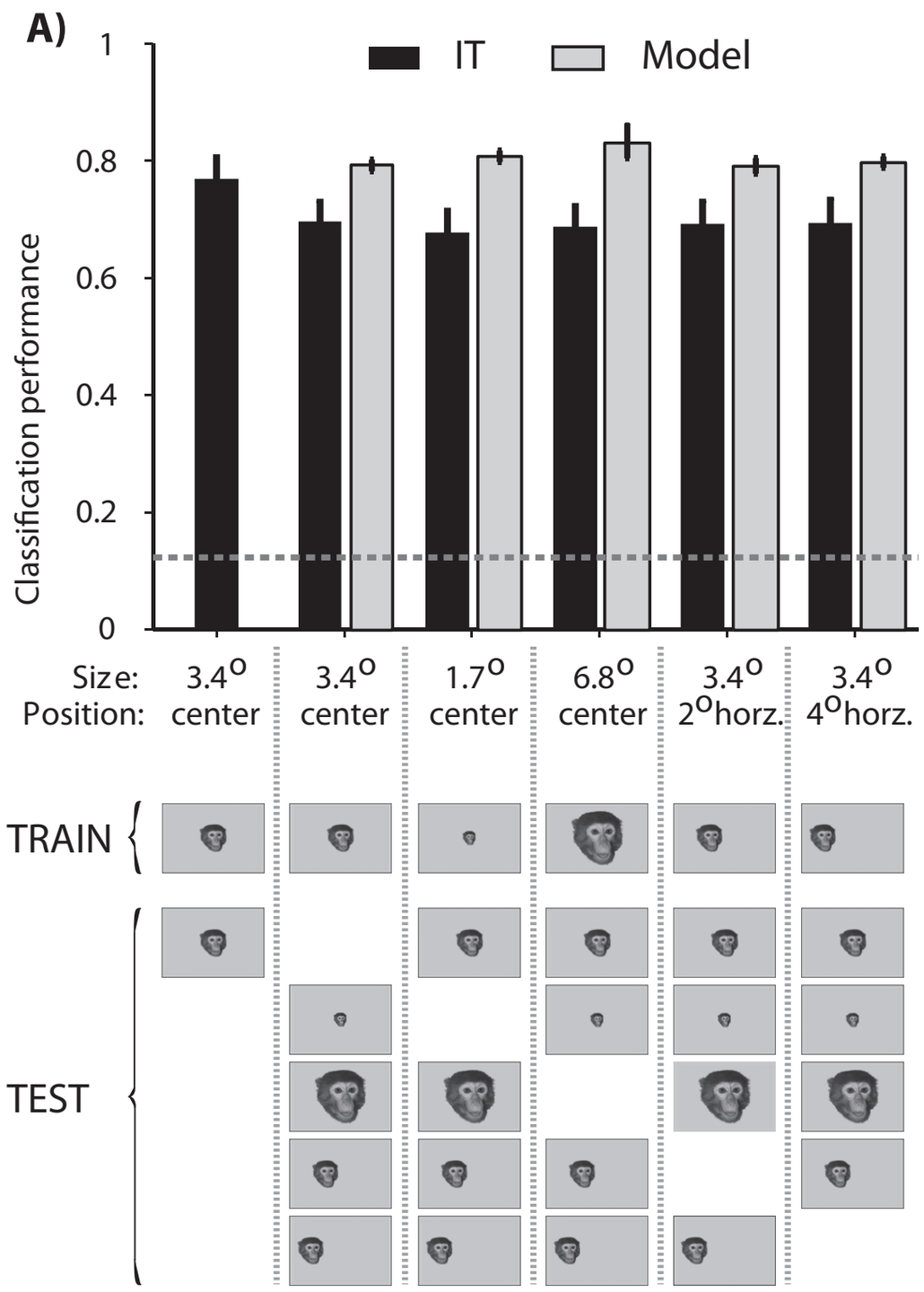

**ρ = 0.78**

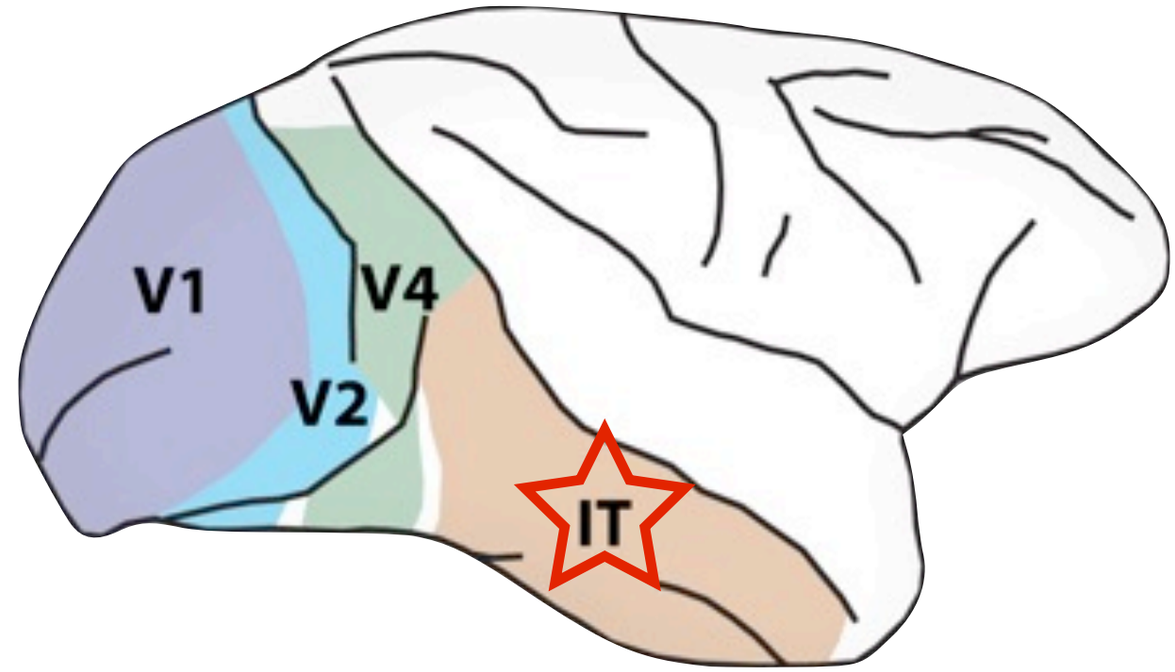
Serre Kouh Cadieu Knoblich Kreiman & Poggio ’05



Model: Serre et al '05
Experimental data: Hung* Kreiman*et al '05

# Invariance in IT

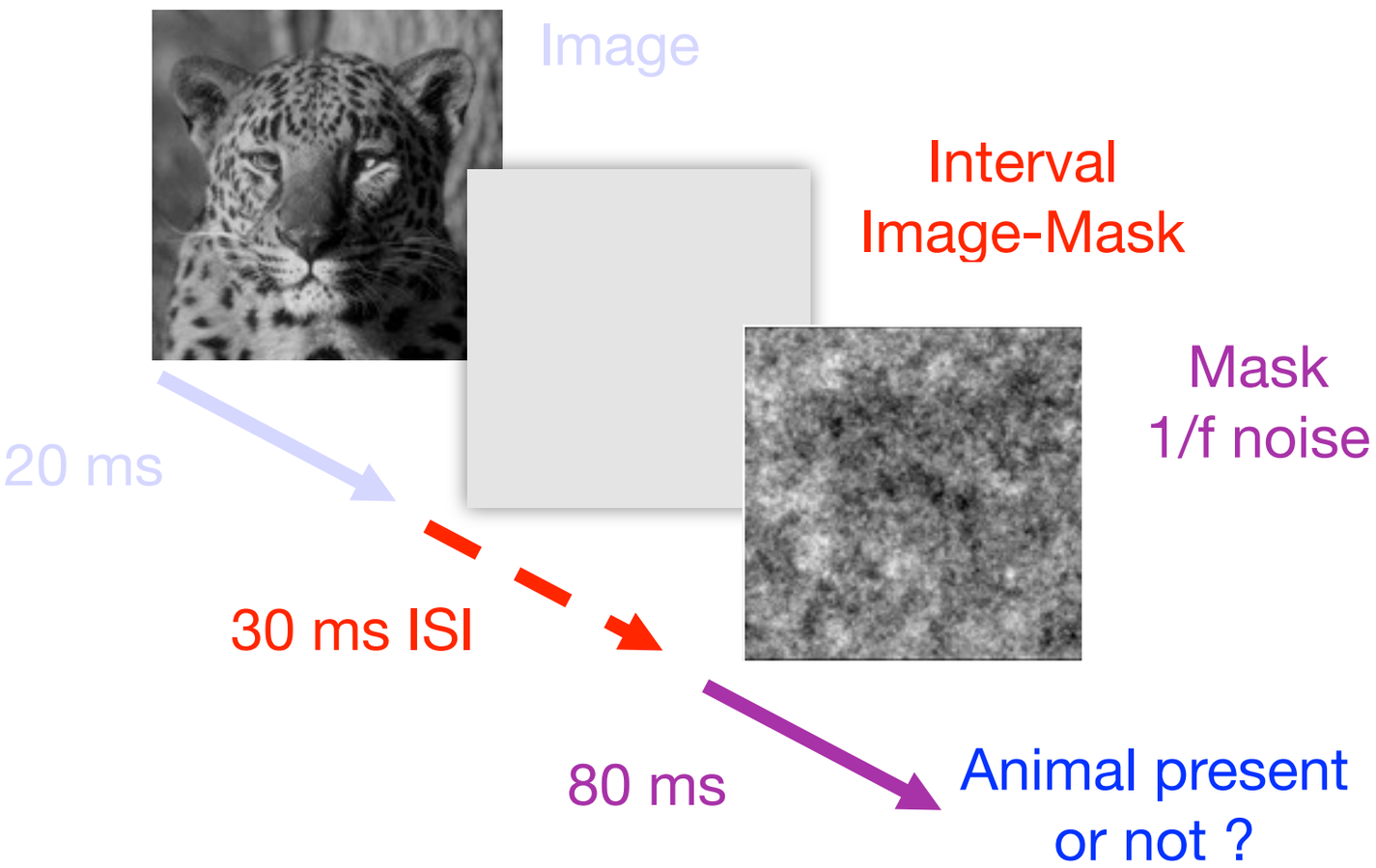


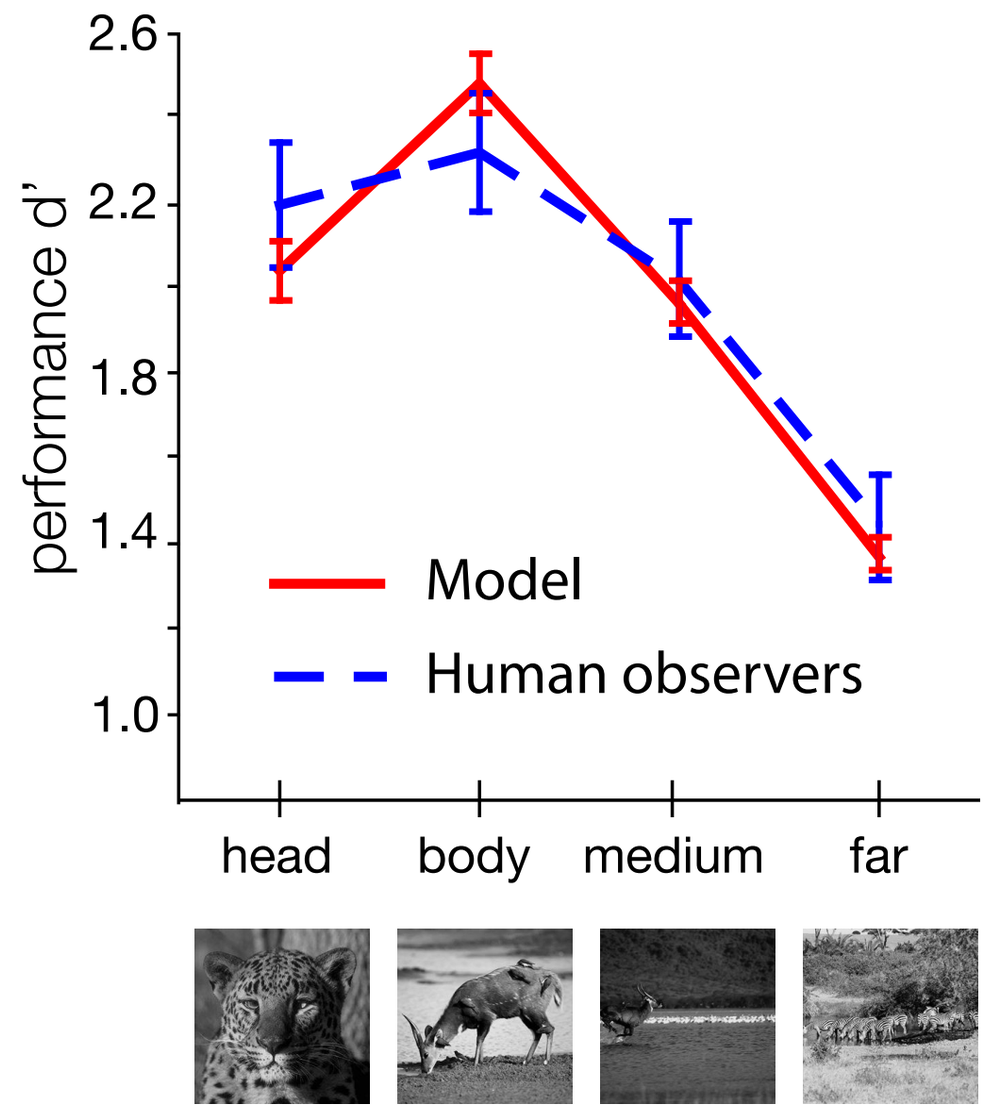




Model: Serre et al '05
Experimental data: Hung* Kreiman*et al '05

# Invariance in IT

# Bottom-up processing and rapid categorization



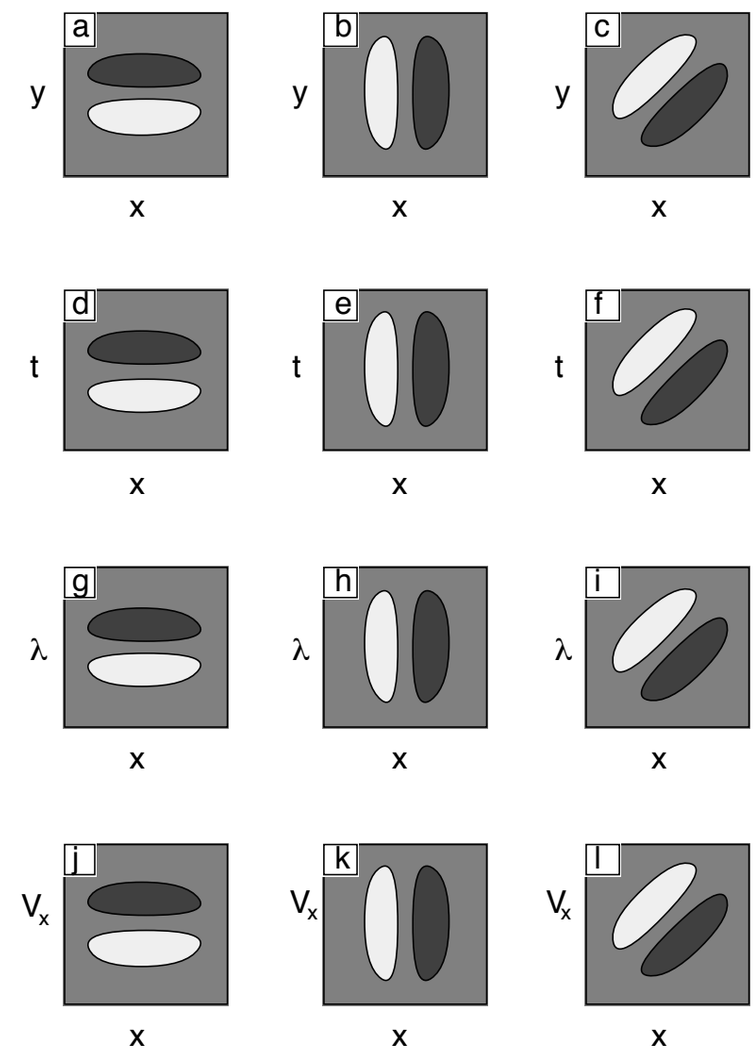






Serre Oliva & Poggio '07

# Ongoing model extensions

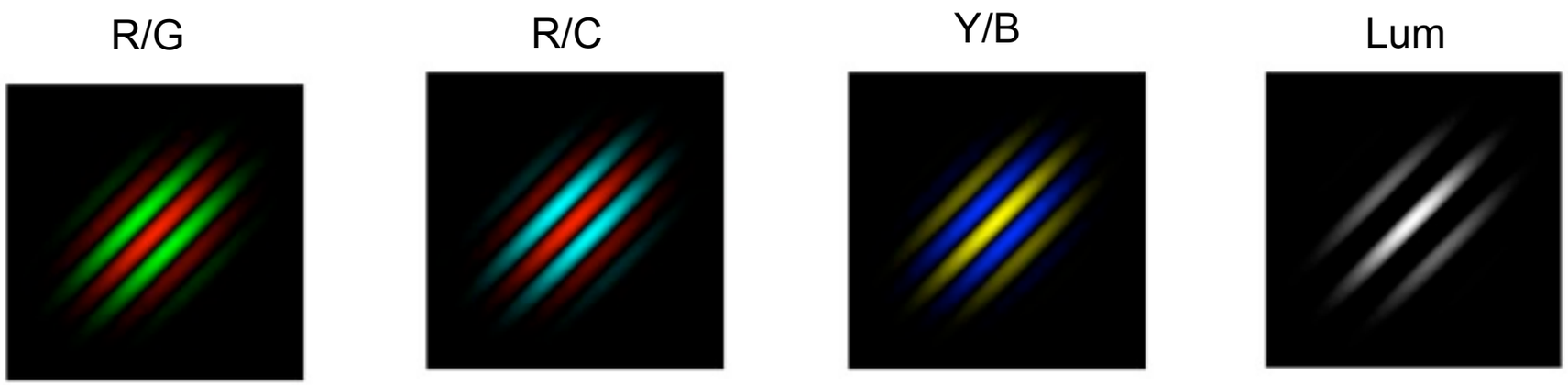
- Edges/object boundaries happen at image discontinuities, e.g,:
  - surface reflectance (hue)
  - depth (binocular vision)
  - surface orientation
  - material properties
  - etc

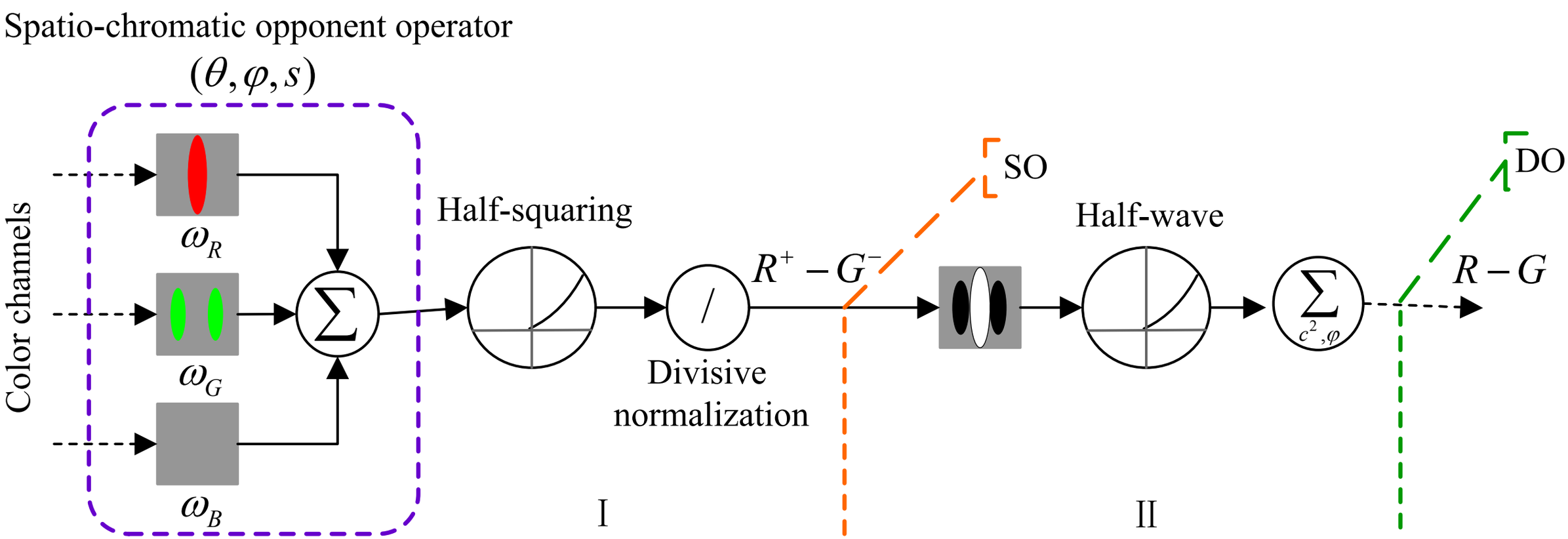



source: Adelson & Bergen (1991)

# Color processing in the ventral stream

Parameters fitted to psychophysics data on color perception



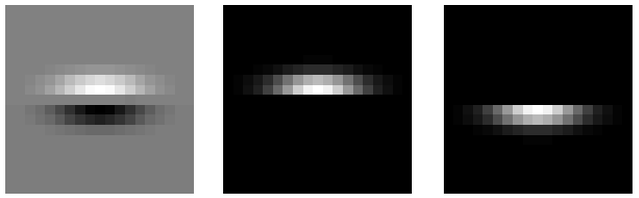





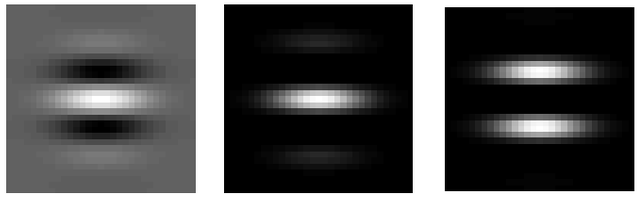

Zhang Barhomi & Serre '12

# Color processing in the ventral stream

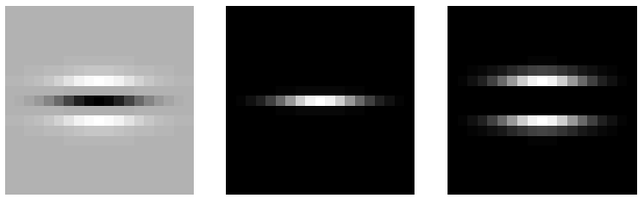
- SO/DO approach improves on all recognition and segmentation datasets tested as compared to existing color representations



A. Gradient used in SIFT



B. Gabor filters used in HMAX



C. Gaussian derivatives used in segmentation

- Color datasets

| | Soccer team | | | Flower | | |
|---|---|---|---|---|---|---|
| Method | Color | Shape | Both | Color | Shape | Both |
| Hue/SIFT | 69 (67) | 43 (43) | 73 (73) | 58 (40) | 65 (65) | 77 (79) |
| Opp/SIFT | 69 (65) | 43 (43) | 74 (72) | 57 (39) | 65 (65) | 74 (79) |
| SOSIFT/DOSIFT | 82 | 66 | 83 | 68 | 69 | 79 |
| SOHMAX/DOHMAX | 87 | 76 | 89 | 77 | 73 | 83 |

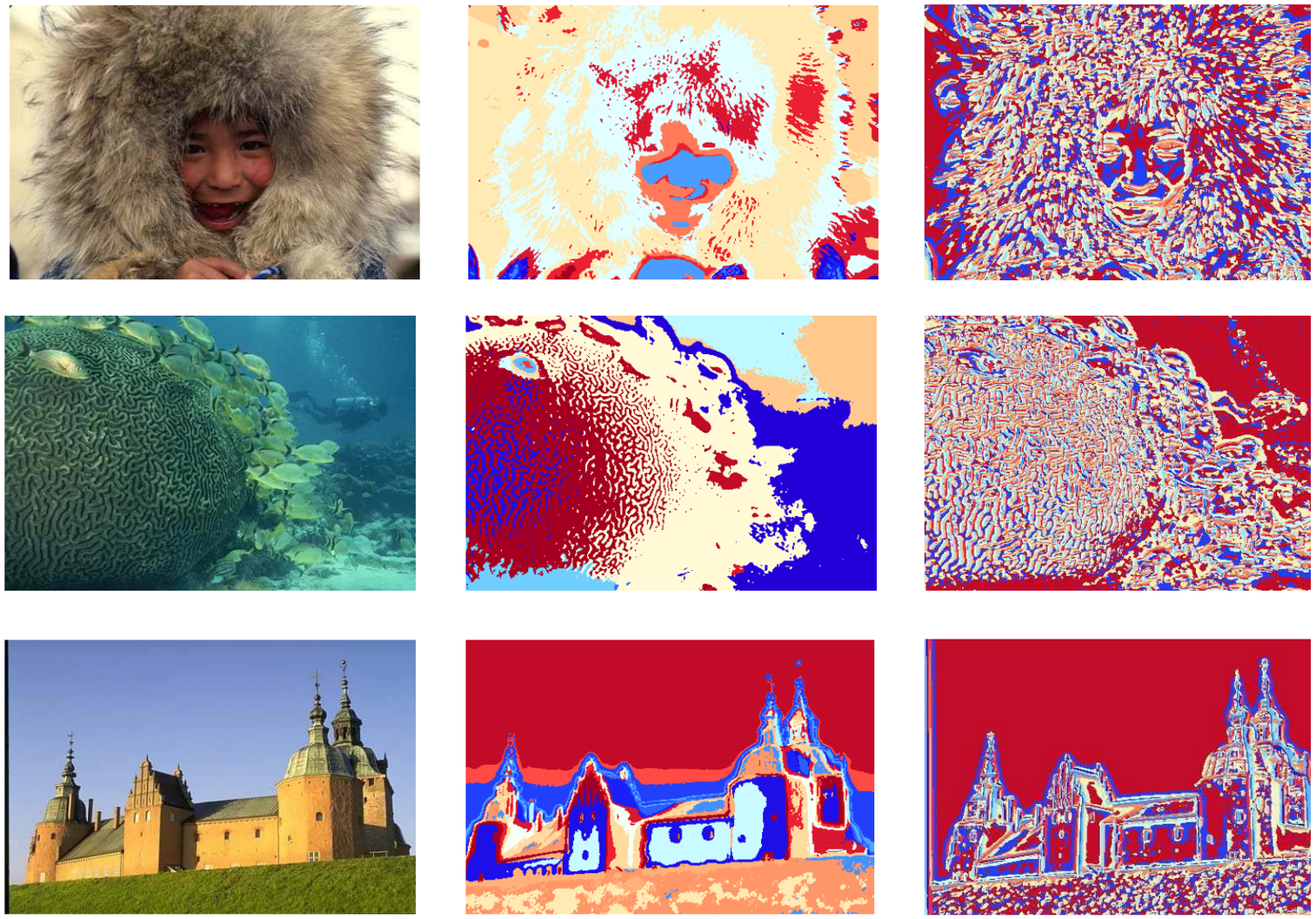
- Pascal challenge

| Method | SIFT | HueSIFT | OpponentSIFT | CSIFT | SODOSIFT | SODOHMAX |
|---|---|---|---|---|---|---|
| AP | 40 (38.4) | 41 | 43 (42.5) | 43 (44.0) | 46.5 **(33.3/39.8)** | 46.8 **(30.1/36.4)** |

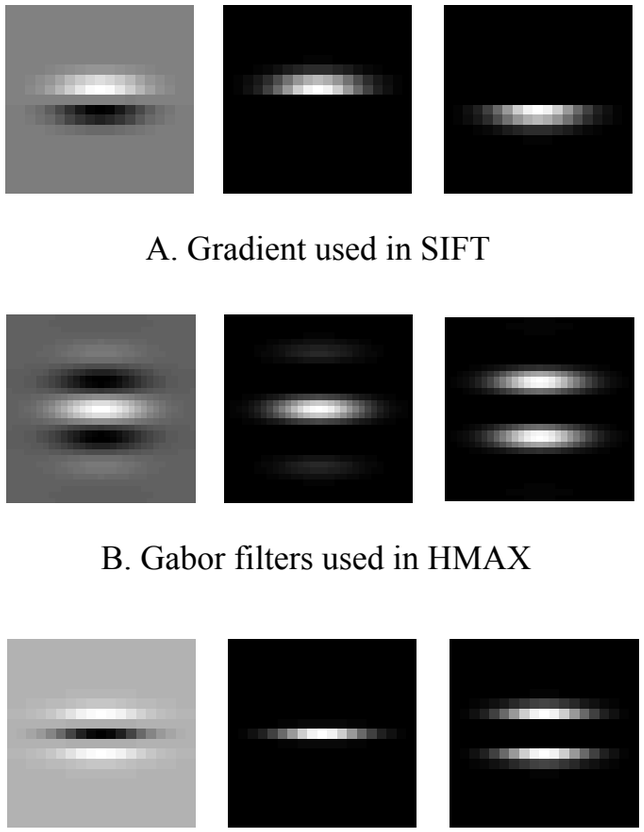


Zhang Barhomi & Serre '12

# Color processing in the ventral stream



Color textons  Original textons



A. Gradient used in SIFT

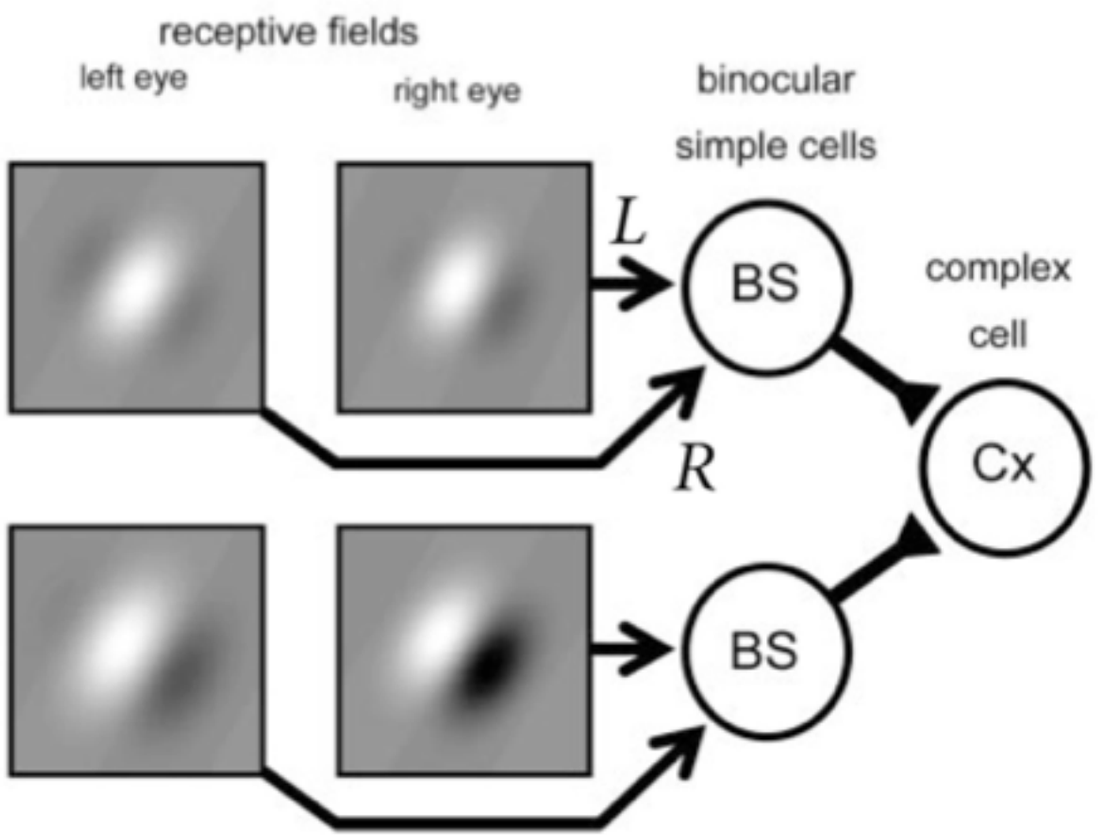
B. Gabor filters used in HMAX

C. Gaussian derivatives used in segmentation

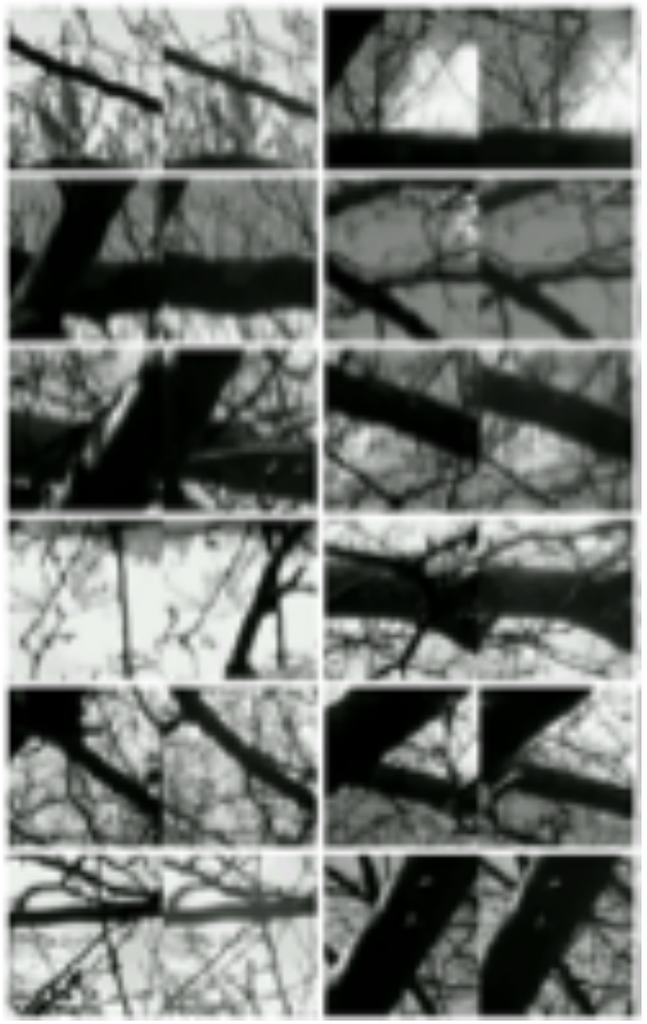


Zhang Barhomi & Serre ’12

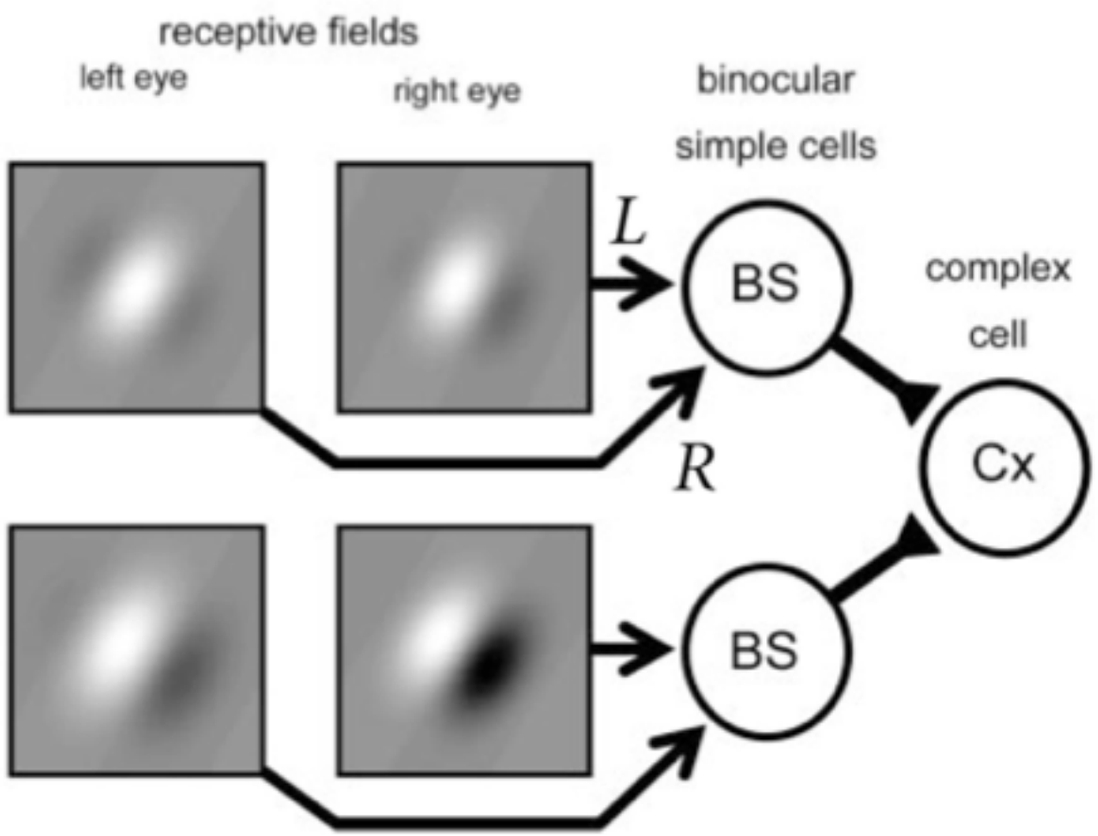
# Disparity processing




Source: Qian ’94



# Disparity processing

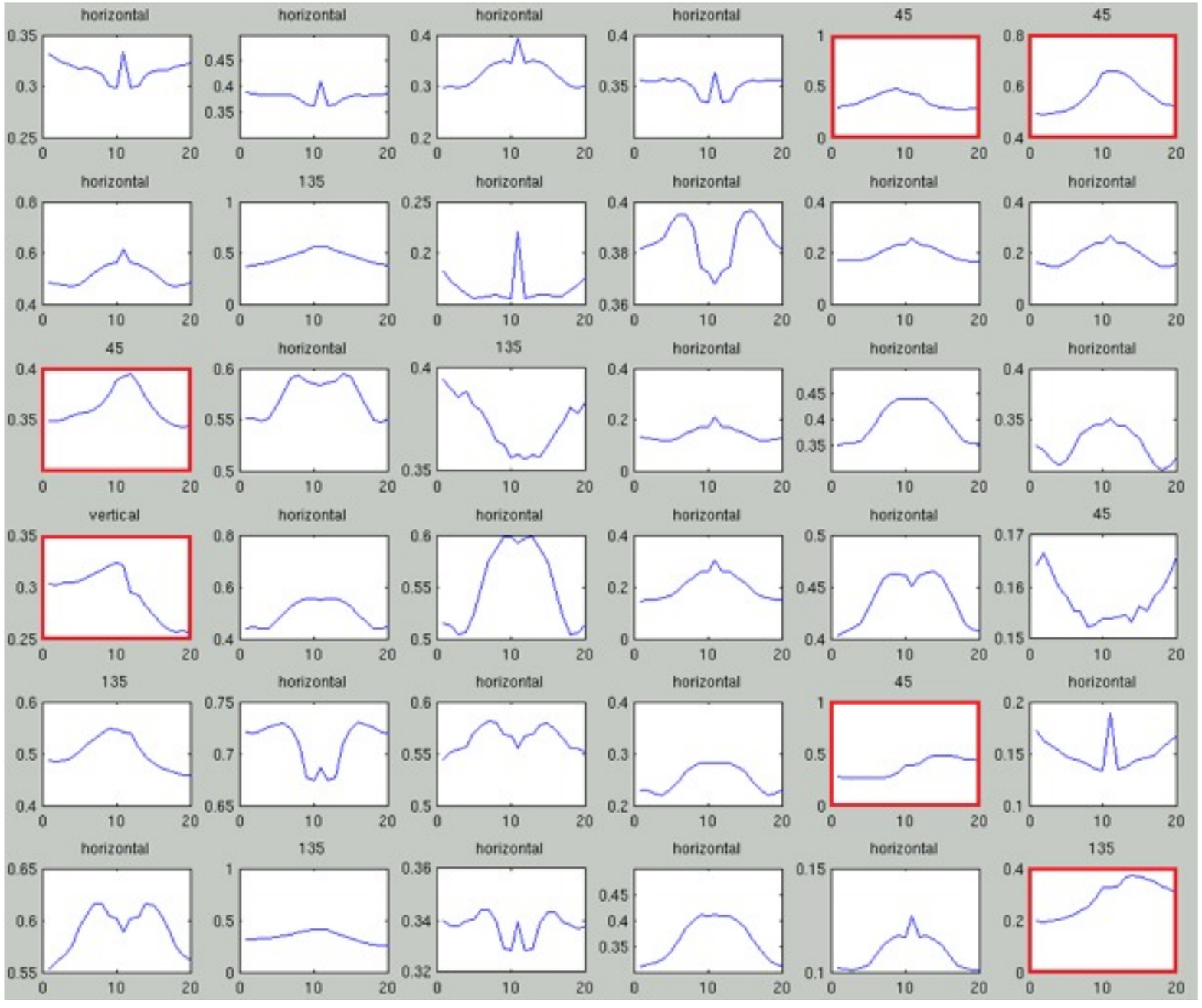



Source: Qian ’94

# Disparity processing

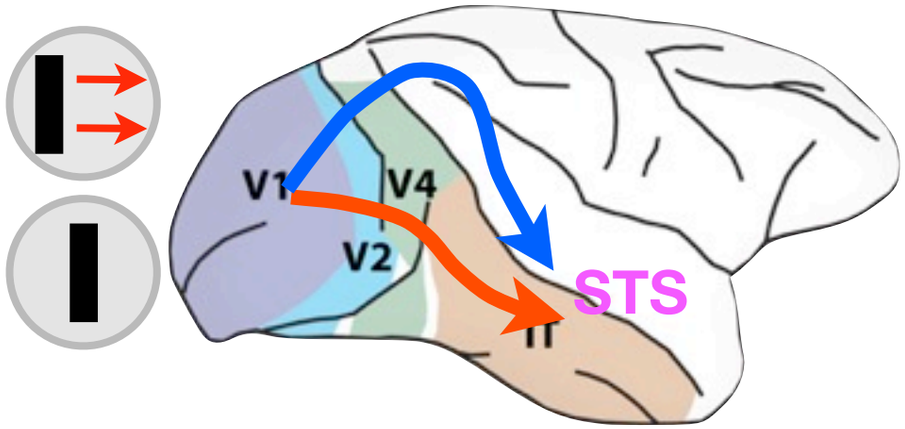
Riesen & Serre unpublished data
See Sasaki et al ‘10 for qualitatively similar results

# Disparity processing

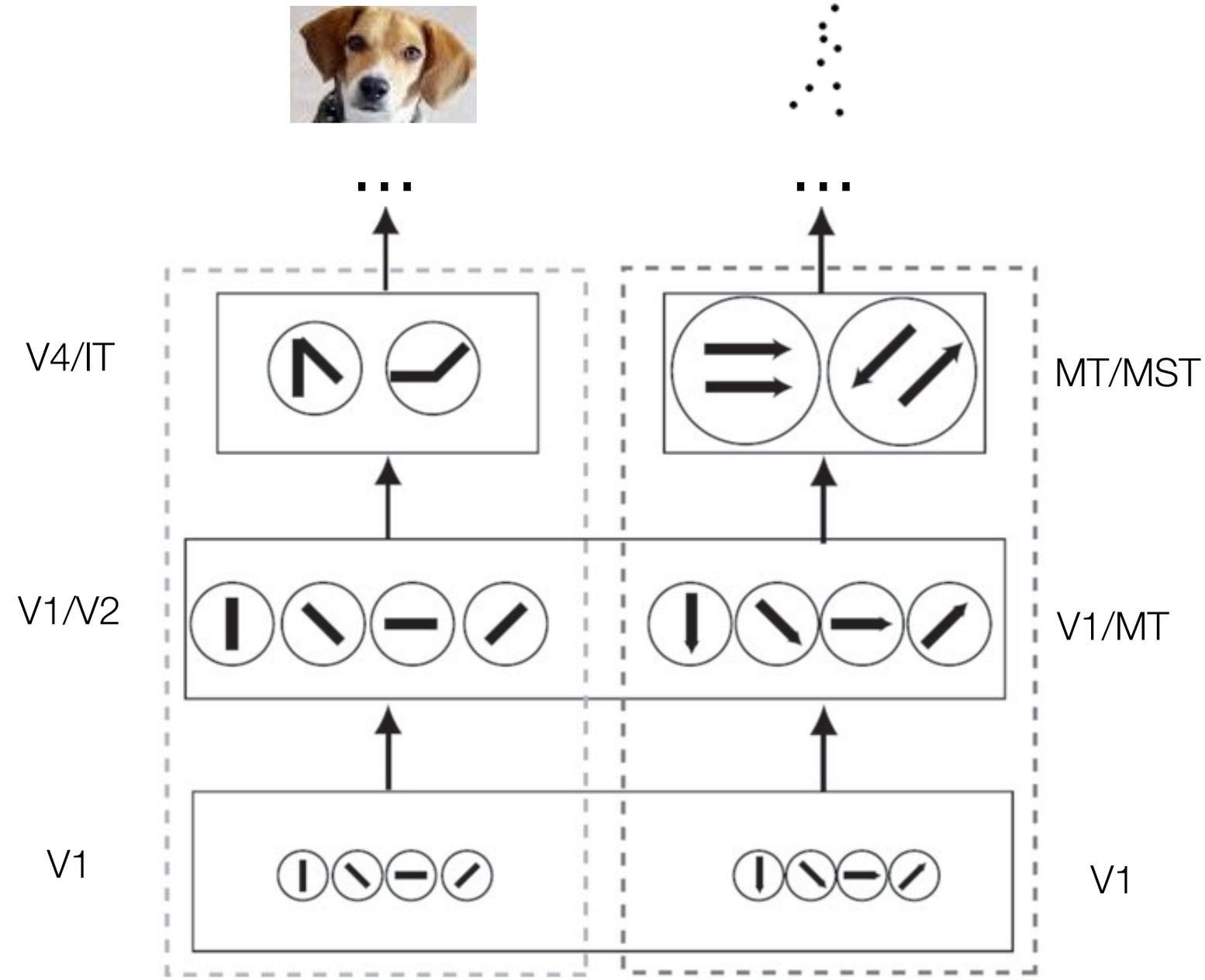



Riesen & Serre unpublished data
See Sasaki et al ‘10 for qualitatively similar results

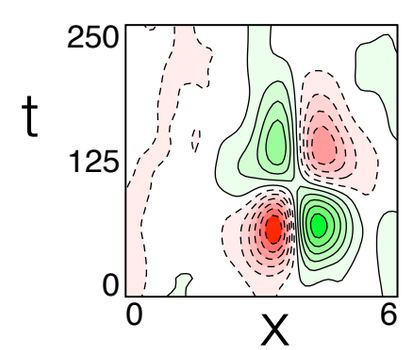
Dorsal "motion" pathway




Ventral "shape" pathway

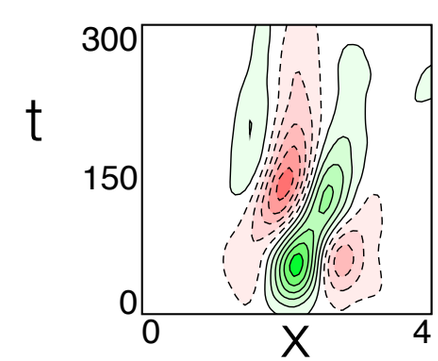



Ventral stream model



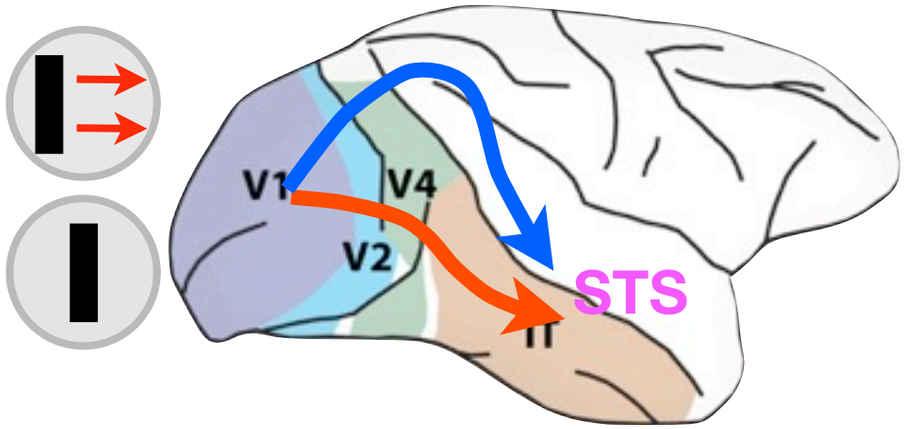

Separable space-time RFs

Dorsal stream model

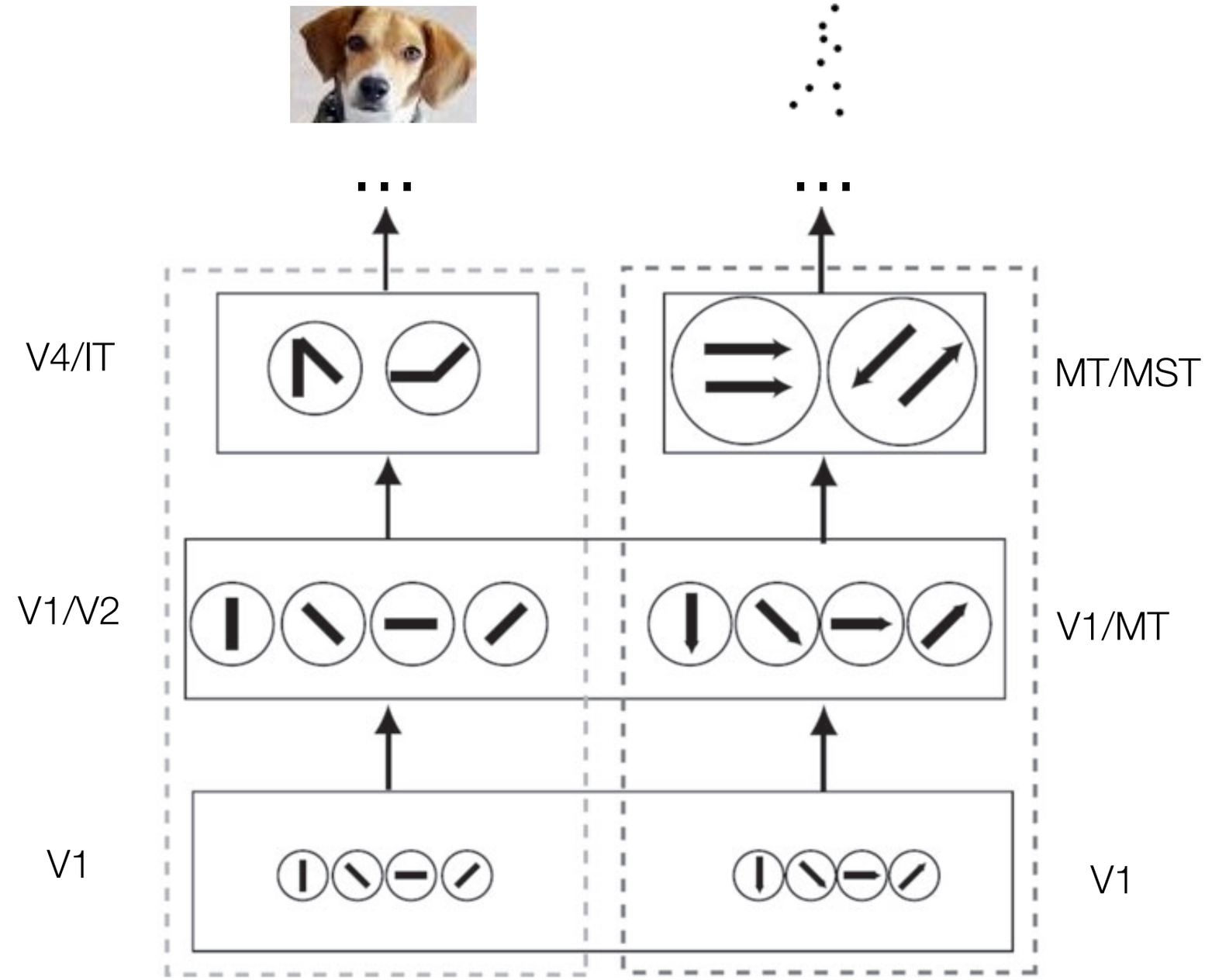



Non-separable space-time RFs

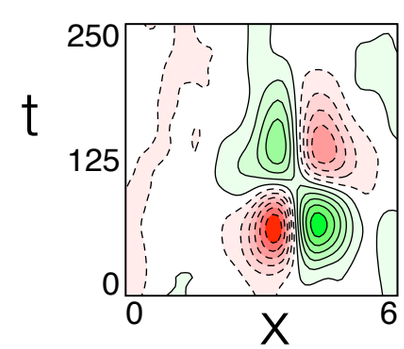
Dorsal "motion" pathway




Ventral "shape" pathway

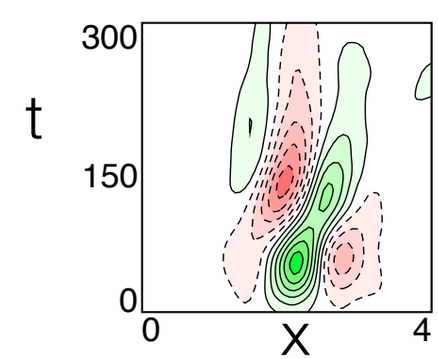



Ventral stream model




Separable space-time RFs

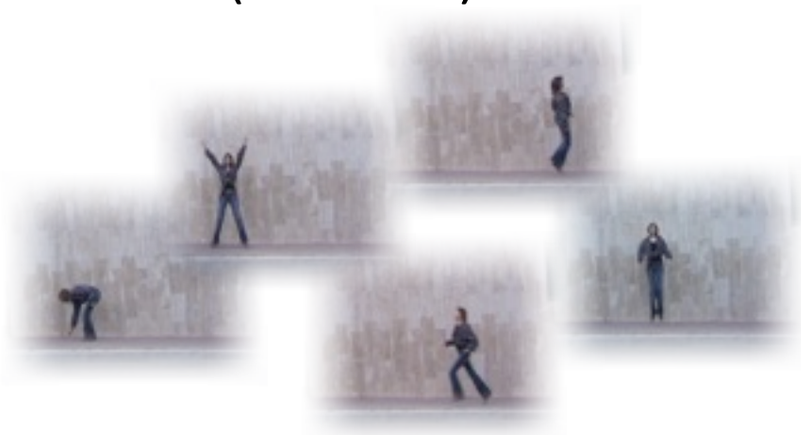
Dorsal stream model




Non-separable space-time RFs

# Action recognition with a dorsal stream model

Weizmann Human action
(9 classes)



| | Dollar et al '05 | Jhuang et al '07 |
|---|---|---|
| KTH Human | 81.3 % | **91.6 %** |
| UCSD Mice | 75.6 % | **79.0 %** |
| Weiz. Human | 86.7 % | **96.3 %** |

Jhuang Serre Wolf & Poggio '07

KTH Human actions
(6 classes)

| Name | Percentile | Ref |
|---|---|---|
| LTP | 100% | |
| bag-of-snip-1 | 100% | [15] |
| Blank | 100% | [2] |
| Jhuang | 98.8% | [8] |
| Wang | 97.8% | [20] |
| Ali | 92.6% | [1] |
| Dollár | 86.7% | [4] |
| Niebles | 72.8% | [12] |

Yeffet & Wolf '09

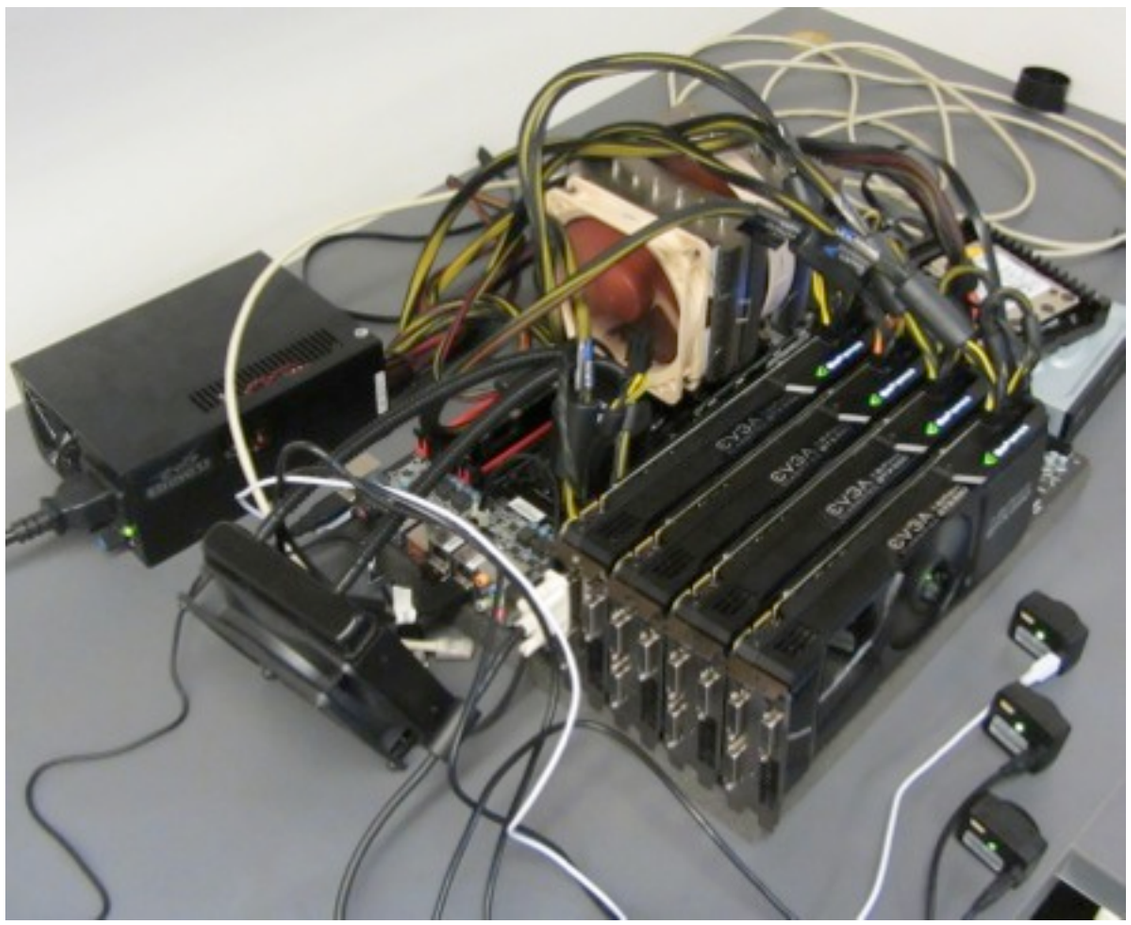
| Name | Percentile | Ref |
|---|---|---|
| LTP | 90.1% | |
| LBPabs | 83.8 | 5.5.1 |
| LBPdir | 79.6 | 5.5.1 |
| Schindler | 92.7% | [15] |
| Laptev | 91.8 | [11] |
| Jhuang | 91.7% | [8] |
| Niebles | 81.5% | [12] |
| Dollár | 81.2% | [4] |
| Schüldt | 71.7% | [16] |

Yeffet & Wolf '09

| methods | accuracy(%) |
|---|---|
| Our method | **87.60** |
| Jhuang et al. [8] | **91.70** |
| Nowozin et al. [15] | 87.04 |
| Niebles et al. [14] | 81.50 |
| Dollár et al. [4] | 81.17 |
| Schuldt et al. [17] | 71.72 |
| Ke et al. [9] | 62.96 |

Wang & Mori '08

# Automated rodent behavioral analysis



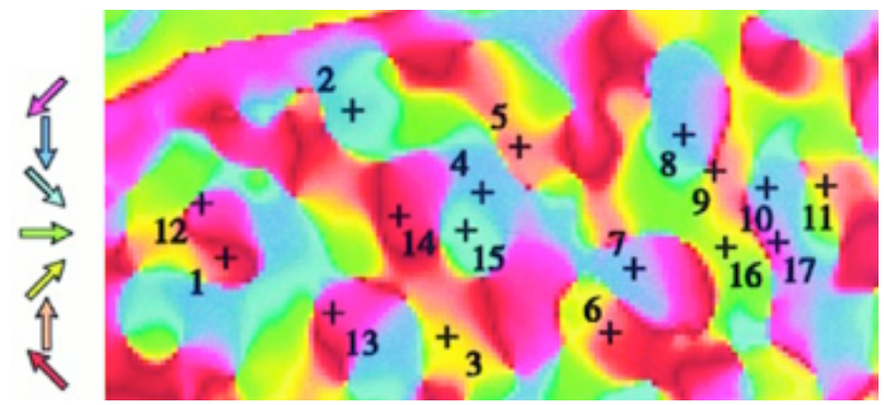
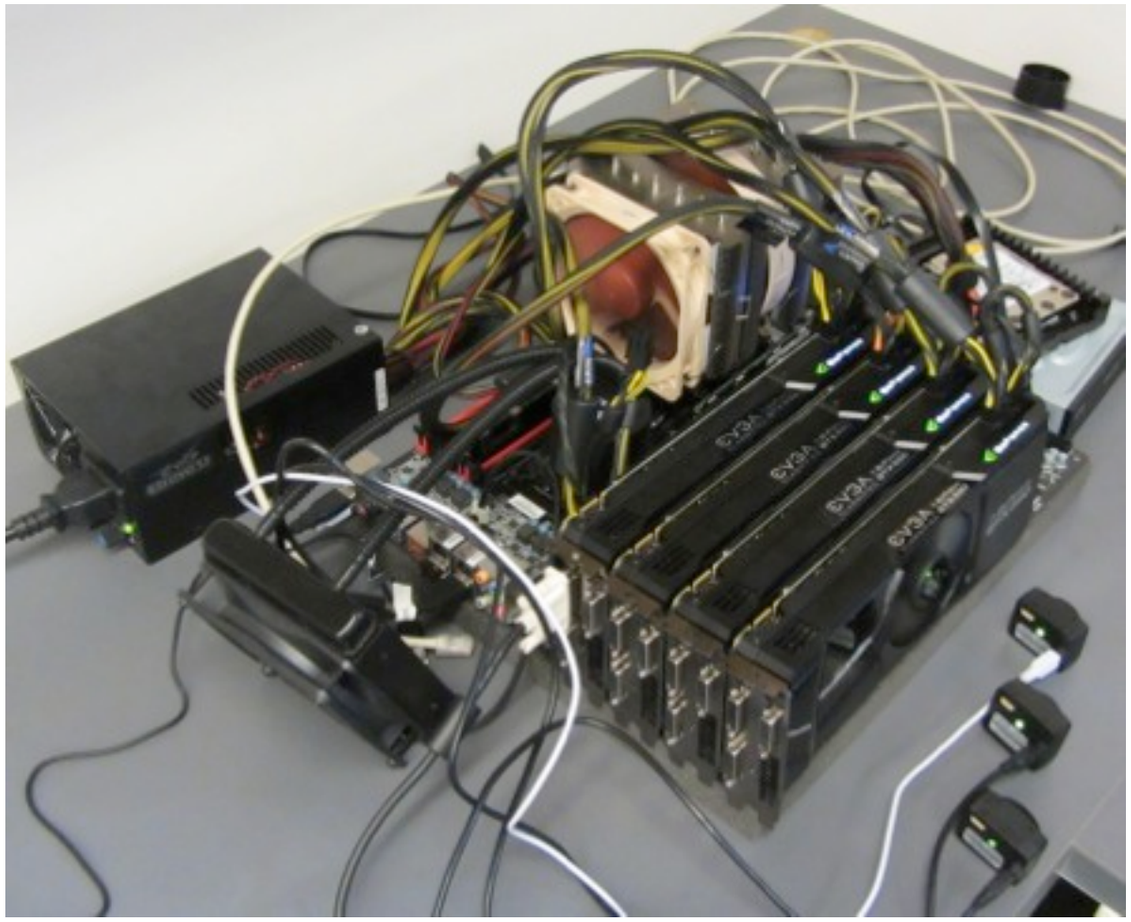


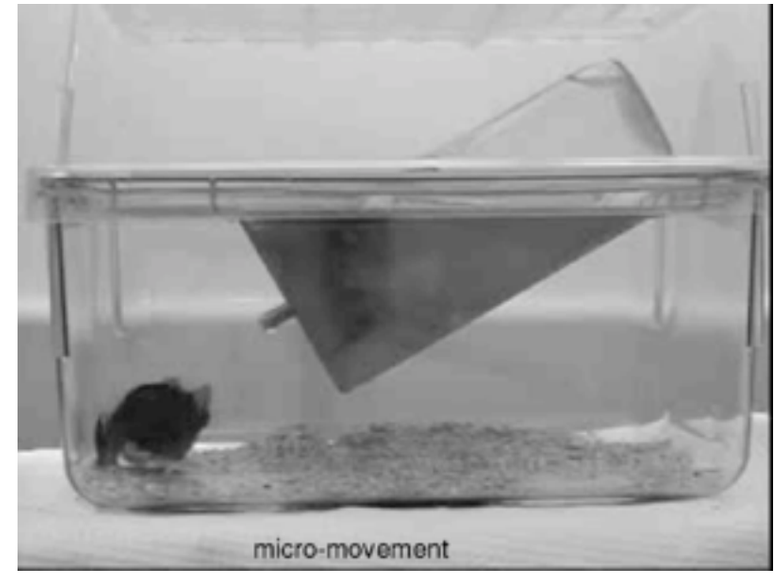

Image source: Shmuel & Grinvald ‘96



Jhuang Serre et al ‘07 ’10; Kuehne Jhuang et al ‘11

# Automated rodent behavioral analysis





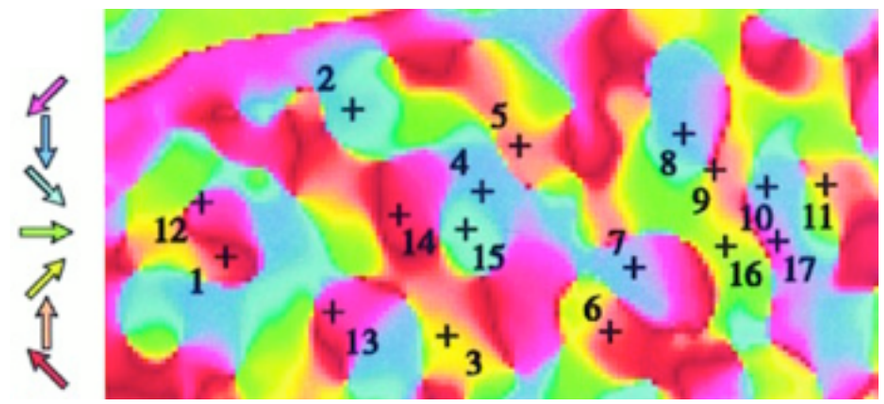
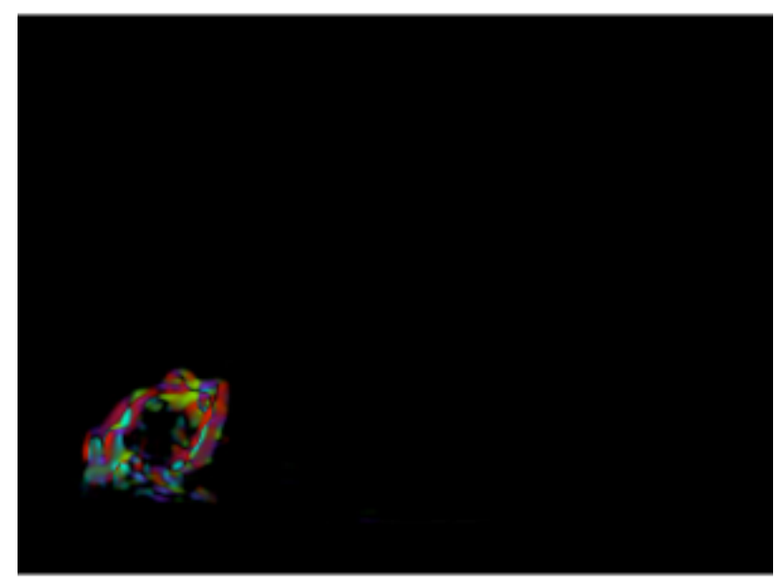




Image source: Shmuel & Grinvald ‘96



Jhuang Serre et al ‘07 ’10; Kuehne Jhuang et al ‘11

# Automated rodent behavioral analysis

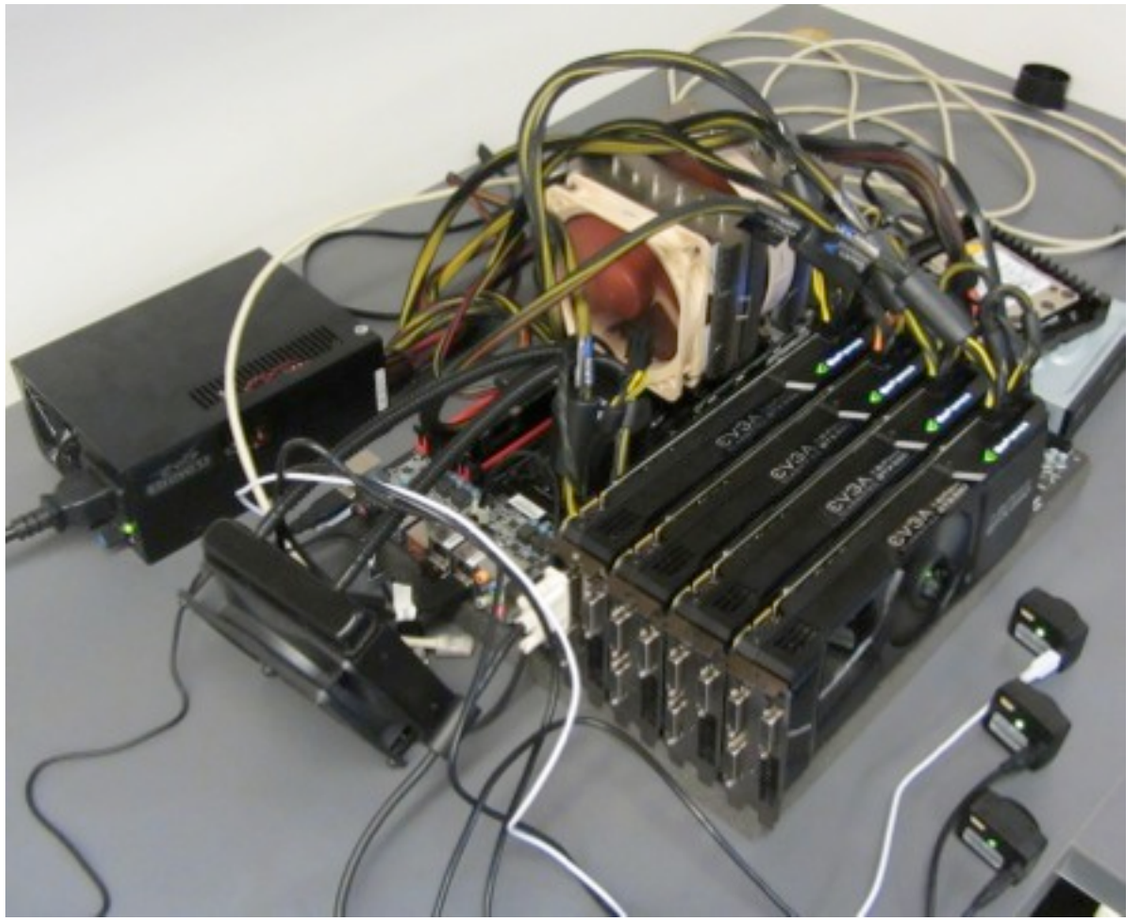


**Table 1 | Accuracy of the system.**

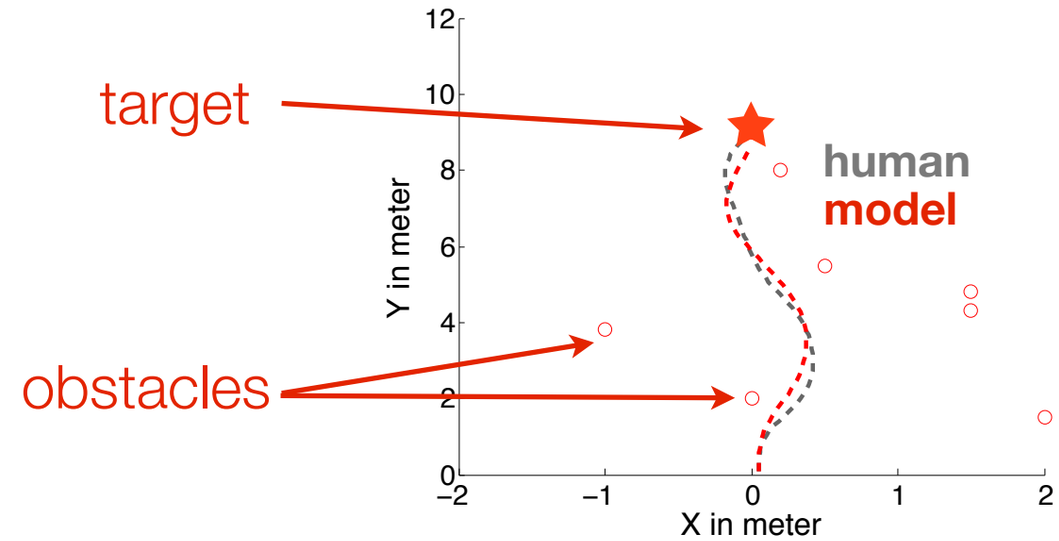
| | Our system | CleverSys commercial system | Human (*'Annotator group 2'*) |
|---|---|---|---|
| 'Set B' (1.6 h of video) | 77.3%/76.4% | 60.9%/64.0% | 71.6%/75.7% |
| 'Full database' (over 10 h of video) | 78.3%/77.1% | 61.0%/65.8% | |

Accuracies are reported as averaged across frames/across behaviours (underlined numbers, computed as the average of the diagonal entities in Figure 3 confusion matrix; chance level is 12.5% for an eight-class classification problem).

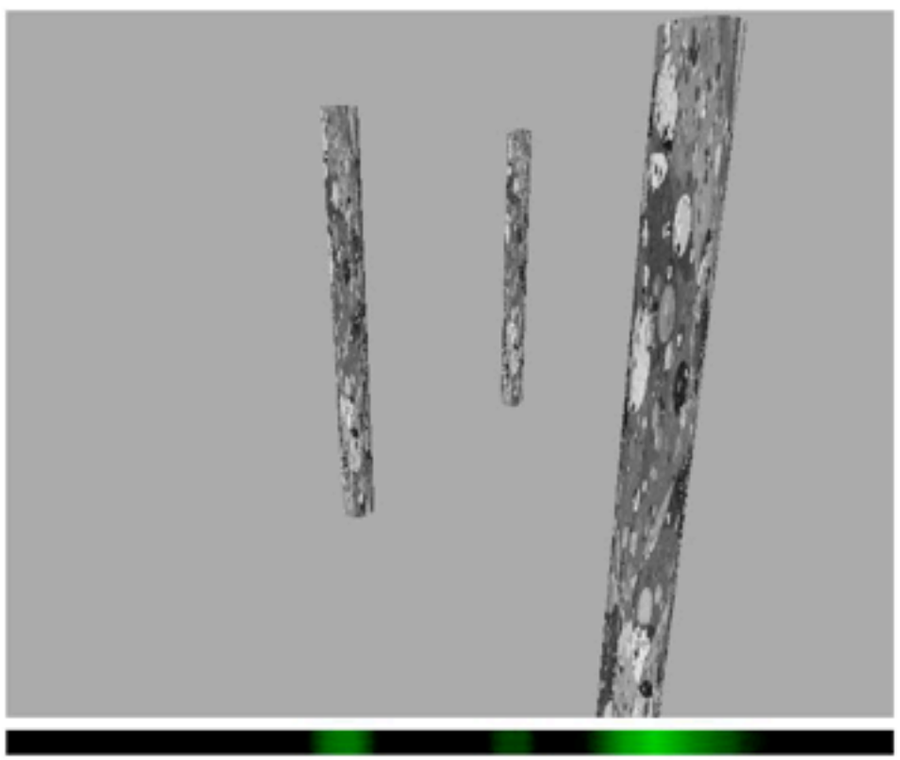


Jhuang Serre et al '07 '10; Kuehne Jhuang et al '11

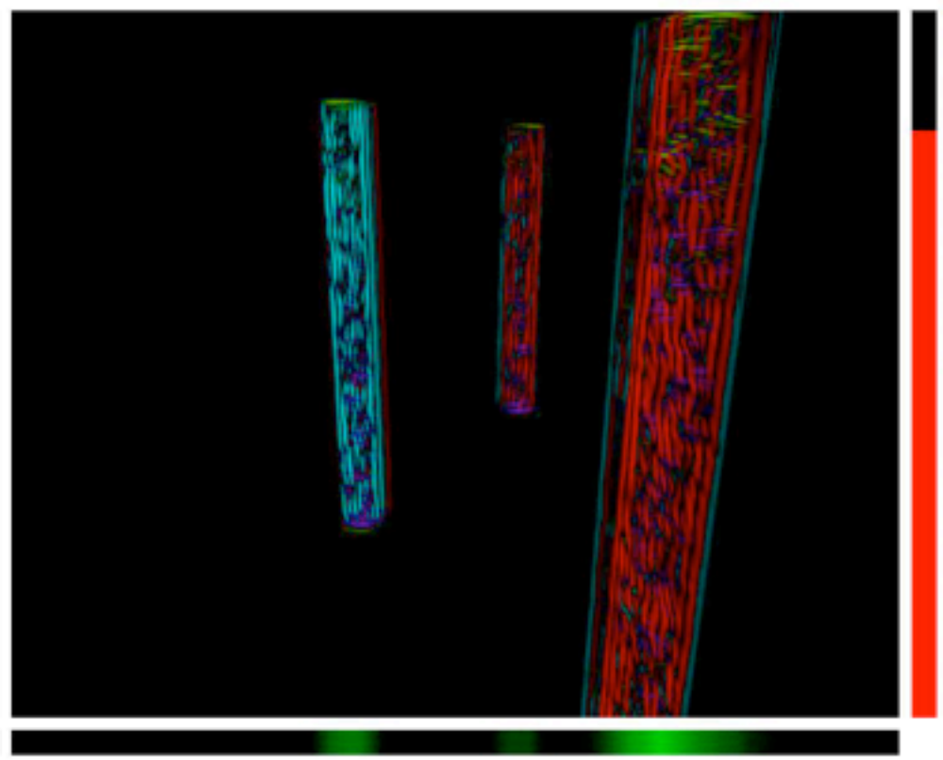
# Visual control of navigation



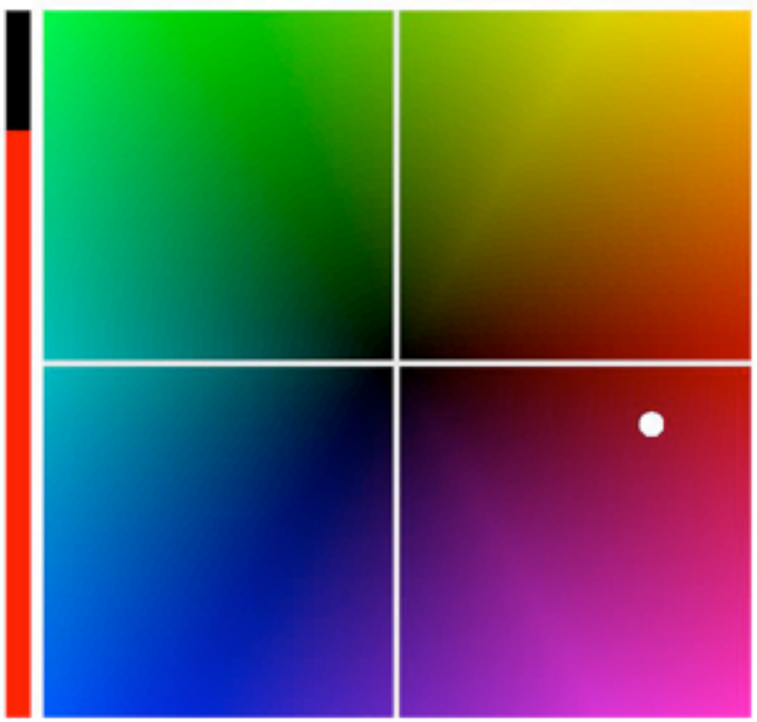

**Viewer-centered visual scene**



**V1-like unit responses**



**Motion direction decoding**

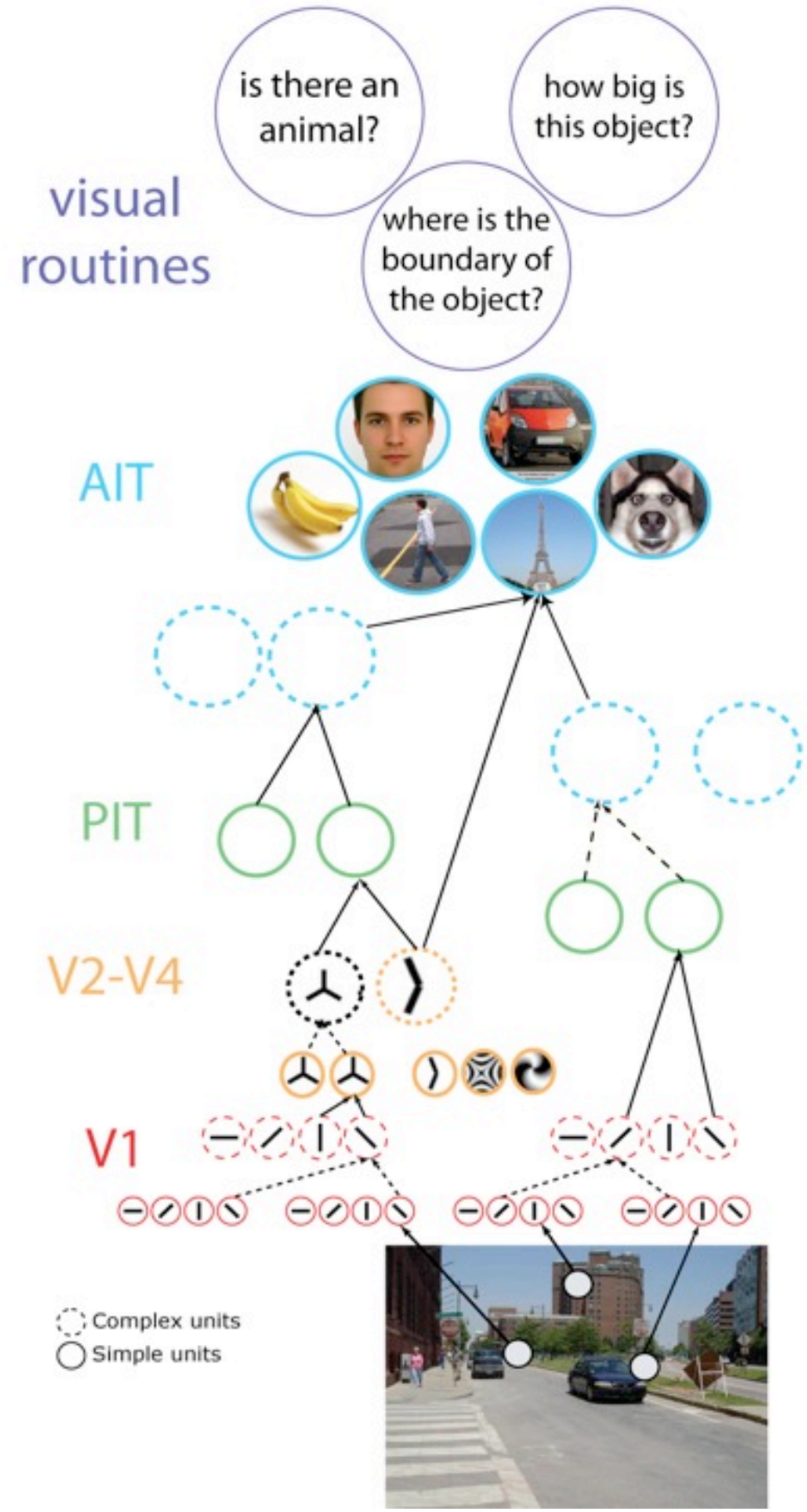




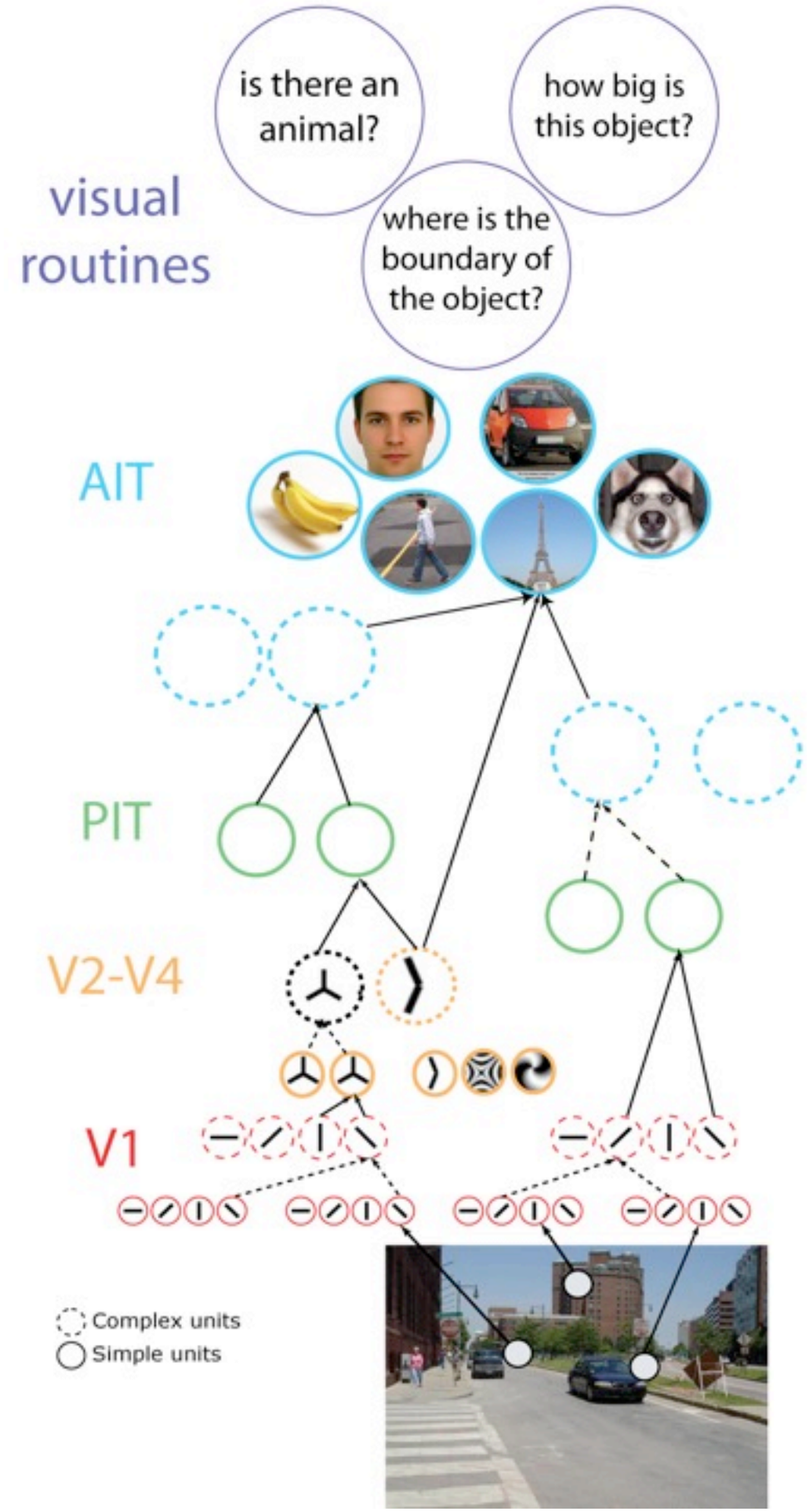
Bonneaud Bruggeman Olfers Irwin & Serre in prep

- Feedforward hierarchical learning architectures seem consistent:
  - With anatomy and physiology of visual cortex
  - With human psychophysics during rapid categorization tasks
- But incomplete models of visual processing
  - Suffer from ‘clutter problem’ and cannot parse and interpret visual scenes
  - Possible role for cortical feedback and shifts of attention

# Summary



- What should you care about biology?
  - i.e., why not learning everything?
- Parameter space is prohibitively large
  - Biology might give you at least a starting point




# Summary

# Deep learning in the visual cortex

Thomas Serre
Brown University

I. Fundamentals of primate vision

II. Computational mechanisms of rapid recognition and feedforward processing

III. Beyond feedforward processing: Attentional mechanisms and cortical feedback