9.520

Vision and visual neuroscience II

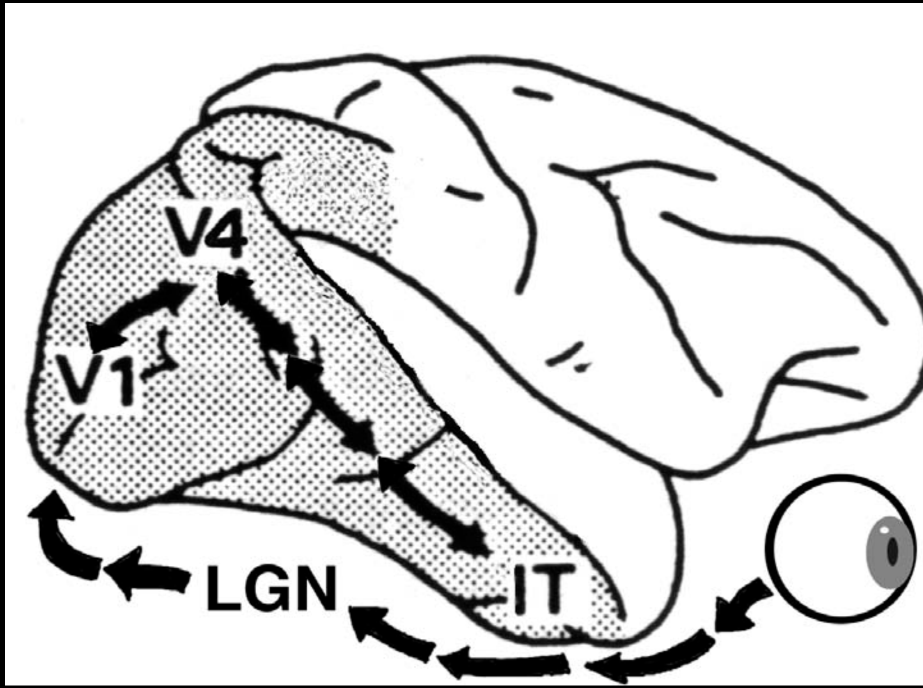
Thomas Serre & Tomaso Poggio

McGovern Institute for Brain Research
Center for Biological and Computational Learning
Department of Brain & Cognitive Sciences

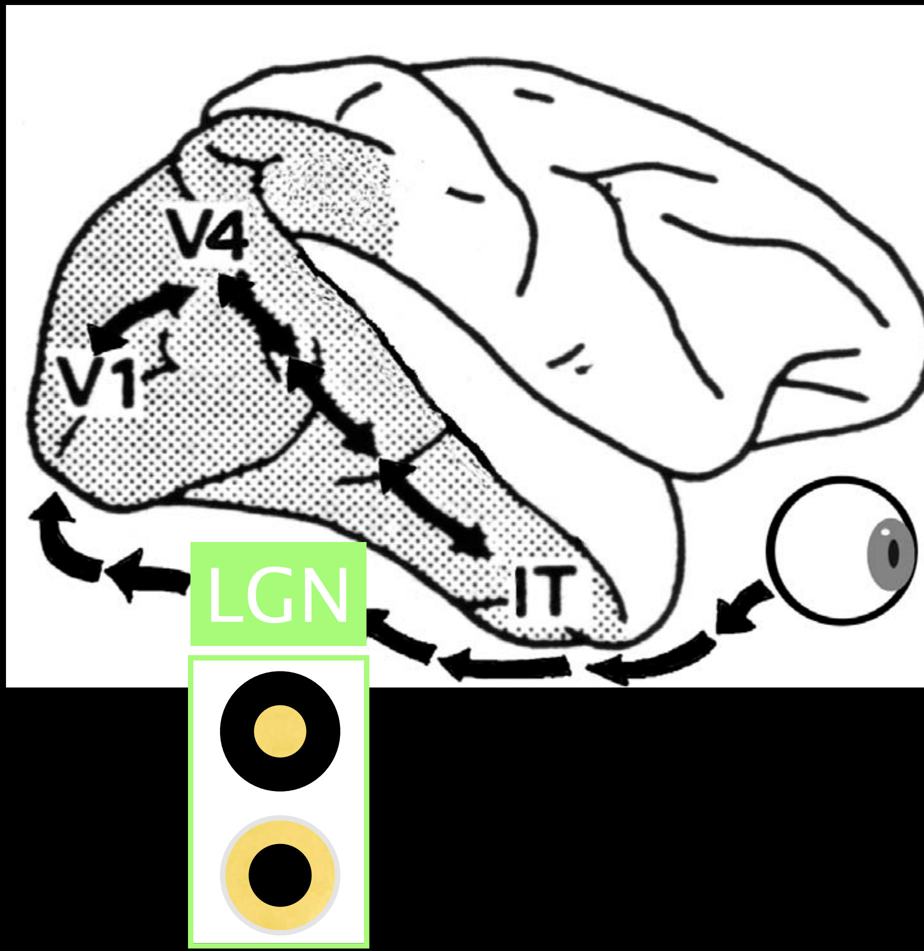
Last class

- ✦ Problem of visual recognition
- ✦ Historical background
- ✦ Neurons and areas in the visual system
- ✦ Data and hierarchical feedforward models

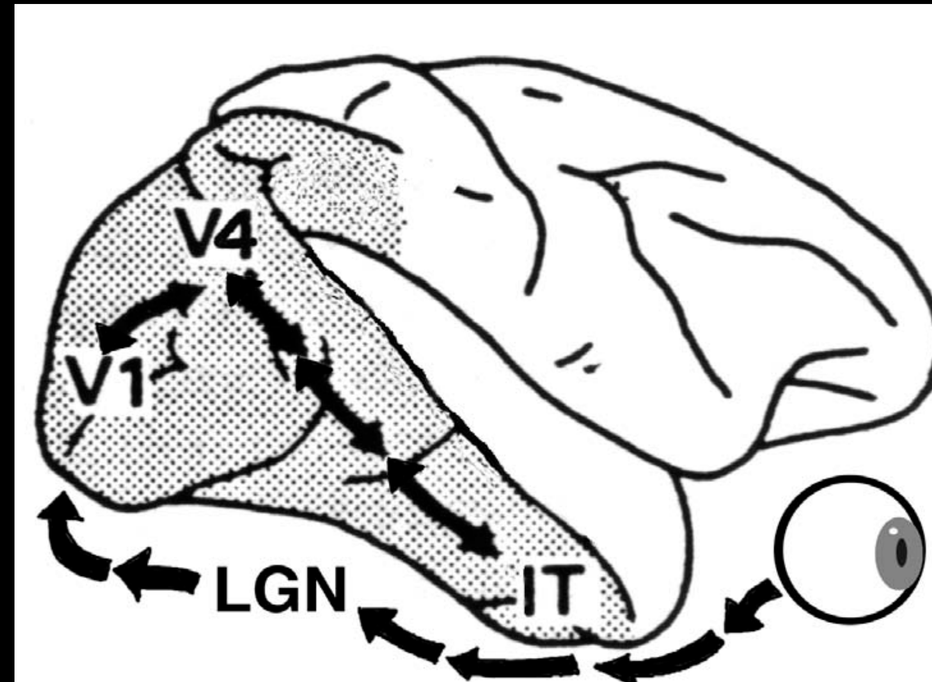
modified from (Ungerleider and Haxby 1994)



modified from (Ungerleider and Haxby 1994)

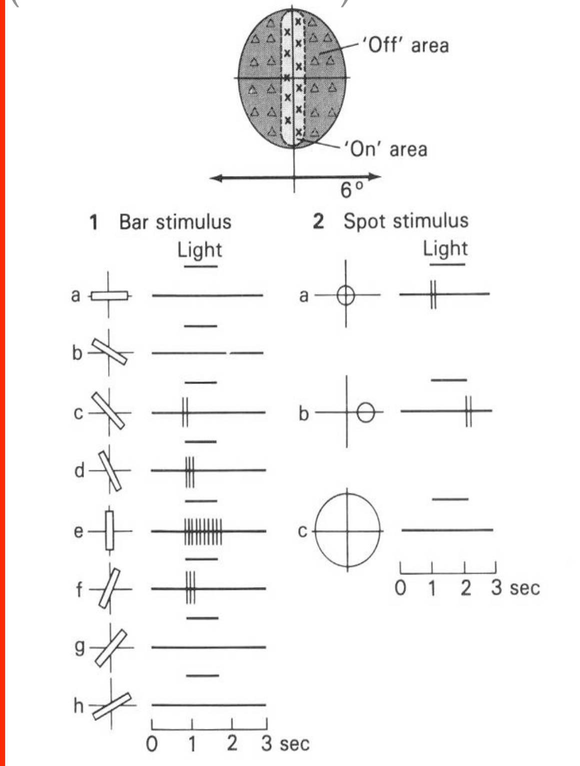


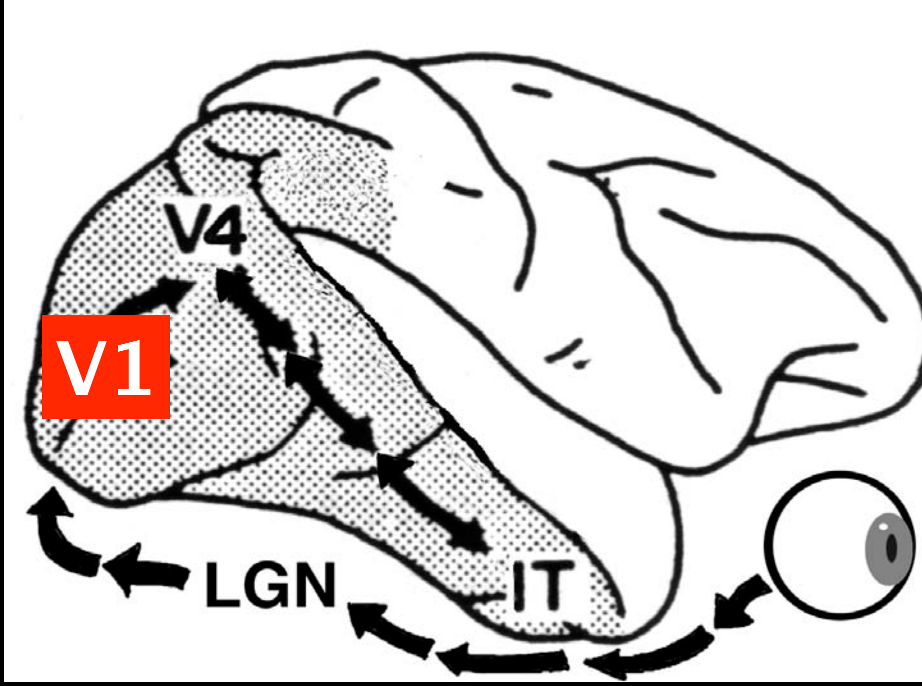
modified from (Ungerleider and Haxby 1994)



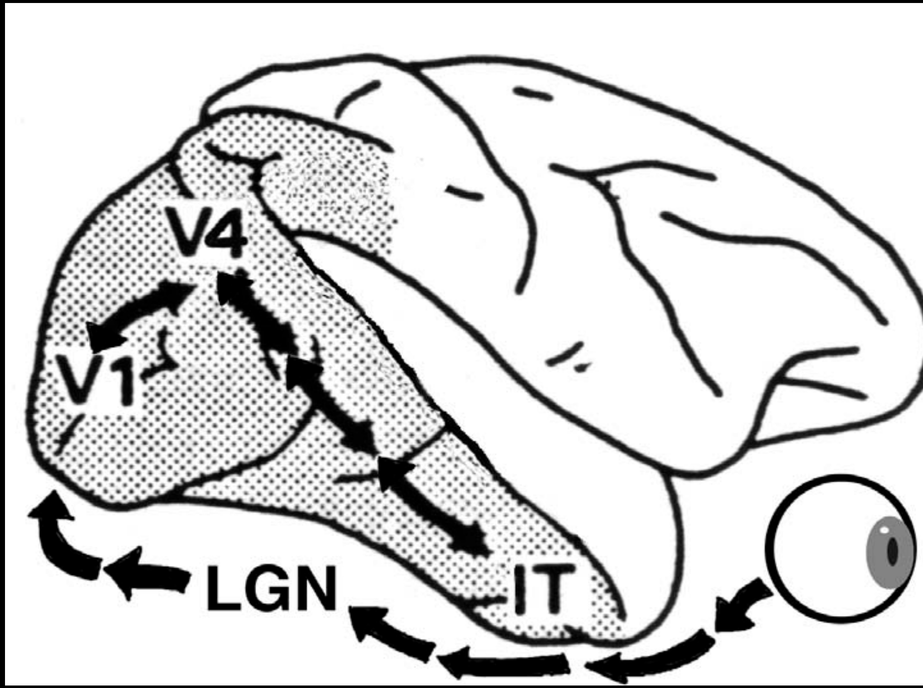
modified from (Ungerleider and Haxby 1994)

(Hubel & Wiesel 1959)

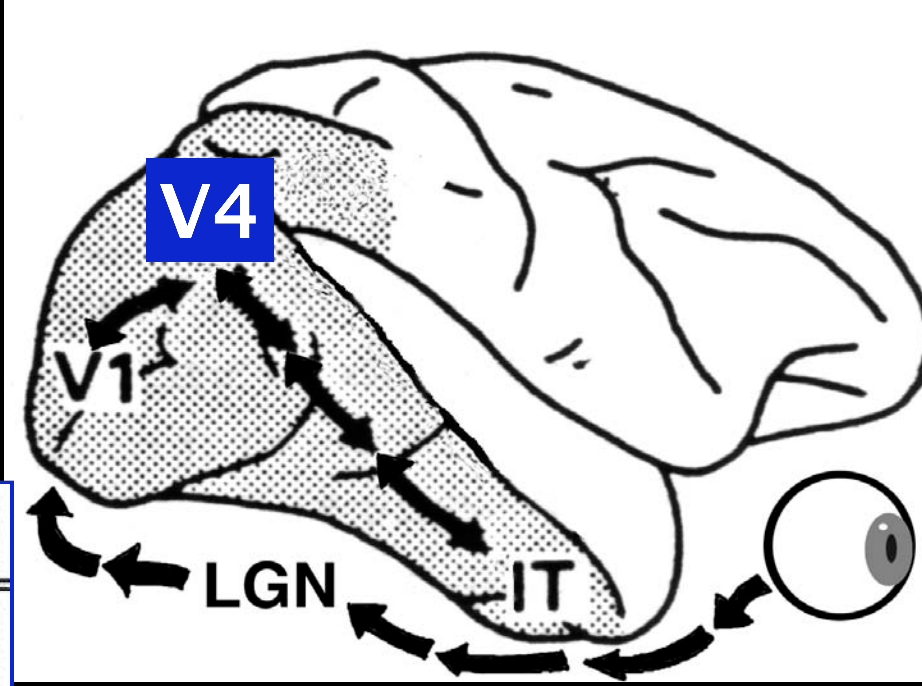


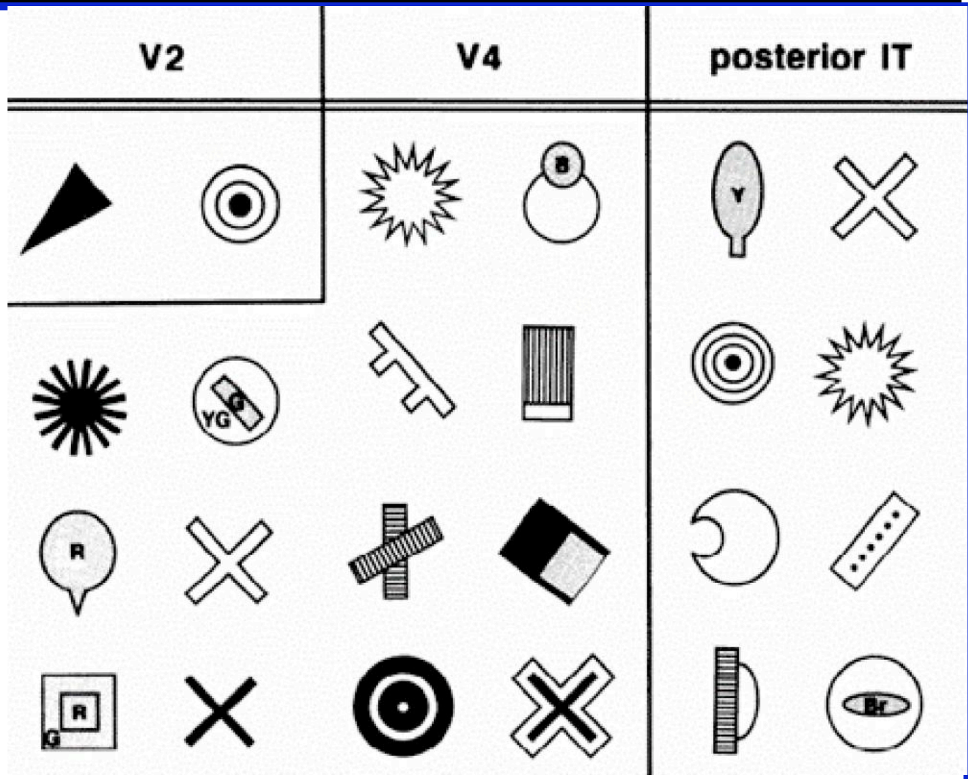


modified from (Ungerleider and Haxby 1994)



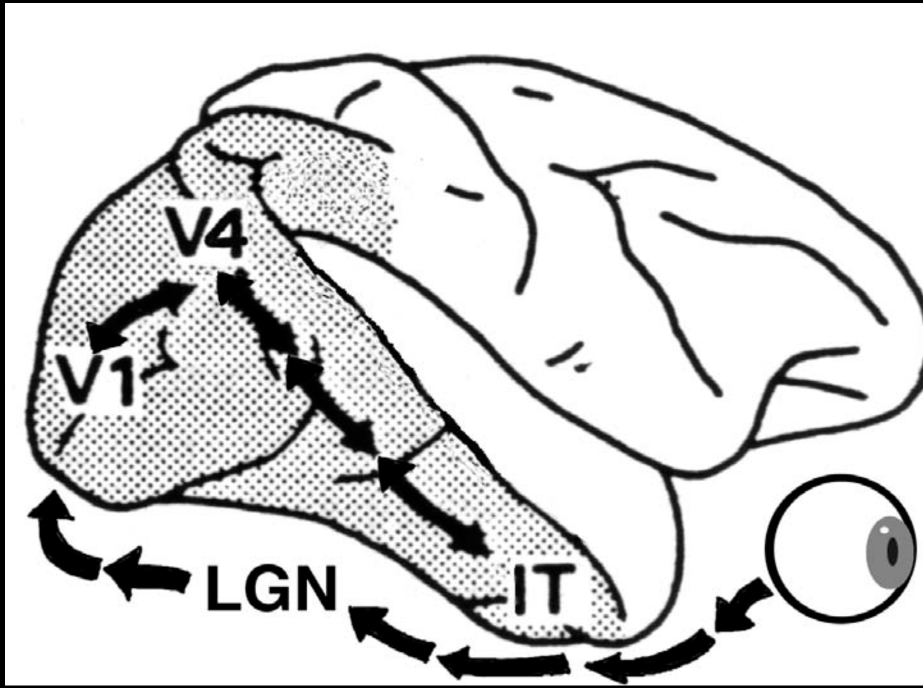
modified from (Ungerleider and Haxby 1994)



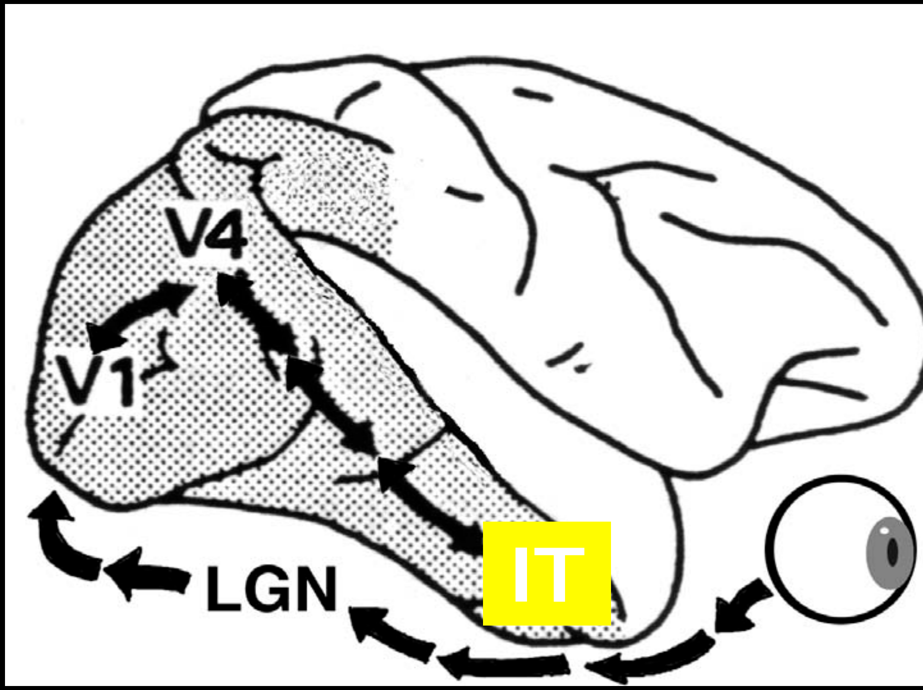


(Kobatake and Tanaka, 1994)

modified from (Ungerleider and Haxby 1994)

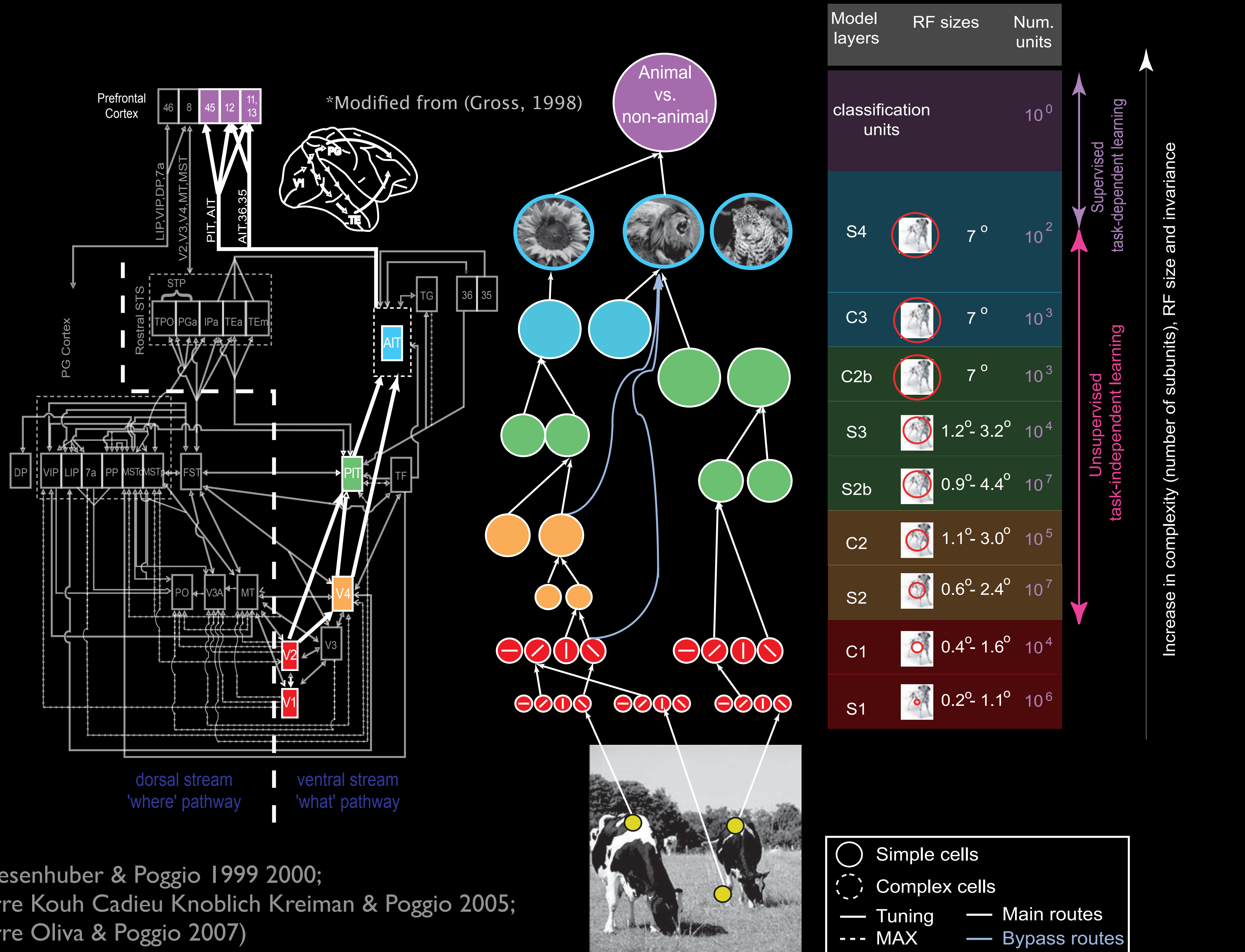


modified from (Ungerleider and Haxby 1994)



Rapid categorization

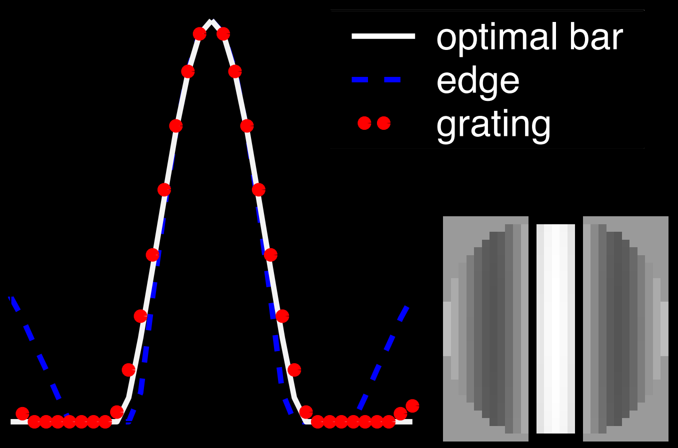
(Biederman 1972; Potter 1975; Thorpe 1996)



(Riesenhuber & Poggio 1999 2000;
Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005;
Serre Oliva & Poggio 2007)

Example: V1

	Model	Cortex	References
Receptive field sizes			
simple cells	$0.2^\circ - 1.1^\circ$	$\approx 0.1^\circ - 1.0^\circ$	[Schiller et al., 1976e; Hubel and Wiesel, 1965]
complex cells	$0.4^\circ - 1.6^\circ$	$\approx 0.2^\circ - 2.0^\circ$	
Peak frequencies (cycles / deg)			
simple cells	range: 1.6 – 9.8 mean/med: 3.7/2.8	bulk ≈ 1.0 – 4.0 mean: ≈ 2.2 range: ≈ 0.5 – 8.0	[DeValois et al., 1982a])
complex cells	range: 1.8 – 7.8 mean/med: 3.9/3.2	bulk ≈ 2.0 – 5.6 mean: 3.2 range ≈ 0.5 – 8.0	
Frequency bandwidth at 50% amplitude (cycles / deg)			
simple cells	range: 1.1 – 1.8 med: ≈ 1.45	bulk ≈ 1.0 – 1.5 med: ≈ 1.45 range ≈ 0.4 – 2.6	[DeValois et al., 1982a]
complex cells	range: 1.5 – 2.0 med: 1.6	bulk ≈ 1.0 – 2.0 med: 1.6 range ≈ 0.4 – 2.6	
Frequency bandwidth at 71% amplitude (index)			
simple cells	range: 44 – 58 med: 55	bulk ≈ 40 – 70	[Schiller et al., 1976d]
complex cells	range 40 – 50 med. 48	bulk ≈ 40 – 60	
Orientation bandwidth at 50% amplitude (octaves)			
simple cells	range: $38^\circ - 49^\circ$ med: 44°	—	[DeValois et al., 1982b]
complex cells	range: $27^\circ - 33^\circ$ med: 43°	bulk ≈ $20^\circ - 90^\circ$ med: 44°	
Orientation bandwidth at 71% amplitude (octaves)			
simple cells	range: $27^\circ - 33^\circ$ med: 30°	bulk ≈ $20^\circ - 70^\circ$	[Schiller et al., 1976c]
complex cells	range: $27^\circ - 33^\circ$ med: 31°	bulk ≈ $20^\circ - 90^\circ$	



(Serre & Riesenhuber 2004)

This class

This class

- Feedforward hierarchical models of the visual cortex

This class

- Feedforward hierarchical models of the visual cortex
  - Detailed implementation + learning

This class

- ✦ Feedforward hierarchical models of the visual cortex
  - ⋆ Detailed implementation + learning
  - ⋆ Comparison w| neural data

This class

- ✦ Feedforward hierarchical models of the visual cortex
  - ⋆ Detailed implementation + learning
  - ⋆ Comparison w| neural data
  - ⋆ Agreement with psychophysics

This class

- Feedforward hierarchical models of the visual cortex
  - Detailed implementation + learning
  - Comparison w| neural data
  - Agreement with psychophysics
  - Application to computer vision

This class

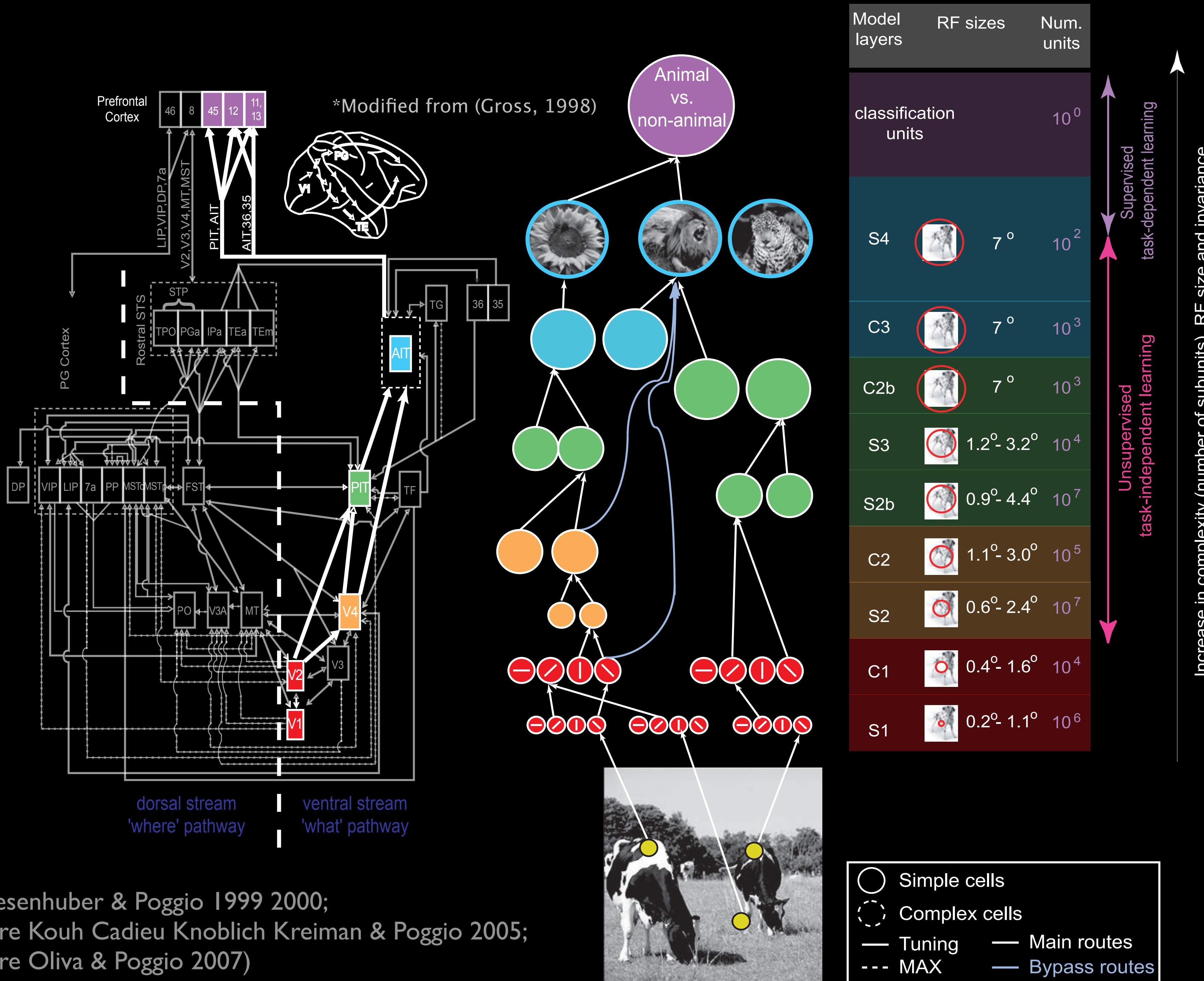
- ✦ Feedforward hierarchical models of the visual cortex
  - ★ Detailed implementation + learning
  - ★ Comparison w| neural data
  - ★ Agreement with psychophysics
  - ★ Application to computer vision
- ✦ Beyond (static) feedforward processing

This class

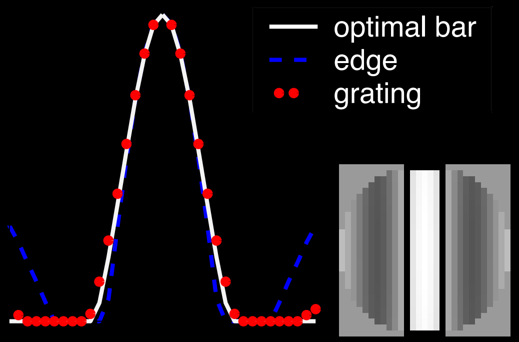
- ✦ Feedforward hierarchical models of the visual cortex
  - ★ Detailed implementation + learning
  - ★ Comparison w| neural data
  - ★ Agreement with psychophysics
  - ★ Application to computer vision
- ✦ Beyond (static) feedforward processing
  - ★ Extension to action recognition in the dorsal stream

This class

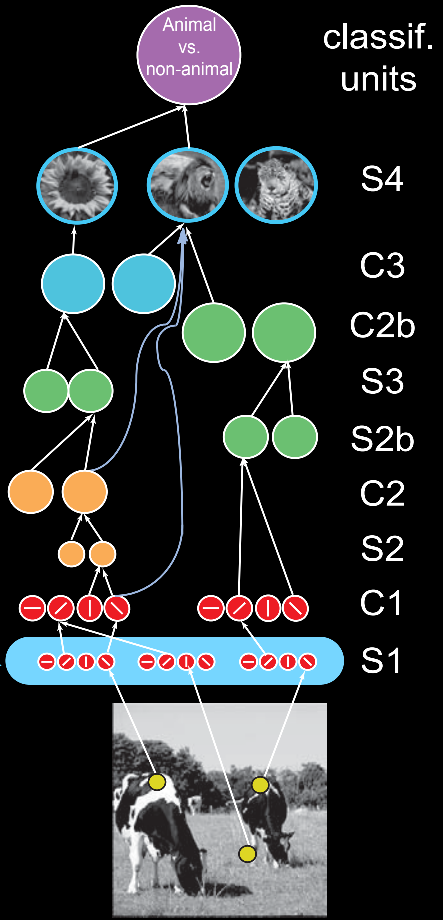
- ✦ Feedforward hierarchical models of the visual cortex
  - ★ Detailed implementation + learning
  - ★ Comparison w| neural data
  - ★ Agreement with psychophysics
  - ★ Application to computer vision
- ✦ Beyond (static) feedforward processing
  - ★ Extension to action recognition in the dorsal stream
  - ★ Attention and cortical feedbacks



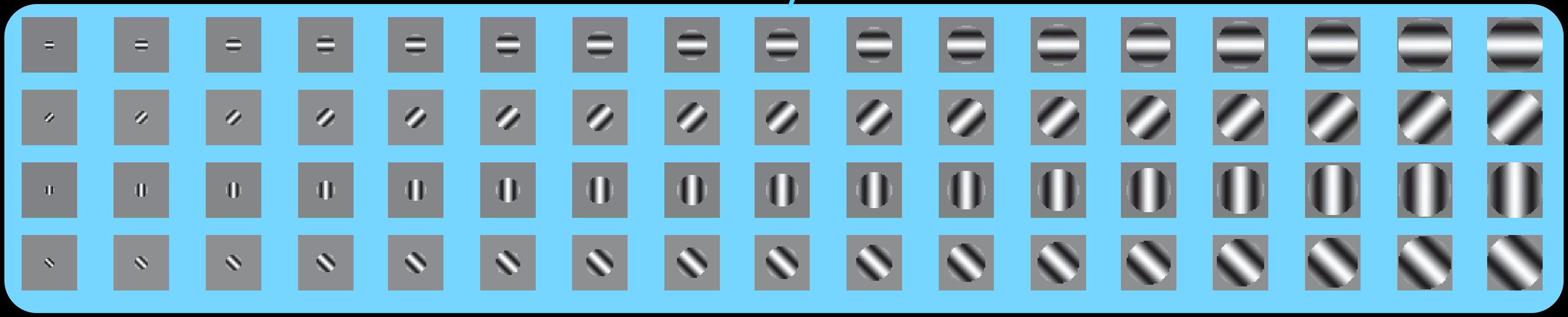
(Riesenhuber & Poggio 1999 2000;
Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005;
Serre Oliva & Poggio 2007)



S1 units

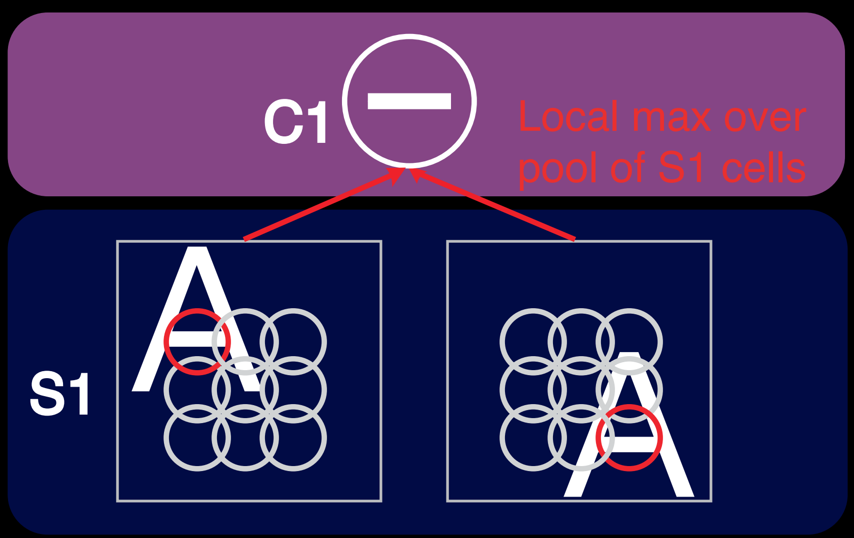


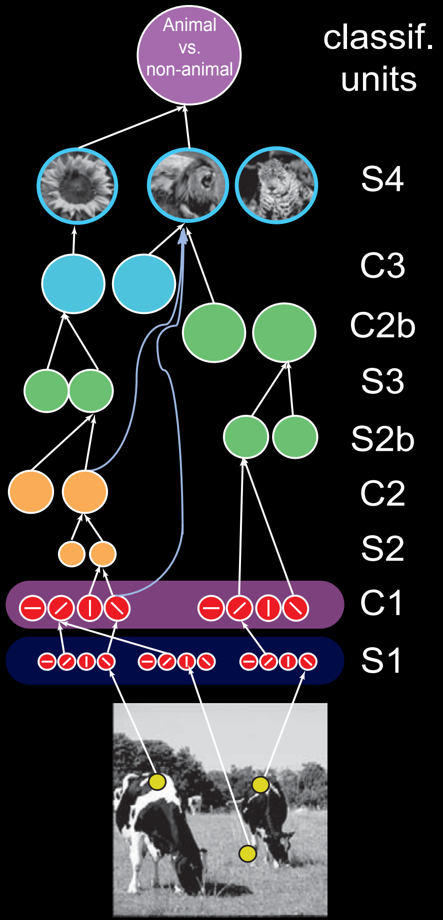
- ✦ Gabor filters
- ✦ Parameters fit to V1 data (Serre & Riesenhuber 2004)
  - 17 spatial frequencies (=scales)
  - 4 orientations



C1 units

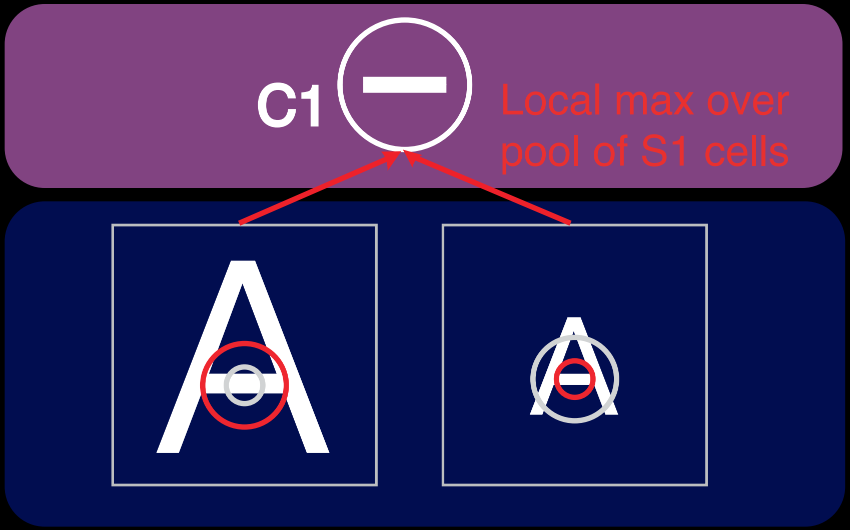
Increase in tolerance to **position** (and in RF size)

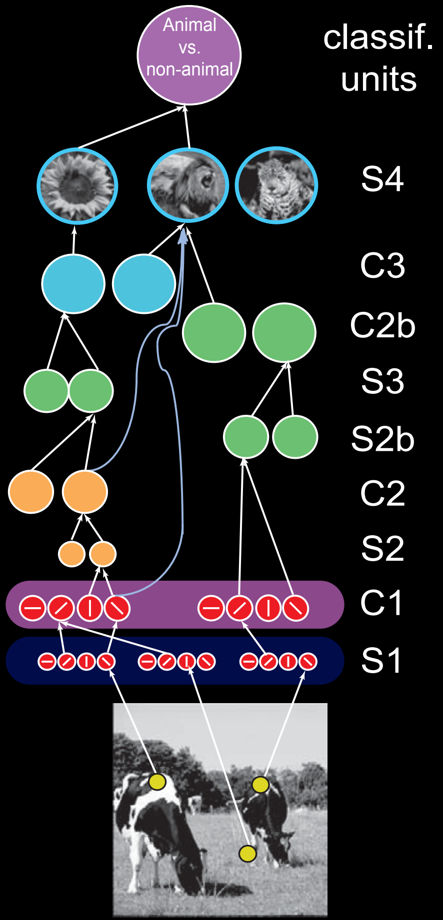




C1 units

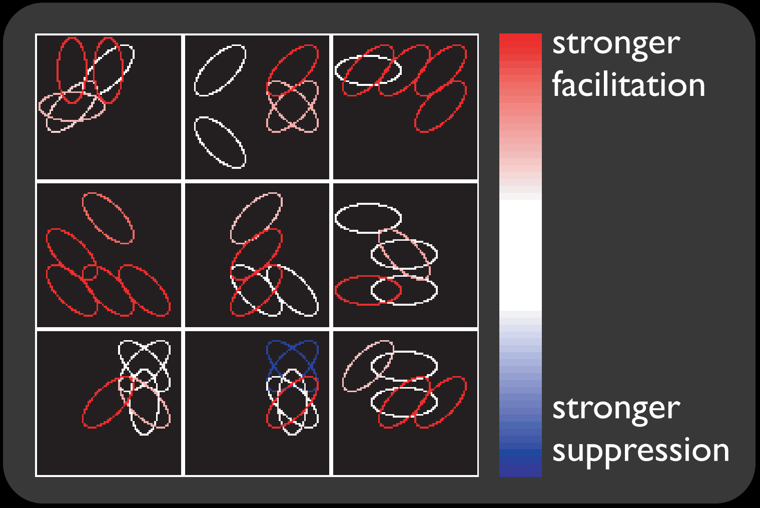
Increase in tolerance to **scale**

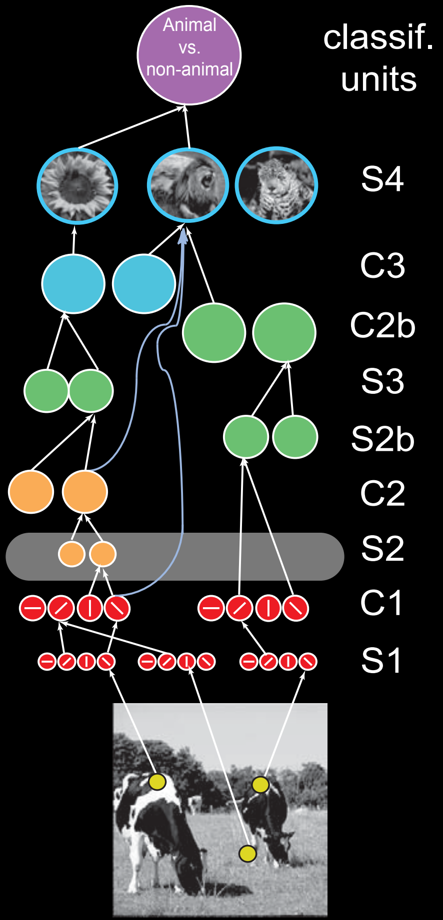




S2 units

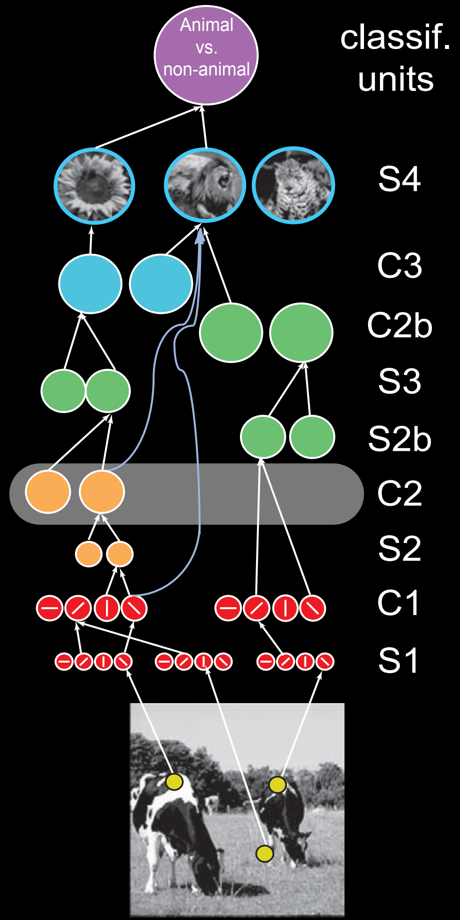
- ✦ Features of moderate complexity (n~1,000 types)
- ✦ Combination of V1-like complex units at different orientations
  - Synaptic weights ***w*** learned from natural images
  - 5-10 subunits chosen at random from all possible afferents (~100-1,000)

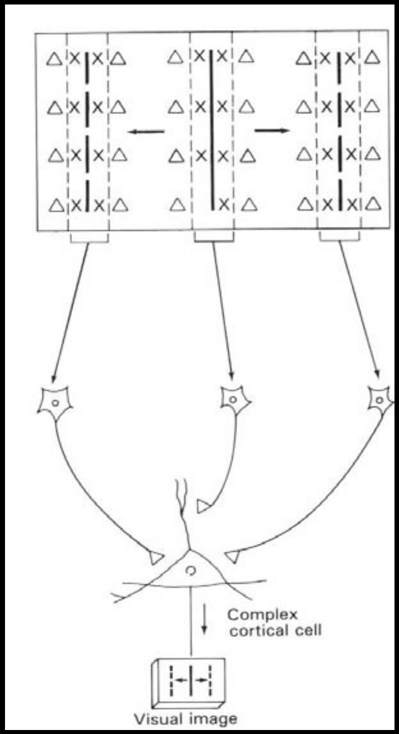




C2 units

- ✦ Same selectivity as S2 units but increased tolerance to position and size of preferred stimulus
- ✦ Local pooling over S2 units with same selectivity but slightly different positions and scales
- ✦ S2 units in V2 and C2 in V4?





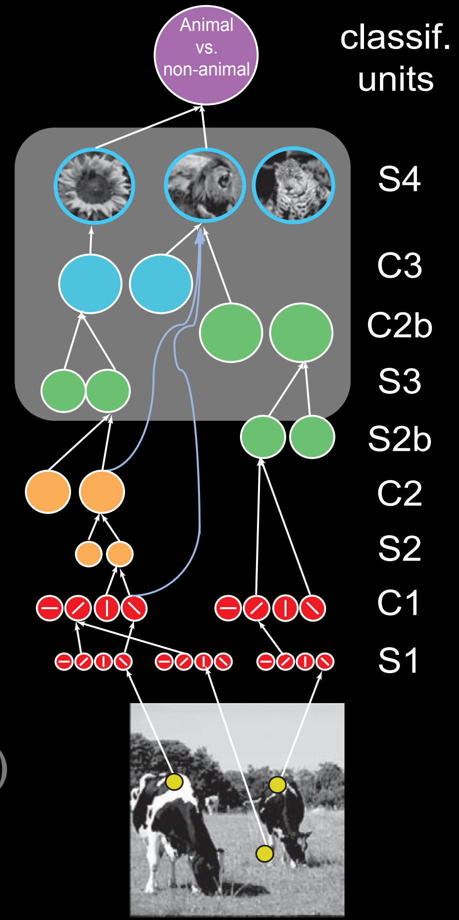
(Hubel & Wiesel 1959)

Beyond C2 units

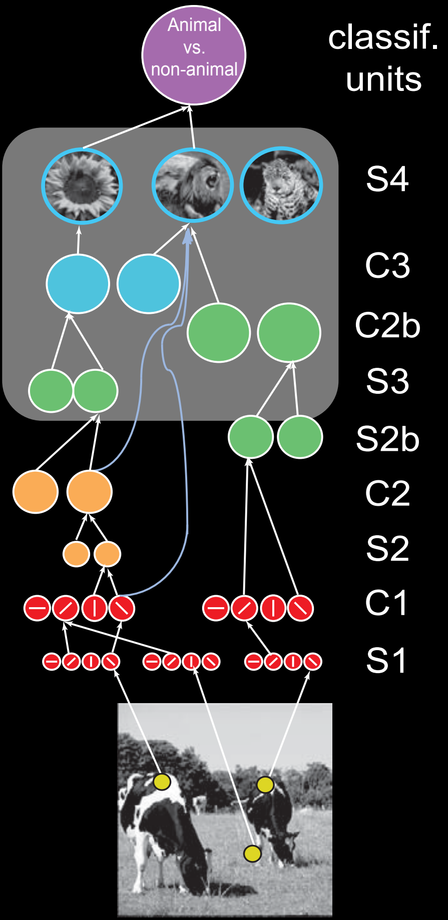
✦Units increasingly complex and invariant

✦ S3/C3 units:

- Combination of V4-like units with different selectivities
- Dictionary of ~1,000 features = num. columns in IT (Fujita 1992)



Beyond C2 units



✦Units increasingly complex and invariant

✦ S3/C3 units:

- Combination of V4-like units with different selectivities
- Dictionary of ~1,000 features = num. columns in IT (Fujita 1992)

✦ S4 units:

- View-tuned units (imprinted with part of the training set, e.g. animal and non-animal images but still unsupervised)
- Tuning and invariance properties agrees with IT data (Logothetis, Pauls & Poggio 1995)

Related to Edelman & Poggio (Edelman & Poggio 1990)

Related to Ullman's visual features of intermediate complexity (Ullman et al 2002)

Gabor filters (Jones & Palmer 1987)

2 key learning stages:

Related to Edelman & Poggio (Edelman & Poggio 1990)

Related to Ullman's visual features of intermediate complexity (Ullman et al 2002)

Gabor filters (Jones & Palmer 1987)

2 key learning stages:

Related to Edelman & Poggio (Edelman & Poggio 1990)

Related to Ullman's visual features of intermediate complexity (Ullman et al 2002)

Gabor filters (Jones & Palmer 1987)

✳ Large dictionary of reusable features:

- "unbound" features (Treisman & Gelade 1980; Wolfe & Bennett 1997; Schyns & Oliva 1994)
- Different levels of invariance and complexity
- Unsupervised learning from natural images ~developmental-like learning stage

2 key learning stages:

✳ Task-specific circuits:
- Supervised learning from ~100-1000 labeled examples
- Linear classifier on top of VTUs (S4 units) [~RBF] (see Fredman Riesenhuber Poggio Miller, 2001, 2003)

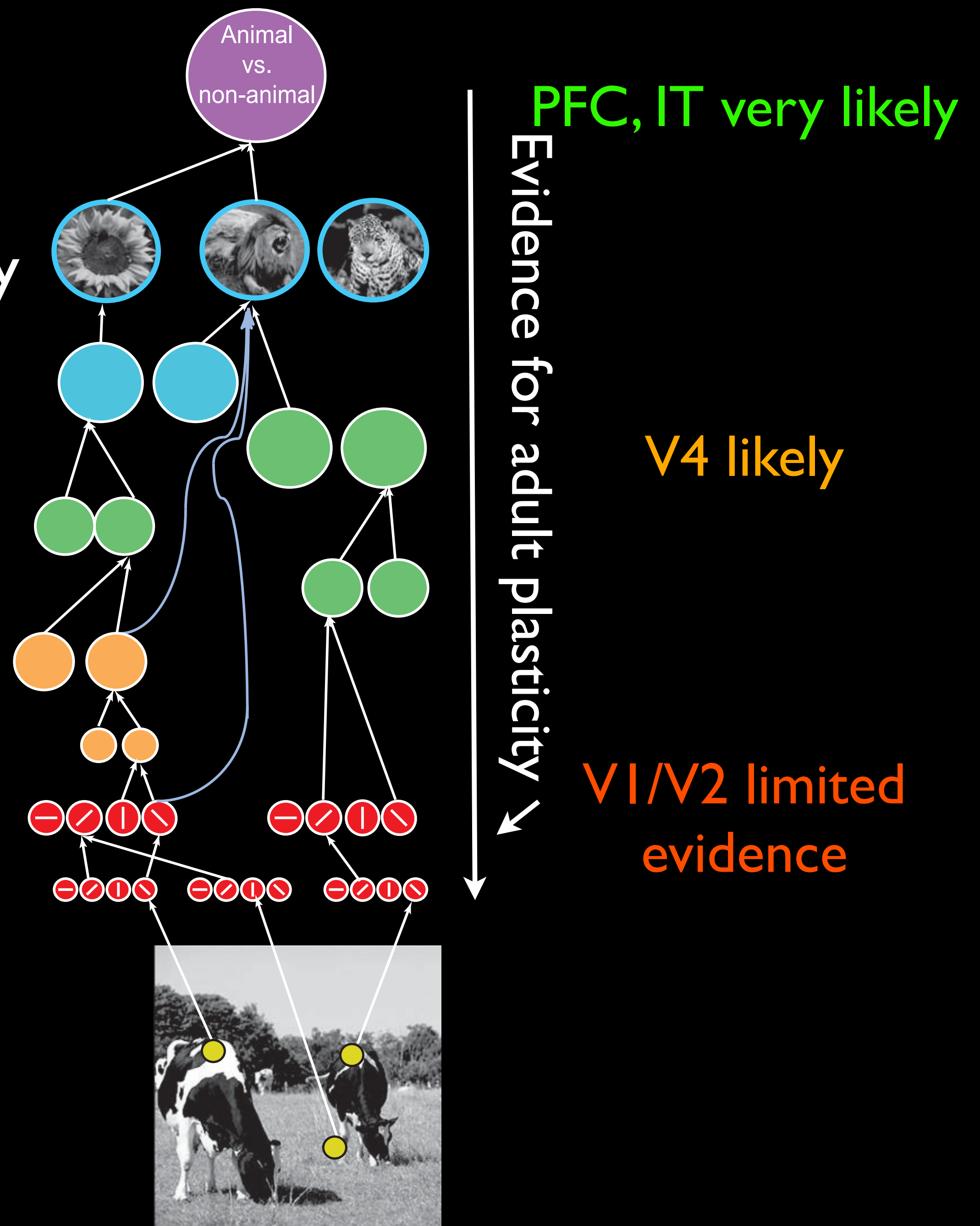
✳ Large dictionary of reusable features:
- "unbound" features (Treisman & Gelade 1980; Wolfe & Bennett 1997; Schyns & Oliva 1994)
- Different levels of invariance and complexity
- Unsupervised learning from natural images ~developmental-like learning stage

Related to Edelman & Poggio (Edelman & Poggio 1990)

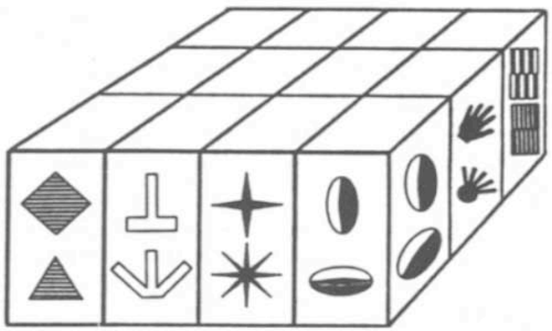
Related to Ullman's visual features of intermediate complexity (Ullman et al 2002)

Gabor filters (Jones & Palmer 1987)

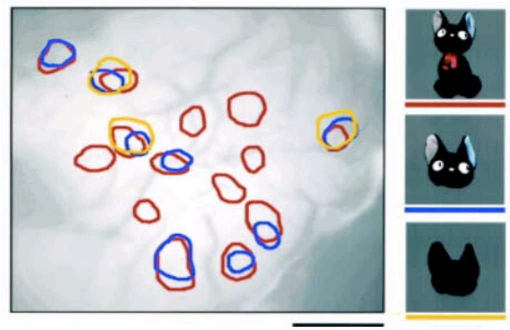
- ✦ Learning likely to play key role in recognition
- ✦ Details still open-ended (lack of neural data to constrain)
- ✦ Learning described in a more “algorithmic” way



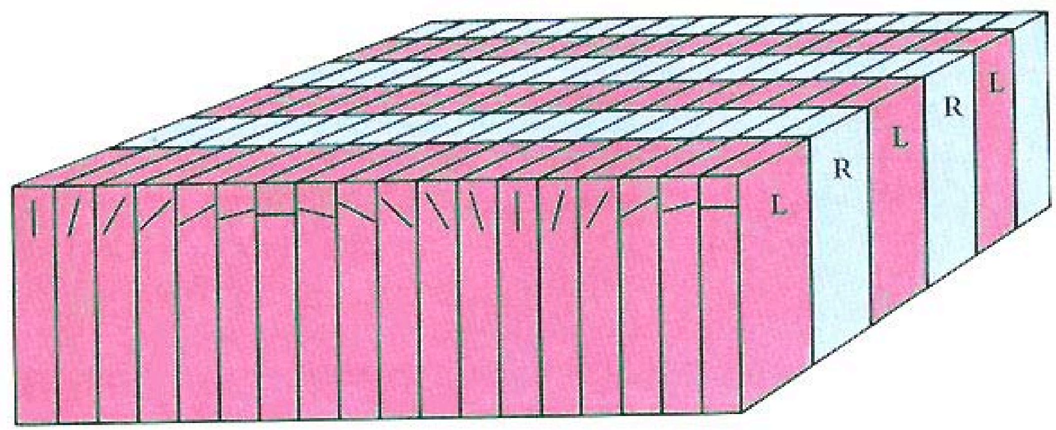
Columns in the cortex



Tanaka et al.



Tsunoda et al.

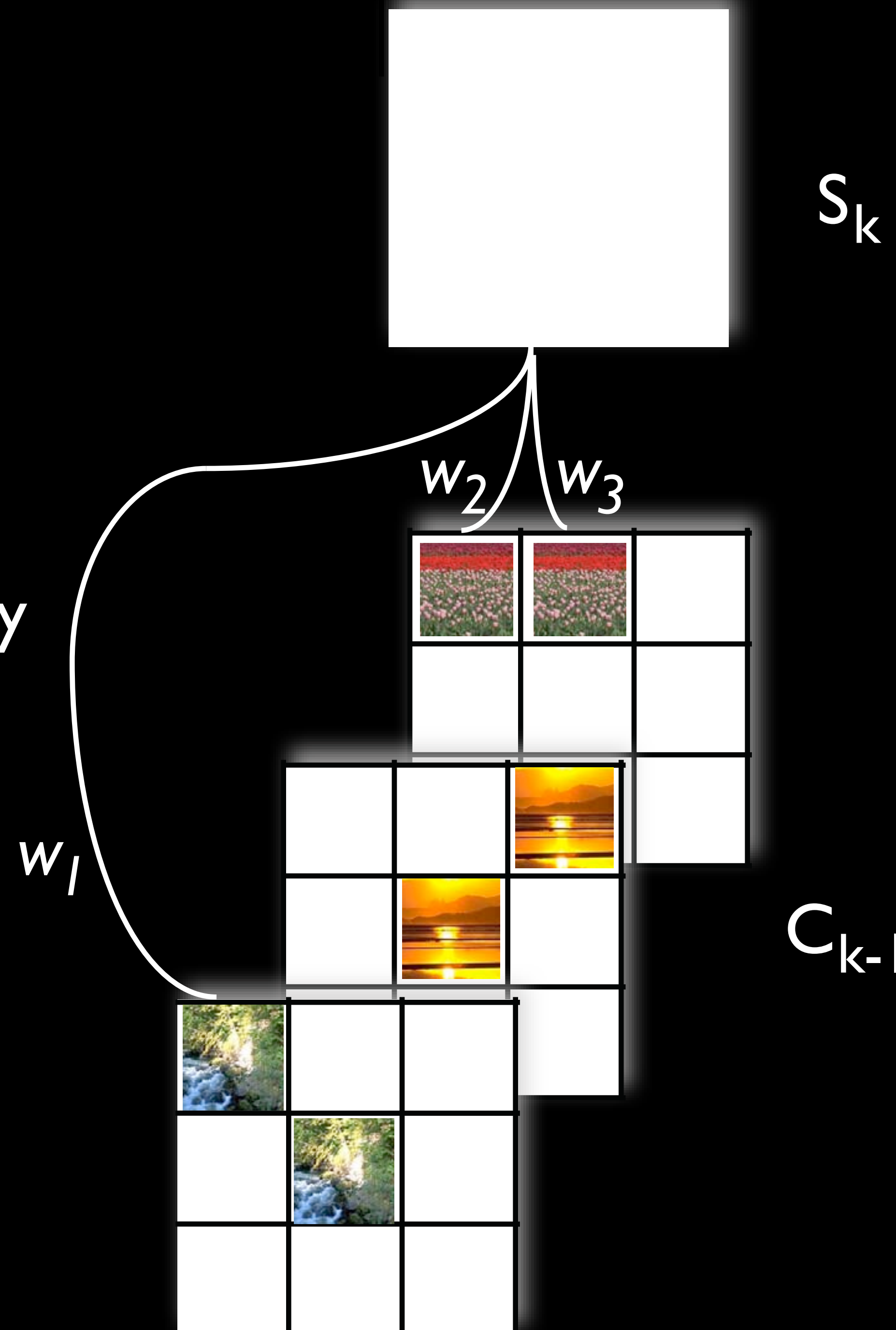


Orientation and ocular dominance columns

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.

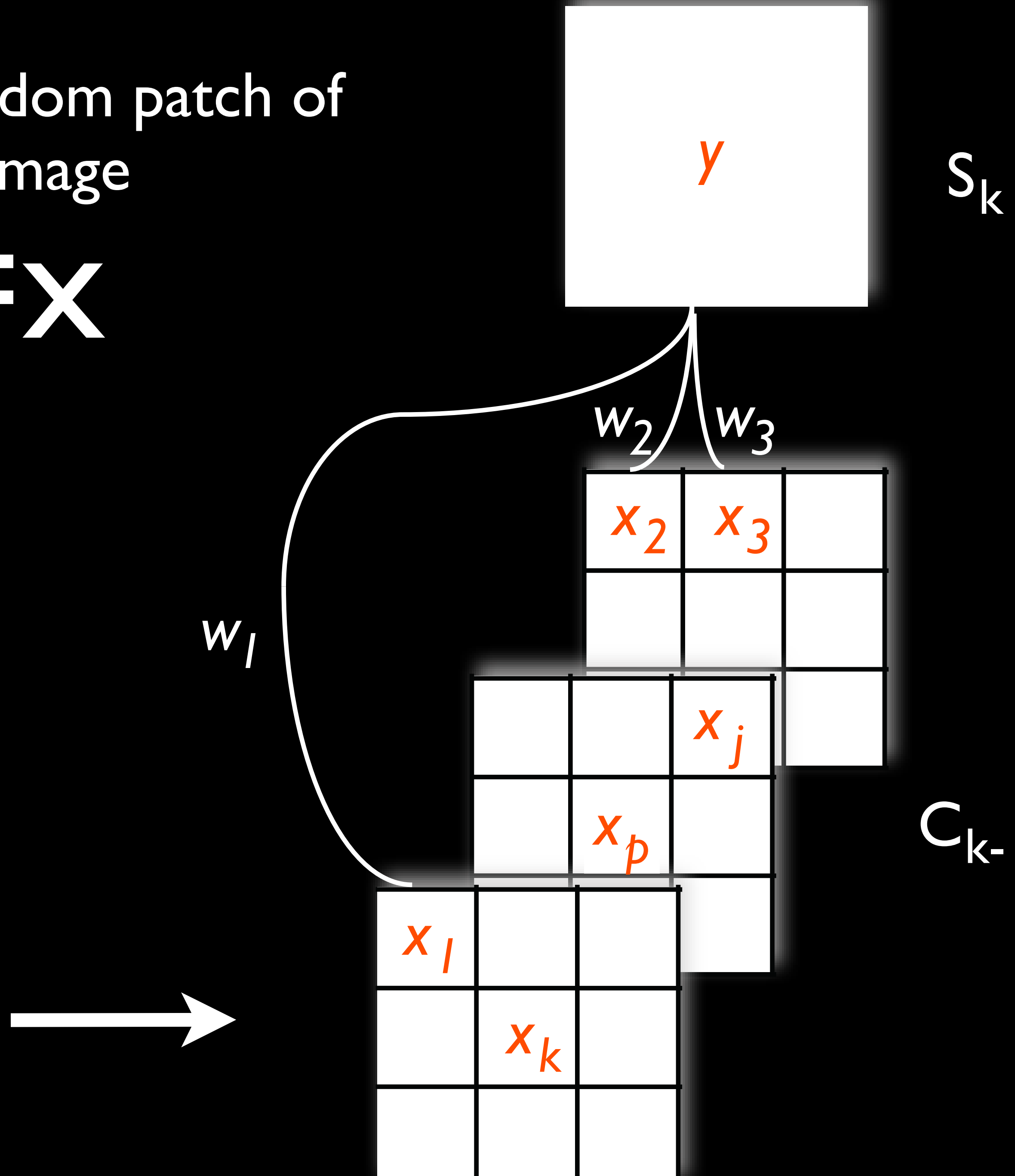
- Layers of the model are organized in columns
- Each model unit is equivalent to ~100 IF (~1 column of cortex)
- Each hypercolumn contains the same basic dictionary of features and is replicated at all positions and scales

- Learning is sequential
- Start with layer S2/C2 then S2b/C2b and S3/C3
- Pick one unit in layer Sk
- Select random set of inputs from retinotopically organized afferents

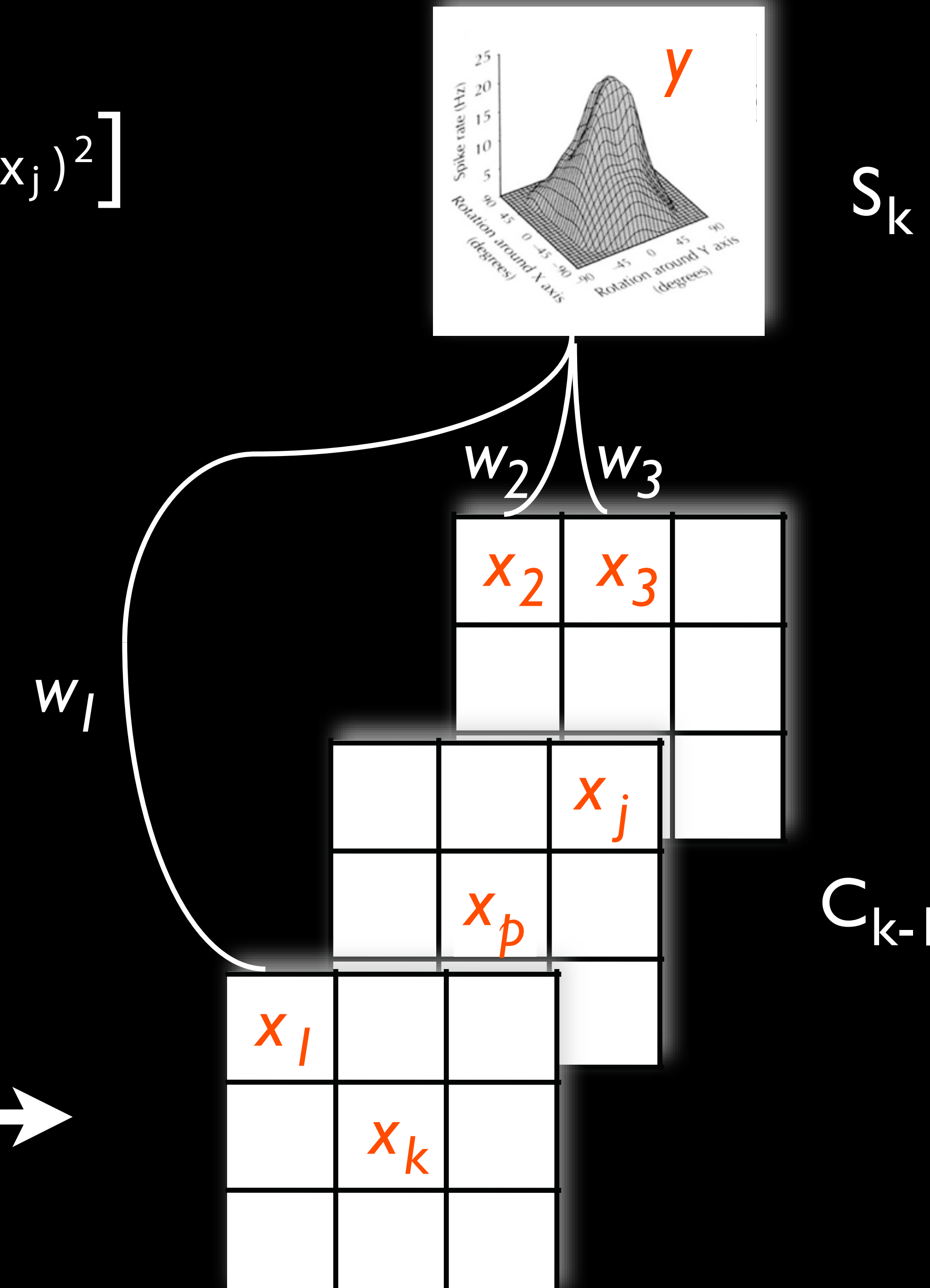


Imprint with random patch of natural image

w=x



$$y = \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(w_j - x_j)^2\right]$$

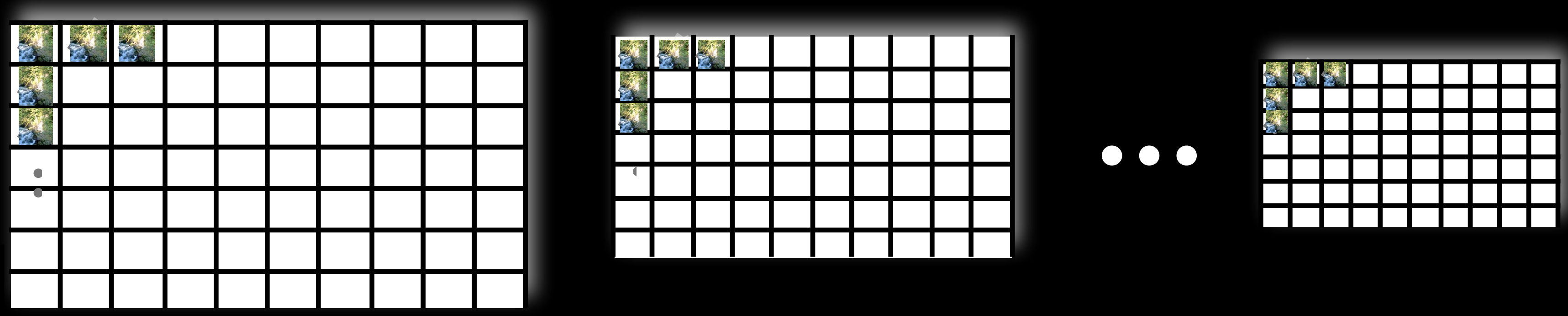


✦ We assume the input image moves (shifting and looming) so that the selectivity of the imprinted units gets replicated at all positions and scales

✦ We learn ~1,000 units this way and then move to the next layer

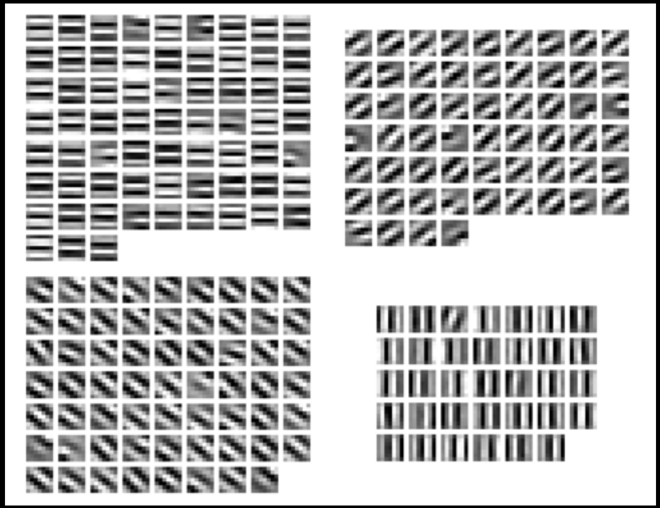
✦ Learning follows a long tradition of researchers who have argued that the visual system may be adapted to the statistics of the natural environment

(Attneave 1954; Barlow 1961; Atick 1992; Ruderman 1994; Simoncelli & Olshausen 2001)



Learning the invariance from temporal continuity

w| T. Masquelier & S. Thorpe (CNRS, France)



✦ Simple cells learn correlation in space (at the same time)

✦ Complex cells learn correlation in time

movie courtesy of Wolfgang Einhauser

see also (Foldiak 1991; Perrett et al 1984; Wallis & Rolls, 1997; Einhauser et al 2002; Wiskott & Sejnowski 2002; Spratling 2005)

Agreement w| experimental data

✦ V1:

- Simple and complex cells tuning properties (Schiller et al 1976; Hubel & Wiesel 1965; Devalois et al 1982)
- MAX operation in subset of complex cells (Lampl et al 2004)

✦ V2:

- Combination of orientations in V2 (Anzai et al,2007)

✦ V4:

- Tuning for two-bar stimuli (Reynolds Chelazzi & Desimone 1999)
- MAX operation (Gawne et al 2002)
- Two-spot interaction (Freiwald et al 2005)
- Tuning for boundary conformation (Pasupathy & Connor 2001)
- Tuning for Cartesian and non-Cartesian gratings (Gallant et al 1996)

✦ IT:

- Tuning and invariance properties (Logothetis et al 1995)
- Differential role of IT and PFC in categorization (Freedman et al 2001 2002 2003)
- Read out data (Hung Kreiman Poggio & DiCarlo 2005)
- Average effect in IT (Zoccolan Cox & DiCarlo 2005; Zoccolan Kouh Poggio & DiCarlo in press)

✦ Human:

- Face processing (fMRI + psychophysics) (Riesenhuber et al 2004; Jiang et al 2006)
- Rapid object categorization (Serre, Oliva & Poggio 2007)

fwd >>

(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

Example: V1

	Receptive field sizes		
	Model	Cortex	References
simple cells	0.2° – 1.1°	≈ 0.1° – 1.0°	[Schiller et al., 1976e; Hubel and Wiesel, 1965]
complex cells	0.4° – 1.6°	≈ 0.2° – 2.0°	

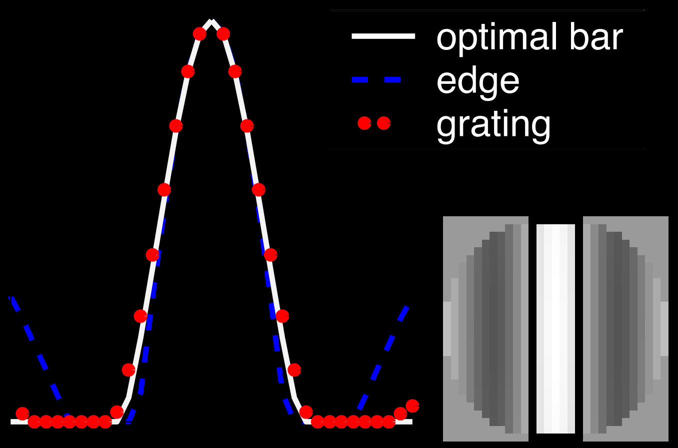
	Peak frequencies (cycles / deg)		
	Model	Cortex	References
simple cells	range: 1.6 – 9.8 mean / med: 3.7 / 2.8	bulk ≈ 1.0 – 4.0 mean: ≈ 2.2 range: ≈ 0.5 – 8.0	[DeValois et al., 1982a])
complex cells	range: 1.8 – 7.8 mean / med: 3.9 / 3.2	bulk ≈ 2.0 – 5.6 mean: 3.2 range ≈ 0.5 – 8.0	

	Frequency bandwidth at 50% amplitude (cycles / deg)		
	Model	Cortex	References
simple cells	range: 1.1 – 1.8 med: ≈ 1.45	bulk ≈ 1.0 – 1.5 med: ≈ 1.45 range ≈ 0.4 – 2.6	[DeValois et al., 1982a]
complex cells	range: 1.5 – 2.0 med: 1.6	bulk ≈ 1.0 – 2.0 med: 1.6 range ≈ 0.4 – 2.6	

	Frequency bandwidth at 71% amplitude (index)		
	Model	Cortex	References
simple cells	range: 44 – 58 med: 55	bulk ≈ 40 – 70	[Schiller et al., 1976d]
complex cells	range 40 – 50 med. 48	bulk ≈ 40 – 60	

	Orientation bandwidth at 50% amplitude (octaves)		
	Model	Cortex	References
simple cells	range: 38° – 49° med: 44°	—	[DeValois et al., 1982b]
complex cells	range: 27° – 33° med: 43°	bulk ≈ 20° – 90° med: 44°	

	Orientation bandwidth at 71% amplitude (octaves)		
	Model	Cortex	References
simple cells	range: 27° – 33° med: 30°	bulk ≈ 20° – 70°	[Schiller et al., 1976c]
complex cells	range: 27° – 33° med: 31°	bulk ≈ 20° – 90°	

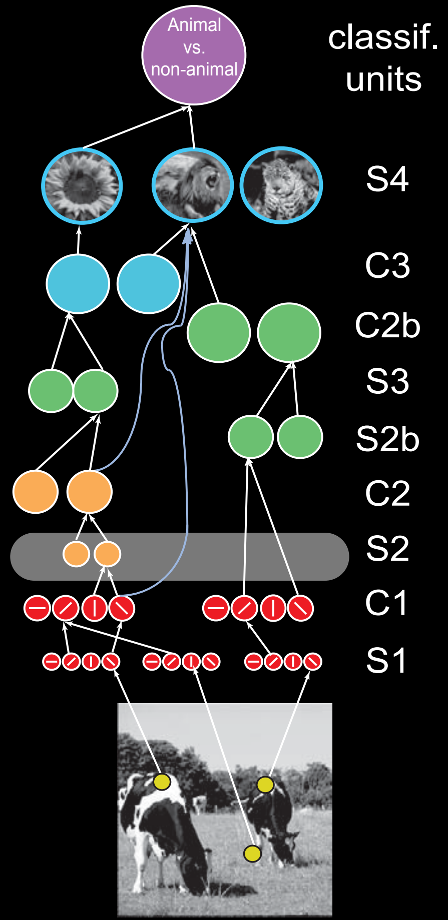


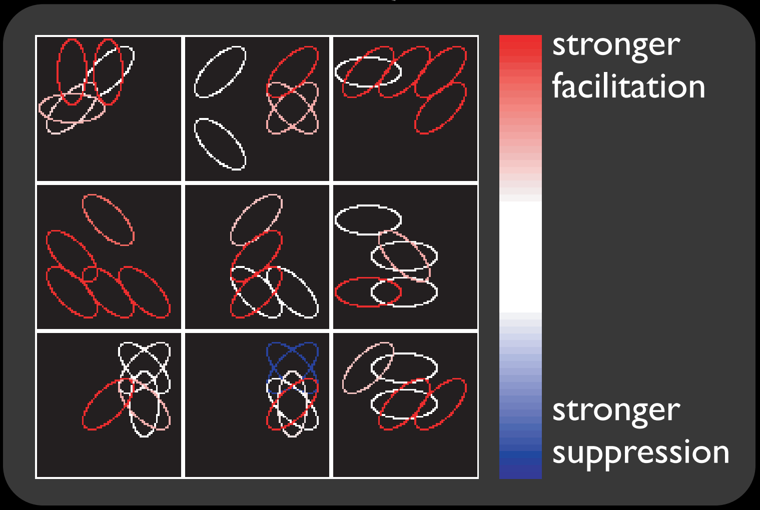
(Serre & Riesenhuber 2004)

<< back fwd >>

S2 units

- ✦ Features of moderate complexity (n~1,000 types)
- ✦ Combination of V1-like complex units at different orientations
  - Synaptic weights ***w*** learned from natural images
  - 5-10 subunits chosen at random from all possible afferents (~100-1,000)

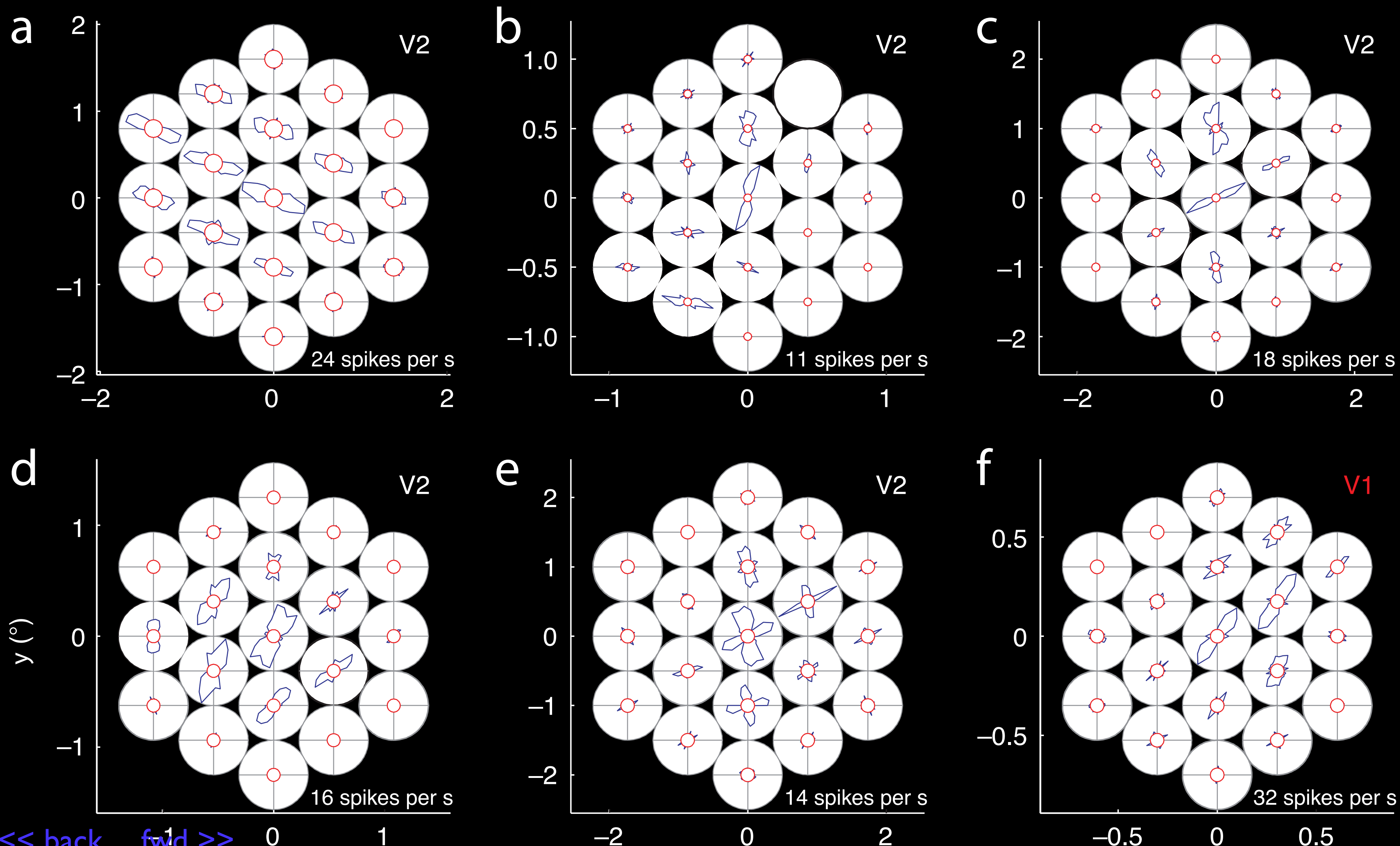




Nature Neuroscience - 10, 1313 - 1321 (2007) / Published online: 16 September 2007 | doi:10.1038/nn1975

Neurons in monkey visual area V2 encode combinations of orientations

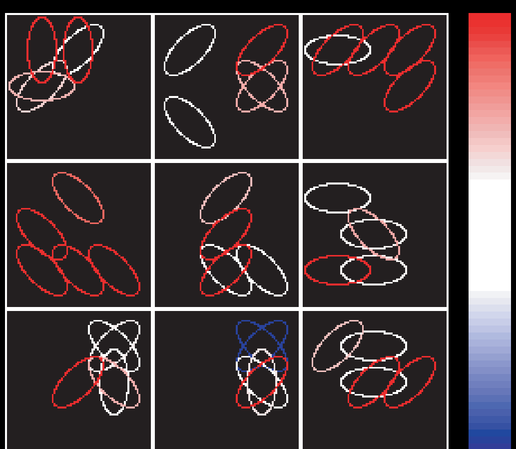
Akiyuki Anzai, Xinmiao Peng & David C Van Essen

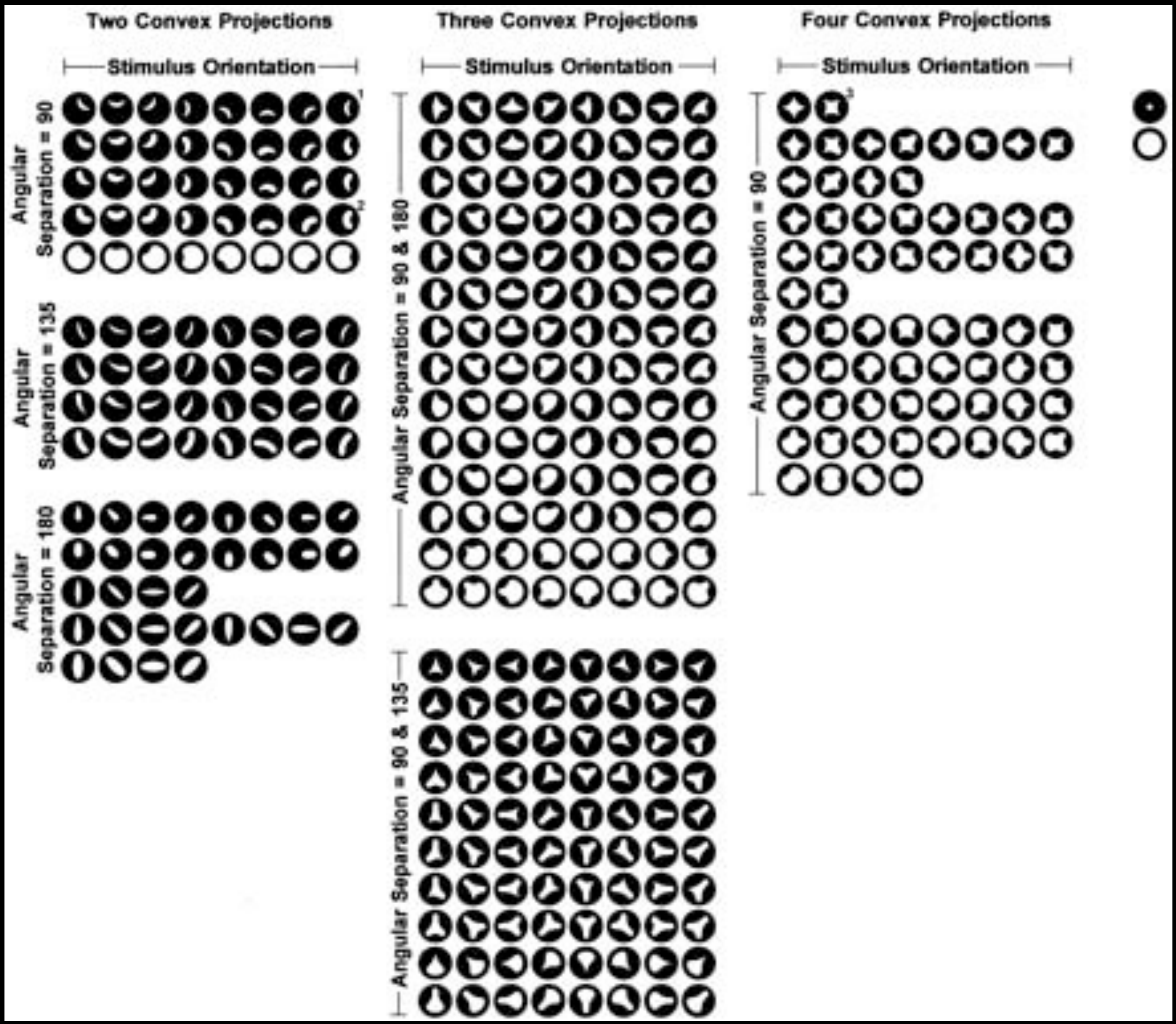


<< back fwd >>

Comparison w| V4

Tuning for curvature and boundary conformations?

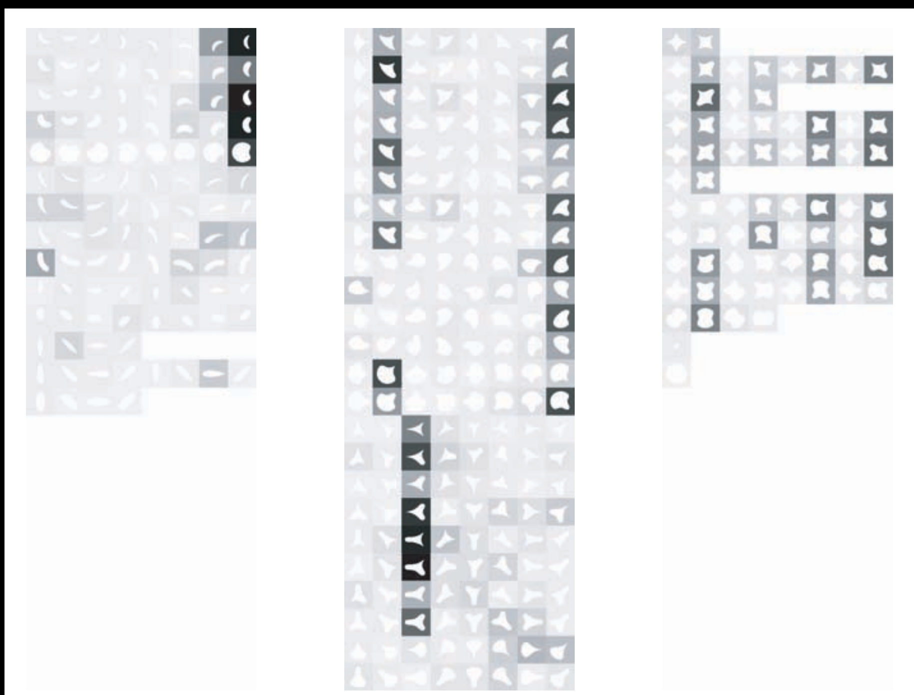




(Pasupathy & Connor 2001)

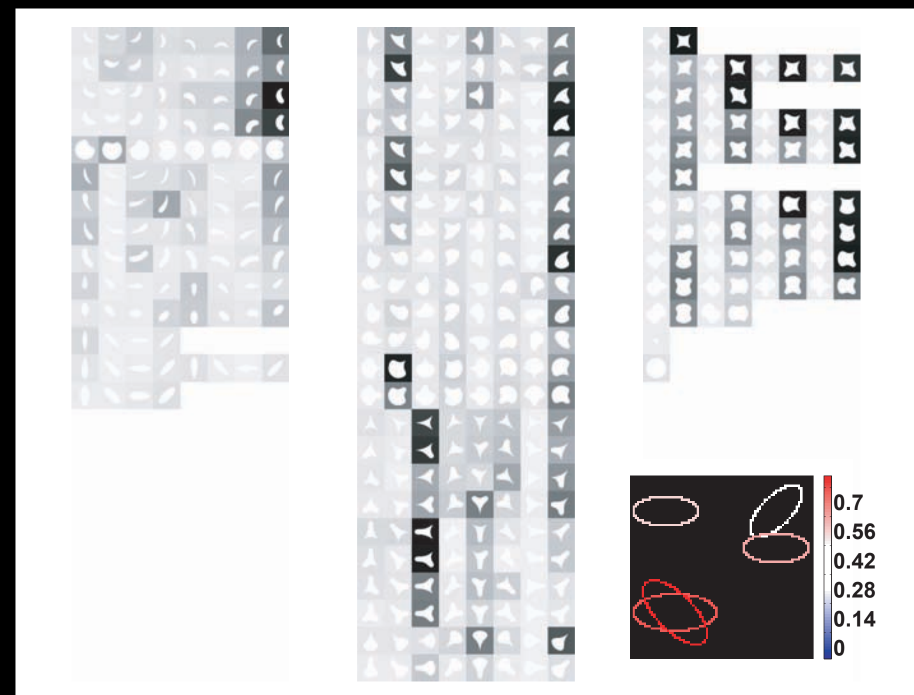
No parameter fitting!

V4 neuron tuned to boundary conformations



modified from (Pasupathy & Connor 1999)

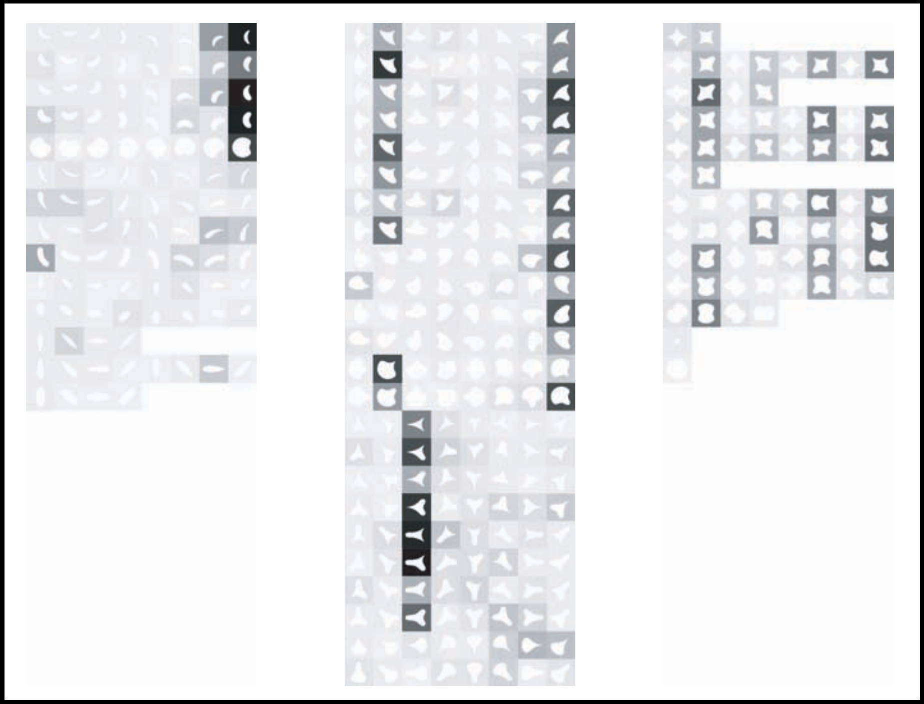
Most similar model C2 unit



(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

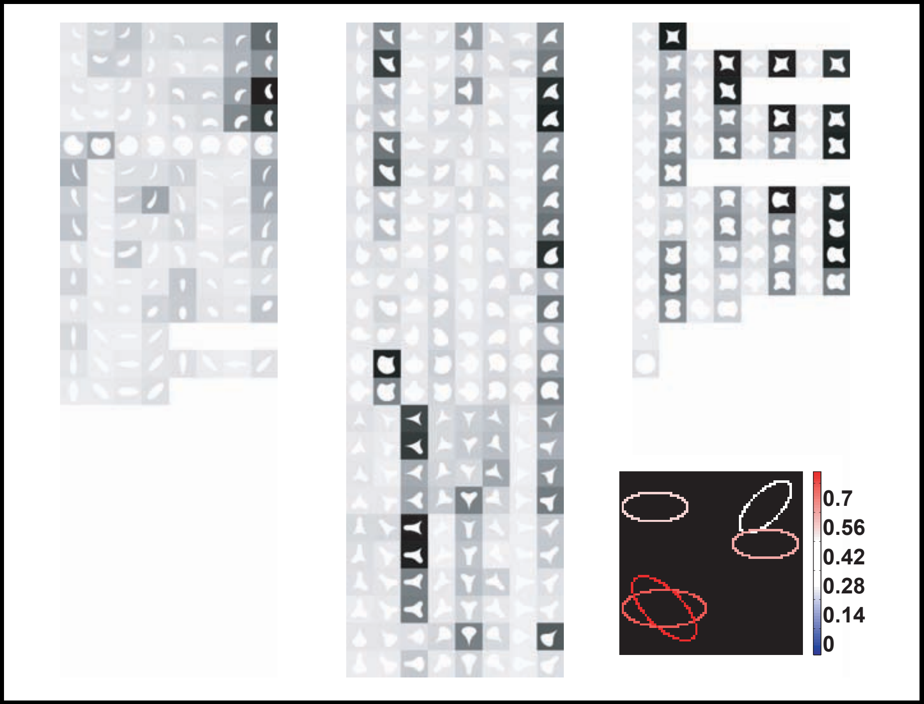
No parameter fitting!

V4 neuron tuned to boundary conformations



modified from (Pasupathy & Connor 1999)

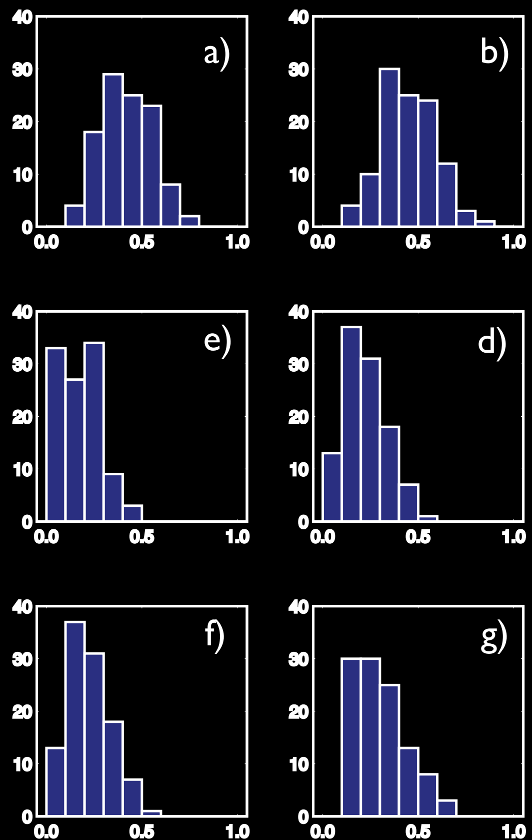
Most similar model C2 unit



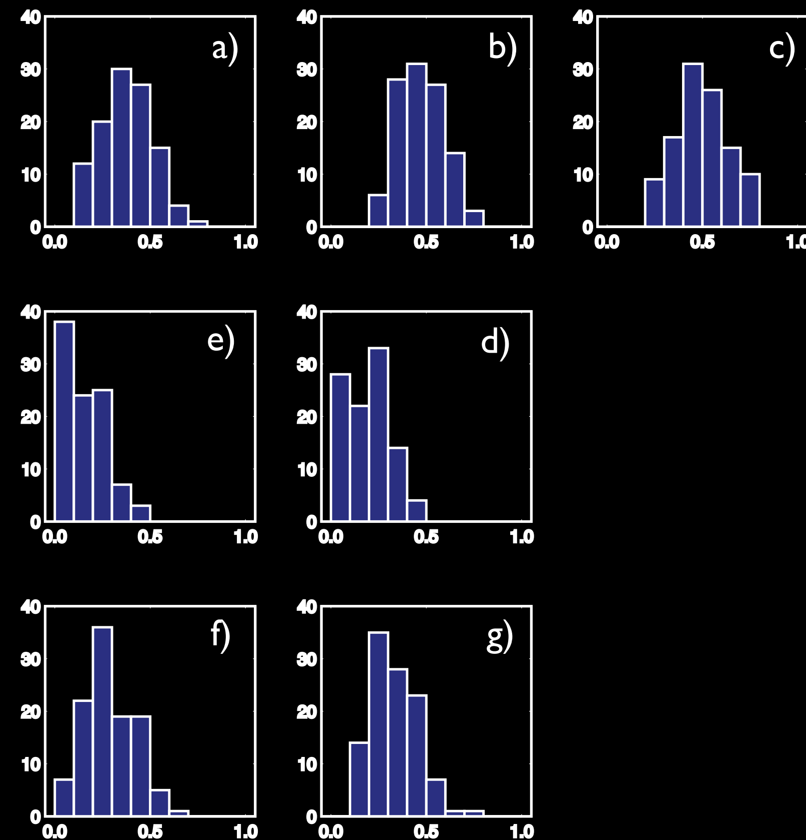
ρ = 0.78

(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

Population of 109 V4 neurons



Population of 109 model C2 units



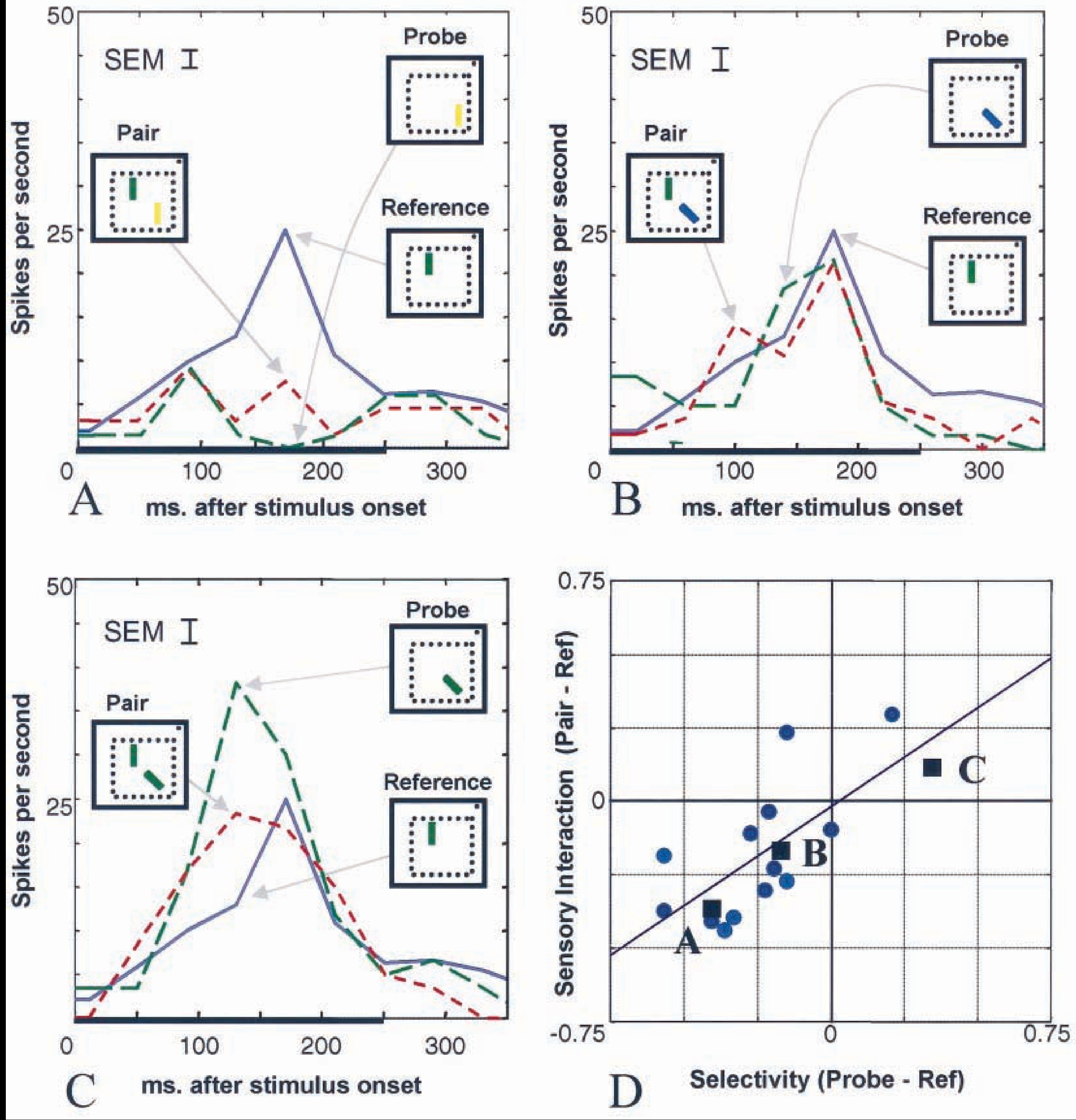
	Tuning functions	Model units	V4 neurons
a)	2-D boundary conformation	0.38	0.41
b)	4-D boundary conformation	0.47	0.46
c)	2-Gaussian boundary conformation	0.50	0.46
d)	Edge orientation	0.11	0.15
e)	Edge orientation + contrast polarity	0.18	0.21
f)	2-D axial orientation × elongation tuning functions	0.28	0.18
g)	3-D axial orientation × length × width tuning functions	0.32	0.28

(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

More comparison w| V4

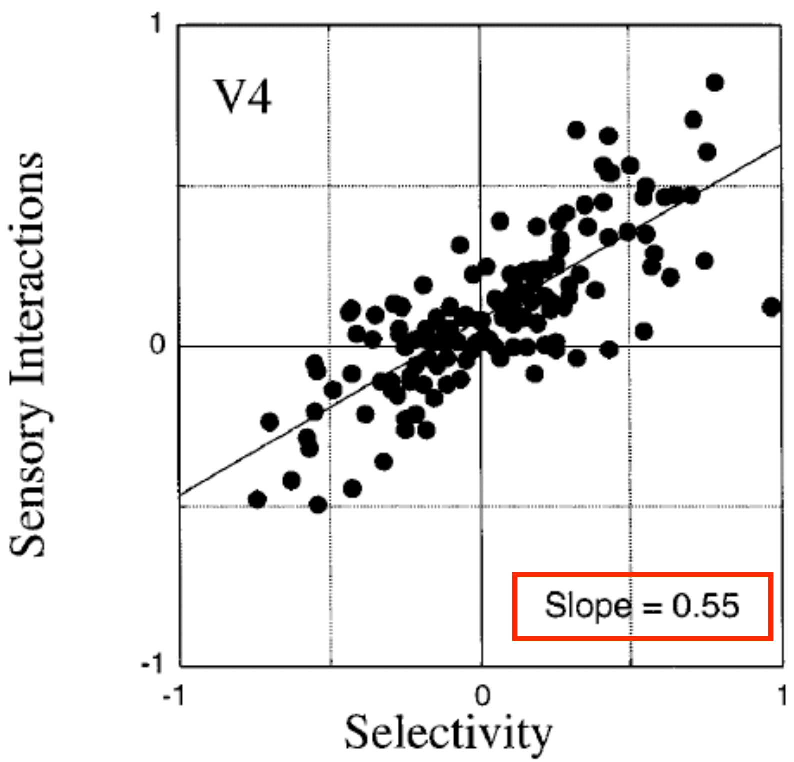
"average" effect in model C2 units?

(Reynolds Chelazzi & Desimone 1999)



V4 neurons

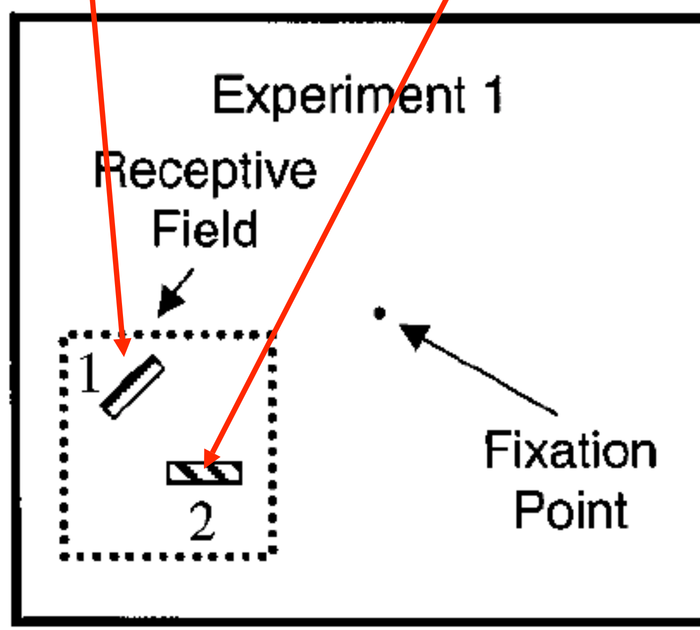
(with attention directed away from receptive field)



(Reynolds et al 1999)

Reference (fixed)

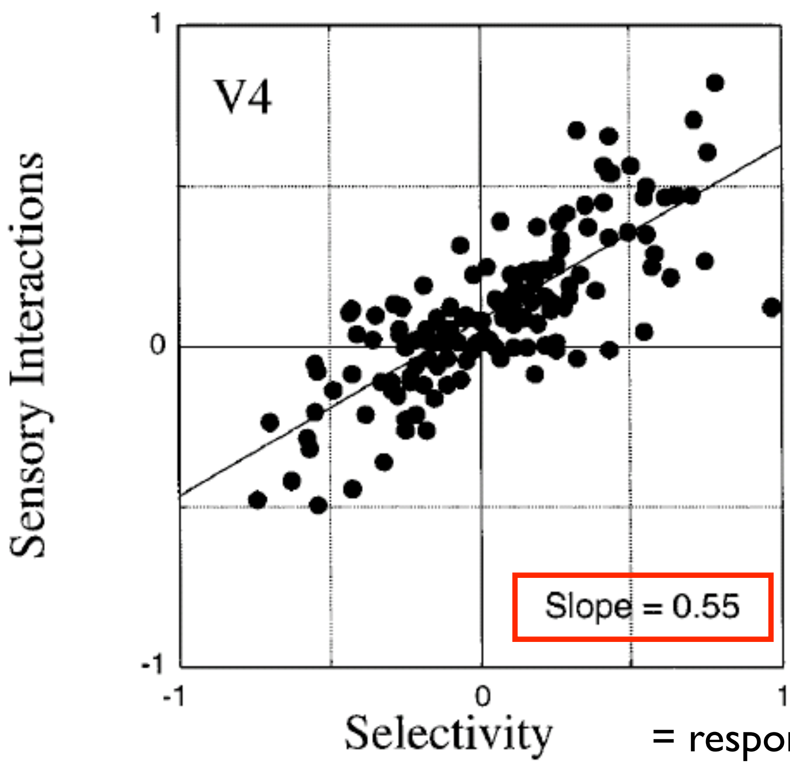
Probe (varying)



(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

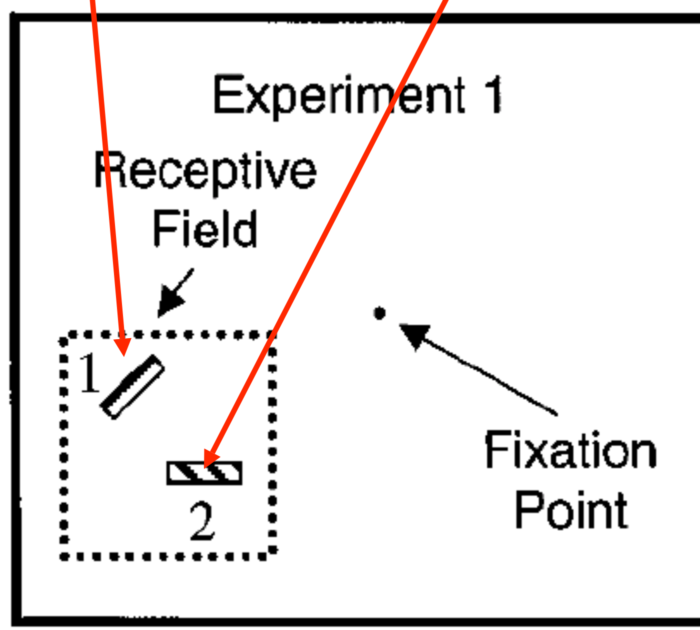
V4 neurons

(with attention directed away from receptive field)



Reference (fixed)

Probe (varying)



= response(probe) –response(reference)

(Reynolds et al 1999)

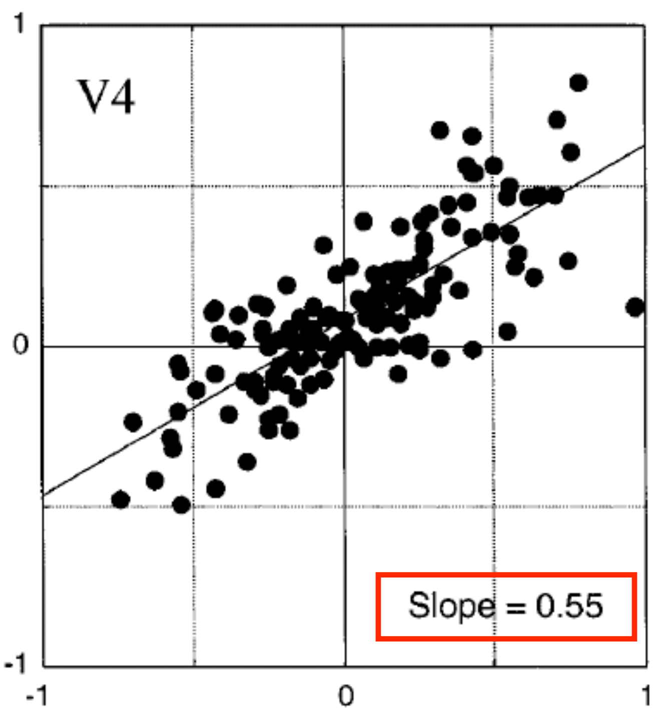
<< back fwd >>

(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

V4 neurons

(with attention directed away from receptive field)

Sensory Interactions = resp (pair) −resp (reference)

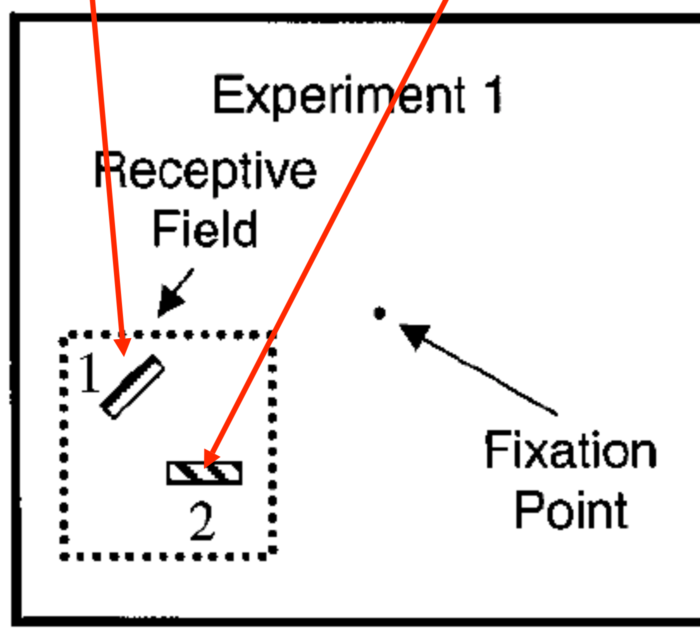


Selectivity = response(probe) −response(reference)

(Reynolds et al 1999)

Reference (fixed)

Probe (varying)



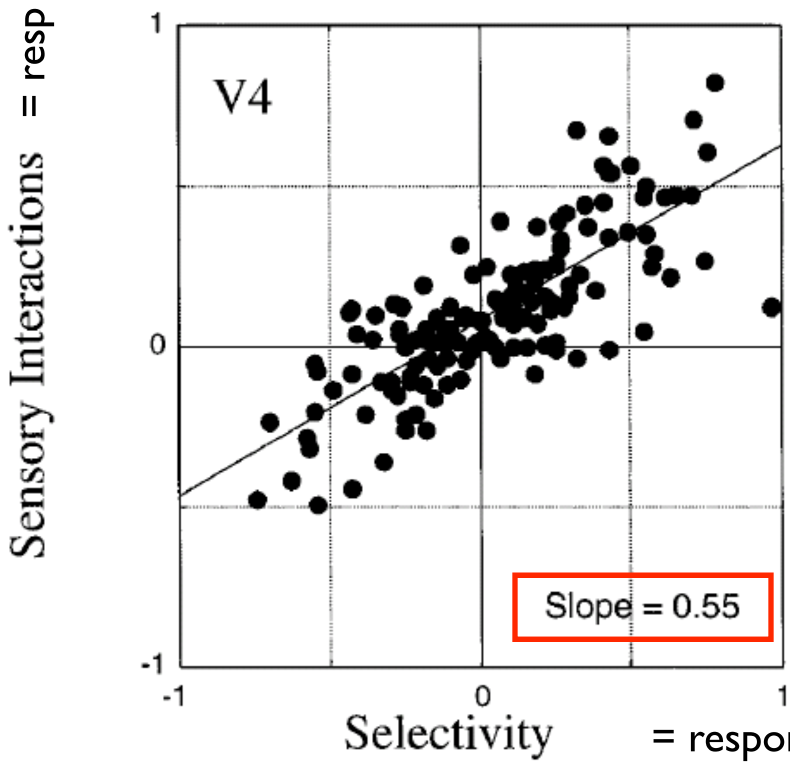
(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

Prediction: Response of the pair is predicted to fall between the responses elicited by the stimuli alone

Sensory Interactions = resp (pair) −resp (reference)

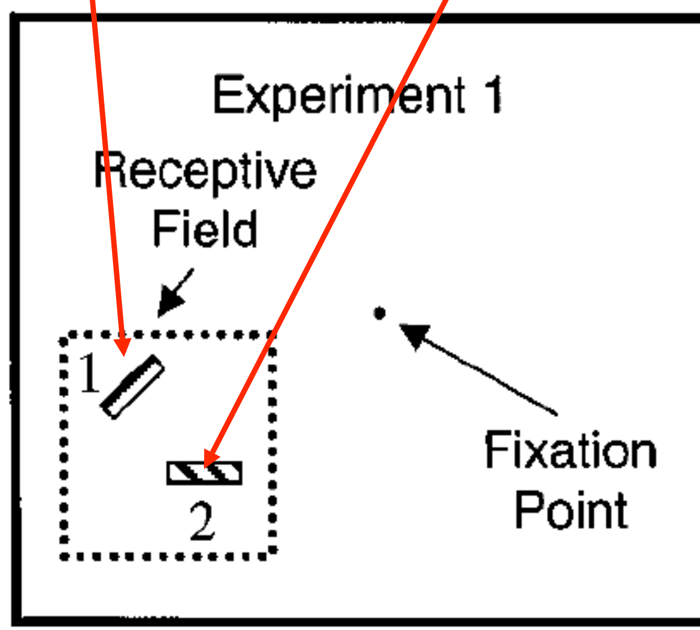
V4 neurons

(with attention directed away from receptive field)



Reference (fixed)

Probe (varying)



= response(probe) −response(reference)

(Reynolds et al 1999)

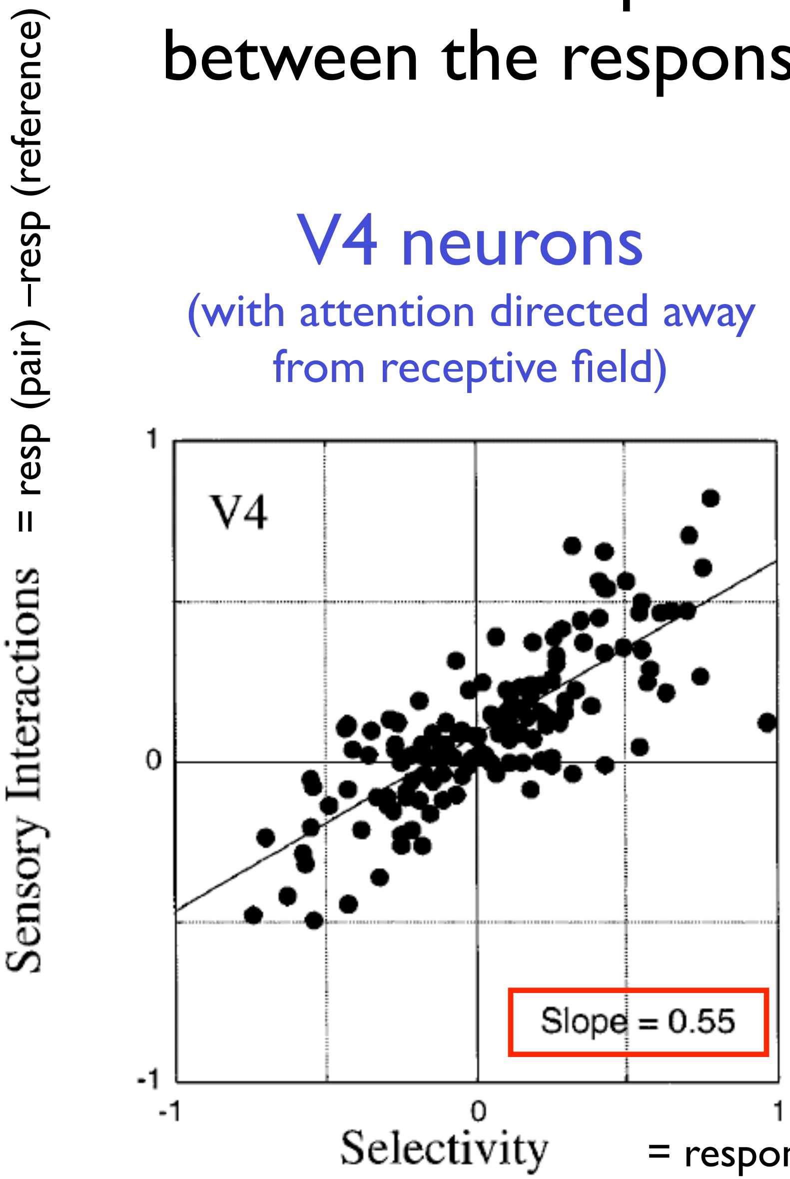
<< back fwd >>

(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

Prediction: Response of the pair is predicted to fall between the responses elicited by the stimuli alone

V4 neurons

(with attention directed away from receptive field)



= response(probe) –response(reference)

(Reynolds et al 1999)

(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

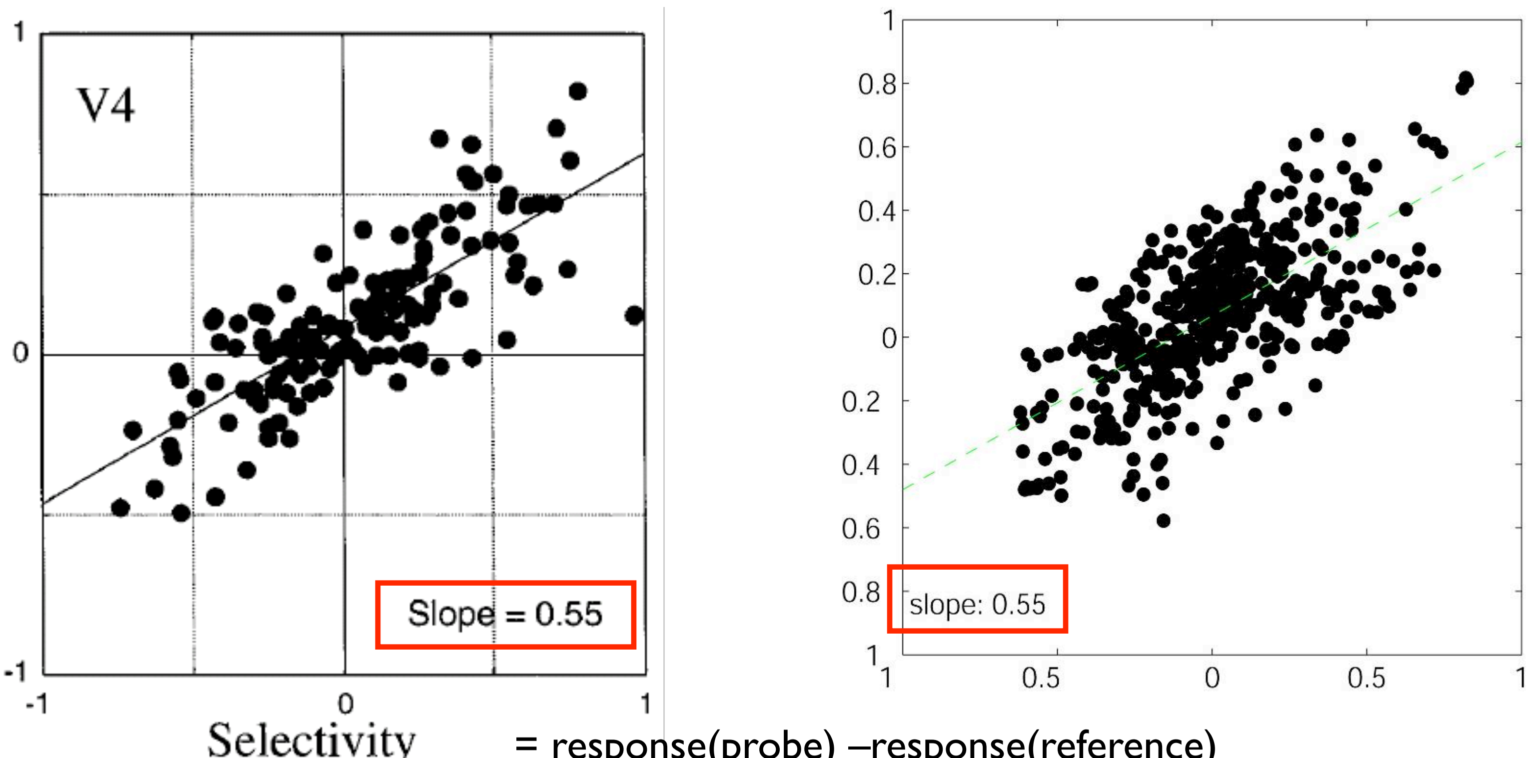
Prediction: Response of the pair is predicted to fall between the responses elicited by the stimuli alone

Sensory Interactions = resp (pair) –resp (reference)

V4 neurons

(with attention directed away from receptive field)

C2 units

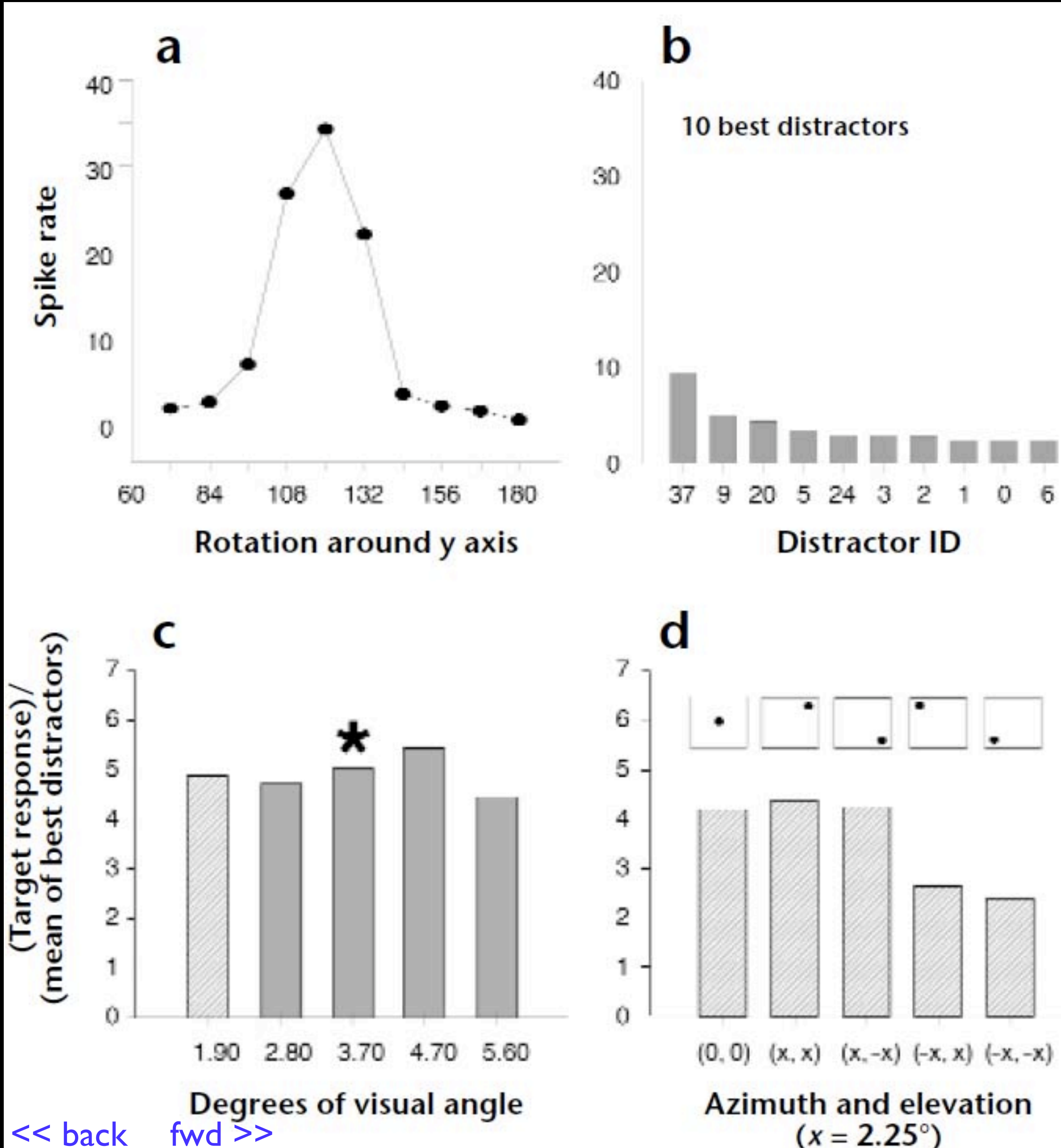


= response(probe) –response(reference)

(Reynolds et al 1999)

(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

Example: IT

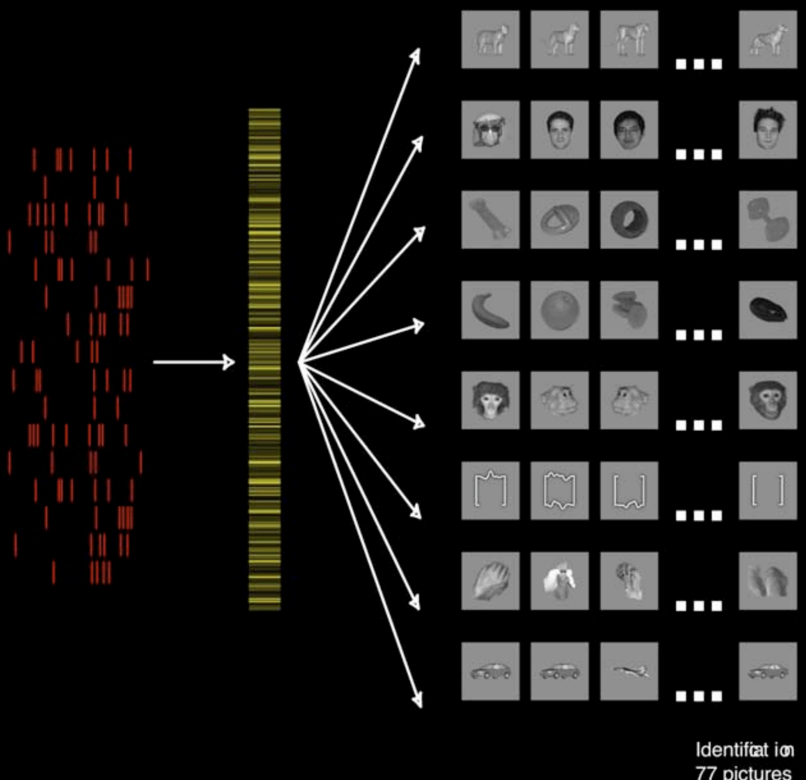


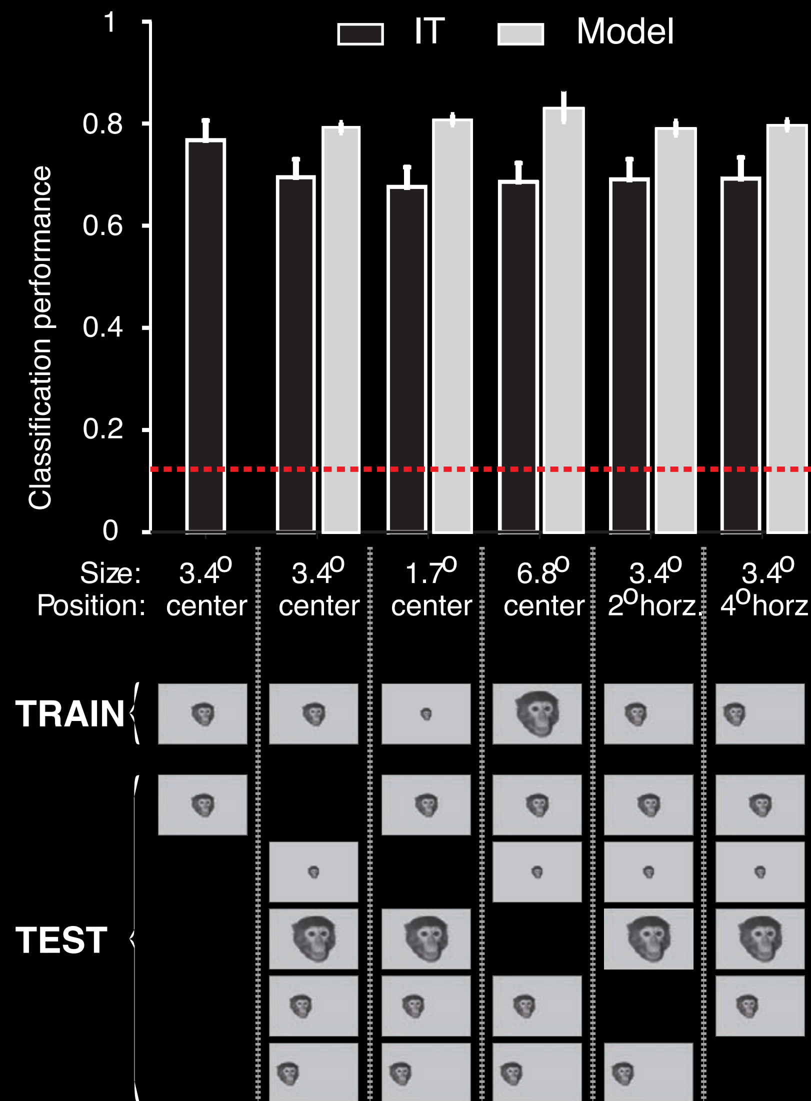
Logothetis Pauls Poggio 1995

<< back fwd >>

Agreement w| IT Readout data

(Hung Kreiman Poggio DiCarlo 2005)

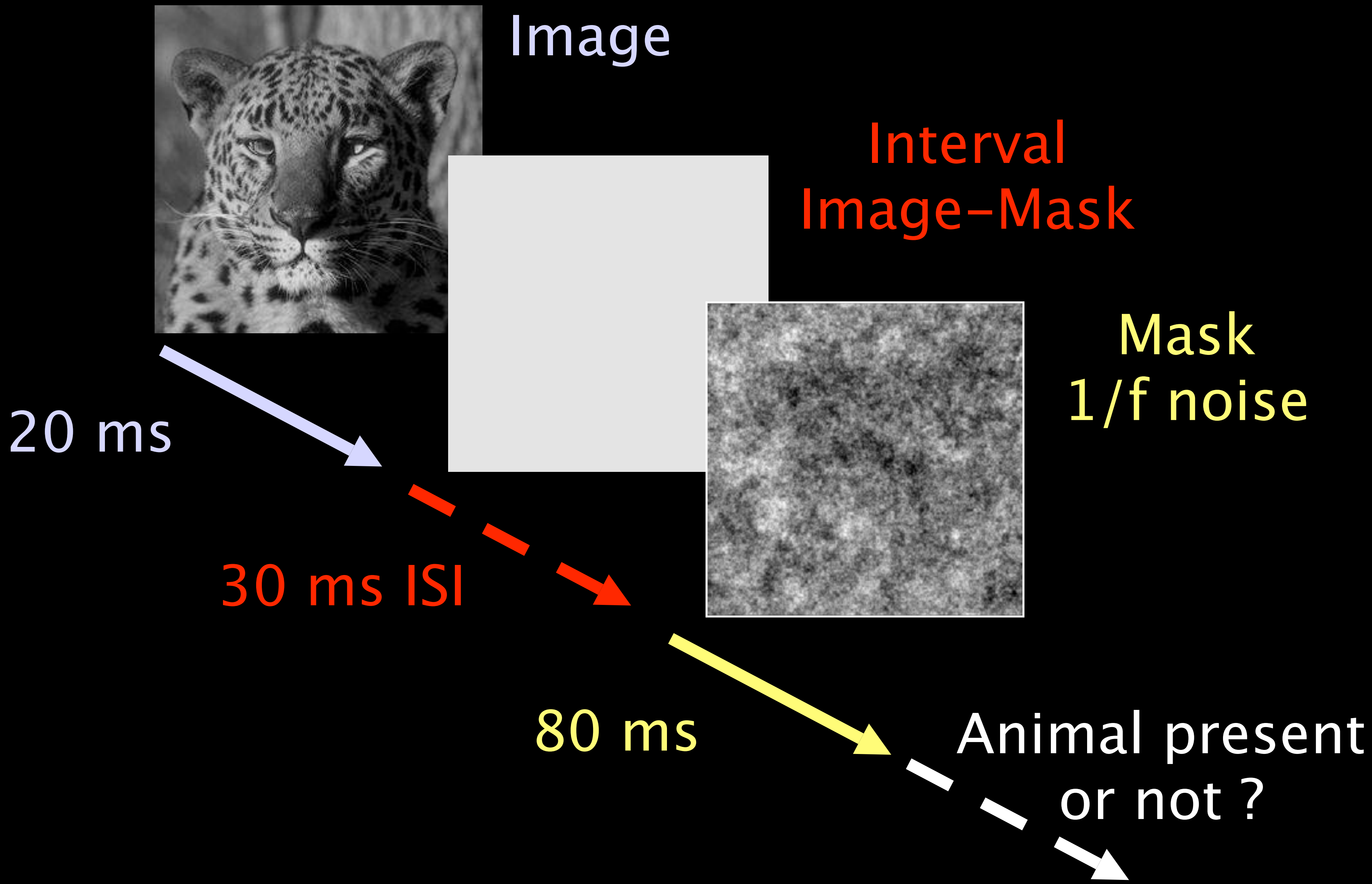




(Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005)

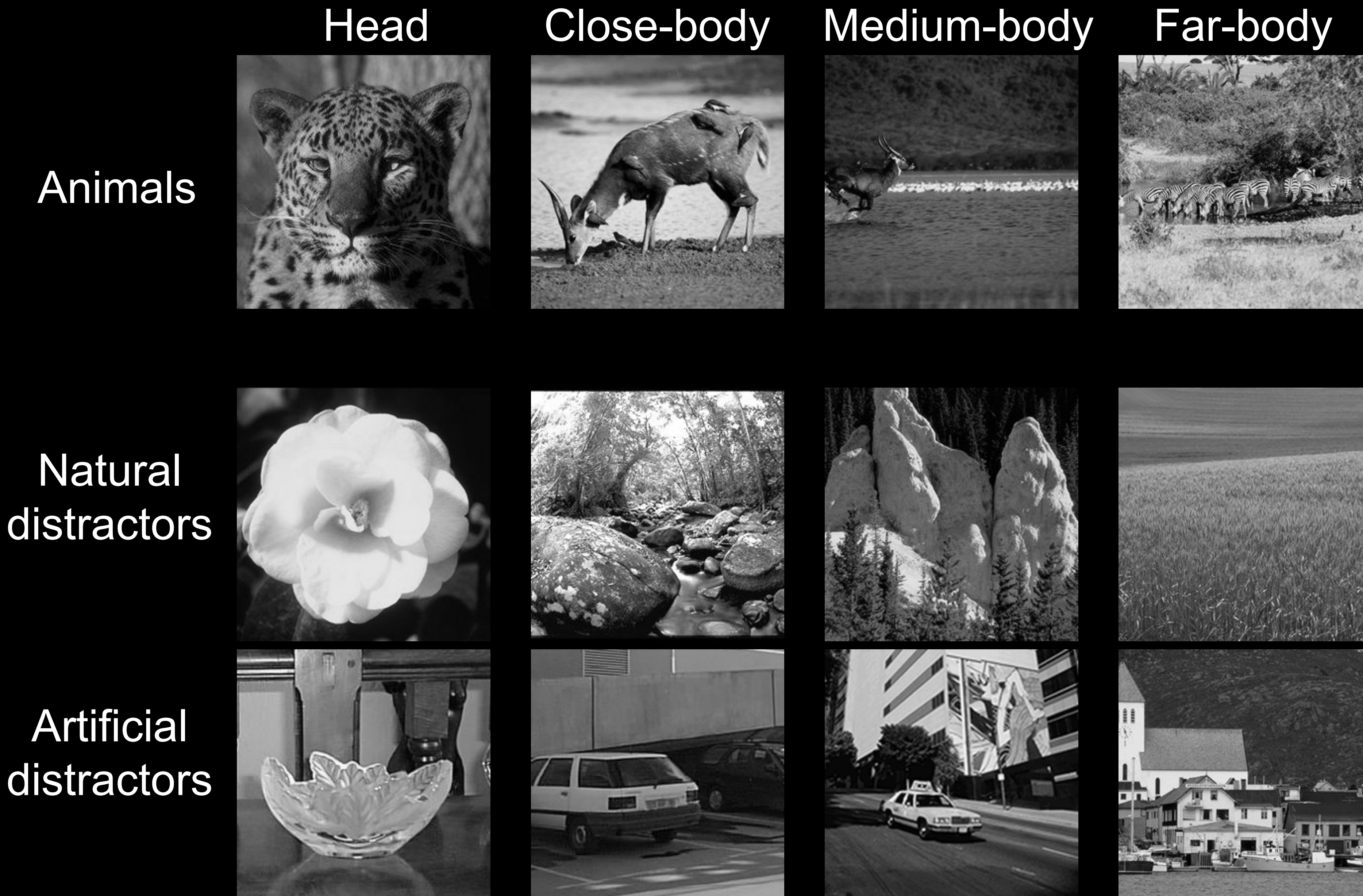
<< back fwd >>

How does the model compare to human observers?

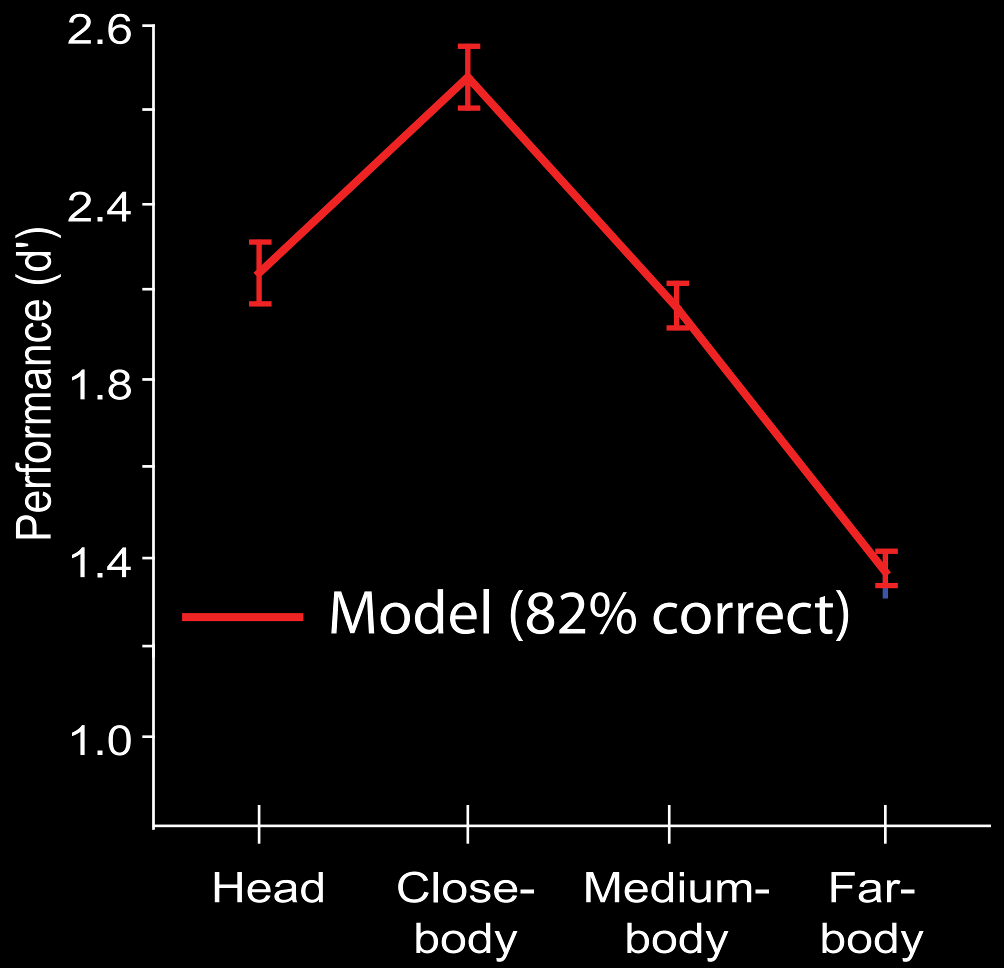


(Thorpe et al 1996; Van Rullen & Koch 2003; Bacon-Mace et al 2005)

Show demo



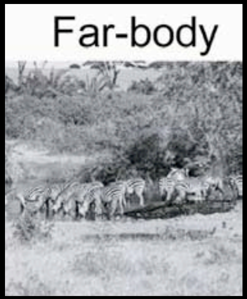
Database collected by Torralba & Oliva (2003)

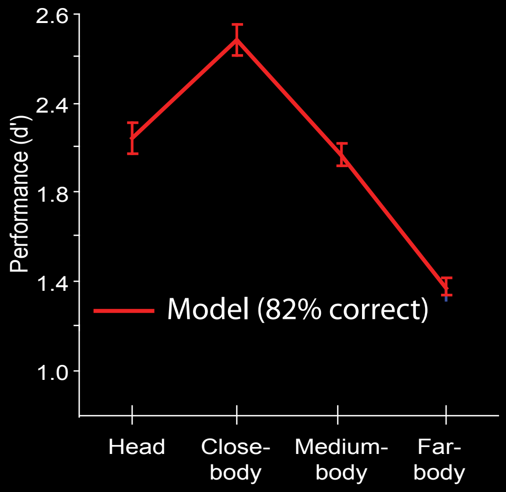


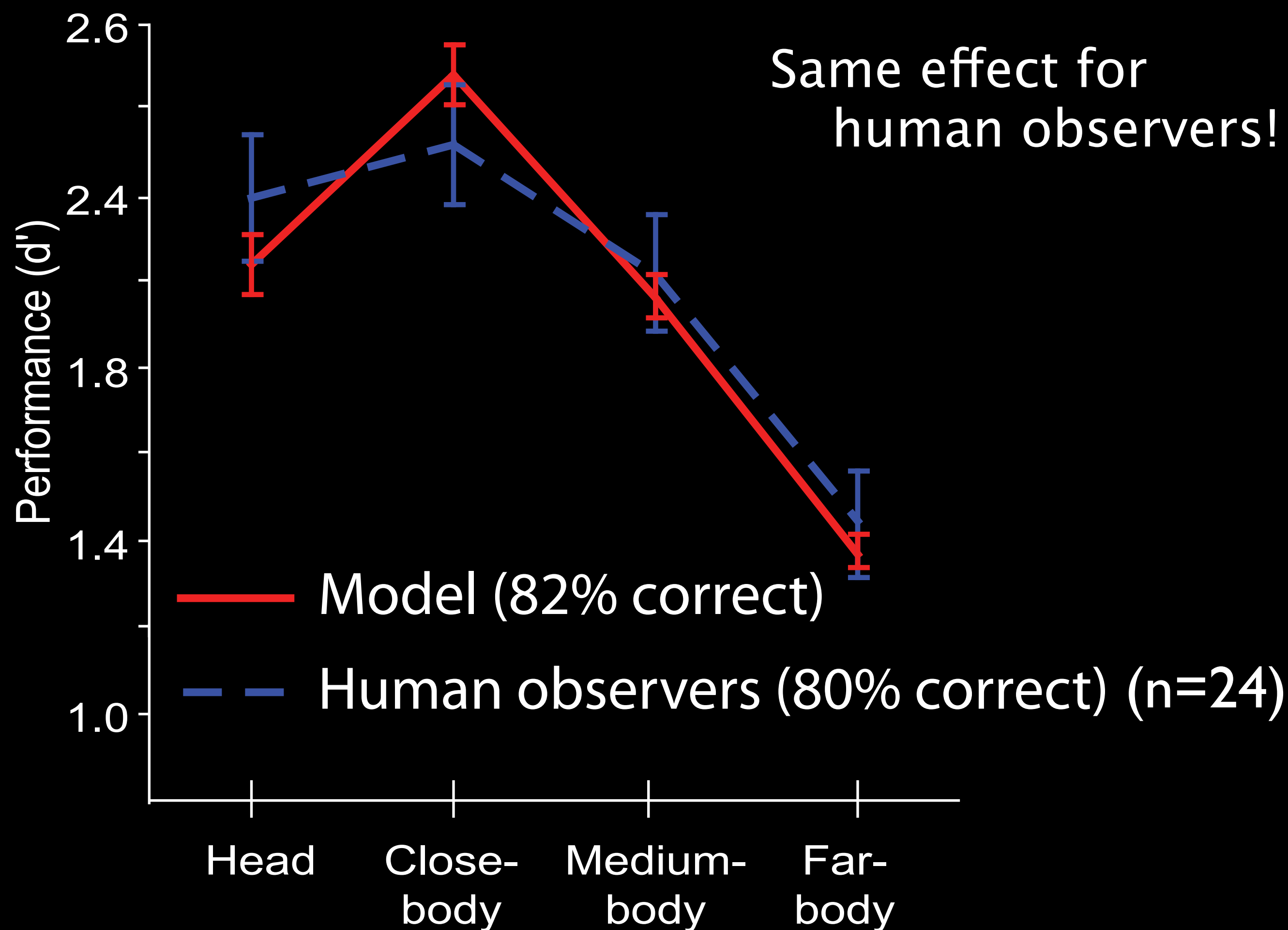
(Serre Oliva & Poggio 2007)

“Clutter effect”

- High performance (~90%) when
  - maximal amount of information present
  - in the absence of clutter
- Performance decreases (~74%) with increasing amount of clutter
- Limitation of feedforward model compatible with decrease in response in V4 (Reynolds Chelazzi & Desimone 1999) and IT in the presence of clutter (Zoccolan, Cox, DiCarlo, 2005; Zoccolan, Kouh, Poggio, DiCarlo, in sub; Rolls, Aggelopoulos, Zheng, 2003)

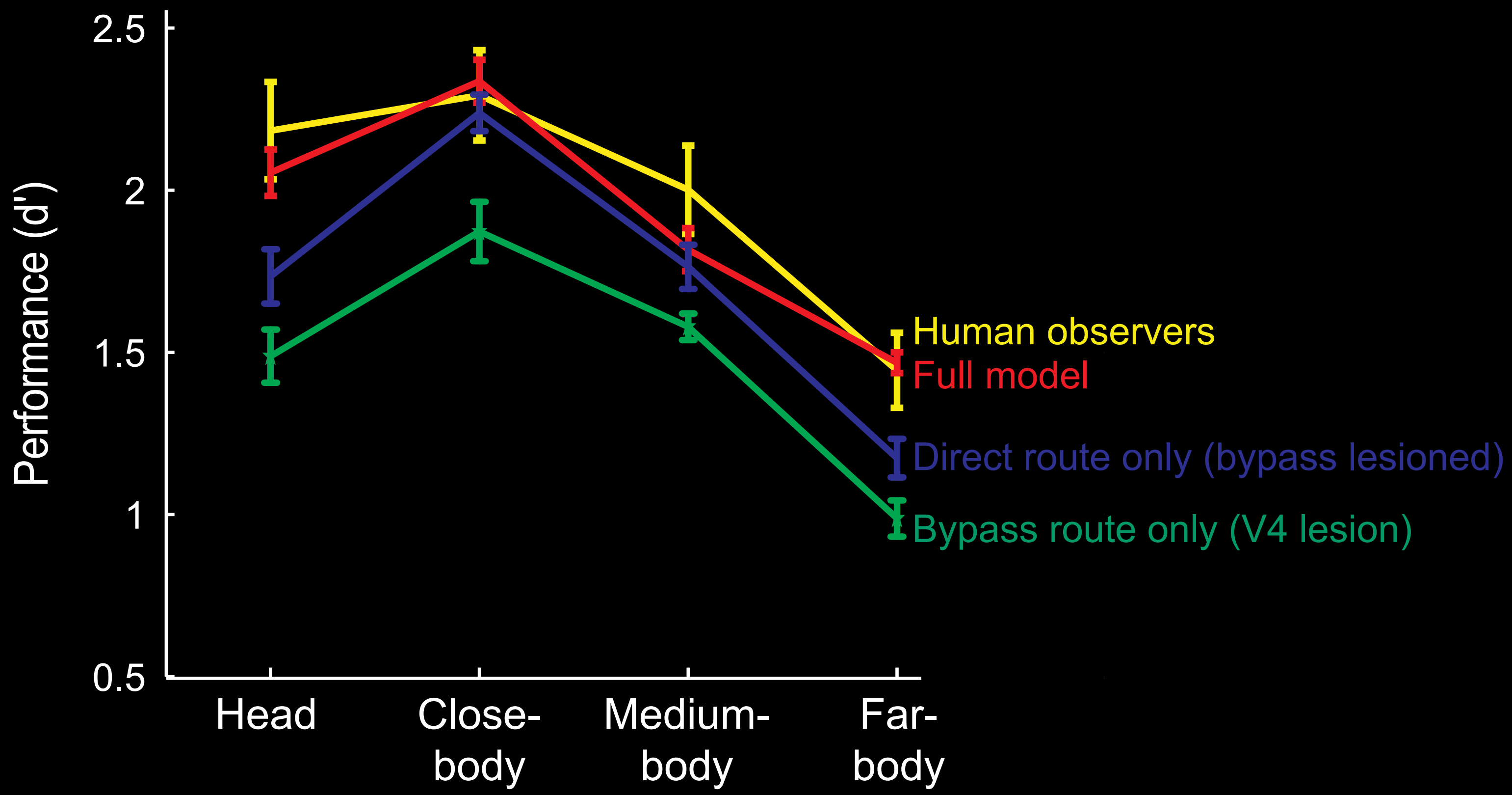






Same effect for human observers!

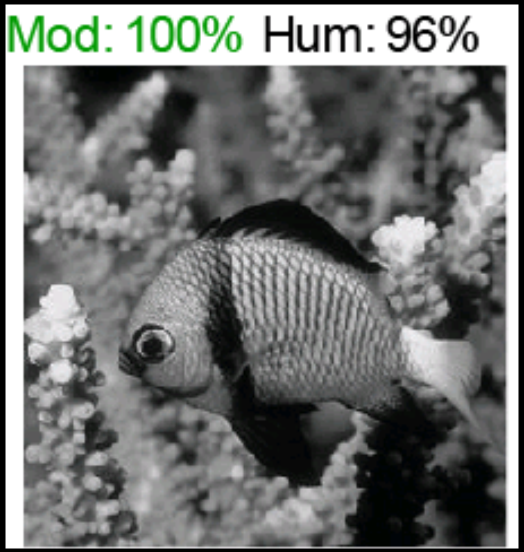
(Serre Oliva & Poggio 2007)



2.5
2
1.5
1
0.5
Performance (d')
Head
Close-body
Medium-body
Far-body
Human observers
Full model
Direct route only (bypass lesioned)
Bypass route only (V4 lesion)

Further comparisons

- ✦ Image-by-image correlation:
  - Heads: $\rho=0.71$
  - Close-body: $\rho=0.84$
  - Medium-body: $\rho=0.71$
  - Far-body: $\rho=0.60$



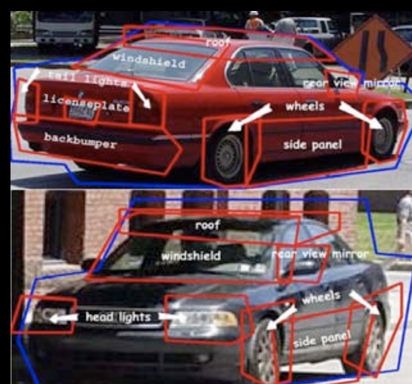
- ✦ Model predicts level of performance on rotated images (90 deg and inversion)

<< back

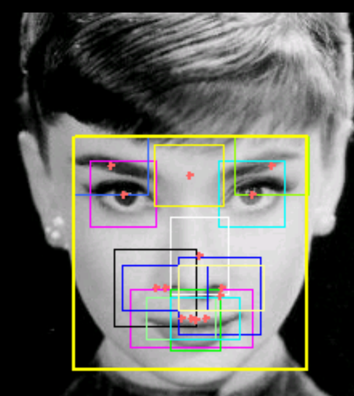
(Serre, Oliva & Poggio, PNAS 2007)

How does the model compare to state-of-the-art machine vision systems?

Datasets	Bench.	Model
MIT-CBCL Faces	90.4	95.9
MIT-CBCL Cars	75.4	95.1



(Leung 2004)



(Heisele Serre Pontil Vetter & Poggio 2002)

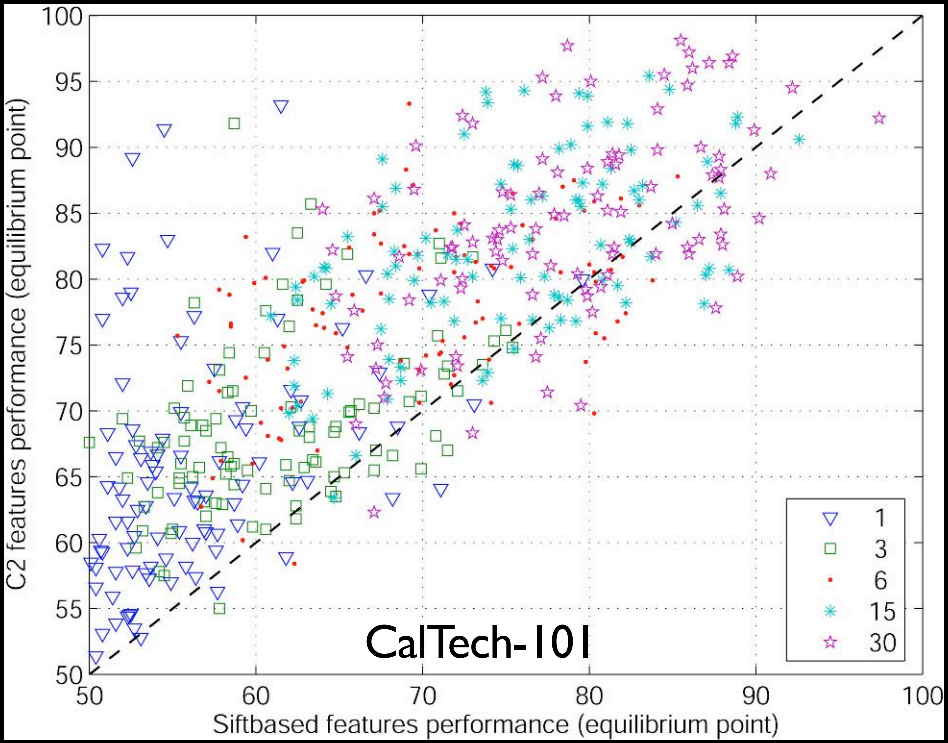
(Serre Wolf & Poggio 2005)

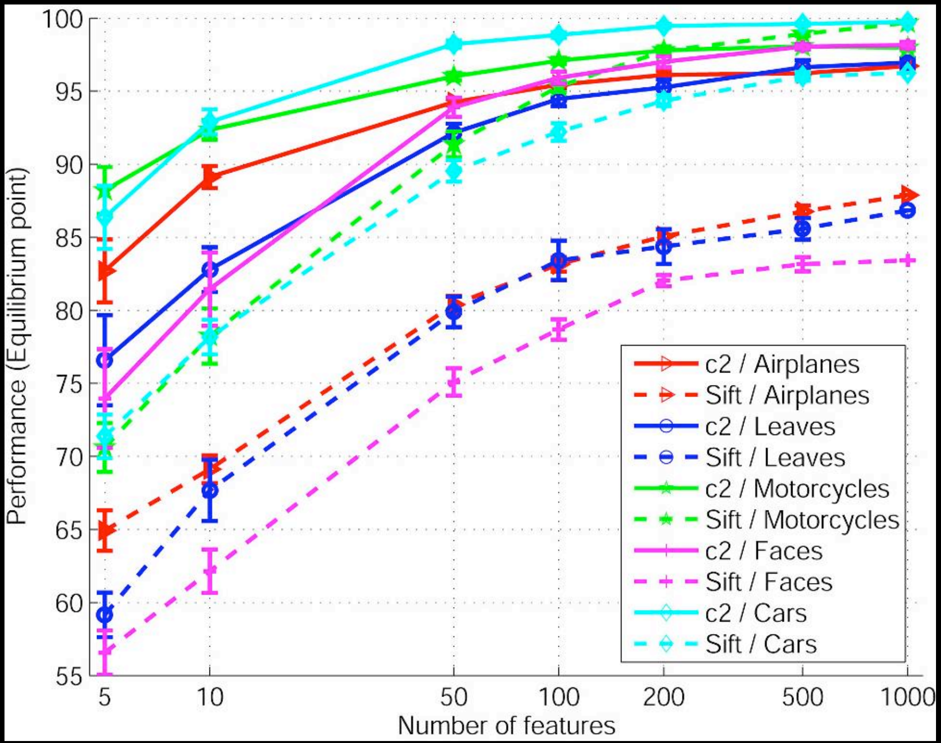
Datasets	Bench.*	Model
CalTech Leaves	84.0	97.0
CalTech Cars	84.8	99.7
CalTech Faces	96.4	98.2
CalTech Airplanes	94.0	96.7
CalTech Motorcycles	95.0	98.0

*constellation model by Perona and colleagues

(Serre Wolf & Poggio 2005)

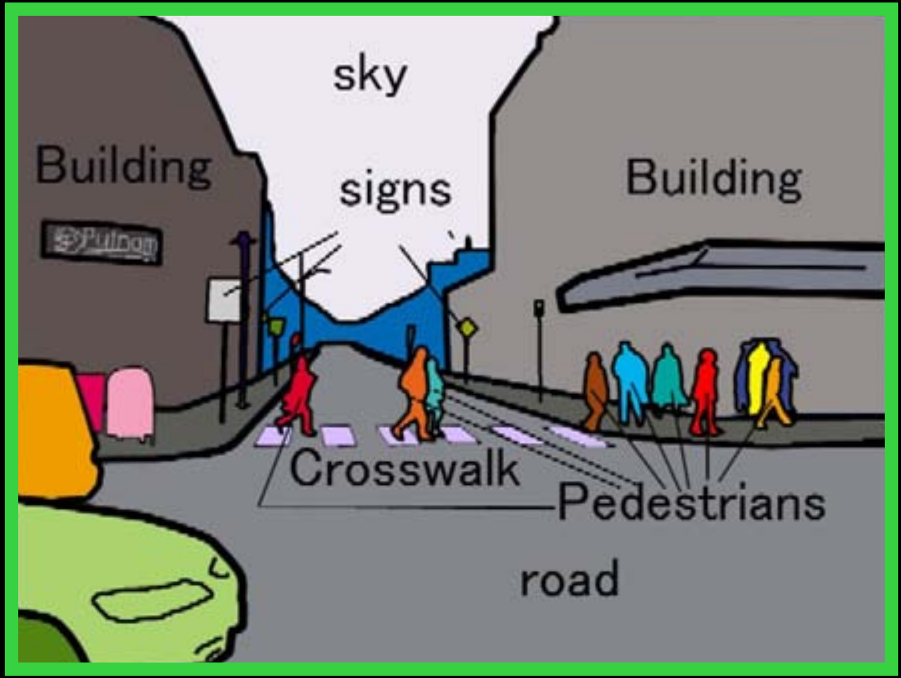
Comparison w| SIFT features





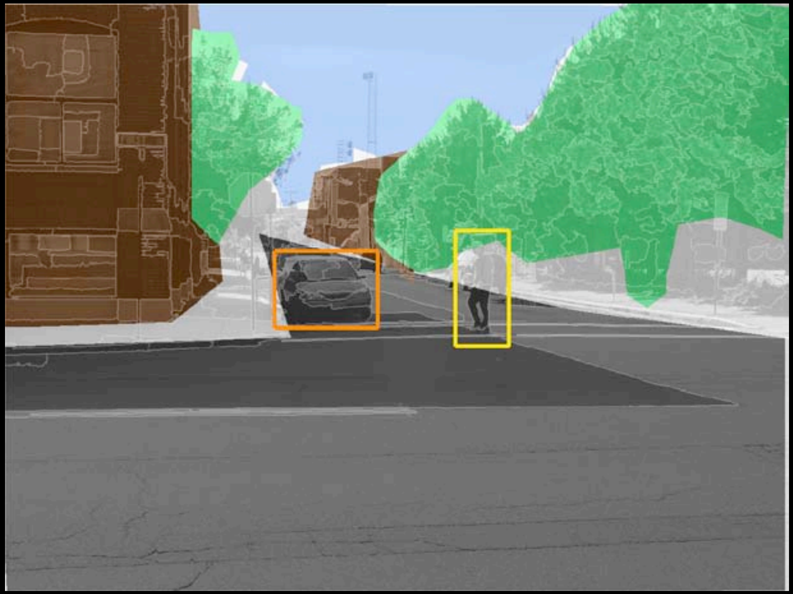
(Serre Wolf & Poggio 2005)

The street scene project



Source: Bileschi & Wolf

The StreetScenes Database

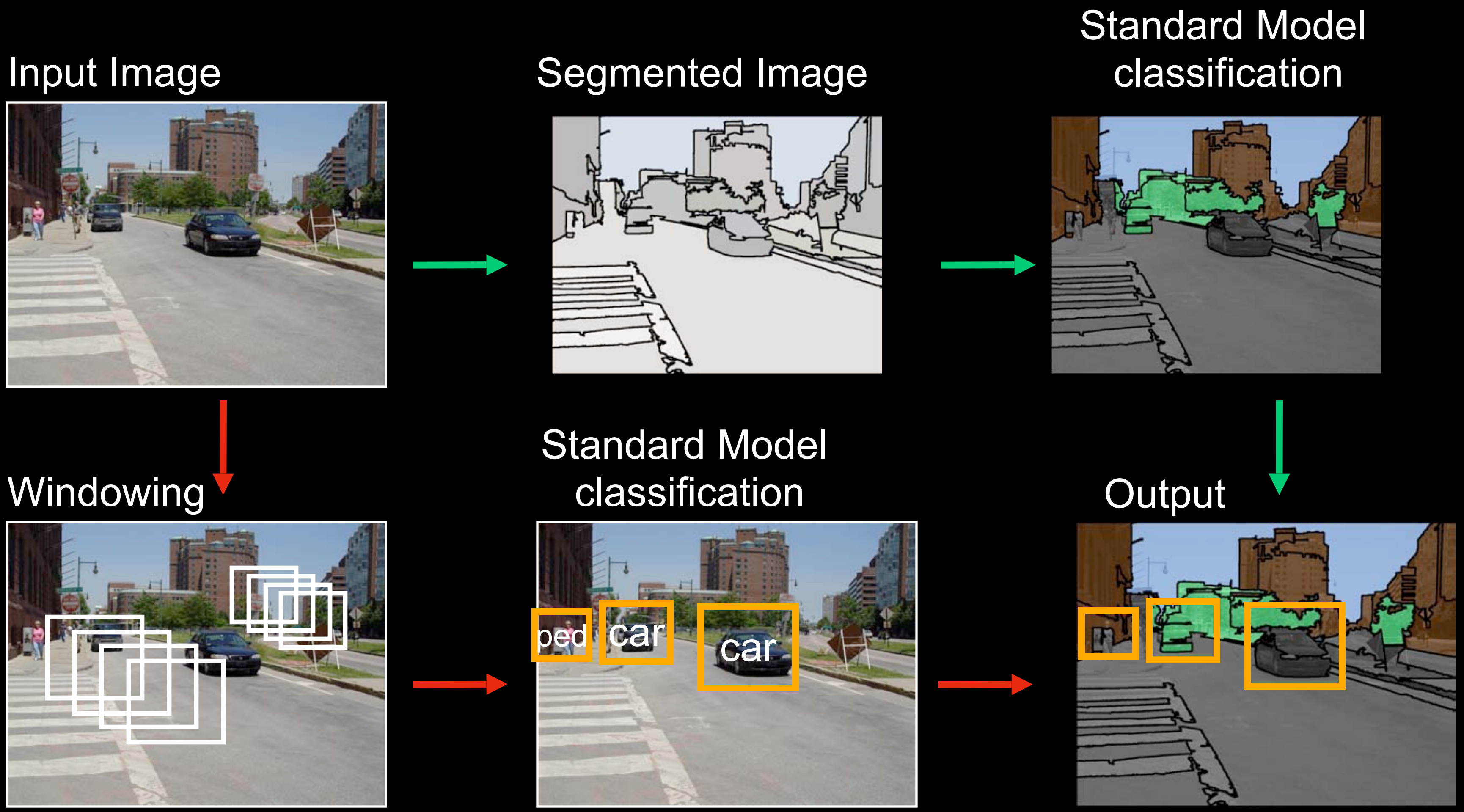


3,547 Images, all taken with the same camera, of the same type of scene, and hand labeled with the same objects, using the same labeling rules.

Object	car	pedestrian	bicycle	building	tree	road	sky
# Labeled Examples	5799	1449	209	5067	4932	3400	2562

http://cbcl.mit.edu/software-datasets/streetscenes/

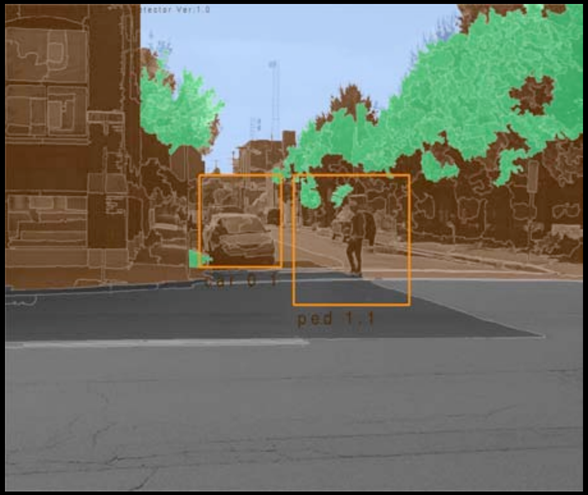
The system

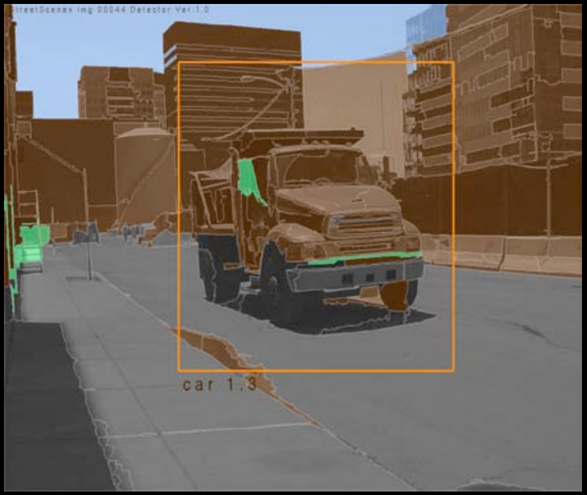


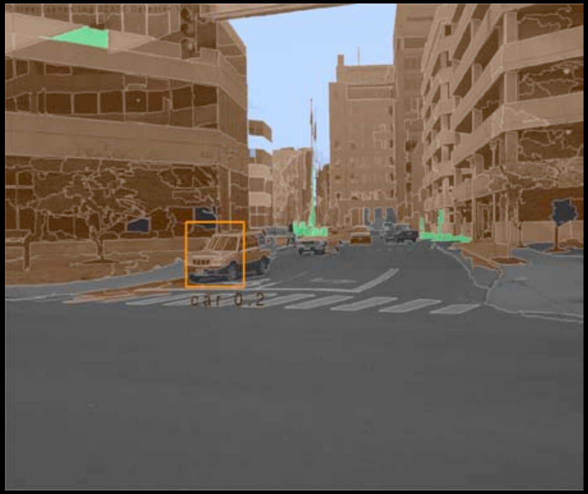
Texture-based objects pathway (e.g., trees, road, sky, buildings)

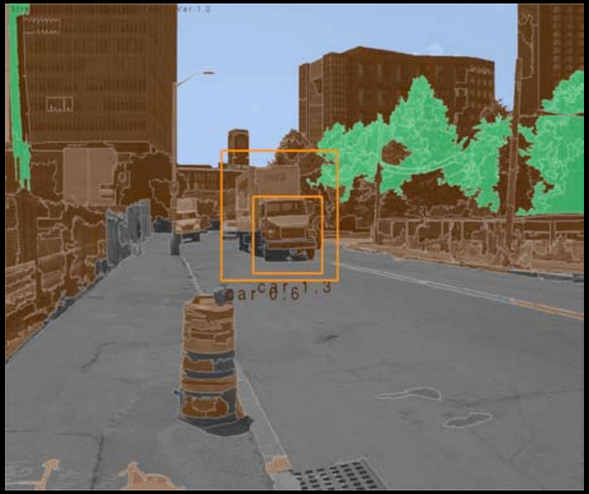
Rigid-objects pathway (e.g., pedestrians, cars)

Examples

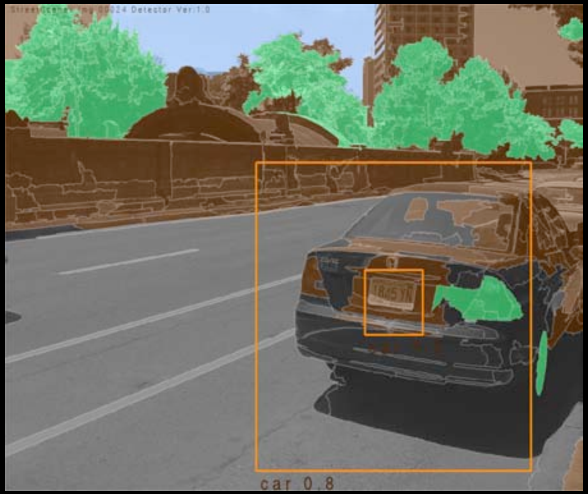


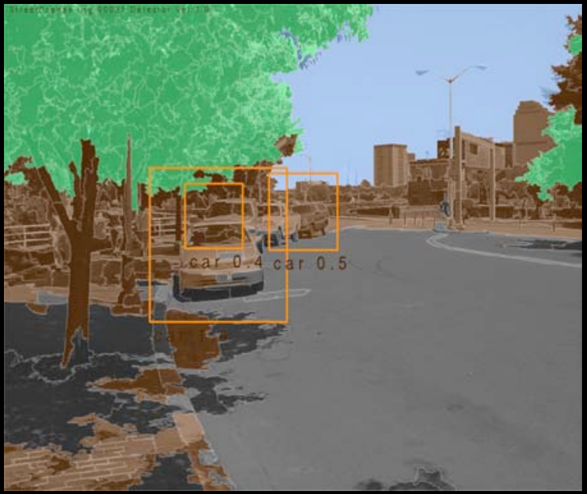


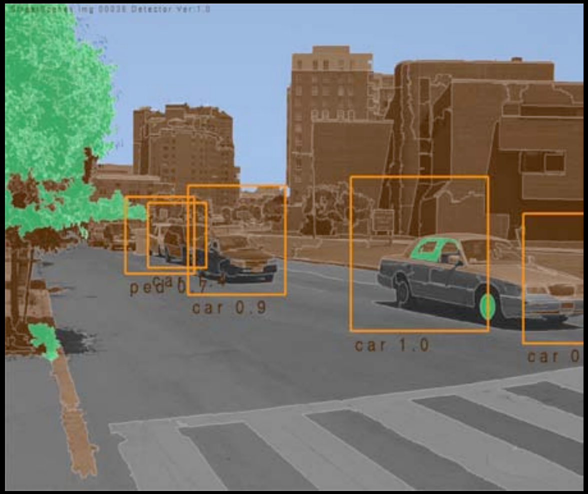


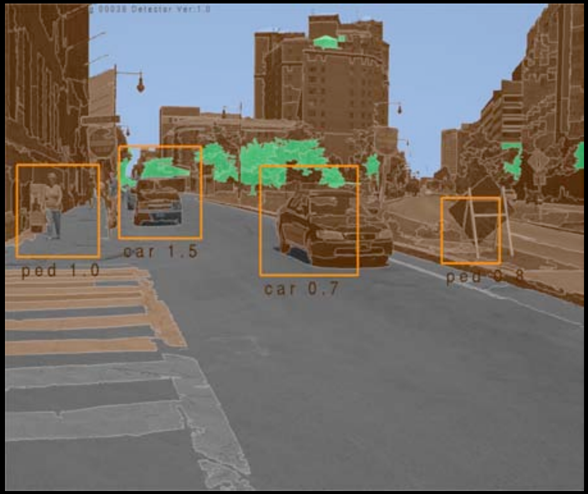


Examples

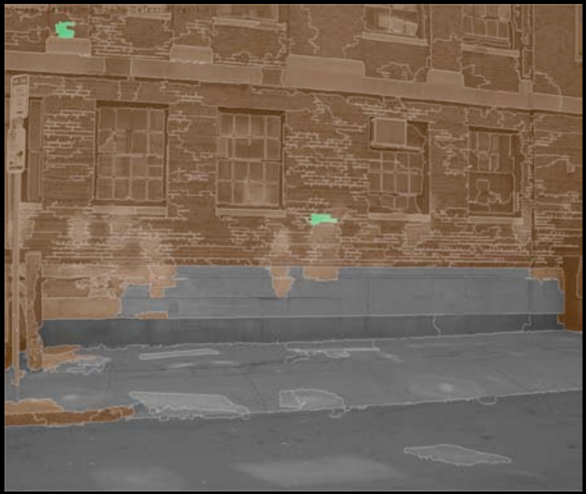


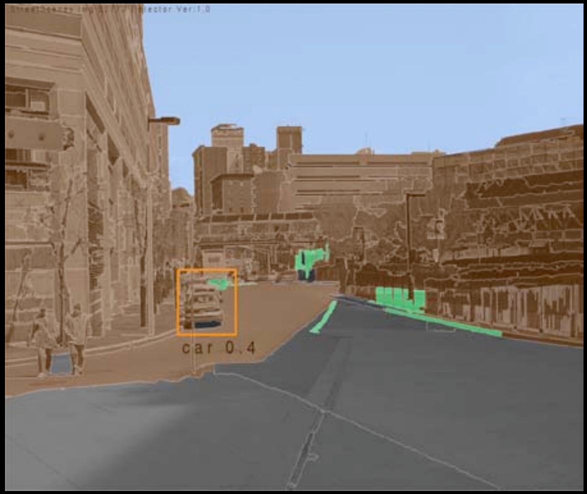




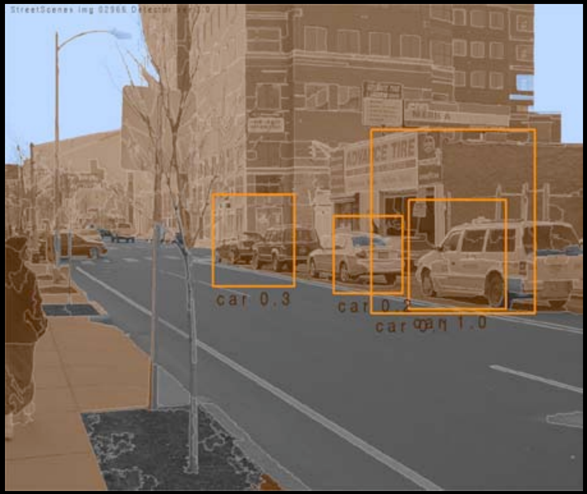


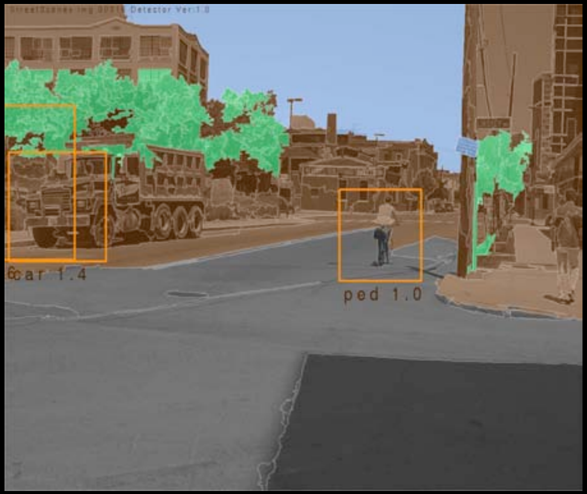
Examples

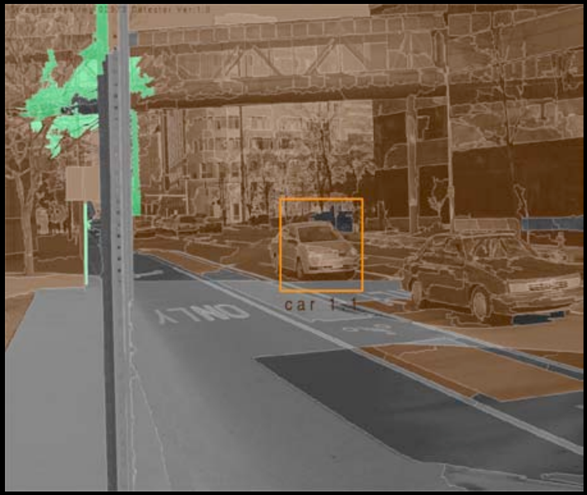


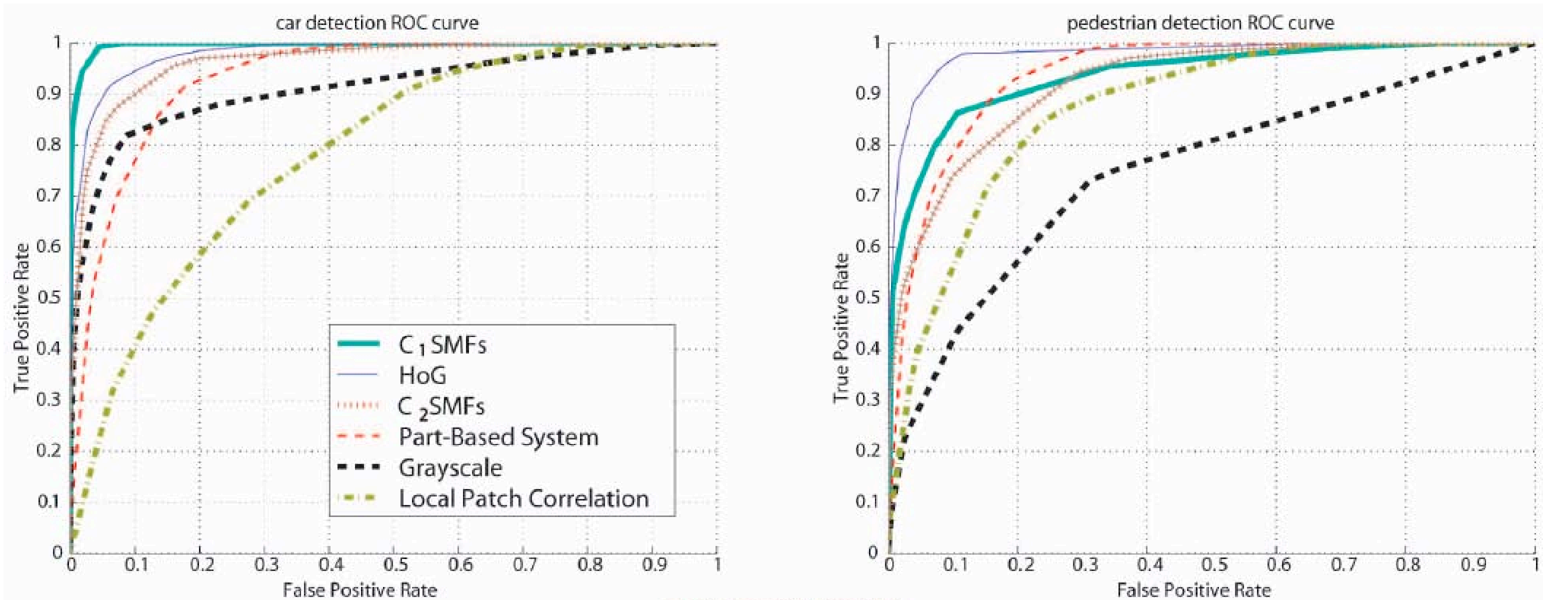


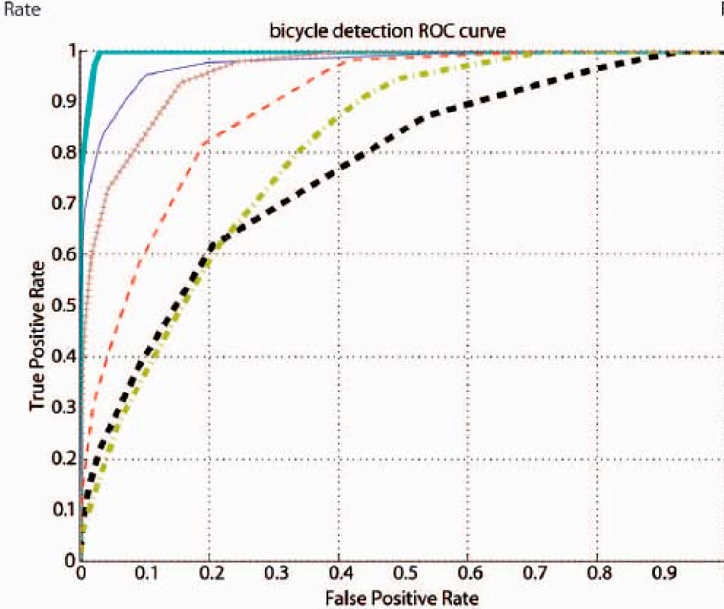
Examples









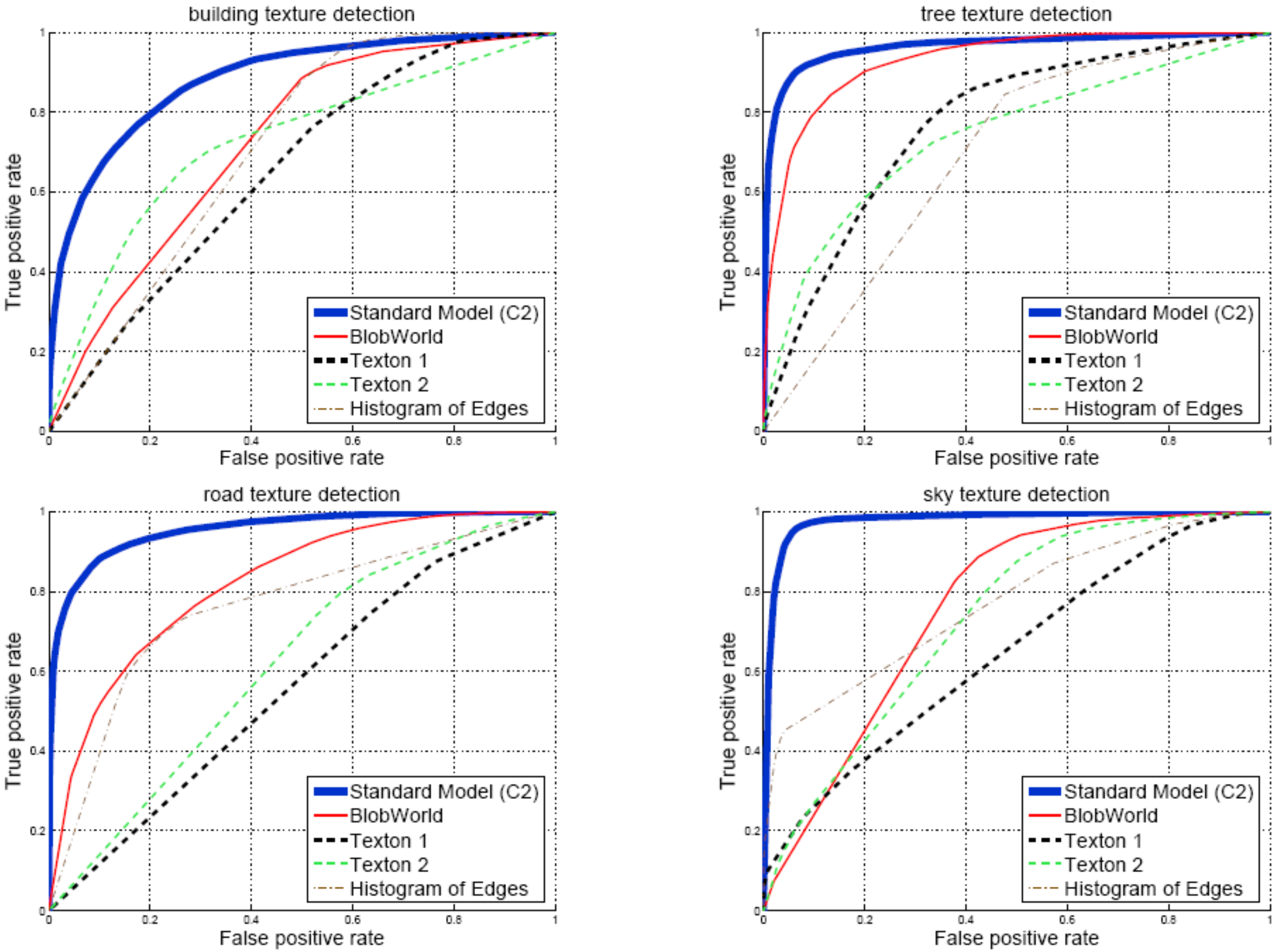


★HoG:
(Dalal & Triggs 2005)

★Part-based system:
(Leibe et al 2004)

★Local patch correlation:
(Torralba et al 2004)

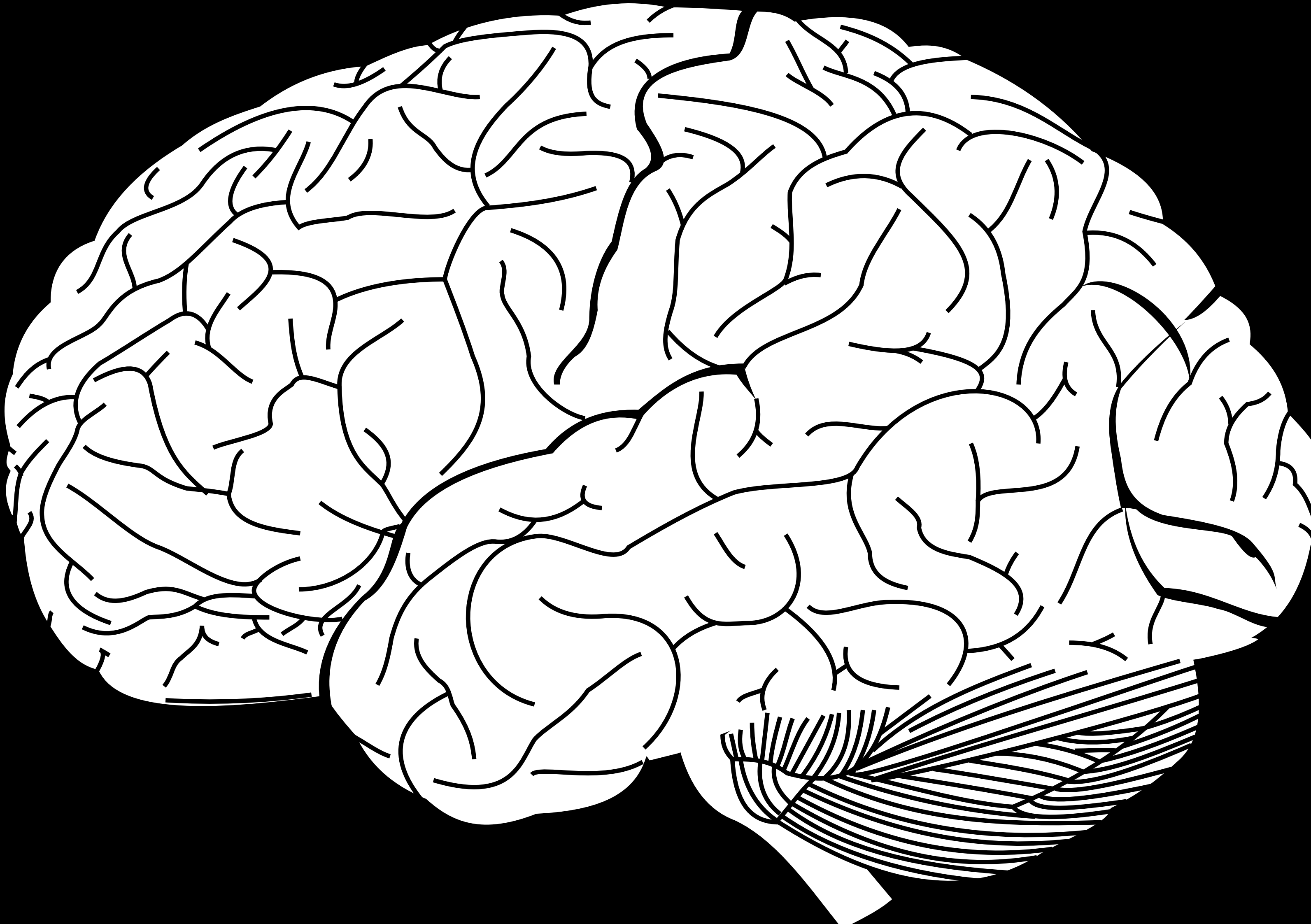
(Serre ***Wolf Bileschi*** Riesenhuber & Poggio PAMI 2007)



(Serre ***Wolf Bileschi*** Riesenhuber & Poggio PAMI 2007)

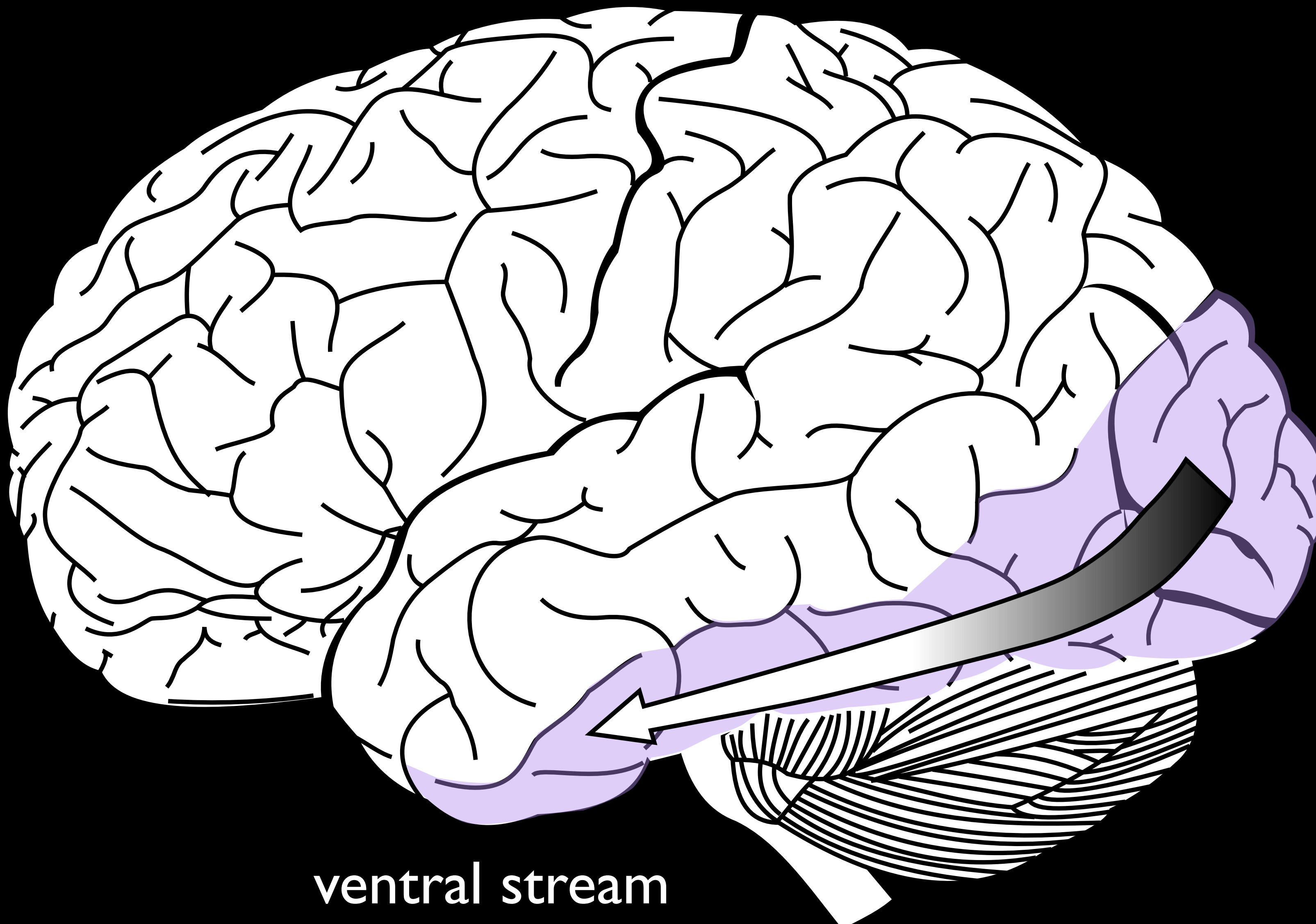
Action recognition with a model of the dorsal stream

Source: Wikipedia, "ventral stream"



(Ungerleider & Mishkin 1984)

Source: Wikipedia, "ventral stream"

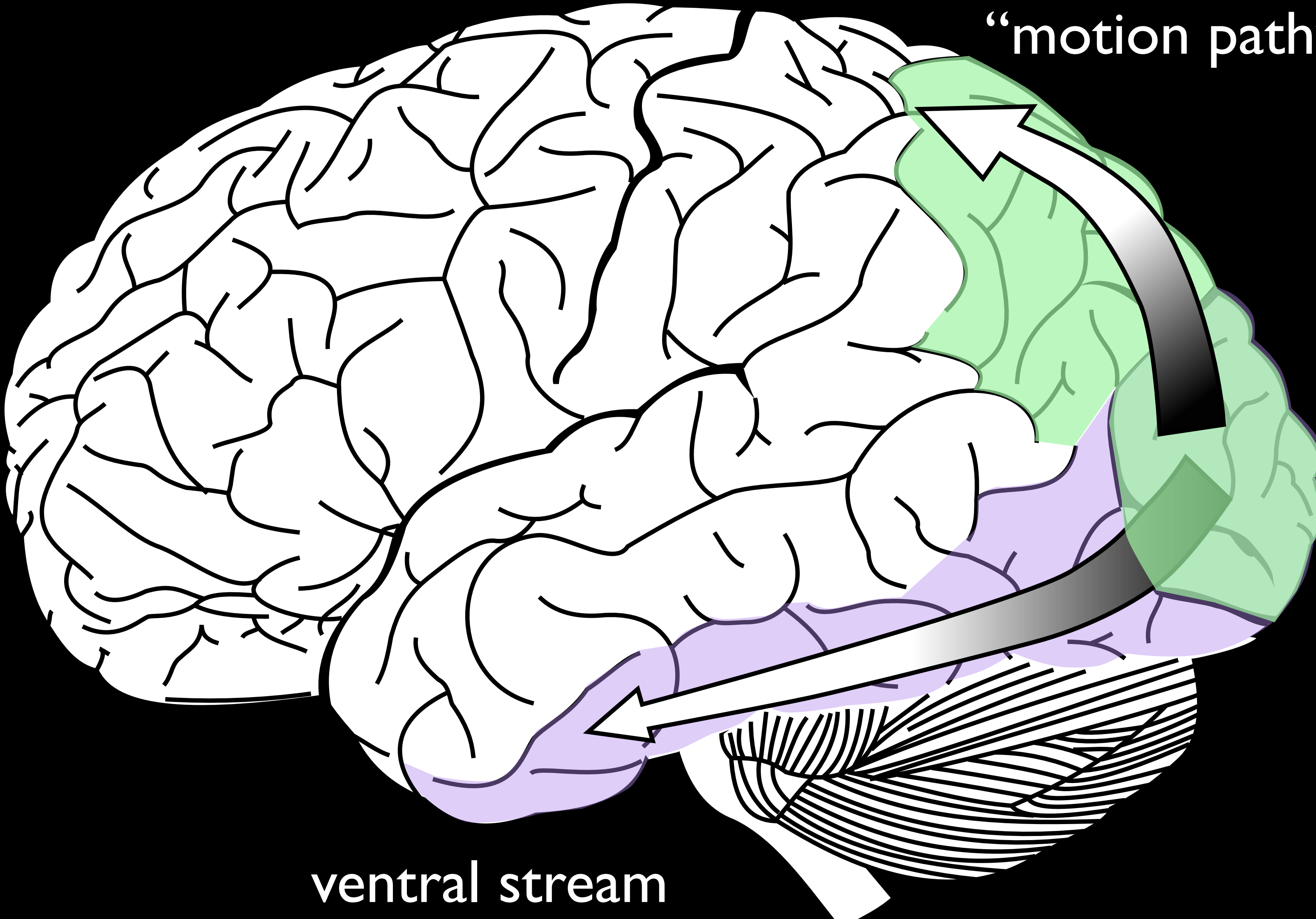


ventral stream
"shape pathway"

(Ungerleider & Mishkin 1984)

Source: Wikipedia, "ventral stream"

dorsal stream
"motion pathway"

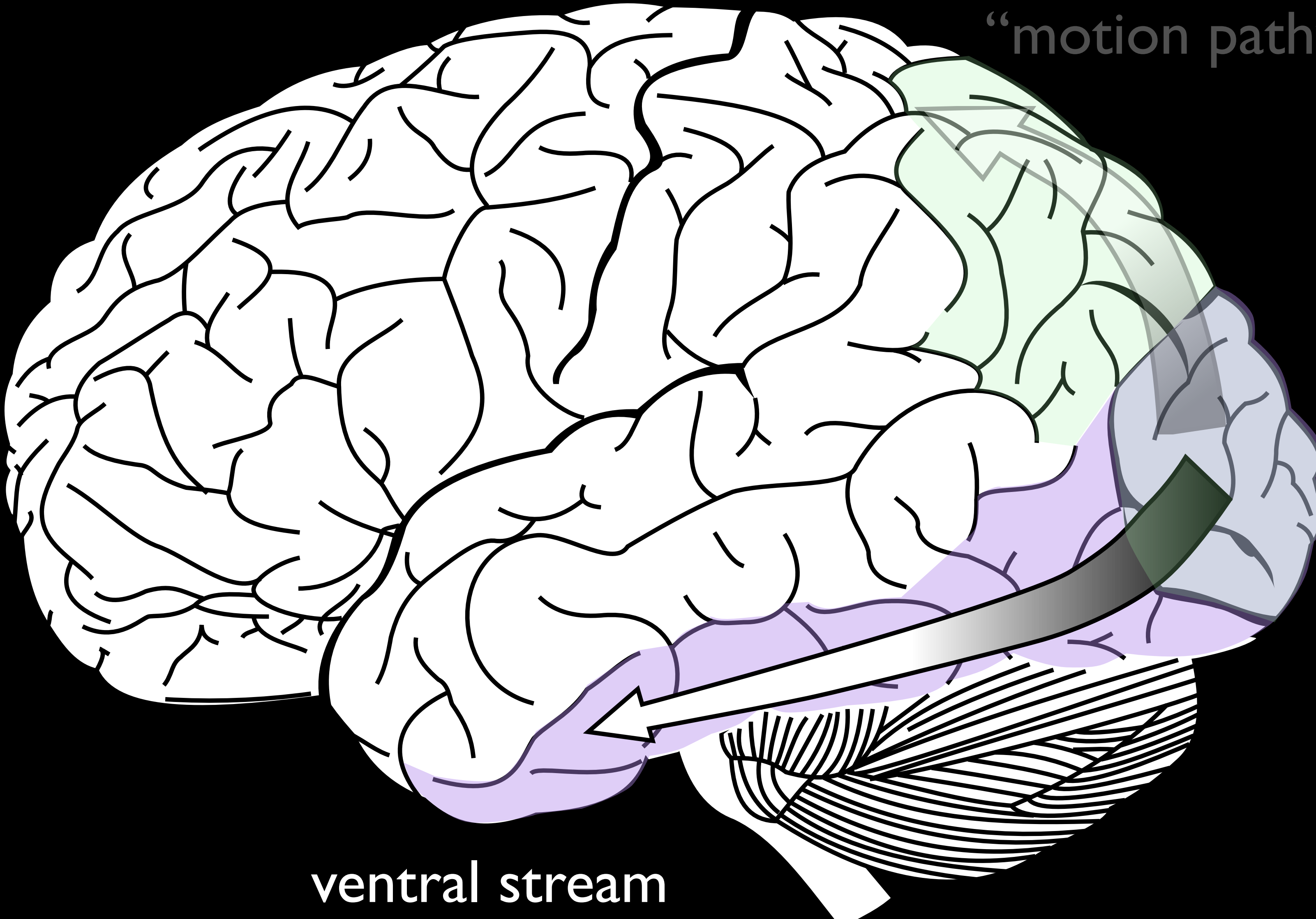


ventral stream
"shape pathway"

(Ungerleider & Mishkin 1984)

Source: Wikipedia, “ventral stream”

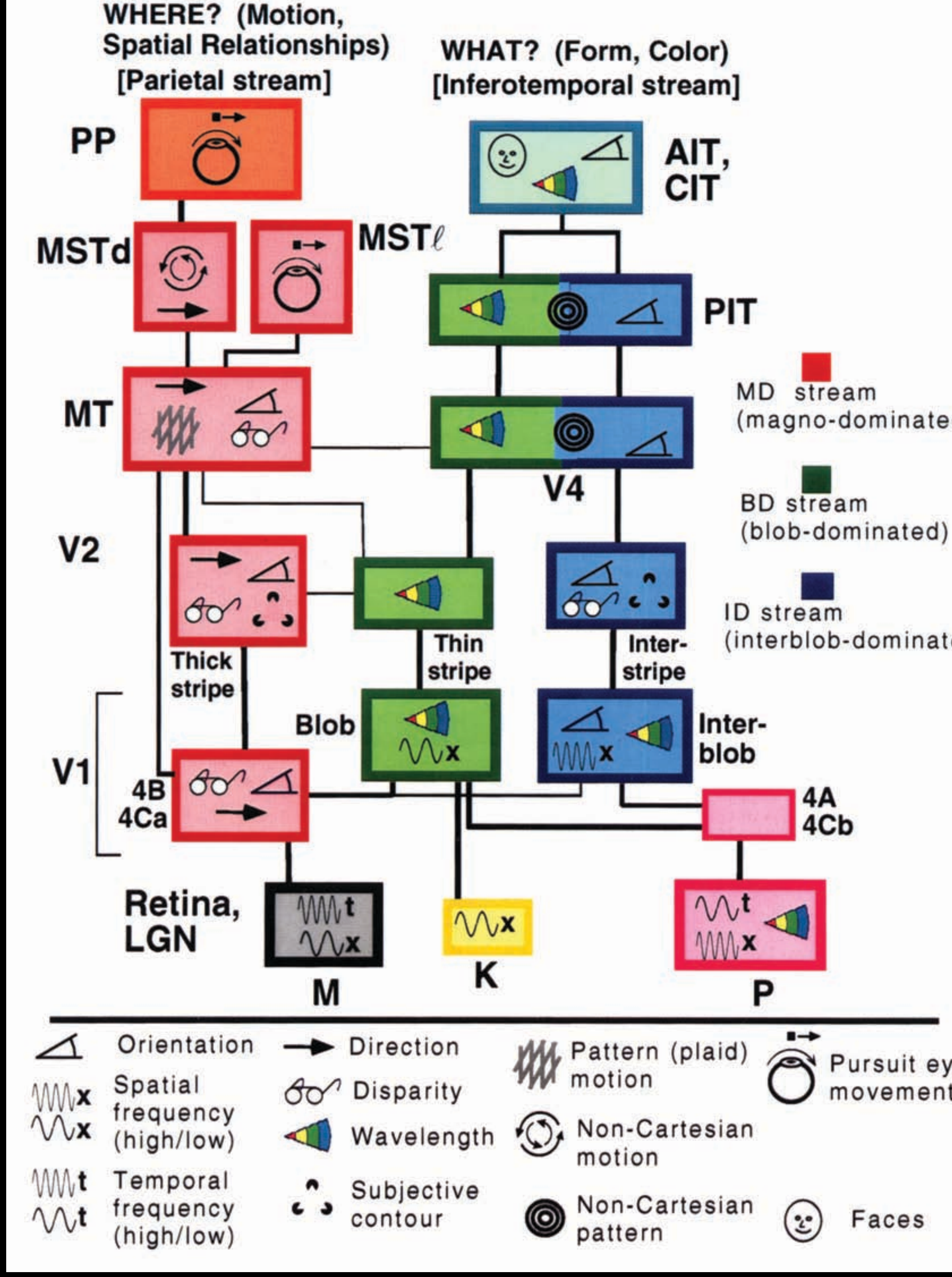
dorsal stream
“motion pathway”



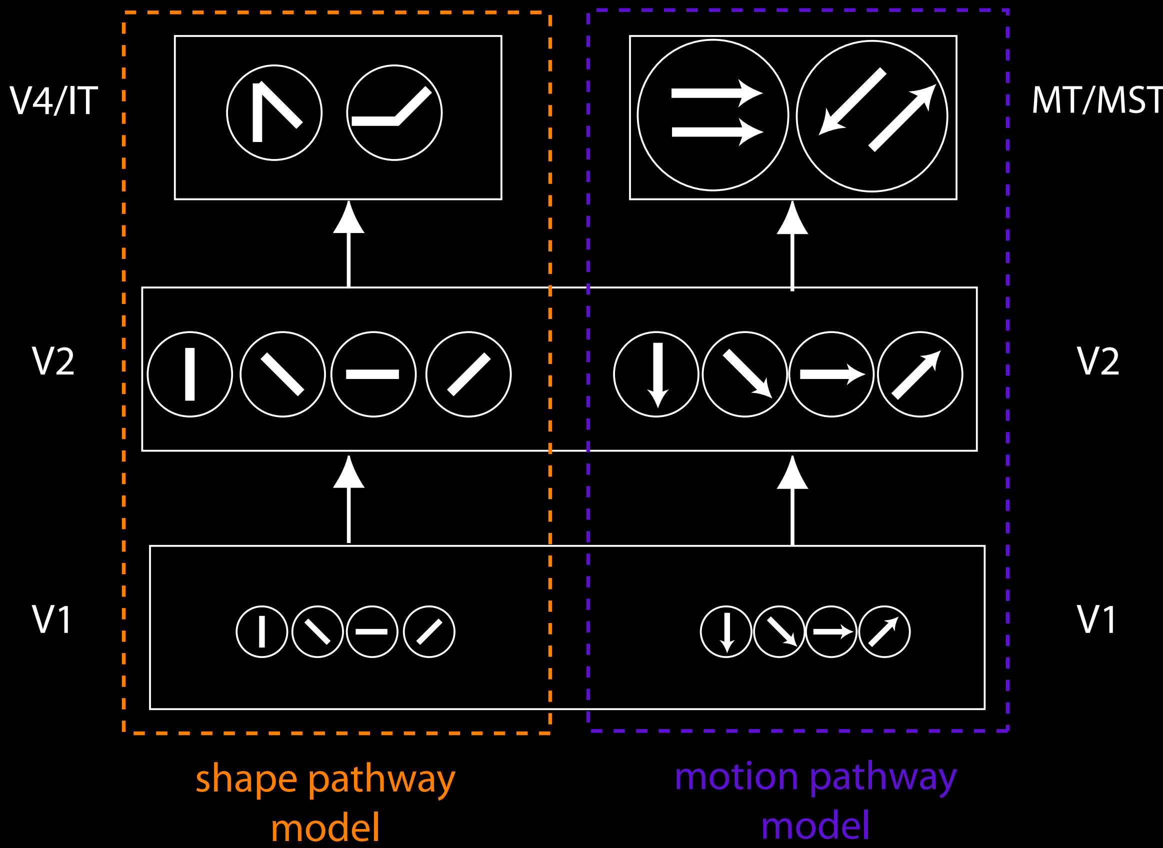
ventral stream
“shape pathway”

(Ungerleider & Mishkin 1984)

Action recognition with a model of the dorsal stream



(Gallant & VanEssen 1994)



(Riesenhuber & Poggio 1999; Serre et al. 2005)

(Giese & Poggio 2003; Casile & Giese 2005; Sigala Serre Poggio & Giese 2005)

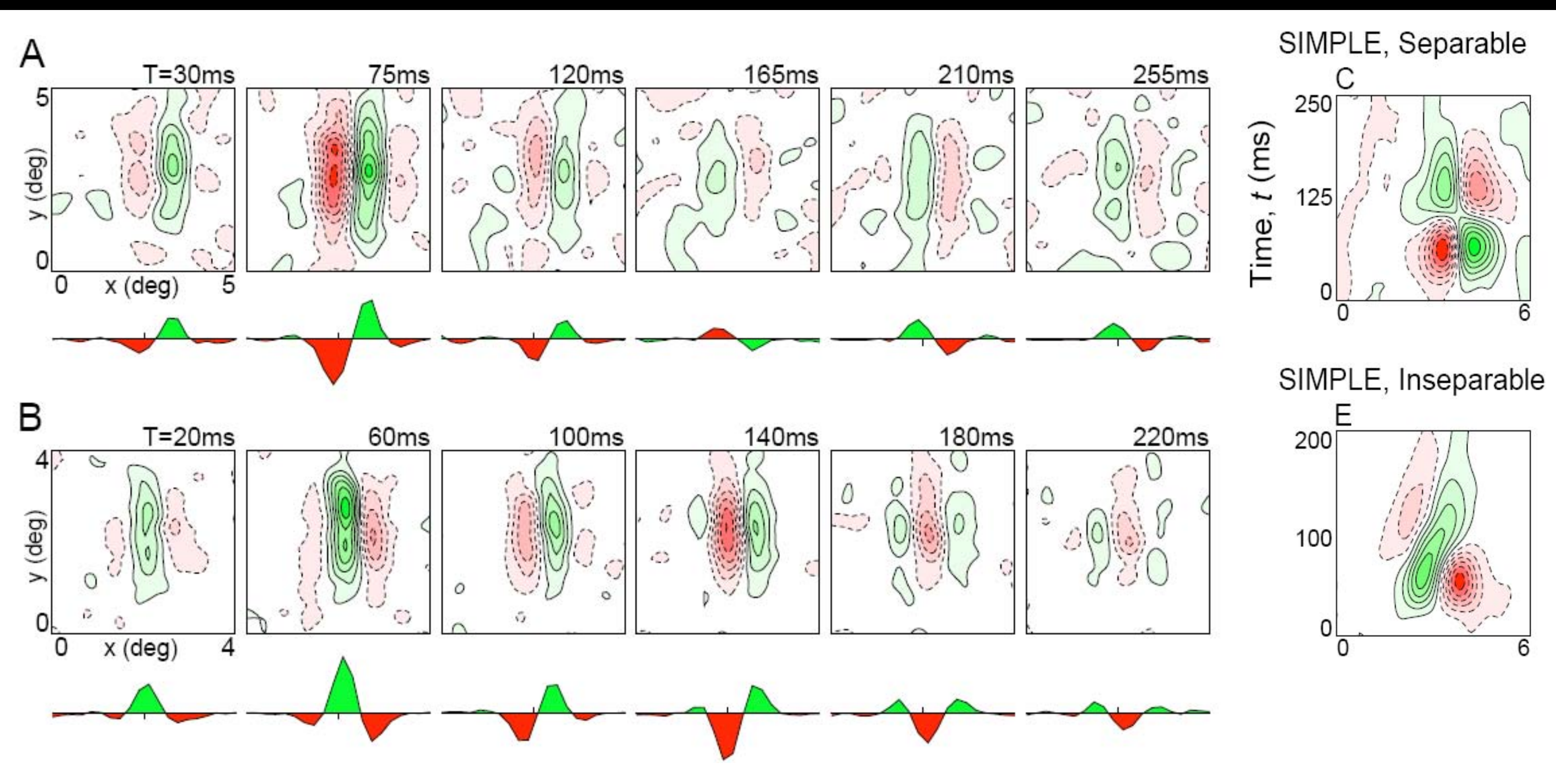
Same "principles", only different parameters:

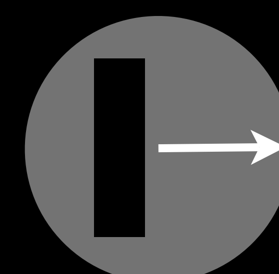
- Same 2 types of functional units [simple and complex]
- Same 2 key operations [tuning and soft-max]
- Same unsupervised learning rule

Jhuang, Serre & Poggio

Action recognition with a model of the dorsal stream of the visual cortex

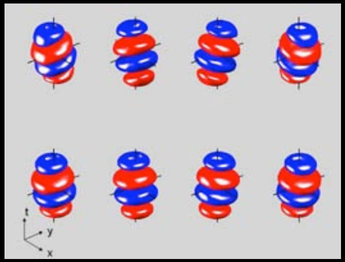
- ✦ Dorsal similar organization as ventral stream
- ✦ Starts with spatio-temporal RFs in V1





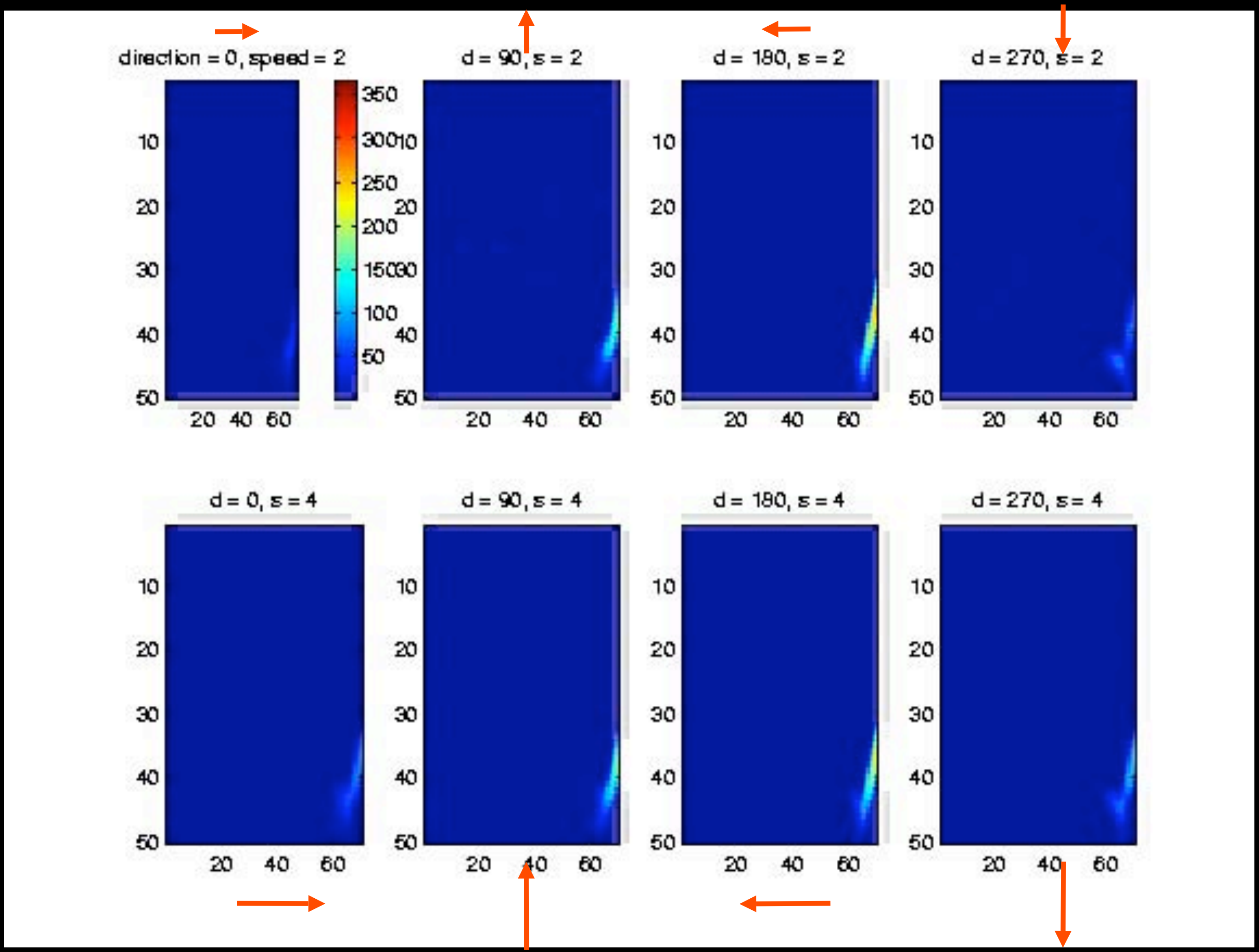
(Oshawa DeAngelis Freeman 1995)

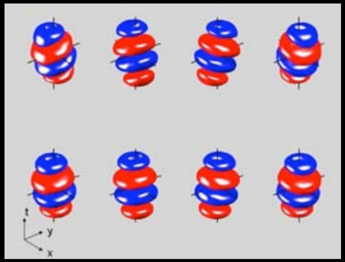
Motion sensitive S1 units as spatio-temporal filters



(Heeger 1987;
Simoncelli & Heeger 1998)

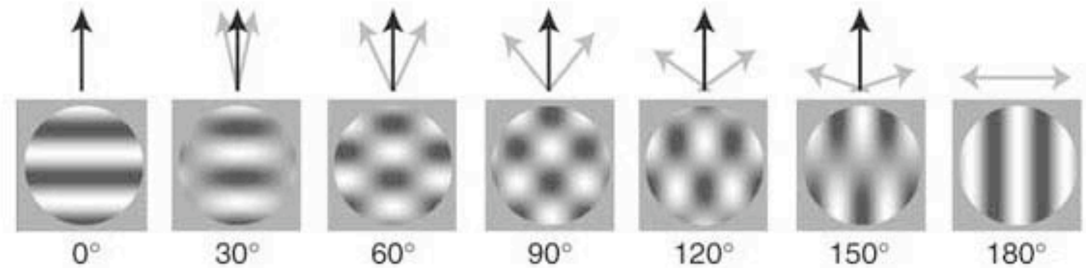
Motion sensitive S1 units as spatio-temporal filters

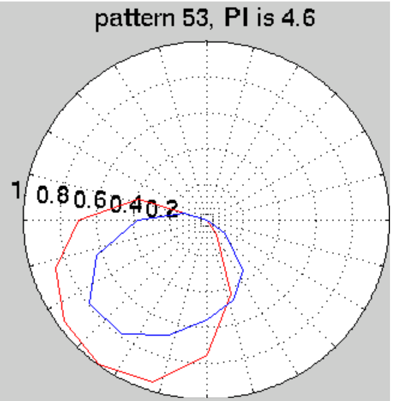


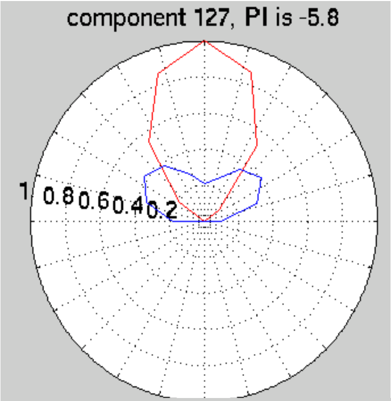


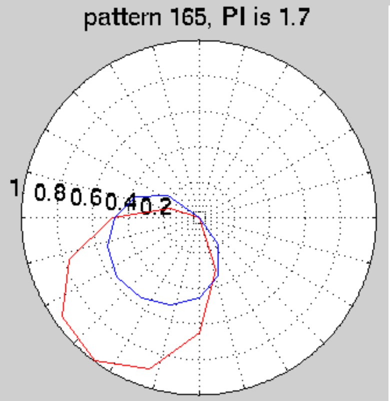
(Heeger 1987;
Simoncelli & Heeger 1998)

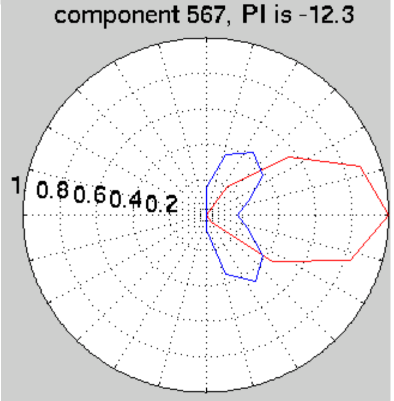
Unsupervised learning in MT produces pattern and component cells



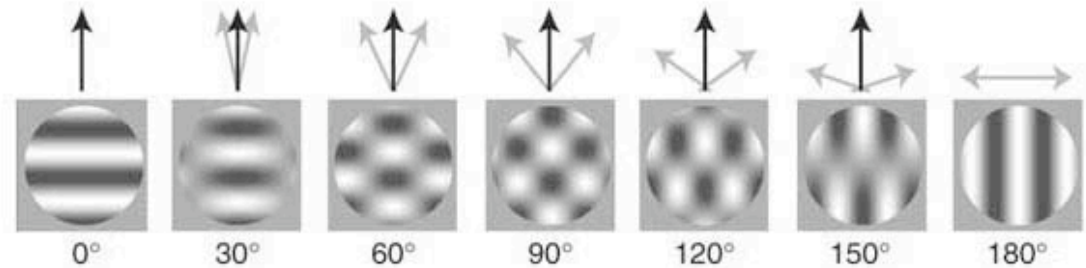


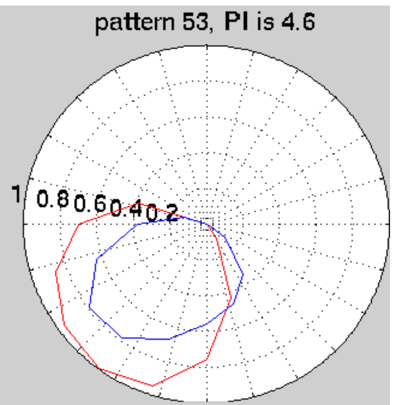


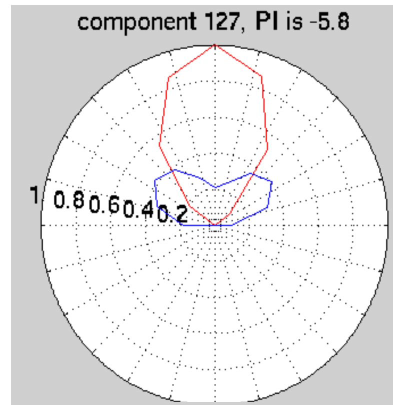


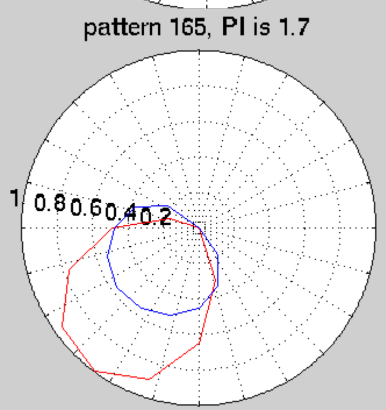


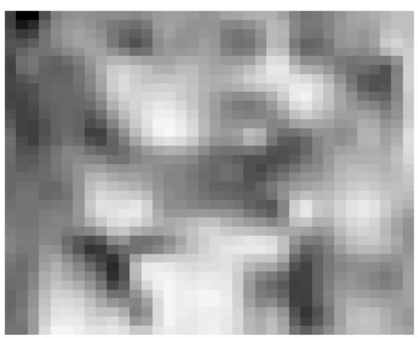
Unsupervised learning in MT produces pattern and component cells

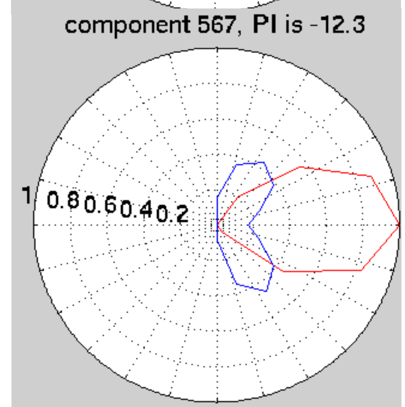












The problem

Training Videos


bend


jack

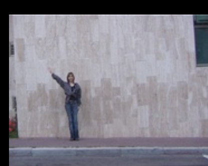

jump 1


jump 2


run


walk


side


wave 1


wave 2

*each video~4s, 50~100 frames

Testing videos

Dataset from (Blank et al, 2005)

The problem

Training Videos


bend


jack


jump 1


jump 2


run


walk


side


wave 1


wave 2

*each video~4s, 50~100 frames

Testing videos

Dataset from (Blank et al, 2005)

Standard action datasets

KTH Human actions (6 classes)

walk jog run box wave clap

UCSD Mice actions (5 classes)

drink eat explore groom sleep

Weizmann Human action (9 classes)

bend jack jump pjump run walk side wave1 wave2

Multi-class recognition accuracy

	Baseline	Our system
KTH Human	81.3%	**91.6%**
UCSD Mice	75.6%	**79.0%**
Weiz. Human	86.7%	**96.3%**

*Accuracy : average over diagonal terms of confusion matrices
*2/3 training, 1/3 testing
* chances: 10%~20%

(Jhuang Serre Wolf & Poggio 2007)

Multi-class recognition accuracy

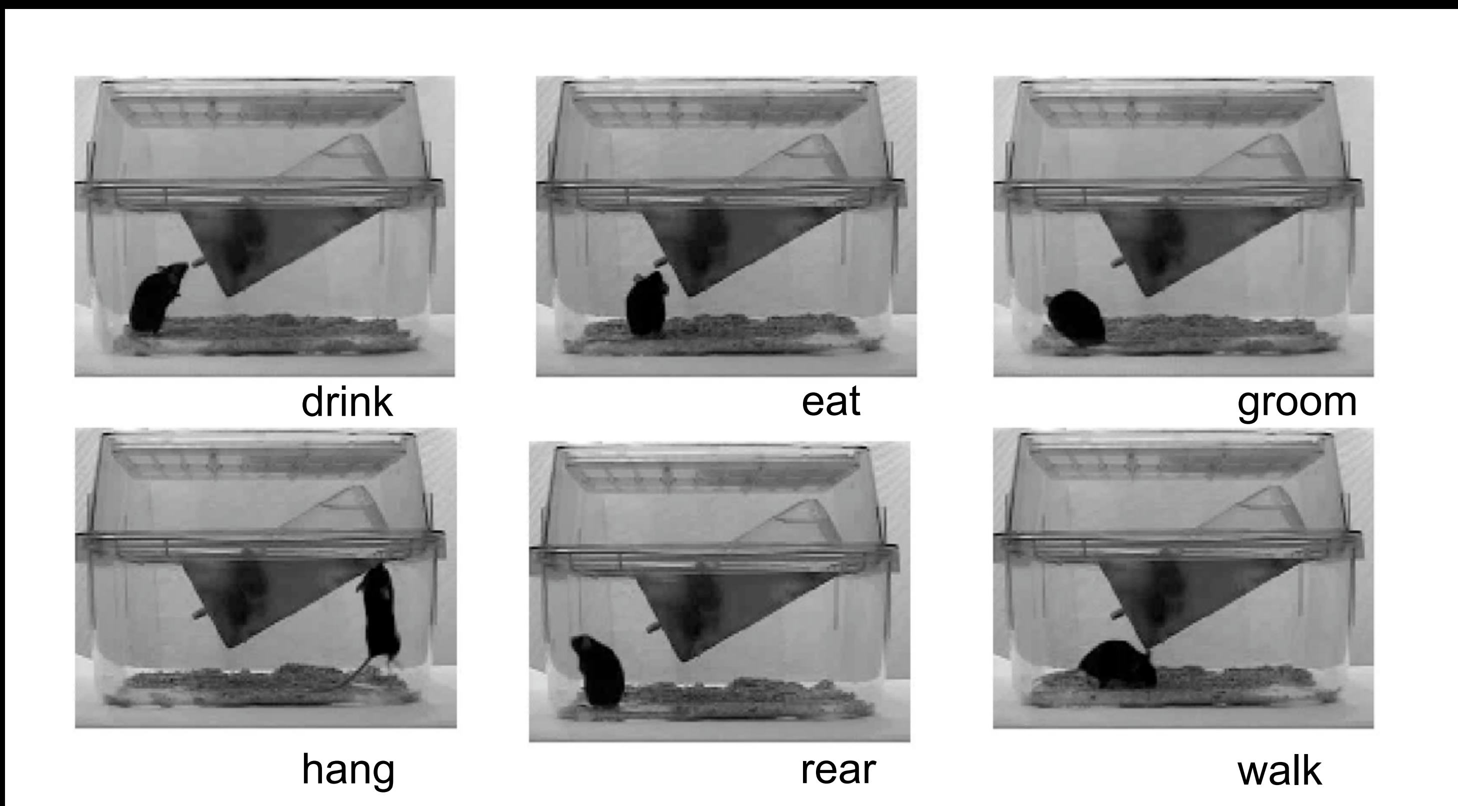
	Baseline	Our system
KTH Human	81.3%	**91.6%**
UCSD Mice	75.6%	**79.0%**
Weiz. Human	86.7%	**96.3%**

*Accuracy : average over diagonal terms of confusion matrices
*2/3 training, 1/3 testing
* chances: 10%~20%

(Jhuang Serre Wolf & Poggio 2007)

Automatic classification of abnormal behavior in mutant vs. wild mice

w| Andrew Steele, Whitehead Institute



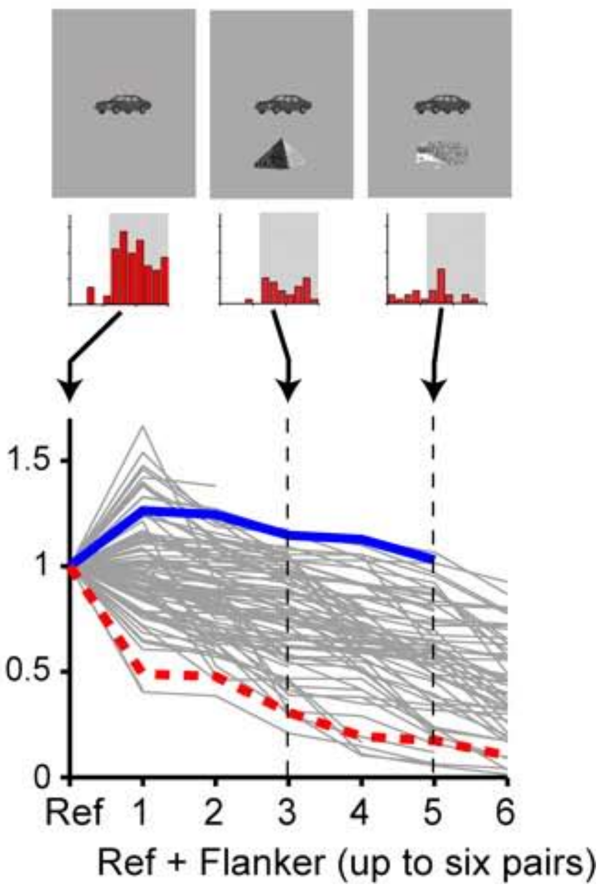
over 95% correct for 6 class-classification

Serre, Steele, Jhuang, Garrote & Poggio

Cortical feedbacks and attention

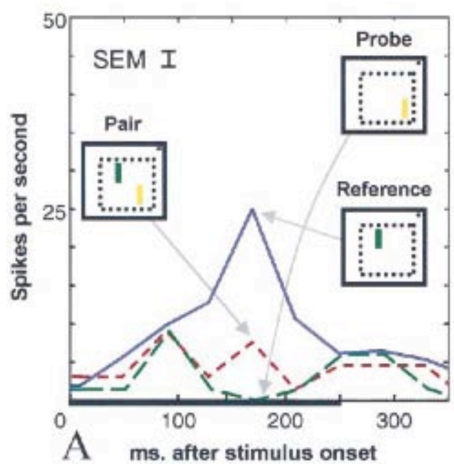
Limitations of feedforward processing: clutter

IT



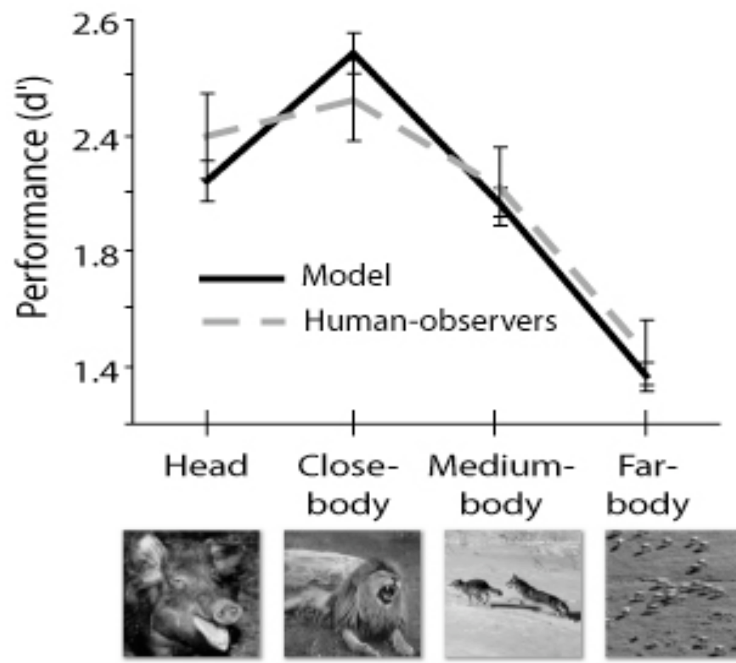
Zoccolan Kouh Poggio DiCarlo 2007

V4



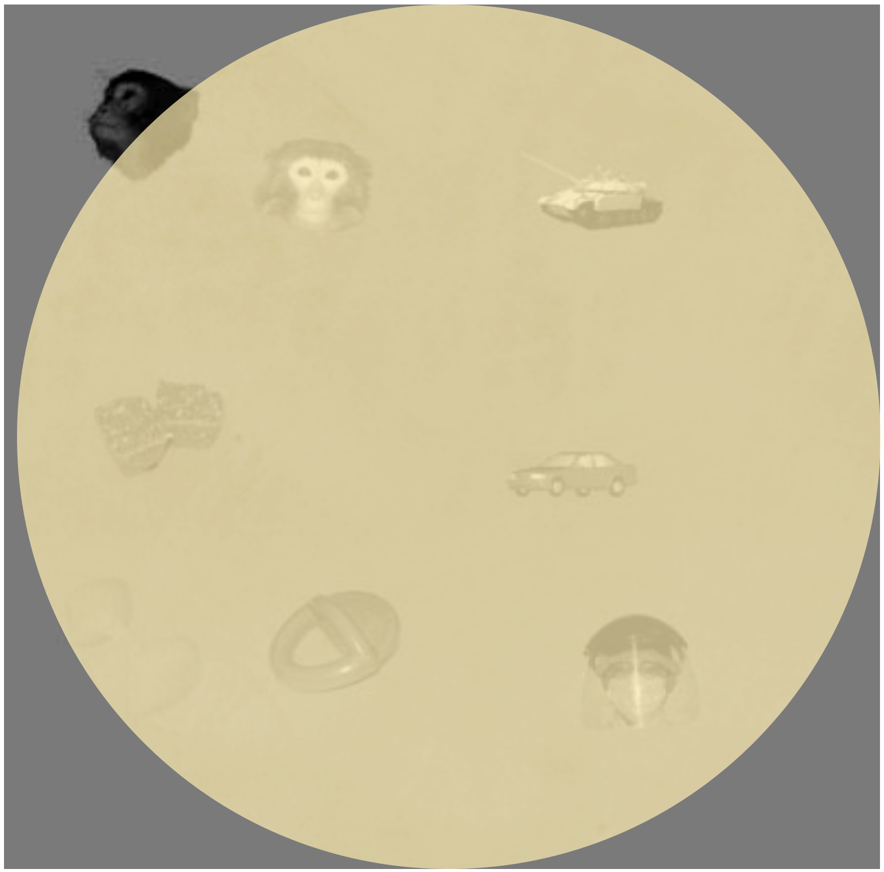
Reynolds Chelazzi & Desimone 1999

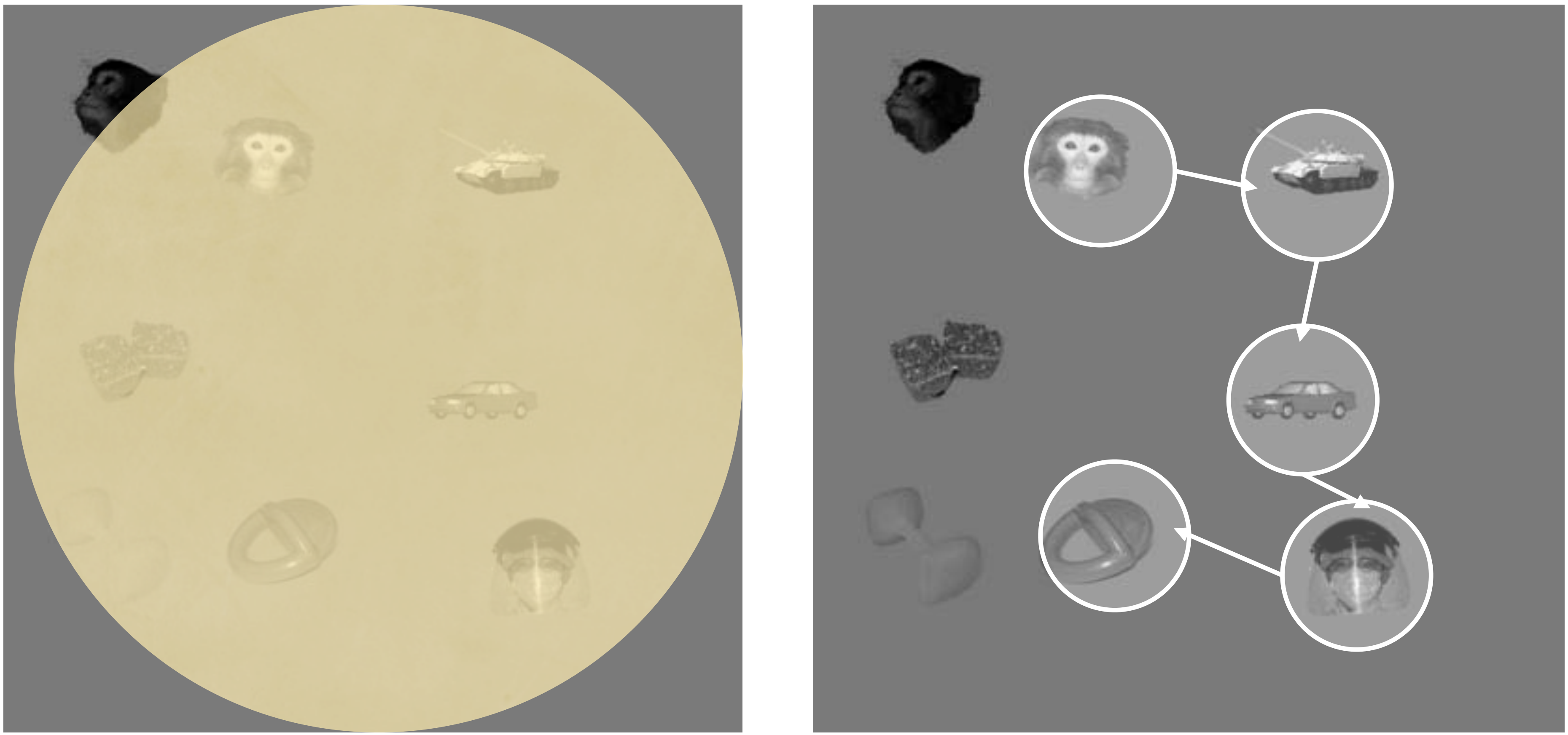
Psychophysics

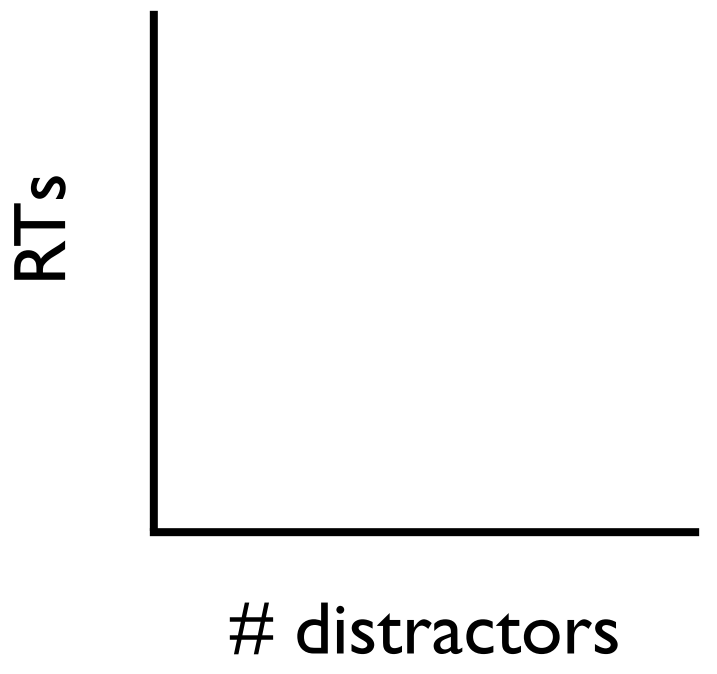


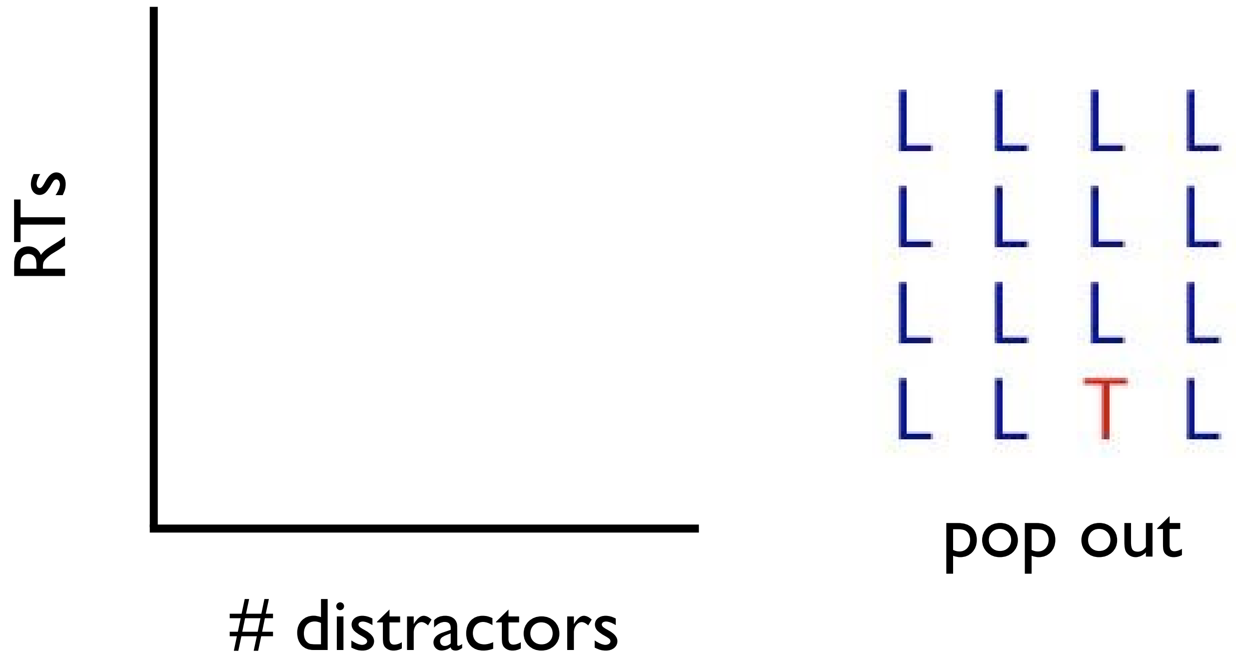
Serre Oliva Poggio 2007

Poggio (MIT)

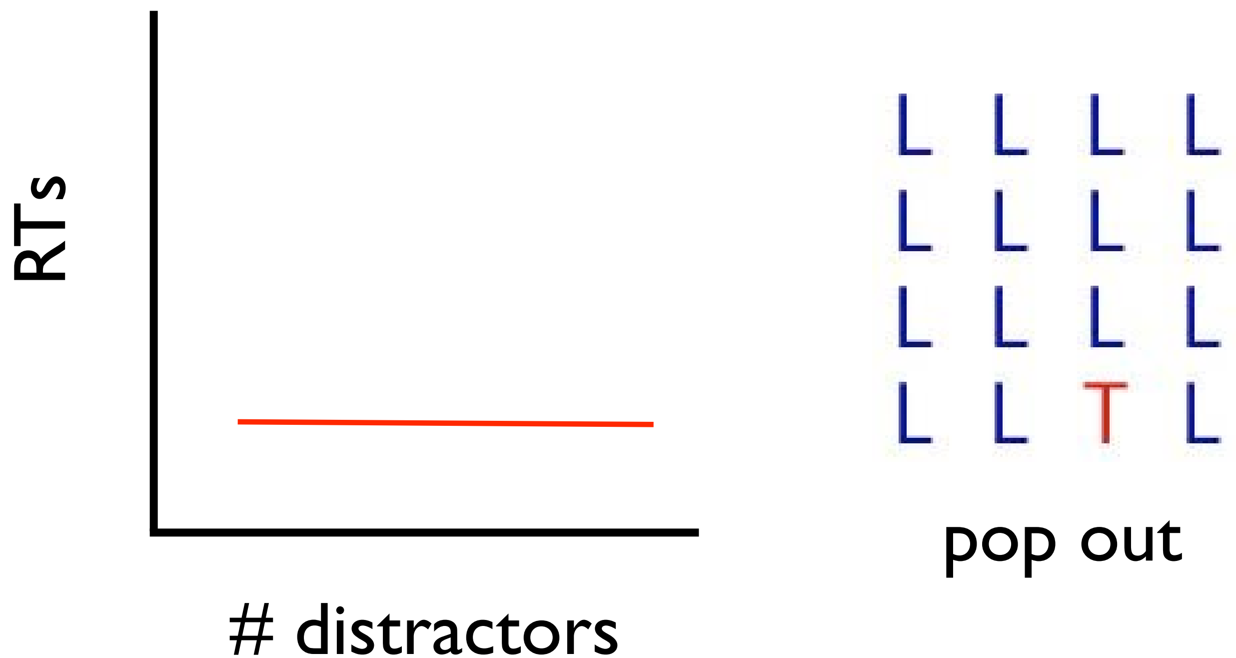




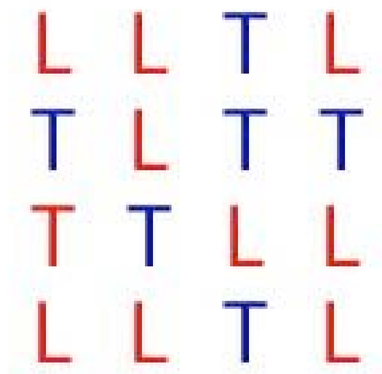

RTs
distractors

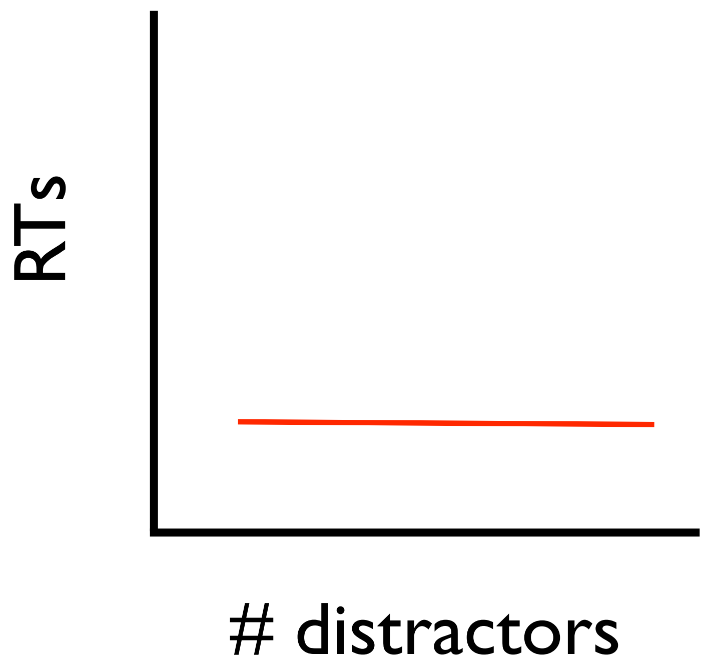


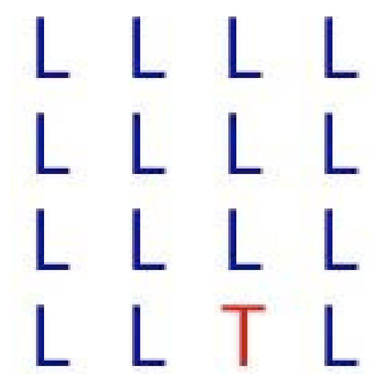
RTs
distractors
L L L L
L L L L
L L L L
L L T L
pop out


RTs
distractors
L L L L
L L L L
L L L L
L L T L
pop out

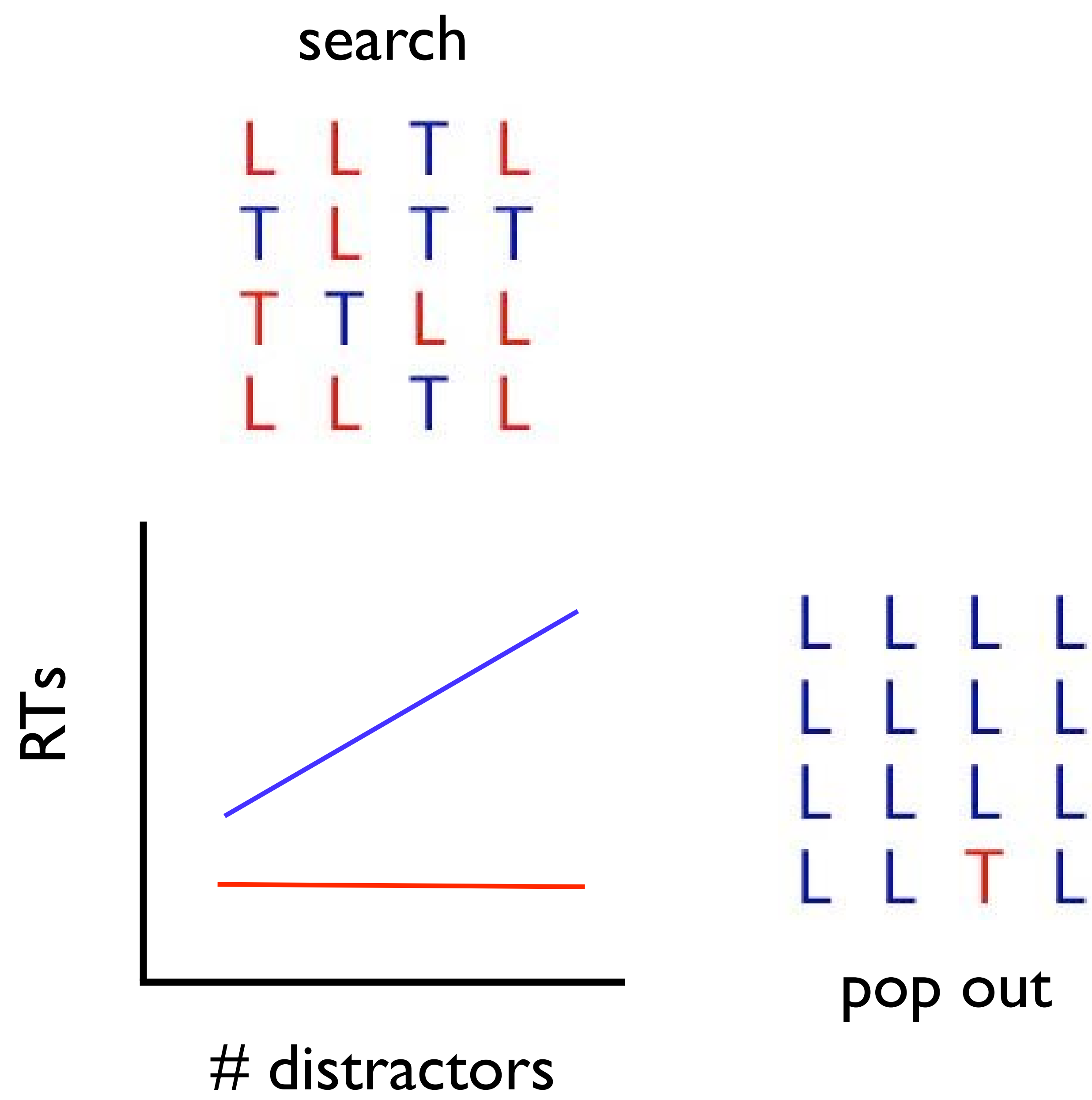
search



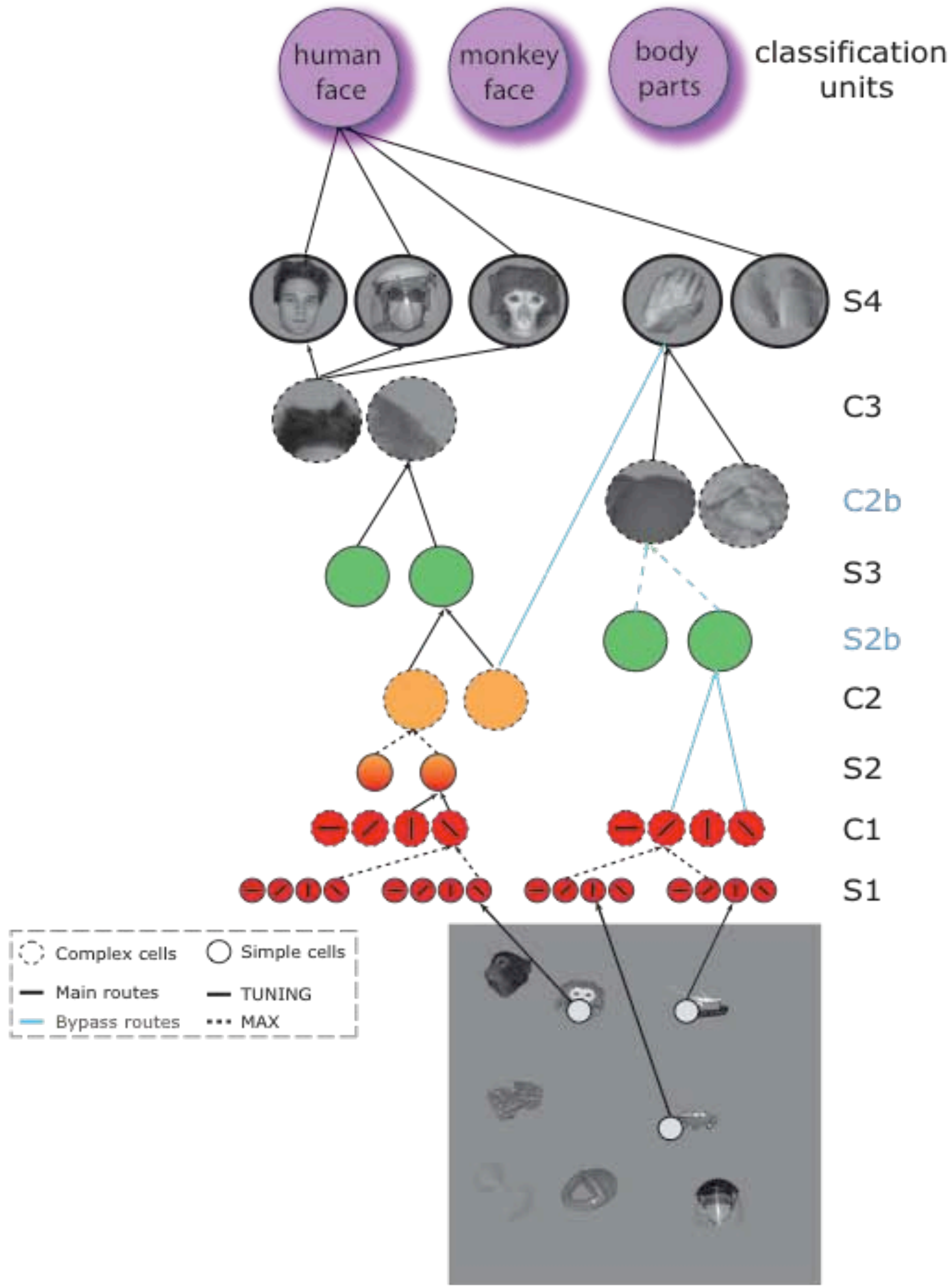


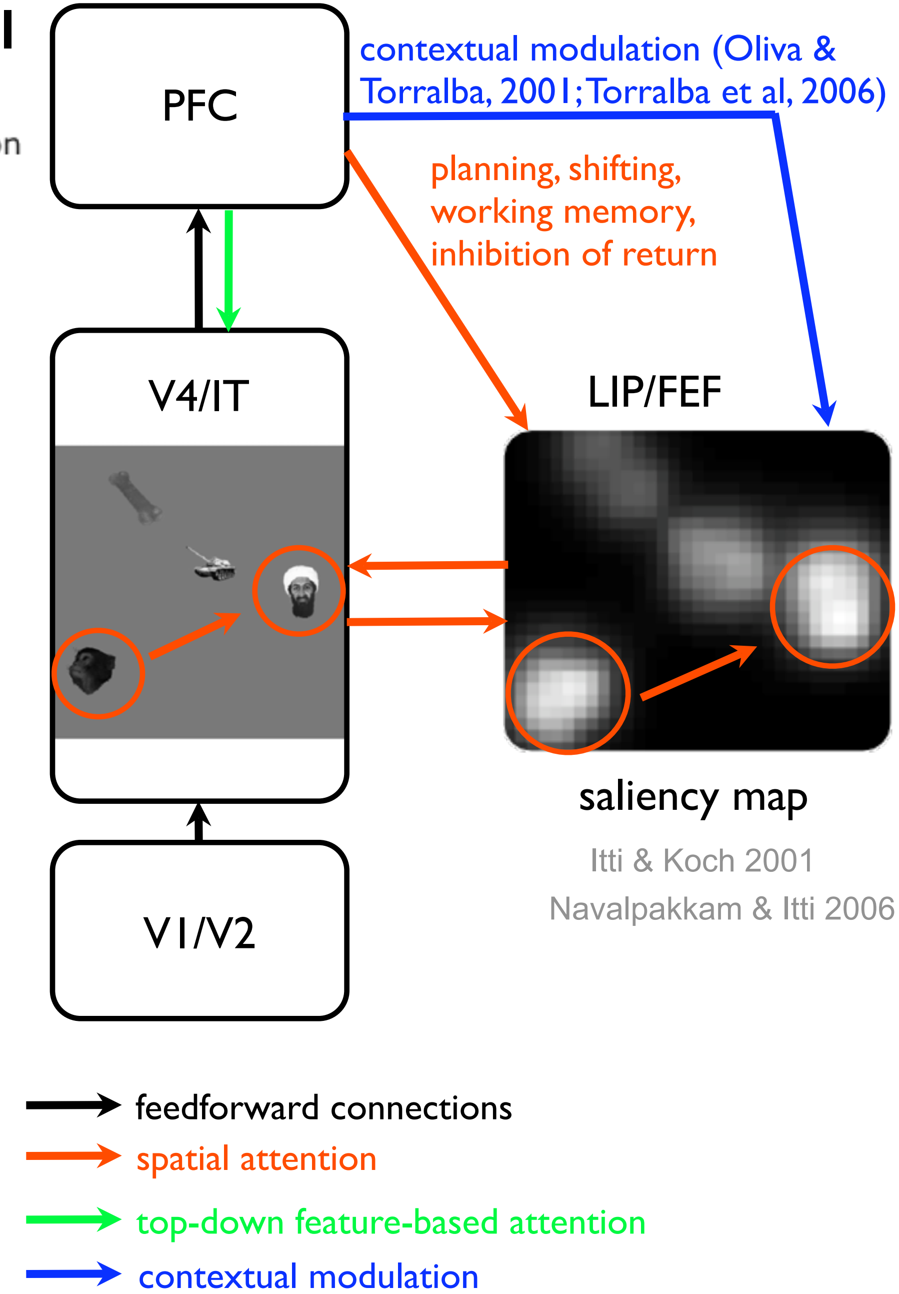


pop out


search
L L T L
T L T T
T T L L
L L T L
RTs
distractors
L L L L
L L L L
L L L L
L L T L
pop out

New top-down attentional model

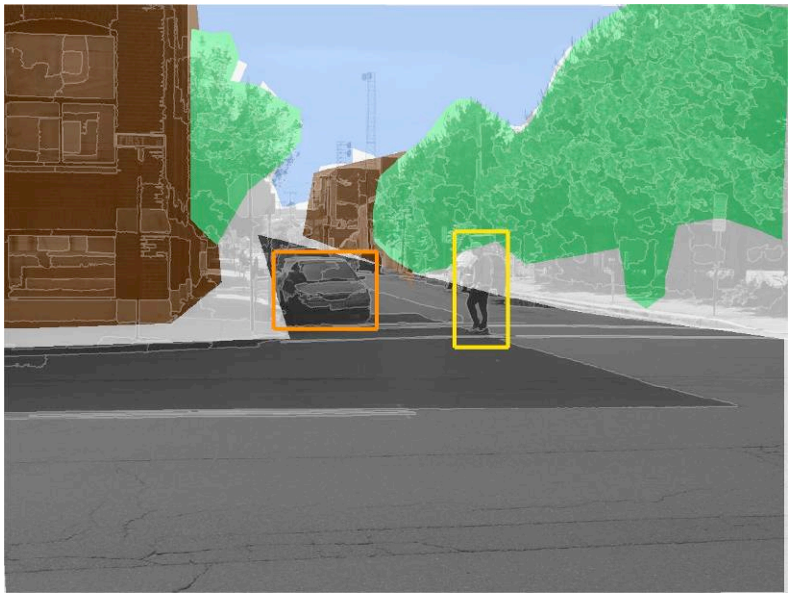




S. Chikkerur, T. Serre, D. Walther, C. Koch and T. Poggio

Comparison with human eye fixations on natural scenes

The StreetScenes Database



3,547 Images, all taken with the same camera, of the same type of scene, and hand labeled with the same objects, using the same labeling rules.

Object	car	pedestrian	bicycle	building	tree	road	sky
# Labeled Examples	5799	1449	209	5067	4932	3400	2562

Poggio (MIT)

Testing the model against human eye movements

Show demo

Pedestrian search

S. Chikkerur, C. Tan, T. Serre and T. Poggio

Pedestrian search

S. Chikkerur, C. Tan, T. Serre and T. Poggio

Pedestrian search

S. Chikkerur, C. Tan, T. Serre and T. Poggio

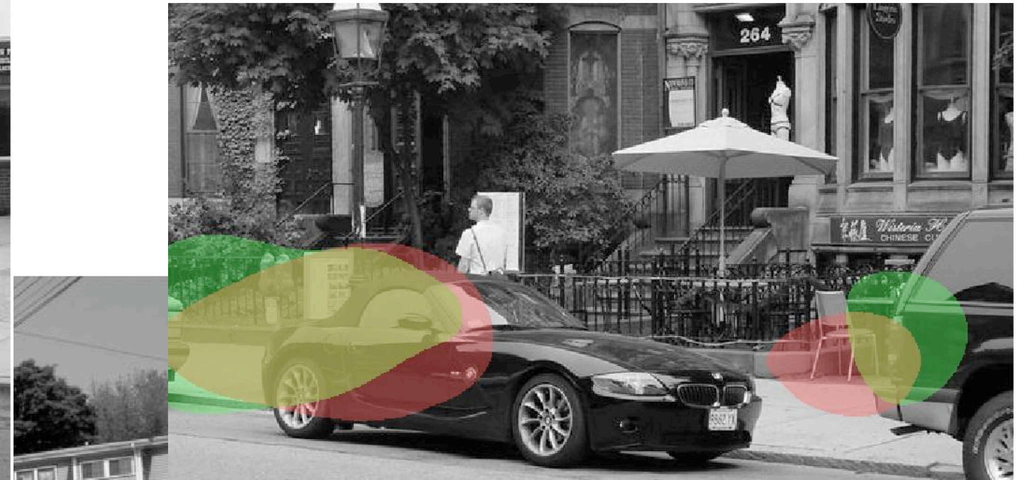
Car search

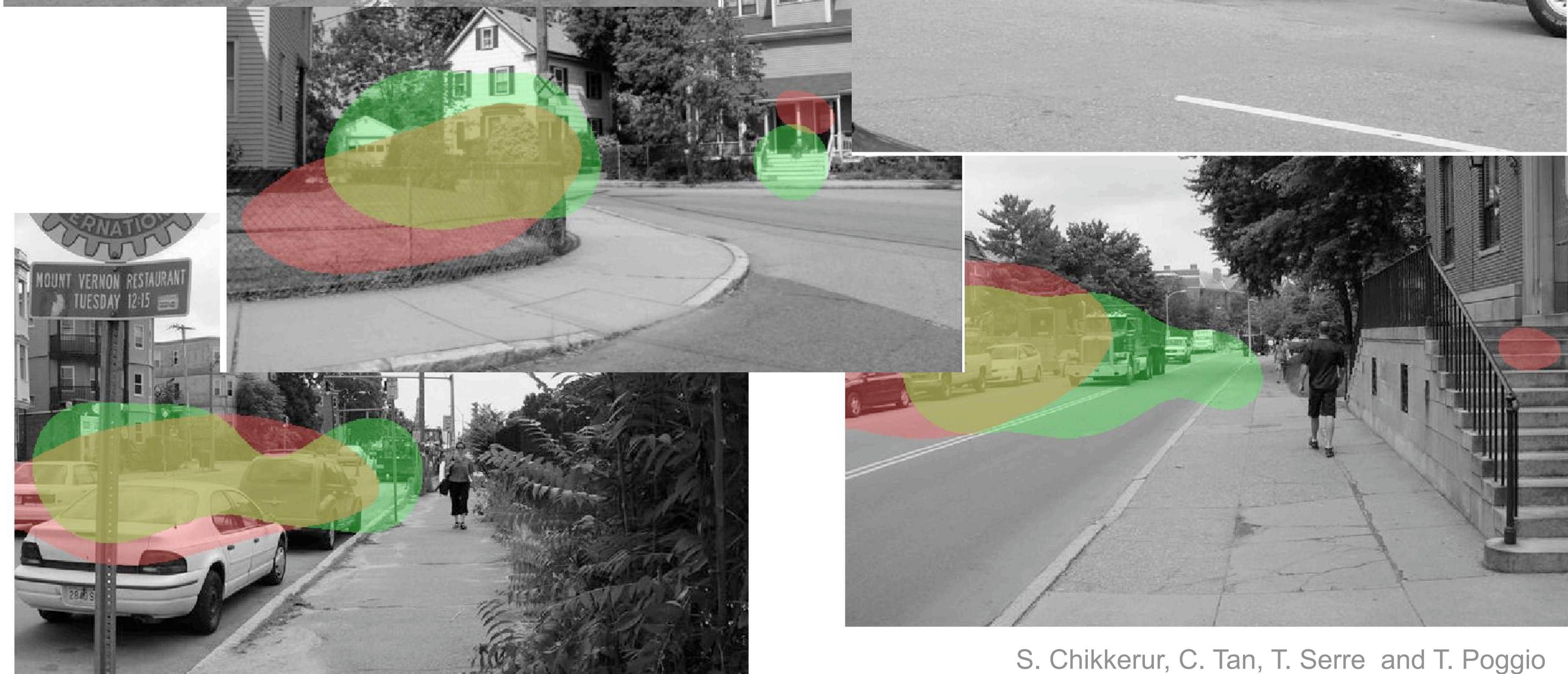
S. Chikkerur, C. Tan, T. Serre and T. Poggio

Car search

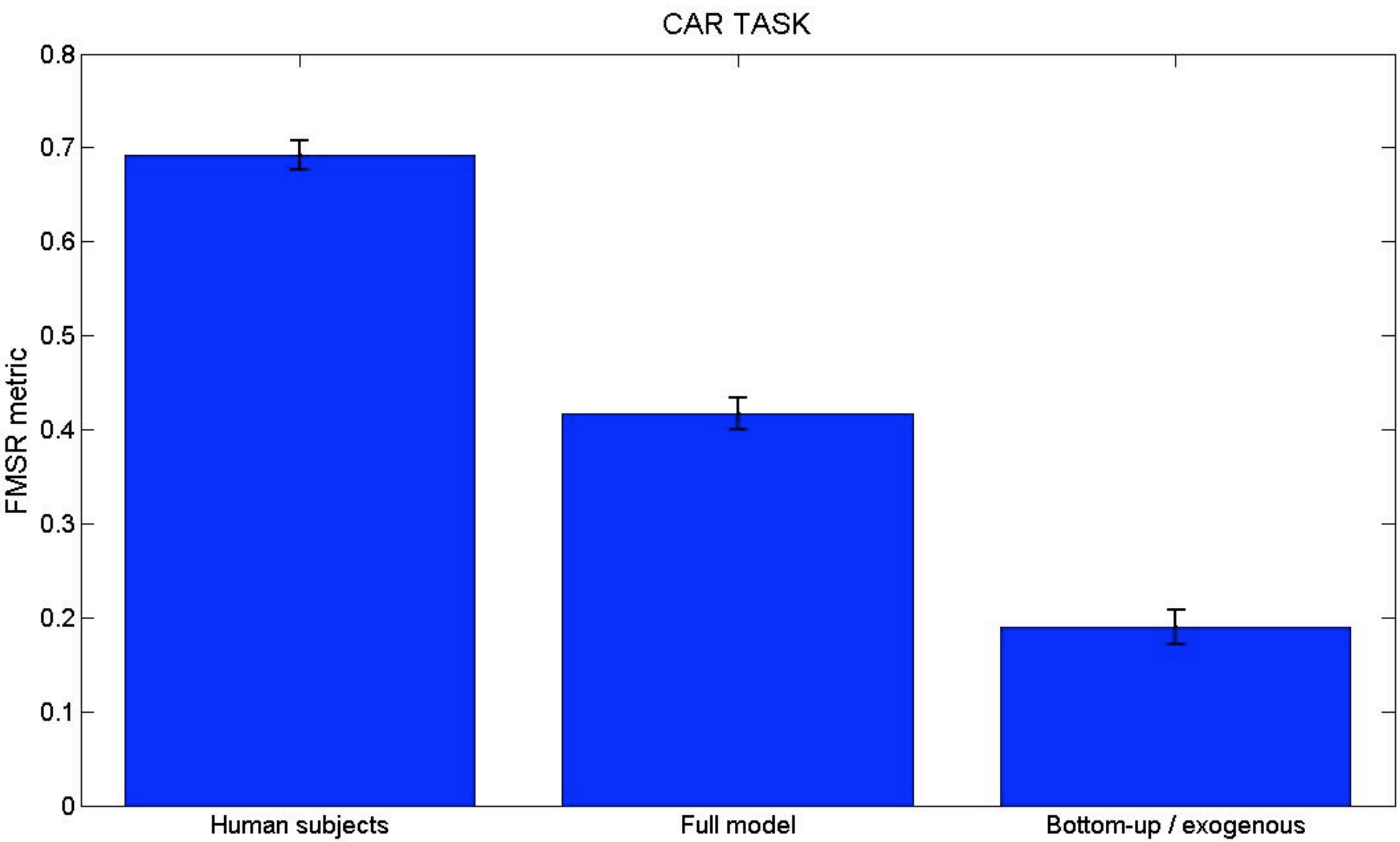
S. Chikkerur, C. Tan, T. Serre and T. Poggio

Car search

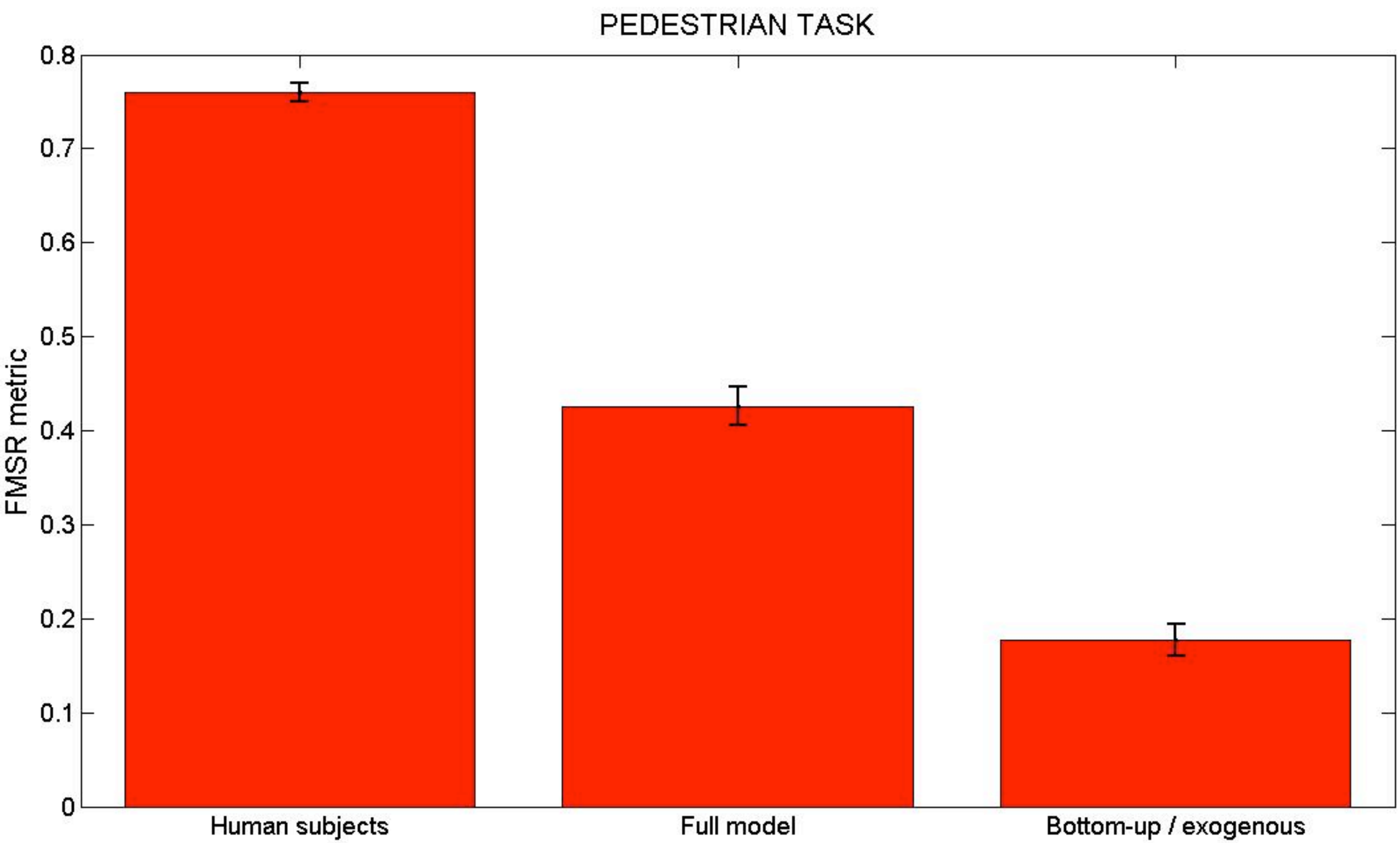




S. Chikkerur, C. Tan, T. Serre and T. Poggio



S. Chikkerur, C. Tan, T. Serre and T. Poggio



S. Chikkerur, C. Tan, T. Serre and T. Poggio

Questions?

serre@mit.edu

slides will be available online
model code available online