The quest for a theory of vision:

from the level framework (revised) to the invariances of the ventral stream

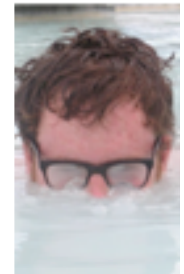
brains and machines

tomaso poggio
McGovern Institute
I2, CBCL, BCS,
CSAIL
MIT

ECCV, Firenze, 2012

Collaborators in recent work




F. Anselmi, J. Mutch , J. Leibo, L. Rosasco, A. Tacchetti

+

L. Isik, S. Ullman, S. Smale, C. Tan

Also: M. Riesenhuber, T. Serre, G. Kreiman, S. Chikkerur, A. Wibisono, J. Bouvrie, M. Kouh, J. DiCarlo, E. Miller, C. Cadieu, A. Oliva, C. Koch, A. Caponnetto ,D. Walther, U. Knoblich, T. Masquelier, S. Bileschi, L. Wolf, E. Connor, D. Ferster, I. Lampl, S. Chikkerur, G. Kreiman, N. Logothetis

Vision as Intelligence

The MIT Press

Vision

A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr
Foreword by Shimon Ullman
Afterword by Tomaso Poggio

David Marr's posthumously published *Vision* (1982) influenced a generation of brain and cognitive scientists, inspiring many to enter the field. In *Vision*, Marr describes a general framework for understanding visual perception and touches on broader questions about how the brain and its functions can be studied and understood. Researchers from a range of brain and cognitive sciences have long valued Marr's creativity, intellectual power, and ability to integrate insights and data from neuroscience, psychology, and computation. This MIT Press edition makes Marr's influential work available to a new generation of students and scientists.

In Marr's framework, the process of vision constructs a set of representations, starting from a description of the input image and culminating with a description of three-dimensional objects in the surrounding environment. A central theme, and one that has had far-reaching influence in both neuroscience and cognitive science, is the notion of different levels of analysis—in Marr's framework, the computational level, the algorithmic level, and the hardware implementation level.

Now, thirty years later, the main problems that occupied Marr remain fundamental open problems in the study of perception. *Vision* provides inspiration for the continui

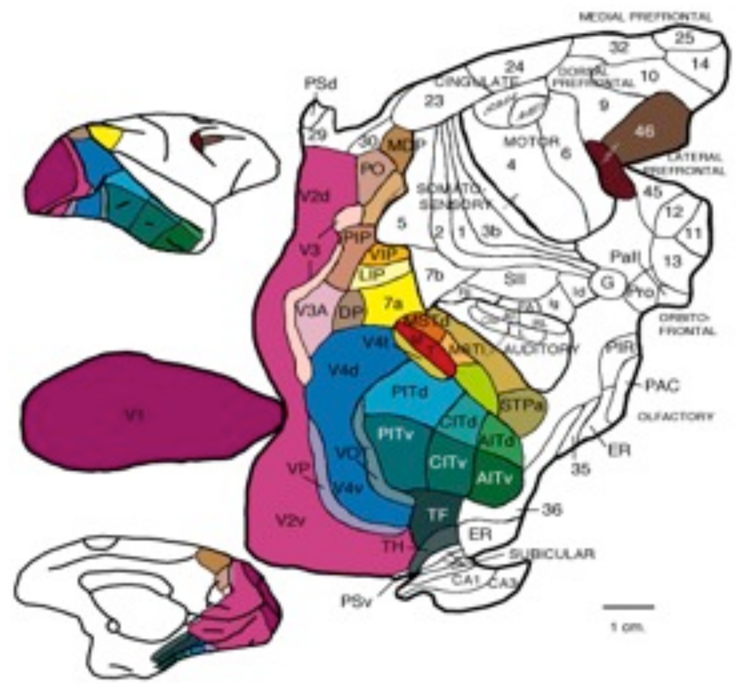
Vision: a very difficult computational problem, at several levels of understanding

Marr Crick, circa 1979)

Things that are easy for us are difficult for computers and viceversa (*Minsky rule*)

After 30 years a revision of Marr's and Poggio's Levels of Understanding Framework has appeared:

Poggio, T. [The Levels of Understanding framework, revised](), *MIT-CSAIL-TR-2012-014, CBCL-308*,



Van Essen & Anderson, 1990

The problem of intelligence (in particular, vision): how it arises in the brain and how to replicate it in machines

The problem of intelligence is one of the great problems in *science*, probably the *greatest*.

Research on intelligence by neuroscience and computer science (AI):

- a great intellectual mission
- will help medicine and develop more intelligent artifacts
- will improve the mechanisms for collective decisions

These advances will be critical to of our society's

- future prosperity
- education, health, security


IBM
RS/6000 SP

Google scholar

Search | Advanced Scholar Search
Scholar Preferences

Articles (include patents) Legal opinions and journals

Stand on the shoulders of giants

About Google Scholar - About Google - My Citations

KINECT
for XBOX 360.


想
PIENSE
THINK
THINK
$3,400
$1,200

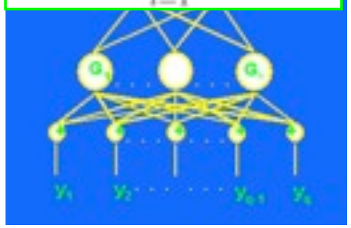

YouTube
iTunes
ตั้งค่า
App Store
แผงหนังสือพิมพ์
รายชื่อ
โทรศัพท์
Mail
Safari
เพลง
What can I help you with?


NAVY
505

Vision @CBCL, ~20 years ago

$$\min_{f \in H} \left[\frac{1}{\ell} \sum_{i=1}^{\ell} V(y_i, f(x_i)) + \mu \|f\|_K^2 \right]$$

$$f(x) = \sum_{i=1}^{l} c_i K(\mathbf{x}_i, \mathbf{x})$$



LEARNING THEORY + ALGORITHMS

Theorems on foundations of learning

Predictive algorithms

Sung & Poggio 1995, also Kanade& Baluja....

COMPUTATIONAL NEUROSCIENCE: models+experiments

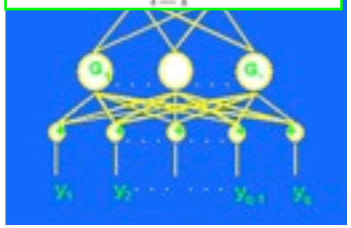
How visual cortex works

Vision @CBCL, ~20 years ago

Vision @CBCL, ~20 years ago

$$\min_{f \in H} \left[\frac{1}{\ell} \sum_{i=1}^{\ell} V(y_i, f(x_i)) + \mu \|f\|_K^2 \right]$$

$$f(x) = \sum_{i=1}^{l} c_i K(\mathbf{x}_i, \mathbf{x})$$



LEARNING THEORY
+
ALGORITHMS

Theorems on foundations of learning

Predictive algorithms

Face detection is now available in digital cameras (commercial systems)

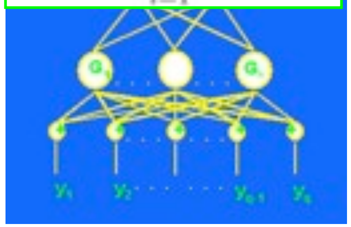
COMPUTATIONAL NEUROSCIENCE: models+experiments

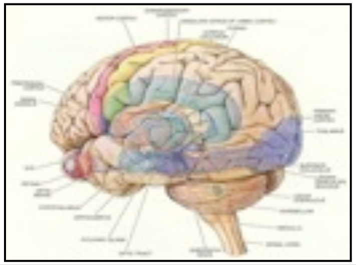
How visual cortex works

Vision @CBCL, ~18 years ago

$$\min_{f \in H} \left[\frac{1}{\ell} \sum_{i=1}^{\ell} V(y_i, f(x_i)) + \mu \|f\|_K^2 \right]$$

$$f(x) = \sum_{i=1}^{l} c_i K(\mathbf{x}_i, \mathbf{x})$$





LEARNING THEORY
+
ALGORITHMS

Theorems on foundations of learning

Predictive algorithms

Papageorgiou&Poggio, 1997, 2000
also Kanade&Scheiderman

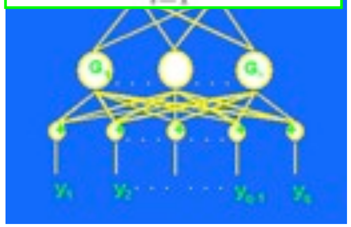
COMPUTATIONAL NEUROSCIENCE: models+experiments

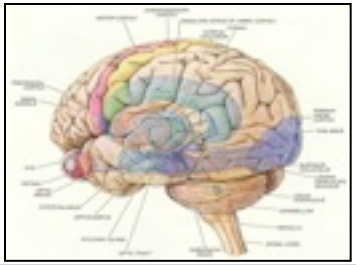
How visual cortex works

Vision @CBCL, ~18 years ago

$$\min_{f\in H}\left[\frac{1}{\ell}\sum_{i=1}^{\ell}V(y_i,f(x_i))+\mu\|f\|_K^2\right]$$

$$f(x)=\sum_{i=1}^{l}c_iK(\mathbf{x}_i,\mathbf{x})$$





LEARNING THEORY
+
ALGORITHMS

Theorems on foundations of learning

Predictive algorithms

Papageorgiou&Poggio, 1997, 2000
also Kanade&Scheiderman

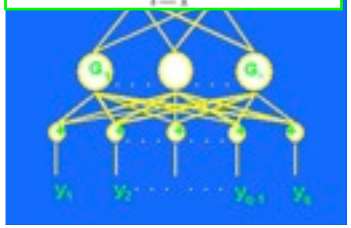
COMPUTATIONAL NEUROSCIENCE: models+experiments

How visual cortex works

Vision @CBCL, ~18 years ago

$$\min_{f\in H}\left[\frac{1}{\ell}\sum_{i=1}^{\ell} V(y_i, f(x_i)) + \mu\|f\|_K^2\right]$$

$$f(x) = \sum_{i=1}^{l} c_i K(\mathbf{x}_i, \mathbf{x})$$



LEARNING THEORY
+
ALGORITHMS

Theorems on foundations of learning

Predictive algorithms

Papageorgiou&Poggio, 1997, 2000
also Kanade&Scheiderman

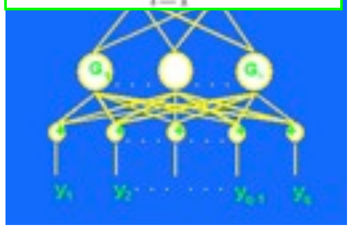
COMPUTATIONAL NEUROSCIENCE: models+experiments

How visual cortex works

Vision, ~ now

$$\min_{f \in H} \left[\frac{1}{\ell} \sum_{i=1}^{\ell} V(y_i, f(x_i)) + \mu \|f\|_K^2 \right]$$

$$f(x) = \sum_{i=1}^{l} c_i K(\mathbf{x}_i, \mathbf{x})$$



LEARNING THEORY
+
ALGORITHMS

Theorems on foundations of learning

Predictive algorithms

Pedestrian and car detection are also “solved” (commercial systems, *MobilEye, Jerusalem*)

COMPUTATIONAL NEUROSCIENCE: models+experiments

How visual cortex works

Mobileye (c) 2004

Mobileye (c) 2004
23.7m
14.9m
5.5m
1.50
440
-0.38

http://www.volvocars.com/us/all-cars/volvo-s60/pages/5-things.aspx?p=5

Recent successes in ML and AI: computer vision

Each of the systems above *can pass* its *narrow* version of the *Turing test*.

However, the full problem of vision

(and intelligence!)

is still *not* solved.

A “fullTuring’’ test for vision?

A “fullTuring’’ test for vision?

A "fullTuring'' test for vision?

A “fullTuring’’ test for vision?

A “fullTuring’’ test for vision?

A “fullTuring’’ test for vision?

A "fullTuring'' test for vision?

A “fullTuring’’ test for vision?

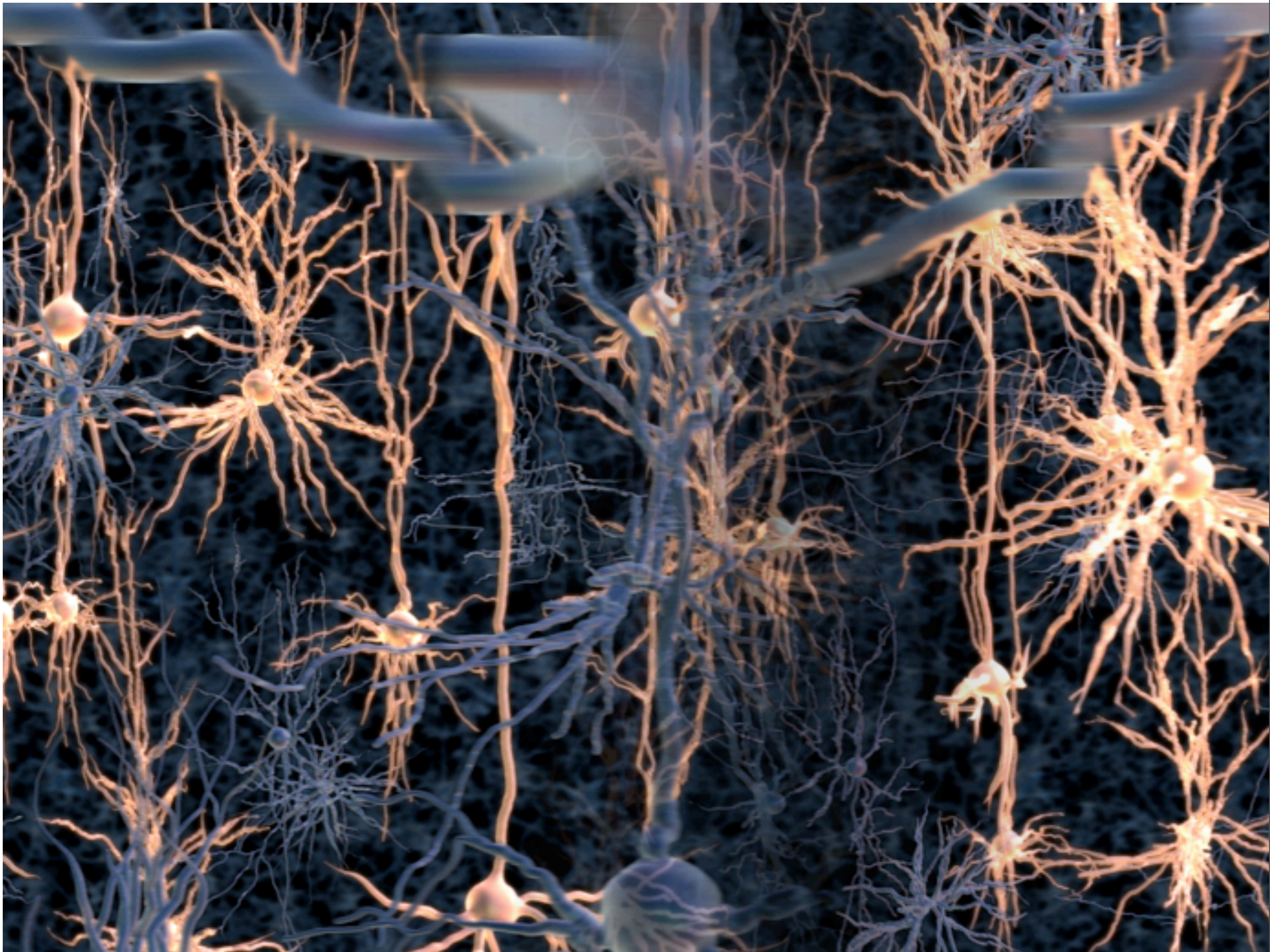
A “fullTuring’’ test for vision?

What is the person wearing the hat holding?

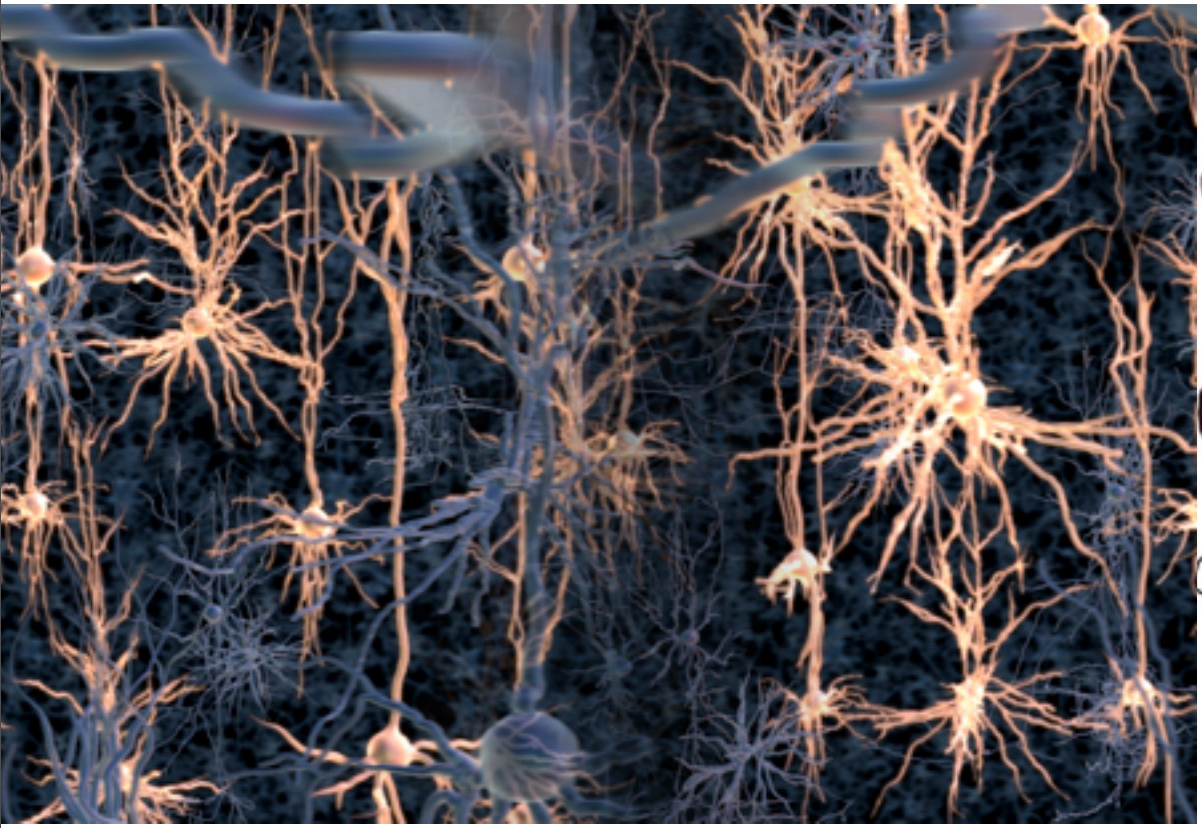
A “fullTuring’’ test for vision?

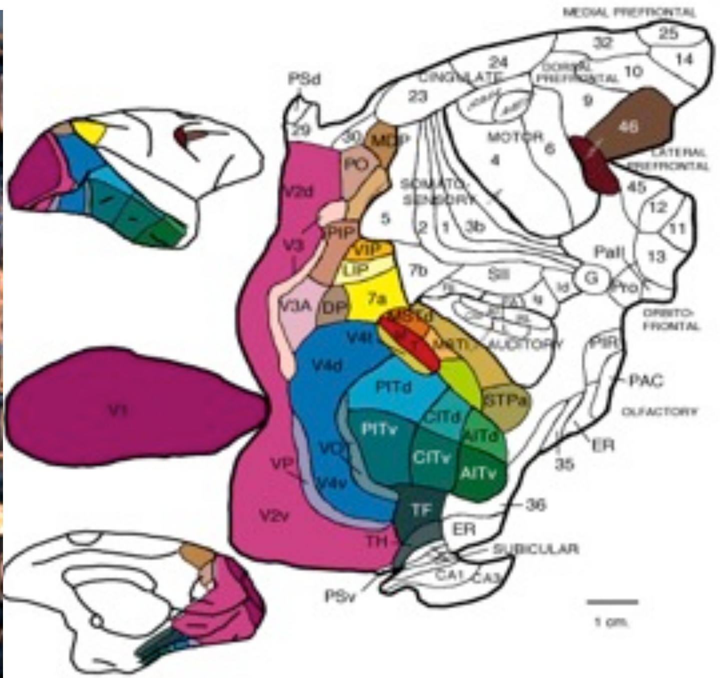
My personal bet: we may need to understand visual cortex (and the brain!) to achieve scene understanding at human level, and thereby develop systems that pass a *full Turing test.*

Thus: *science* of (natural) vision.



Vision in the Brain

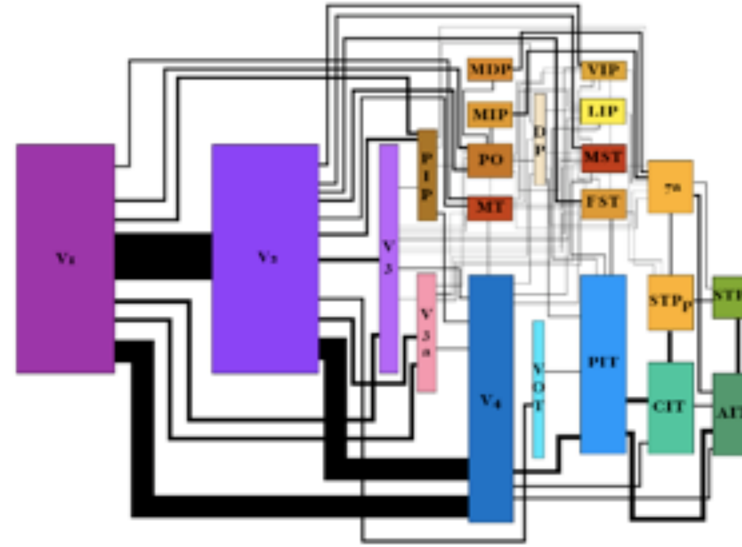


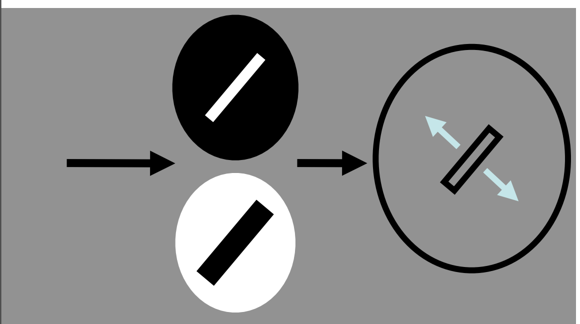


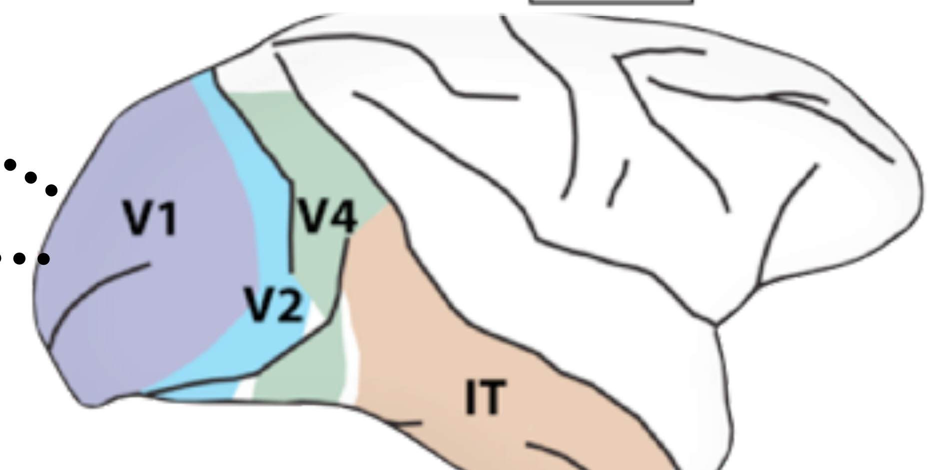
Van Essen & Anderson, 1990

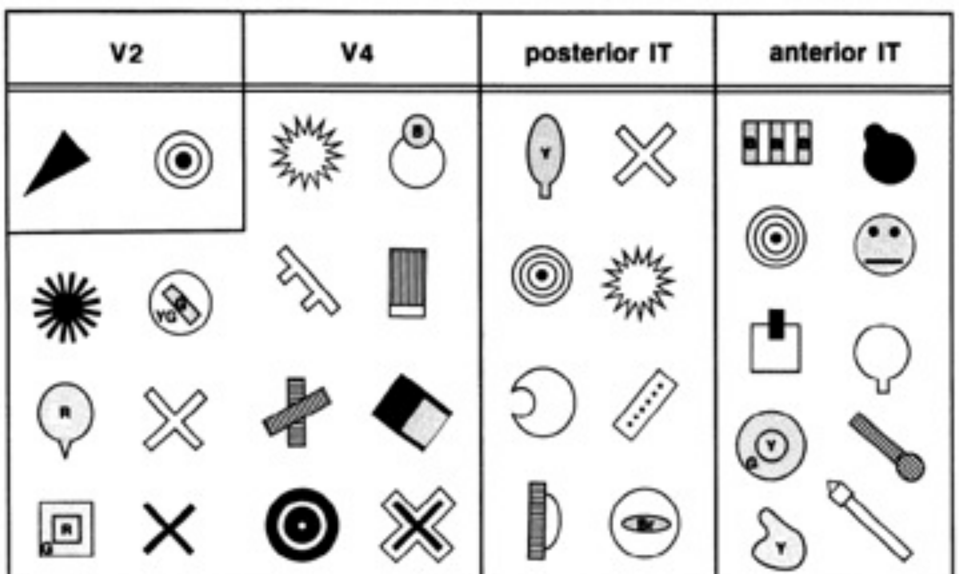
- Human Brain
  - 10^{10}-10^{11} neurons (~1 million flies)
  - 10^{14}- 10^{15} synapses
  - ~ 30% cortex is vision (more than for
  - language and any other modality)

Visual Object Recognition: the ventral stream (macaque)









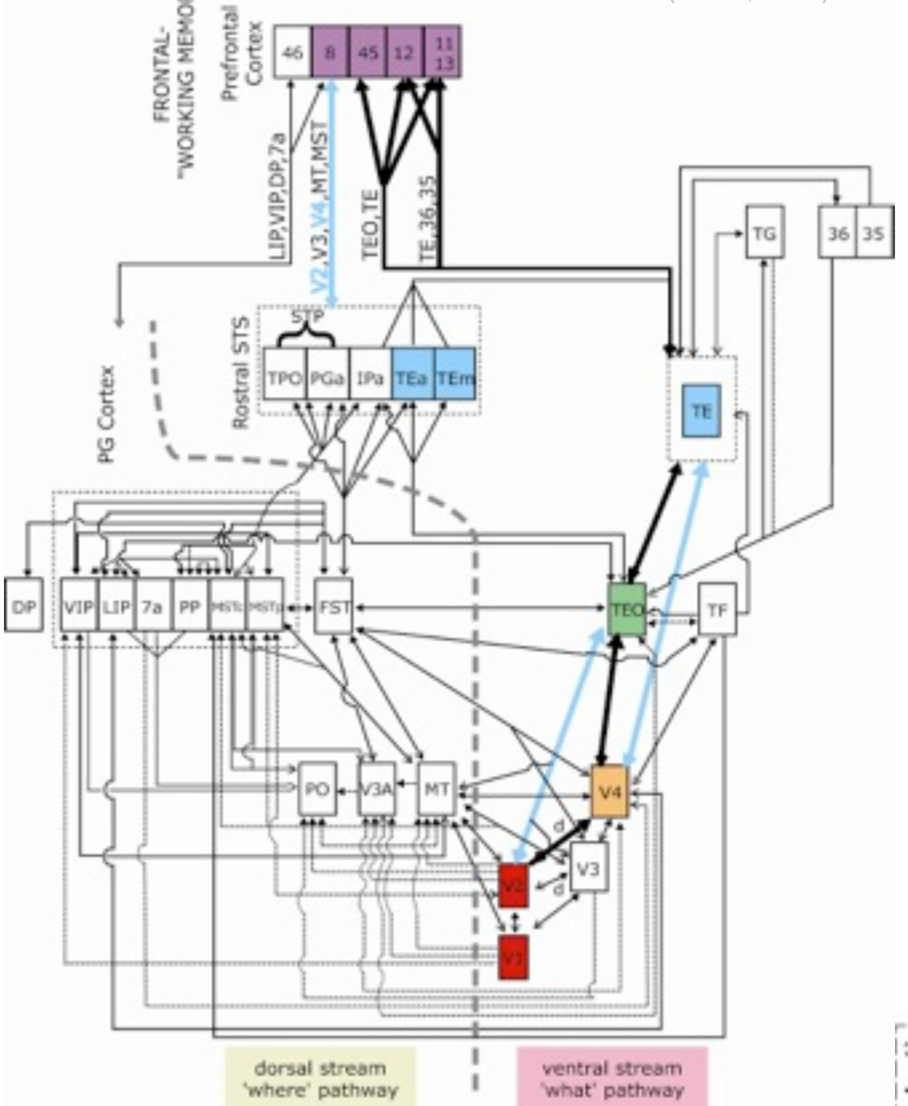
The ventral stream hierarchy: V1, V2, V4, IT

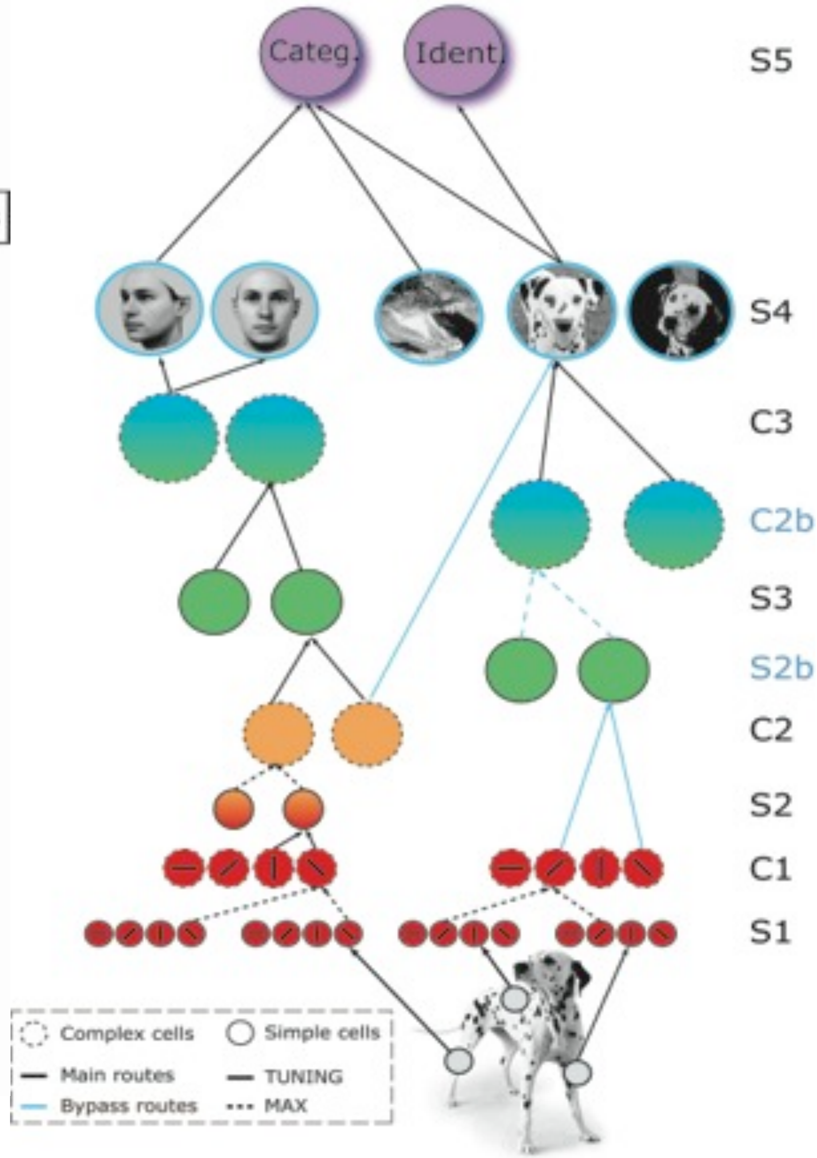
A gradual increase in the receptive field size, in the "**complexity**" of the preferred stimulus, in "**invariance**" to position and scale changes

Kobatake & Tanaka, 1994

Recognition in the Ventral Stream: "standard" feedforward model

*Modified from (Gross, 1998)

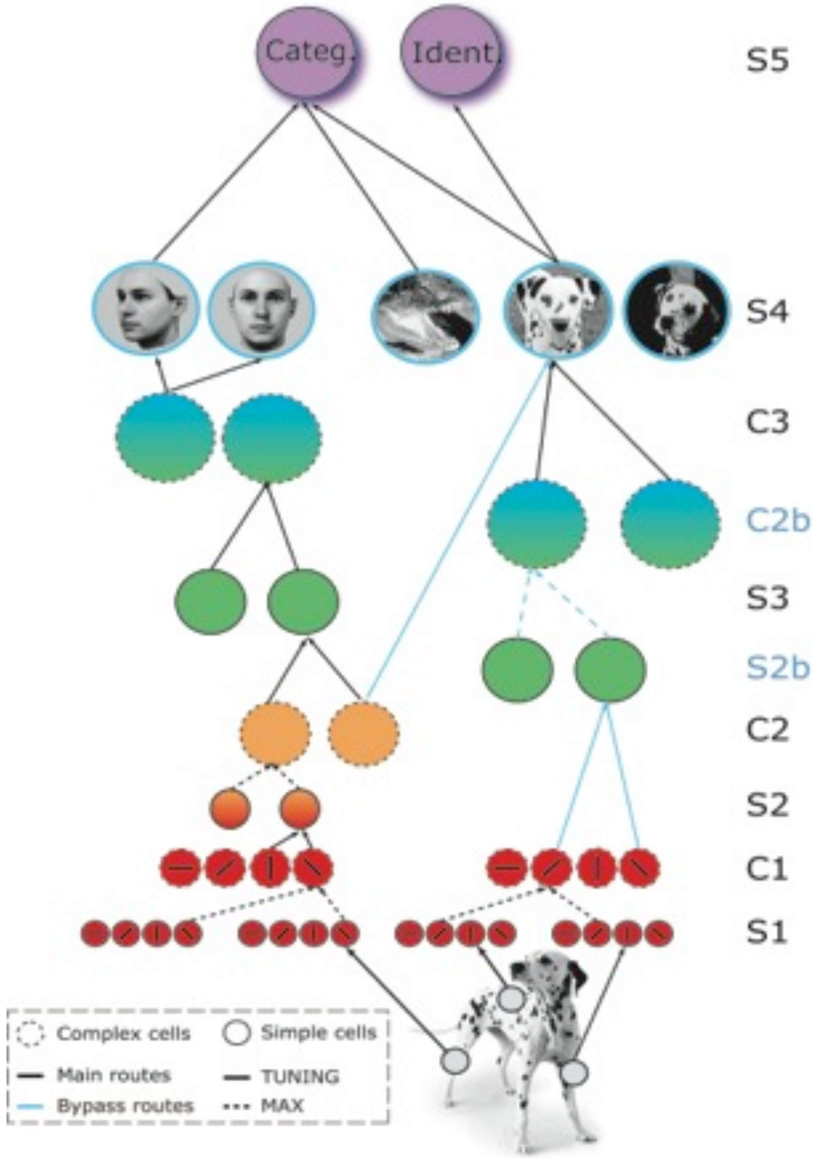




[software available online
with CNS (for GPUs)]

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

Recognition in Visual Cortex: "classical model", selective and invariant



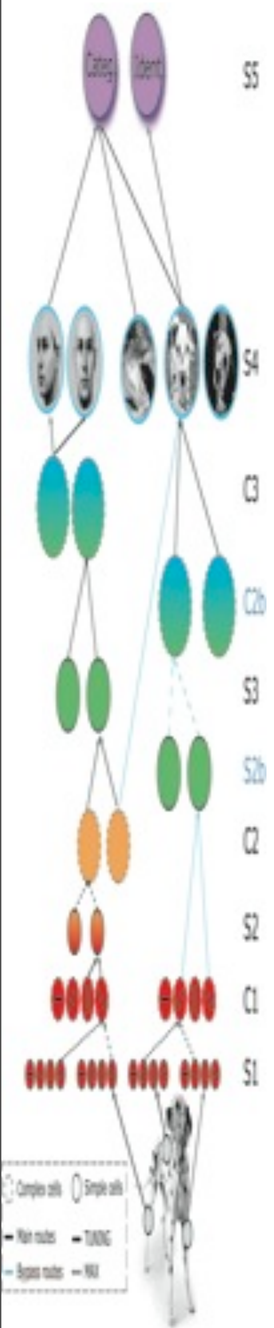
[software available online]

- It is in the family of "Hubel-Wiesel" models (Hubel & Wiesel, 1959: *qual.* **Fukushima**, 1980: *quant*; Oram & Perrett, 1993: *qual*; Wallis & Rolls, 1997; Riesenhuber & Poggio, 1999; Thorpe, 2002; Ullman et al., 2002; Mel, 1997; Wersing and Koerner, 2003; LeCun et al 1998: *not-bio*; Amit & Mascaro, 2003: *not-bio*; Hinton, LeCun, Bengio *not-bio;* Deco & Rolls 2006…)

- As a biological model of object recognition in the ventral stream – from V1 to PFC -- it is *perhaps* the most quantitatively faithful to known neuroscience data

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

Model "works": it accounts for physiology



Hierarchical Feedforward Models: is consistent with or predict neural data

V1:

Simple and complex cells tuning (Schiller et al 1976; Hubel & Wiesel 1965; Devalois et al 1982)

MAX-like operation in subset of complex cells (Lampl et al 2004)

V2:

Subunits and their tuning (Anzai, Peng, Van Essen 2007)

V4:

Tuning for two-bar stimuli (Reynolds Chelazzi & Desimone 1999)

MAX-like operation (Gawne et al 2002)

Two-spot interaction (Freiwald et al 2005)

Tuning for boundary conformation (Pasupathy & Connor 2001, Cadieu, Kouh, Connor et al., 2007)

Tuning for Cartesian and non-Cartesian gratings (Gallant et al 1996)

IT:

Tuning and invariance properties (Logothetis et al 1995, paperclip objects)

Differential role of IT and PFC in categorization (Freedman et al 2001, 2002, 2003)

Read out results (Hung Kreiman Poggio & DiCarlo 2005)

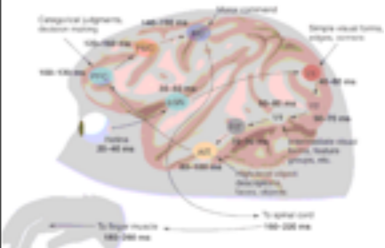
Pseudo-average effect in IT (Zoccolan Cox & DiCarlo 2005; Zoccolan Kouh Poggio & DiCarlo 2007)

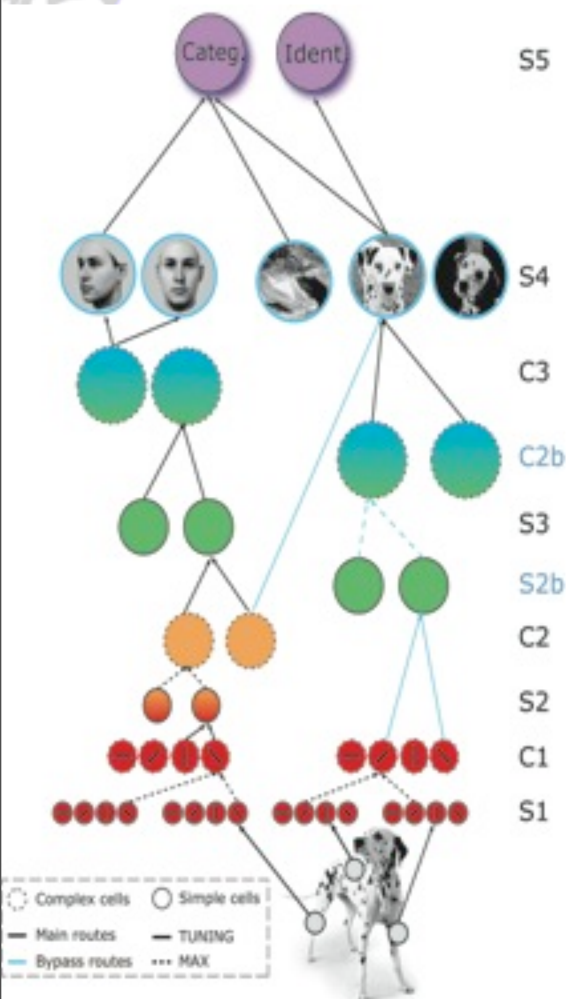
Human:

Rapid categorization (Serre Oliva Poggio 2007)

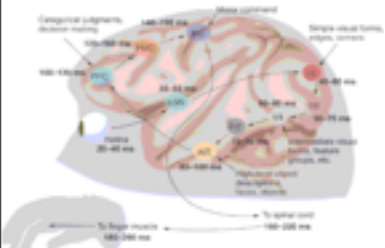
Face processing (fMRI + psychophysics) (Riesenhuber et al 2004; Jiang et al 2006)

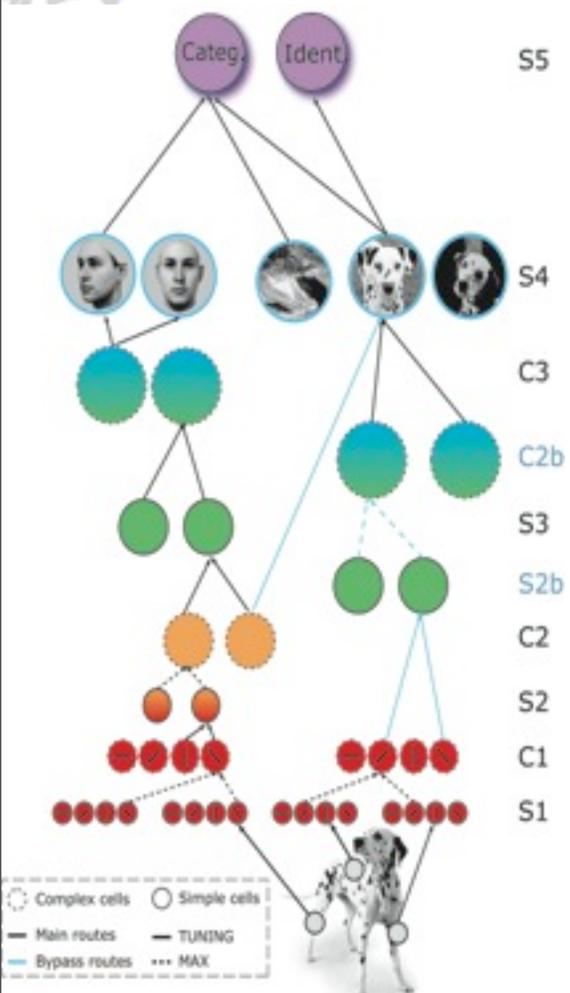
Model "works": it accounts for psychophysics



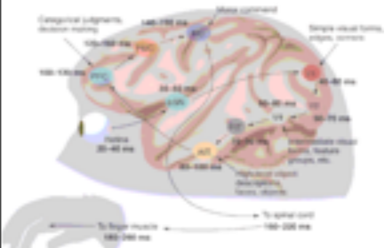


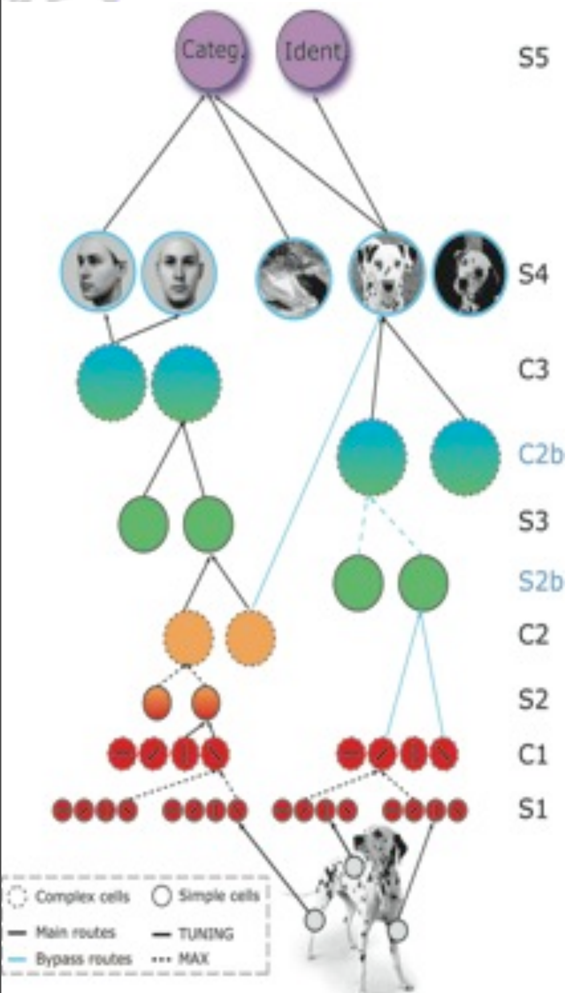
Model "works": it accounts for psychophysics





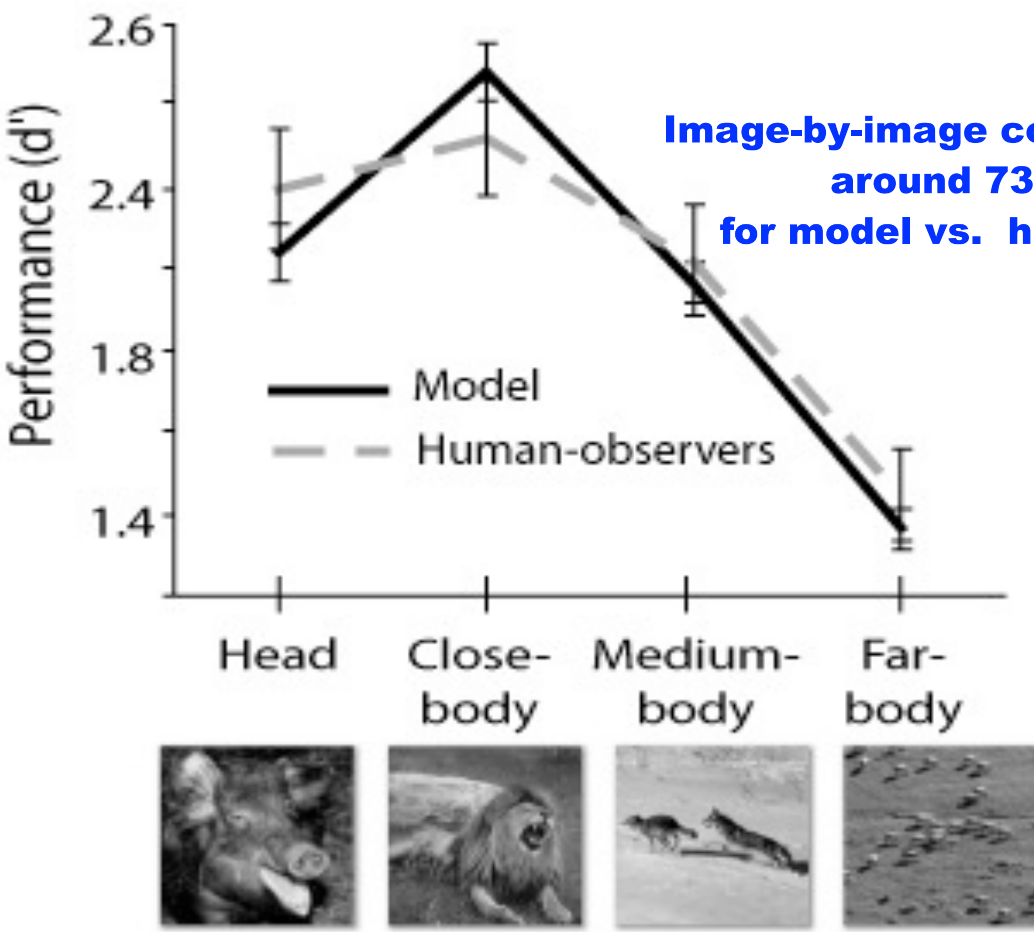
Model "works": it accounts for psychophysics



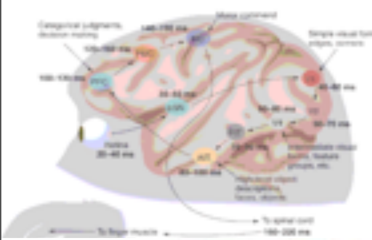


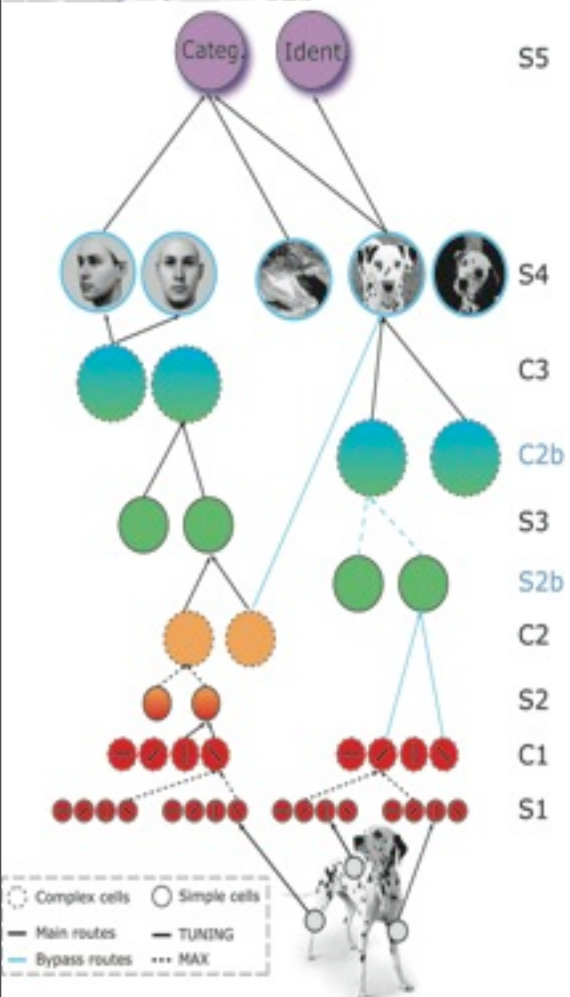
Feedforward Models:
"predict" rapid categorization
(82% model vs. 80% humans)

Image-by-image correlation:
around 73%
for model vs. humans)



Model "works": it performs well at computational level

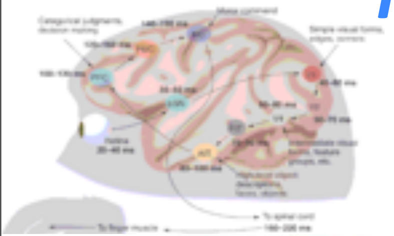


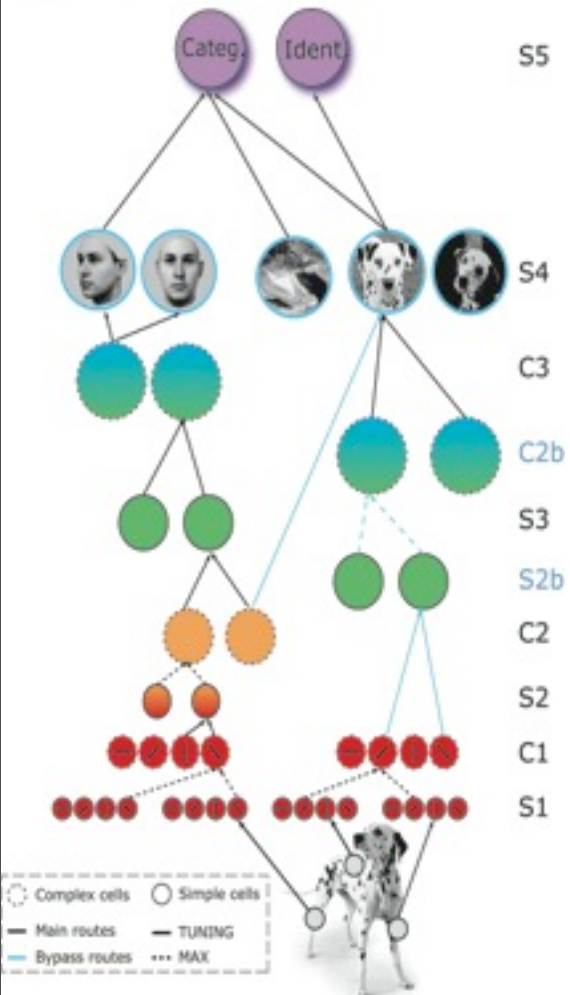


Models of the <u>*ventral stream*</u> in cortex
perform well compared to
engineered computer vision systems (in 2006)
on several databases

Bileschi, Wolf, Serre, Poggio, 2007

Model "works": it performs well at computational level

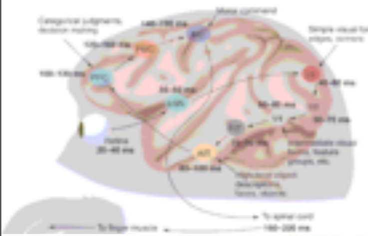


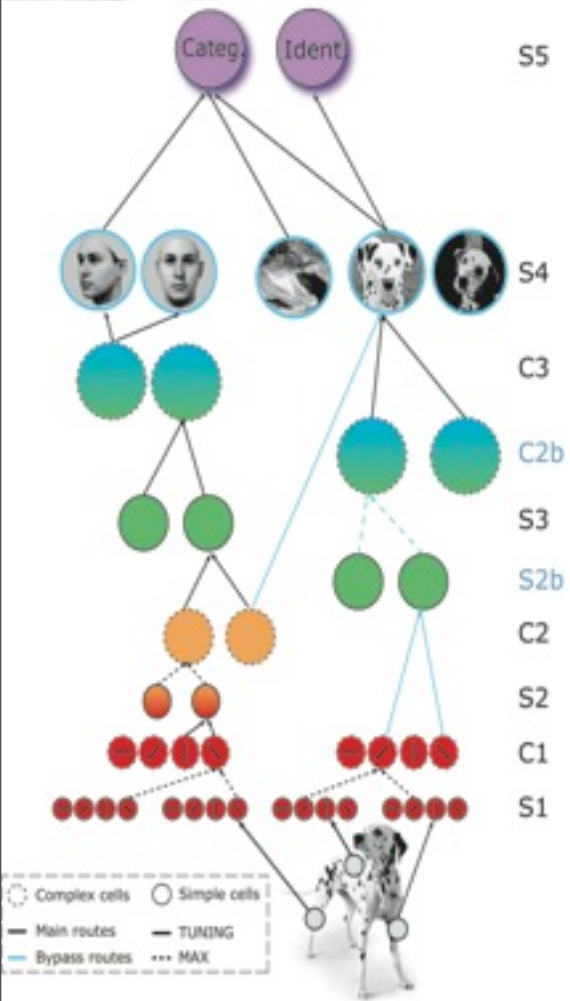


Models of the *ventral stream* in cortex
perform well compared to
engineered computer vision systems (in 2006)
on several databases

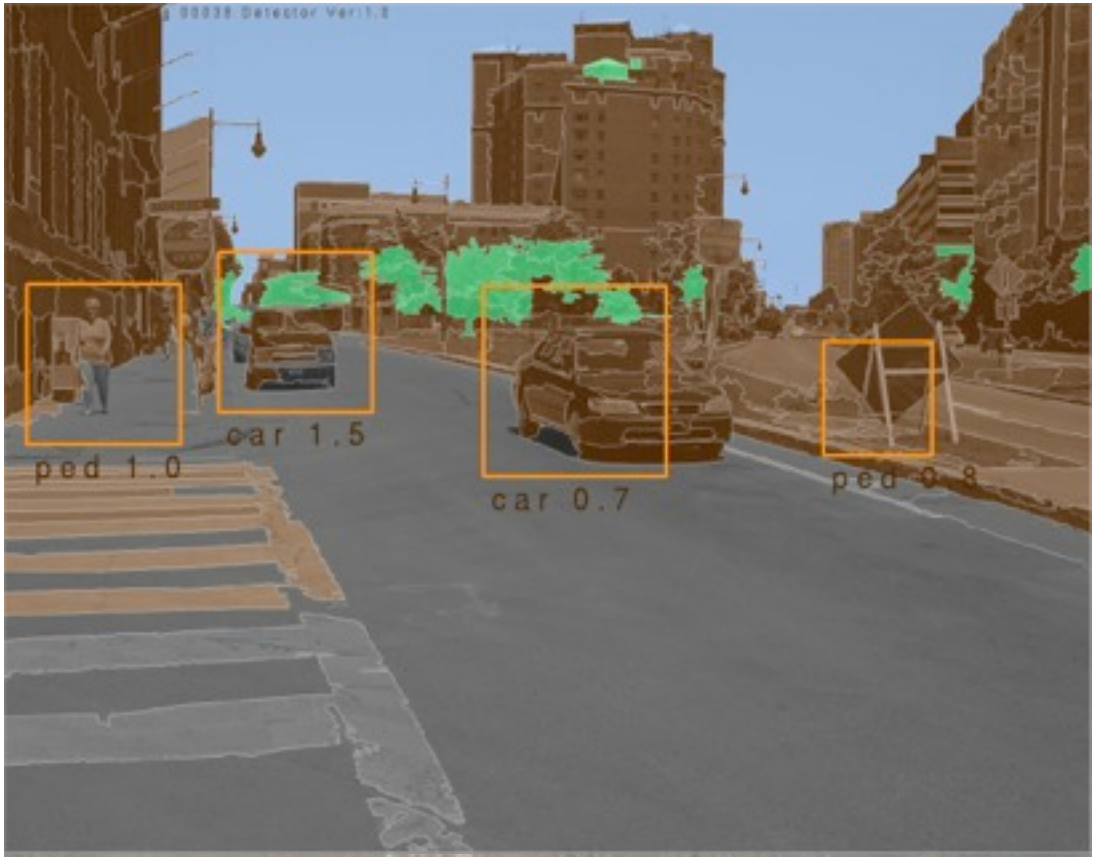
Bileschi, Wolf, Serre, Poggio, 2007

Model "works": it performs well at computational level





Models of the *ventral stream* in cortex perform well compared to engineered computer vision systems (in 2006) on several databases



Bileschi, Wolf, Serre, Poggio, 2007

Model "works": it performs well at computational level

Performance	
human agreement	72%
proposed system	77%
commercial system	61%
chance	12%

Models of cortex lead to better systems for action recognition in videos: automatic phenotyping of mice

Jhuang , Garrote, Yu, Khilnani, Poggio, Mutch Steele, Serre, Nature Communicatons, 2010

Model "works": it performs well at computational level

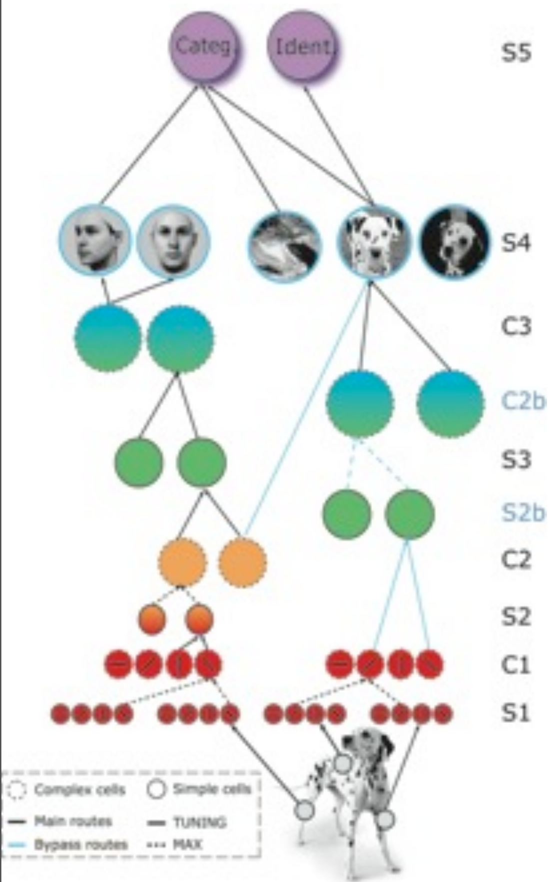
Performance

human agreement	72%
proposed system	77%
commercial system	61%
chance	12%

Models of cortex lead to better systems for action recognition in videos: automatic phenotyping of mice

Jhuang , Garrote, Yu, Khilnani, Poggio, Mutch Steele, Serre, Nature Communicatons, 2010

Visual Cortex: models and theories



Forward, HMAX-type models work well (summarizing+predicting physiology AND in terms of performance in visual recgnition) but...

For 10years+
I did not manage to understand how model works....

So...we need theories -- not only models!

Found Comput Math (2010) 10: 67–91
DOI 10.1007/s10208-009-9049-1

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS

Mathematics of the Neural Response

S. Smale · L. Rosasco · J. Bouvrie · A. Caponnetto · T. Poggio

A theory (unpublished) of the ventral stream: too nice to be true?

THE COMPUTATIONAL MAGIC OF THE VENTRAL STREAM: TOWARDS A THEORY

Tomaso Poggio[⋆,†] (section 4 with Jim Mutch[⋆]; appendix 7.2 with Joel Leibo[⋆] and appendix 7.9 with Lorenzo Rosasco†)
⋆ *CBCL, McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA, USA*
† Istituto Italiano di Tecnologia, Genova, Italy

Nature Precedings, doi:10.1038/npre.2011.6117.1 July 16, 2011: outdated version;

new (today's talk) will be posted in the future.

Motivation

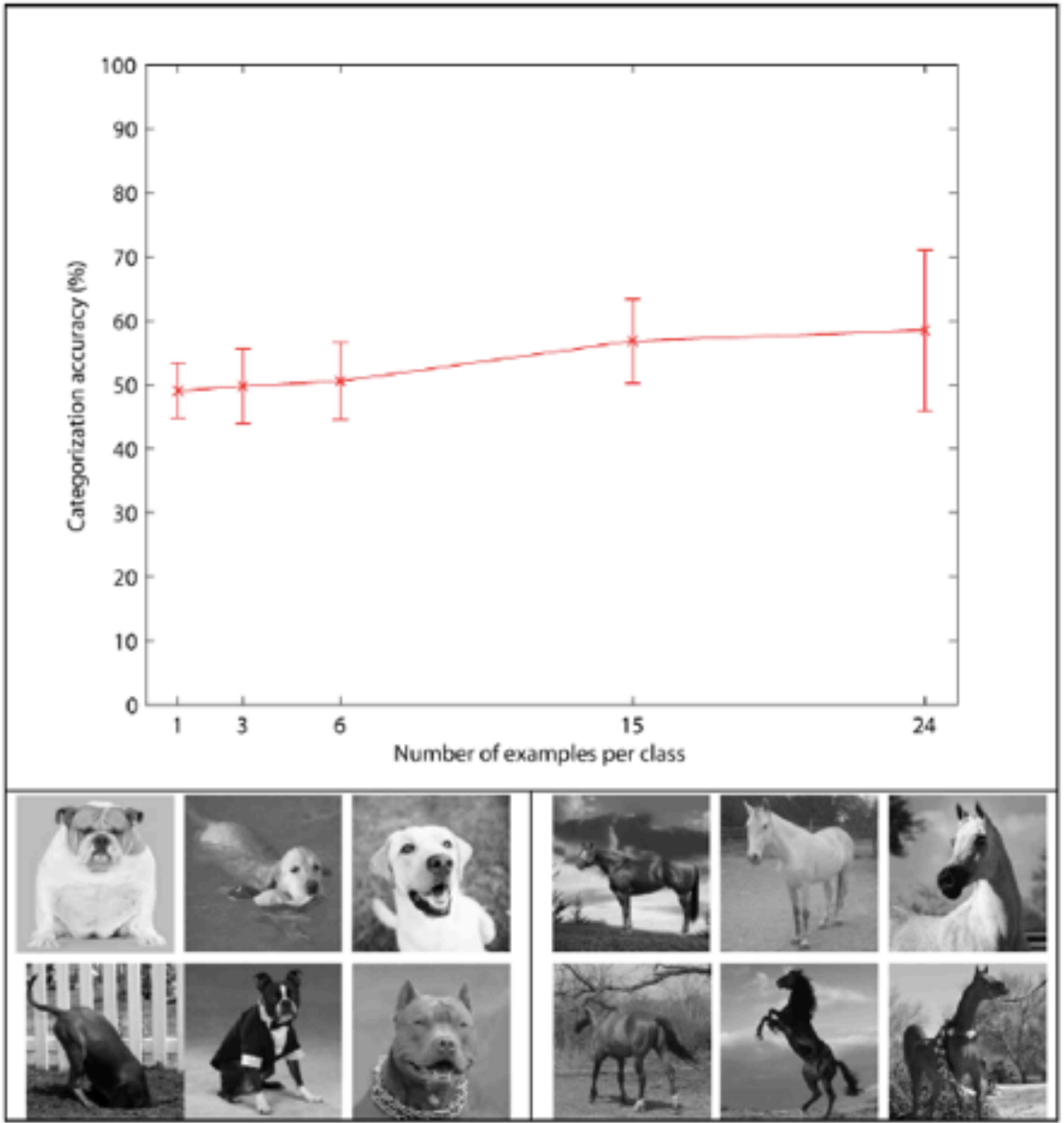
Cardinality of the universe of possible images generated by an object:

- Assuming: a granularity of a few minutes of arc + a visual field of say 10 degrees
- then
  - $10^3 - 10^5$ different images of the same object from x, y translations
  - $10^3 - 10^5$ from rotations in depth
  - a factor of $10 - 10^2$ from rotations in the image plane
  - another factor of $10 - 10^2$ from scaling.

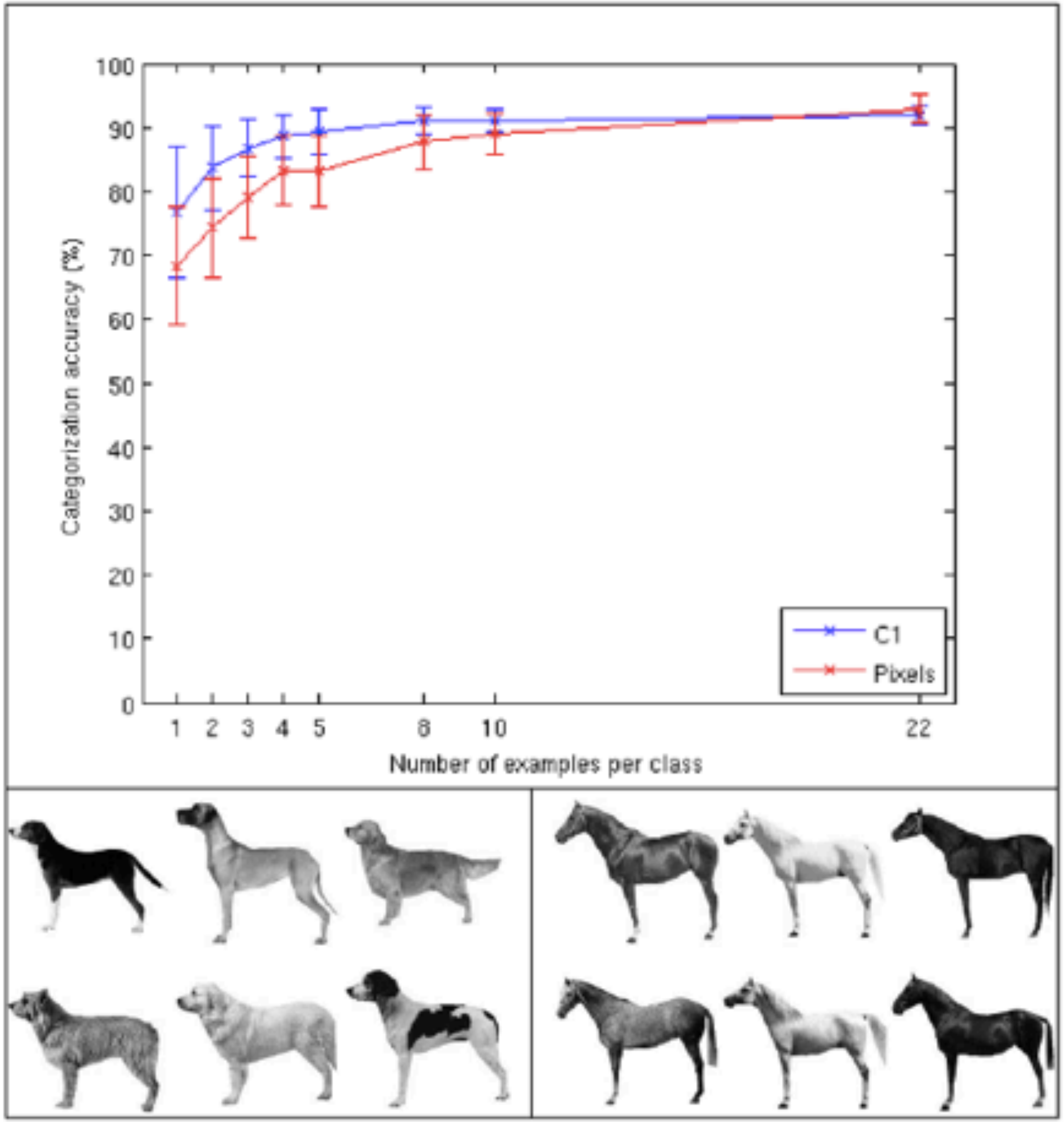
  for a total $10^8 - 10^{14}$ distinguishable images for a single object.

How many different types of dogs exist within the "dog" category? No more than, say, $10^2 - 10^3$. Thus it is greater win to be able to factor out the geometric transformations than the intracategory differences.

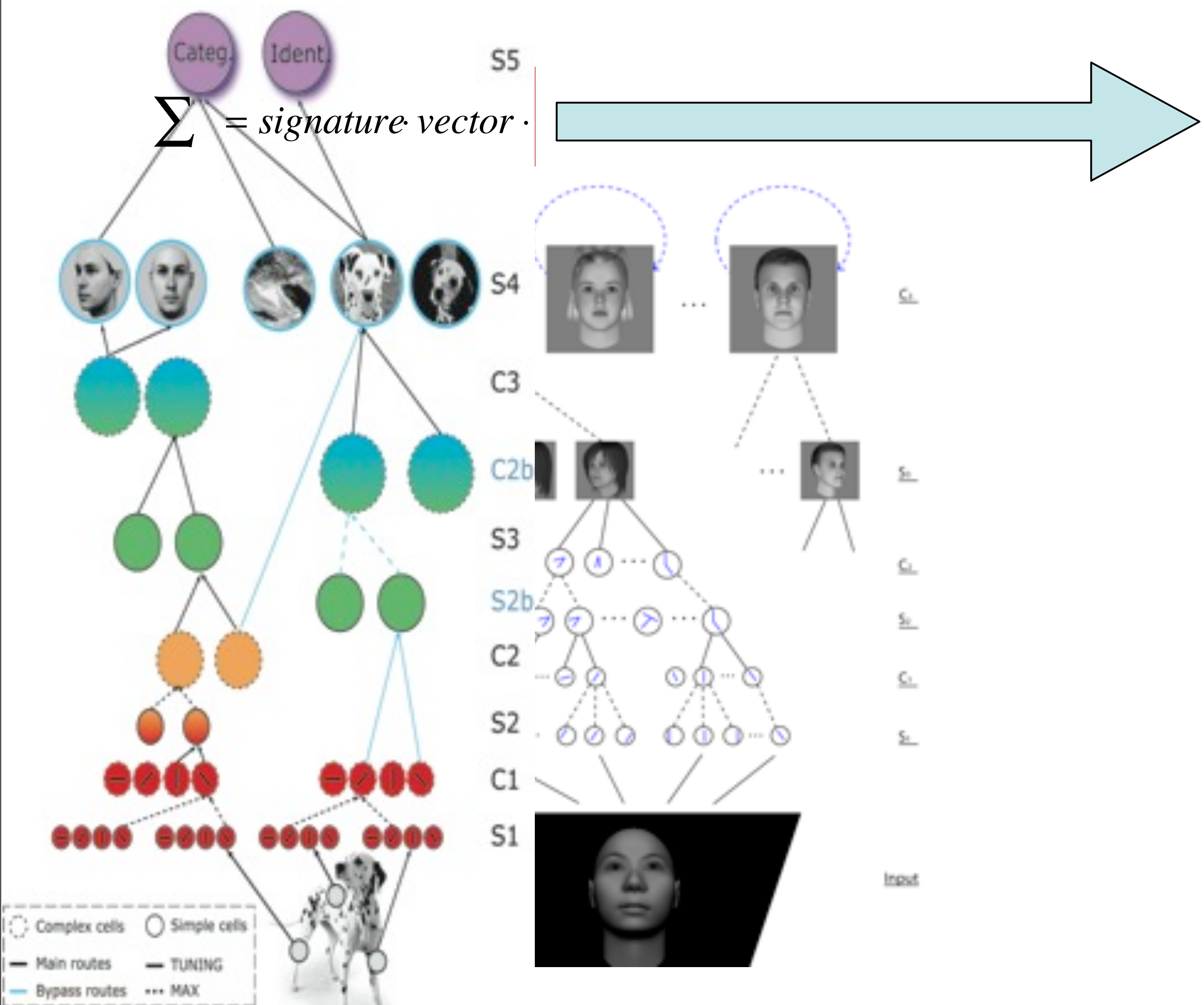
Motivation: transformations may be a main difficulty for (biological) object recognition

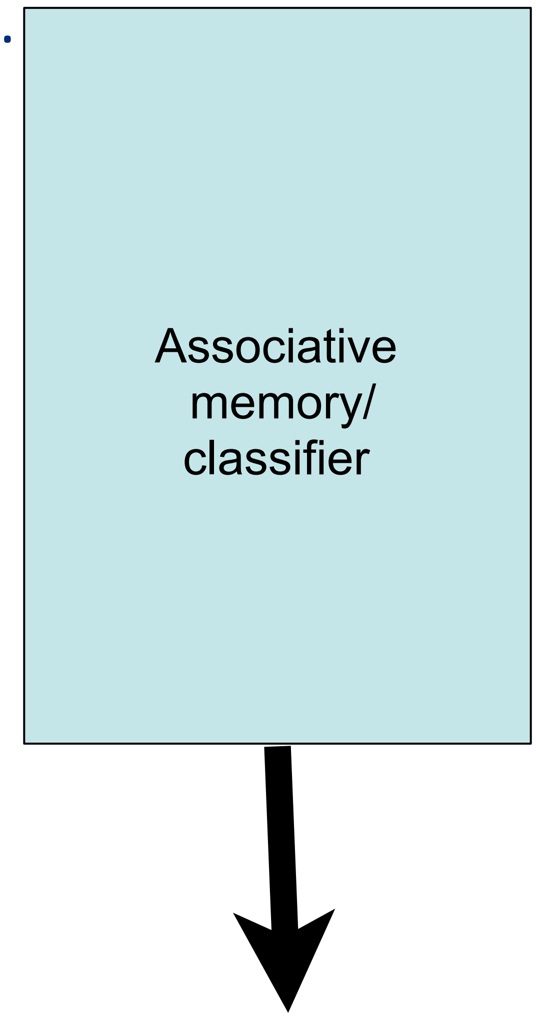


Motivation: transformations may be the main difficulty for (biological) object recognition



Conjecture: the key computational problem “solved” by the ventral stream is object recognition from a single training image, invariant to geometric transformations.





A theory of invariant recognition in the ventral stream

Theory starts with a simple but novel assumption: the computational goal of the feedforward path in the ventral stream – from V1, V2, V4 and to IT – is to discount image transformations after learning them during development.

- Part I:
  - Layer of biologically plausible memory-based modules learn transformations from unsupervised visual experience.
  - Invariance Lemma: 1-layer module provides a *signature* automatically invariant to global affine transformations.
  - *Whole-parts theorem*: a hierarchical architecture provides local invariance for small patches of the image and global invariance for the whole image.
- Part II:
  - Spectral properties of the hierarchical architectures.
  - *Linking Conjecture*: cells compute online the eigenvectors of the covariance of their inputs during development and store them in their synaptic weights.
  - Cortical equation predicts tuning of cells in cortex. Simple cells in V1 are predicted to be oriented Gabor-like wavelets.
- Part III:
  - Class-specific modules.
  - Prediction re: macaque cortex face patches.
  - Mirror symmetric cells in AL

Some of the questions answered by the theory

- What is the main computational task of the ventral stream?

- Why do simple cells in V1 have Gabor tuning curves? What is the reason for the “universal constants” of simple cells?

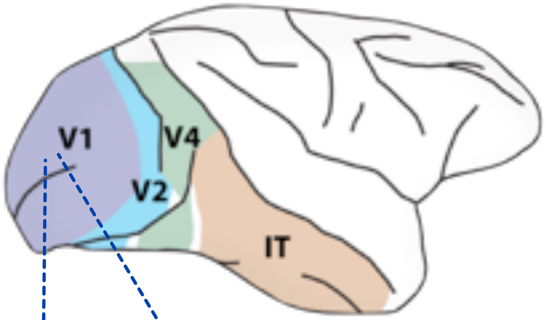
- Why do cells in the AL face patch in the macaque have mirror symmetric tuning curves?

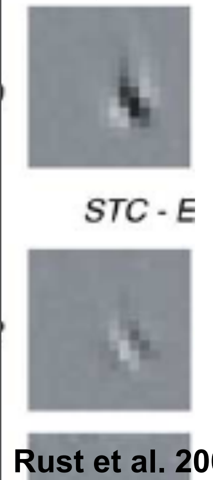
Key theorems

- *Invariance lemma*
- *Extension to groups of Johnson-Lindenstrauss*
- *Uniqueness of polynomials of group averages*
- *Whole-Parts theorem: robustness to diffeomorphisms via hierarchies*
- *Cortical Equation: Gabor-like solutions*
- *Symmetry-breaking by motion during development*

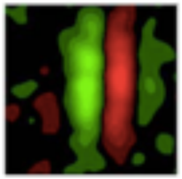
Gabor-like tuning with "universal constants" in simple cells (Jones and Palmer, 1987; Ringach, 2002; Niell and Stryker, 2008):

why?



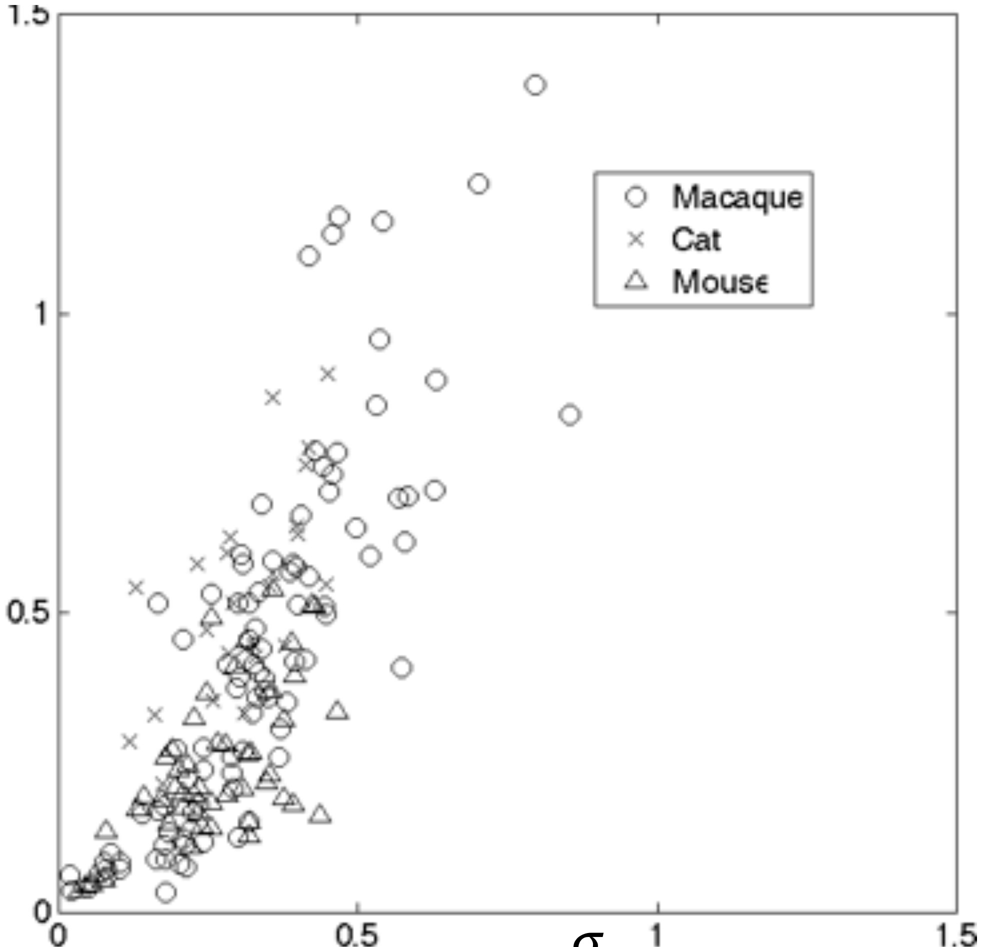


Rust et al. 2005



Carandini

$$n_y == \frac{\sigma_y}{\lambda}$$



$$n_x = \frac{\sigma_x}{\lambda}$$

2 Different stages in the theory

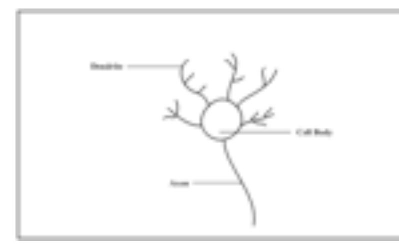
1. **development:** **learning of transformations (and acquiring invariance) via motion sequences**

2. **mature stage:** **acquire an object (single image) and (later) recognize it (from single image)**

Image representation in the ventral stream

- Images can be represented by a set of functionals on the image, eg a set of measurements
- Neuroscience suggests that natural functionals for neurons to compute are dot products between "image patches" and another image patch (called *template)* which is stored in terms of synaptic weights

$$x \bullet t$$



- Later: a natural way to store templates is to update online according to Hebb's rule t according to quasi-Hebb rule

$$t_{k+1} - t_k = x \bullet y + n(t, y) \qquad y = x \bullet t$$

Templates and signature

We look at a finite ($|\mathcal{T}| = D < \infty$) set of measurement on the image such as

$$\langle I, t_i \rangle, \quad i = 1, ..., D$$

Thus an image I is represented by a set of neurons as a *signature vector* of I defined with respect to the templateset $\mathcal{T}$:

$$\Sigma_I = \begin{pmatrix} \langle I, t_1 \rangle \\ \langle I, t_2 \rangle \\ \vdots \\ \langle I, t_D \rangle \end{pmatrix}$$

A motivation for signatures: the Johnson-Lindenstrauss theorem (features do not matter much!)

For any set V of n points in $\mathbb{R}^d$, there exists a map $P : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in V$

$$(1-\epsilon) \parallel u - v \parallel^2 \leq \parallel Pu - Pv \parallel^2 \leq (1+\epsilon) \parallel u - v \parallel^2$$

where the map P is a random projection on $\mathbb{R}^k$ and

$$kC(\epsilon) \geq ln(n), \quad C(\epsilon) = \frac{1}{2}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})$$

JL suggests that good image representations for classification and discrimination of n objects can be provided by k dot products with *random* templates!

Geometric transformations

We define as geometric transformations of the image I transformations $T \circ I$ such that:

$$T \circ I(x, y) = I(x', y')$$

An example of T is the affine case, eg

$$\mathbf{x}' = A\mathbf{x} + \mathbf{t}_{\mathbf{x}}$$

Learning to be invariant for any new object

Suppose that (during development) one template and all its transformations are stored

$$g_0 t, g_1 t \ldots g_n t$$

Then if the group is compact

$$I \cdot g_0 t, I \cdot g_1 t \ldots, I \cdot g_n t \sim g_0^{-1} I \cdot t, g_1^{-1} I \cdot t, \ldots, g_n^{-1} I \cdot t$$

that is the two sets of dot products are the *same* apart from ordering. Thus any *group average* will provide a number which is invariant to transformations of the image even if the image has been seen only once.

Templatebook

Suppose we have a finite and closed set of transformations and templates:

$$g_1, \ldots, g_{|G|} \in G, \quad t_1, \ldots, t_D \in \mathcal{T}$$

The memory-based module stores sequences of transformed templates, called Templatebook,

$$\mathbb{T}_{t_1,\ldots,t_D} = \begin{pmatrix} g_0 t_1, g_1 t_1, & \ldots, g_{|G|} t_1 \\ \vdots & \\ g_0 t_D, g_1 t_D, & \ldots, g_{|G|} t_D \end{pmatrix}$$

Invariance lemma

If we consider a set of finite transformations g_j, $j = 1, \ldots |G|$ the function Average Energy:

$$\phi_{t_i}(I) = \frac{1}{|G|} \sum_{j=1}^{|G|} |\langle I, g_j t_i \rangle|^2$$

is invariant, that is:

$$\phi_{t_i}(g_j I) = \phi_{t_i}(I), \quad \forall g_j \in G$$

Other examples of invariant group functionals are

- Max: $\phi_{t_i}(I) = max_j \langle I, g_j t_i \rangle$
- Average: $\phi_{t_i}(I) = \frac{1}{|G|} \sum_{j=1}^{|G|} \langle I, g_j t_i \rangle$

We define these kind of functions pooling functions.

Invariant Signature

We have therefore that the vector

$$\tilde{\Sigma}_f = \begin{pmatrix} \phi_{t_1}(f) \\ \phi_{t_2}(f) \\ \vdots \\ \phi_{t_D}(f) \end{pmatrix}$$

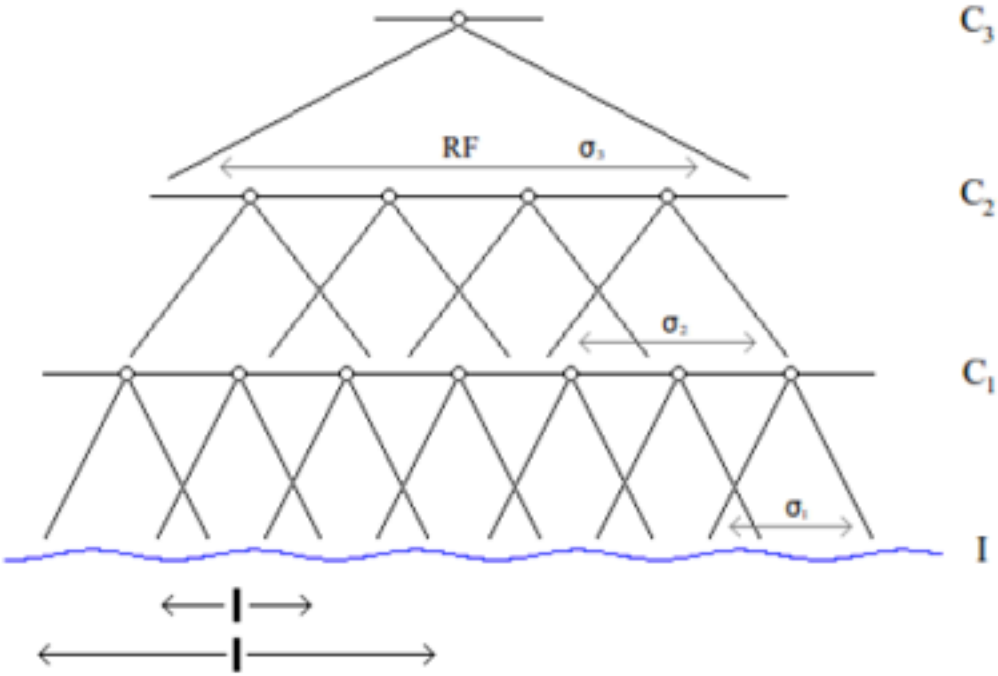
gives an invariant signature for group transformations g_j.

One-layer architectures: pro and contra

- Pro
  - Robustness to clutter: *matched filter* is optimal in L_2 sense.
- Contra
  - Fragility to diffeomorphisms (locally affine transforms)
  - Whole vs parts invariance
  - Memory storage issues for many object classes

Whole-parts invariance

Consider the network composed by different receptive fields Λ:



c^n is the complex cell response at layer n and σ_n may be equal or larger than σ_{n-1}.
Notice that the patch of the image seen at layer n is at least as large as the patch seen at level $n-1$, that is $\sigma^n_{eff} \geq \sigma^{n-1}_{eff}$. In general σ^n_{eff} increases (rapidly) with n. The following properties are true:

Whole-Parts theorem

Theorem

Assume a network of $\bigwedge$ moduli which is shift-invariant. Then if the complex response to a transforming image I is covariant at a layer n, there exists $m > n$ s.t. at layer m, the complex cell response is invariant, that is:

$$c^n(\bar{g}I)(g) = c^n(I)(\bar{g}^{-1}g)$$
$$\Rightarrow c^m(\bar{g}I)(g) = c^m(I)(g)$$

In other words the complex response of a transformed image patch becomes invariant when the transformation is within the receptive field σ_{eff} at level m.

Part II
Linking Conjecture

- The memory in a layer of cells (such as simple cells in V1) is stored in the weights of the connections between the neurons and the inputs (from the previous layers).
- Instead of storing a sequence of discrete frames (the templatebook) as assumed in Part I, online learning is more likely, with synaptic weights being incrementally modified **during development**.
- Hebbian-like synapses exist in visual cortex.
- Hebbian-like learning is equivalent to an online algorithm computing PCAs.
- As a consequence, the tuning of simple cortical cells is dictated by the PCAs of the templatebook.

Tuning and eigenvectors of covariance matrix

The consequence of assuming Hebb synapses is that the tuning of the neuron converges to the top eigenvector of the covariance matrix of the "frames" of the movie of objects transforming. The convergence follow the Oja equation

$$t_{k+1} - t_k = x \bullet y + n(t, y) \qquad y = x \bullet t$$

Different cells "see" translations in different directions.

Gaussian aperture: the cortical equation

Define as templatebook T the matrix where each column represents a template t shifted relative to the previous column and "seen through a Gaussian aperture". The image is assumed to be 1D. The image seen through a Gaussian aperture is then $t(y-x)g(x)$ when the image is shifted by y. We are led to the following problem: find the eigenvectors of the symmetric matrix $G^T T^T T G$ where G is a diagonal matrix with the values of a Gaussian along the diagonal. We consider the continuous version of the problem, that is the eigenvalue problem

$$\int dx g(y)g(x)\psi_n(x)\int ds\bar{t}(y-s)\bar{t}(s-x) = \lambda_n\psi_n(y)$$

which is rewritten as **the cortical equation**

$$\int dx g(y)g(x)t(y-x)\psi_n(x) = \lambda_n\psi_n(y).$$

with $t(x)$ being the autocorrelation function of the template.
This is an equation describing the development of simple cells in V1; it describes development of other cortical layers as well.

The cortical equation: general properties

The equation

$$\int dx g(y) g(x) \psi_n(x) t(y-x) = \lambda_n \psi_n(y) \quad (3)$$

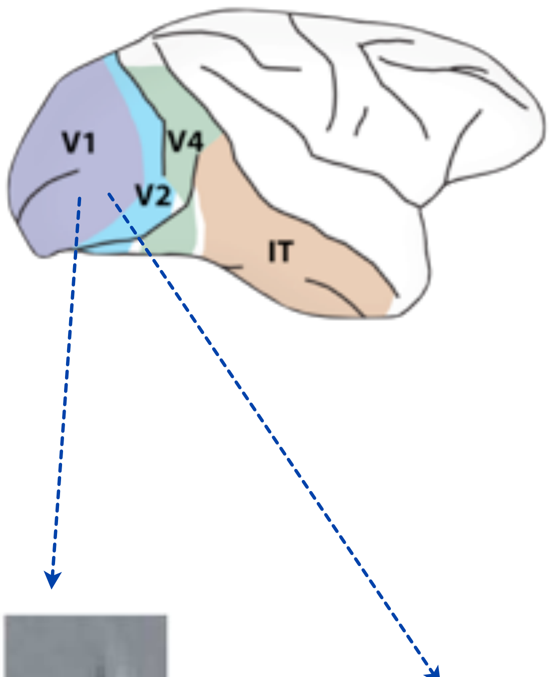
with $t(x)$ being real and even, is an eigenvalue equation for an integral operator with symmetric positive definite kernel $K(x,y) = g(y)g(x)t(y-x)$, $K(x,y) = K(y,x)$. The kernel is measurable on the square $[a,b] \times [a,b]$, since

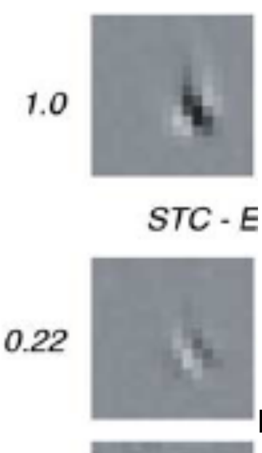
$$\int_a^b \int_a^b |K(x,y)|^2 dxdy < \infty. \quad (4)$$

Then the Hilbert-Schmidt theory of symmetric kernels applied to (3) implies

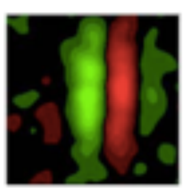
- the eigenfunctions corresponding to distinct eigenvalues are orthogonal
- the eigenvalues are real and positive
- there is at least one eigenvalues and one eigenfunctions (when K is almost everywhere nonzero) and in general a countable set of eigenfunctions.

Cortical equation in 2D:
natural images, Gabor-like receptive fields

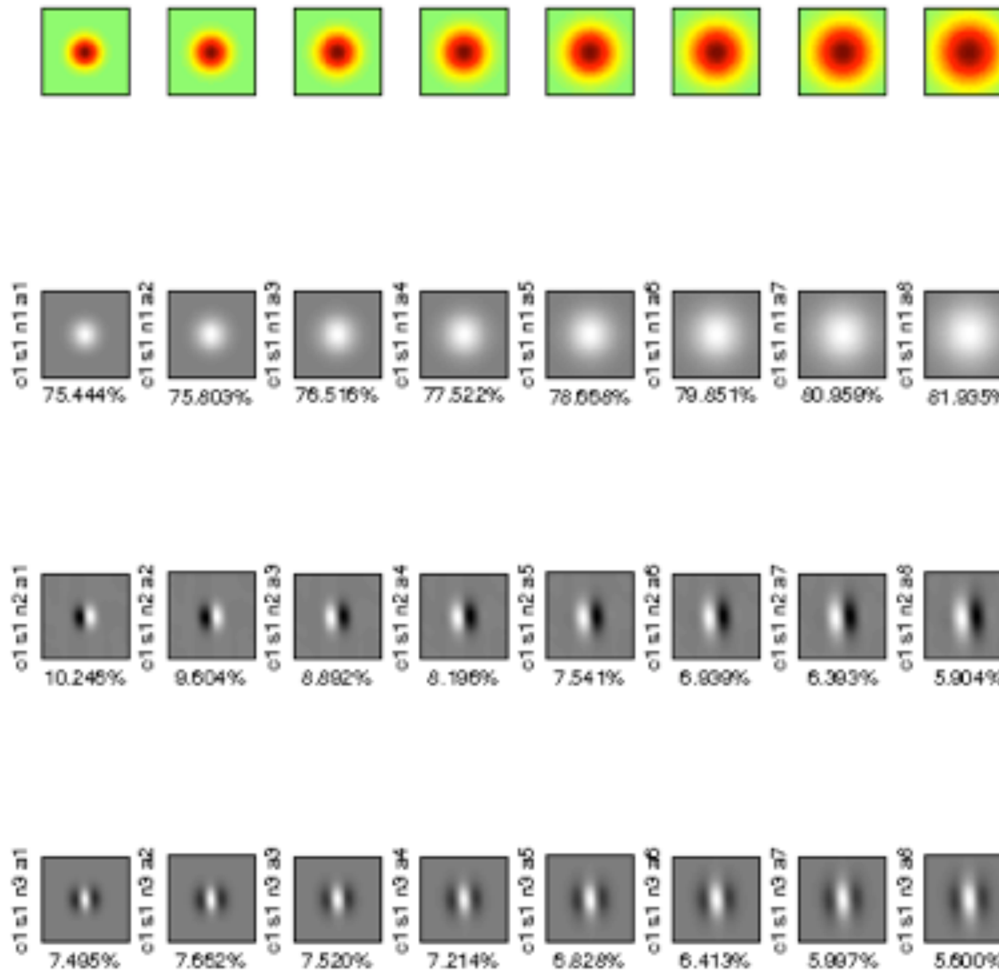




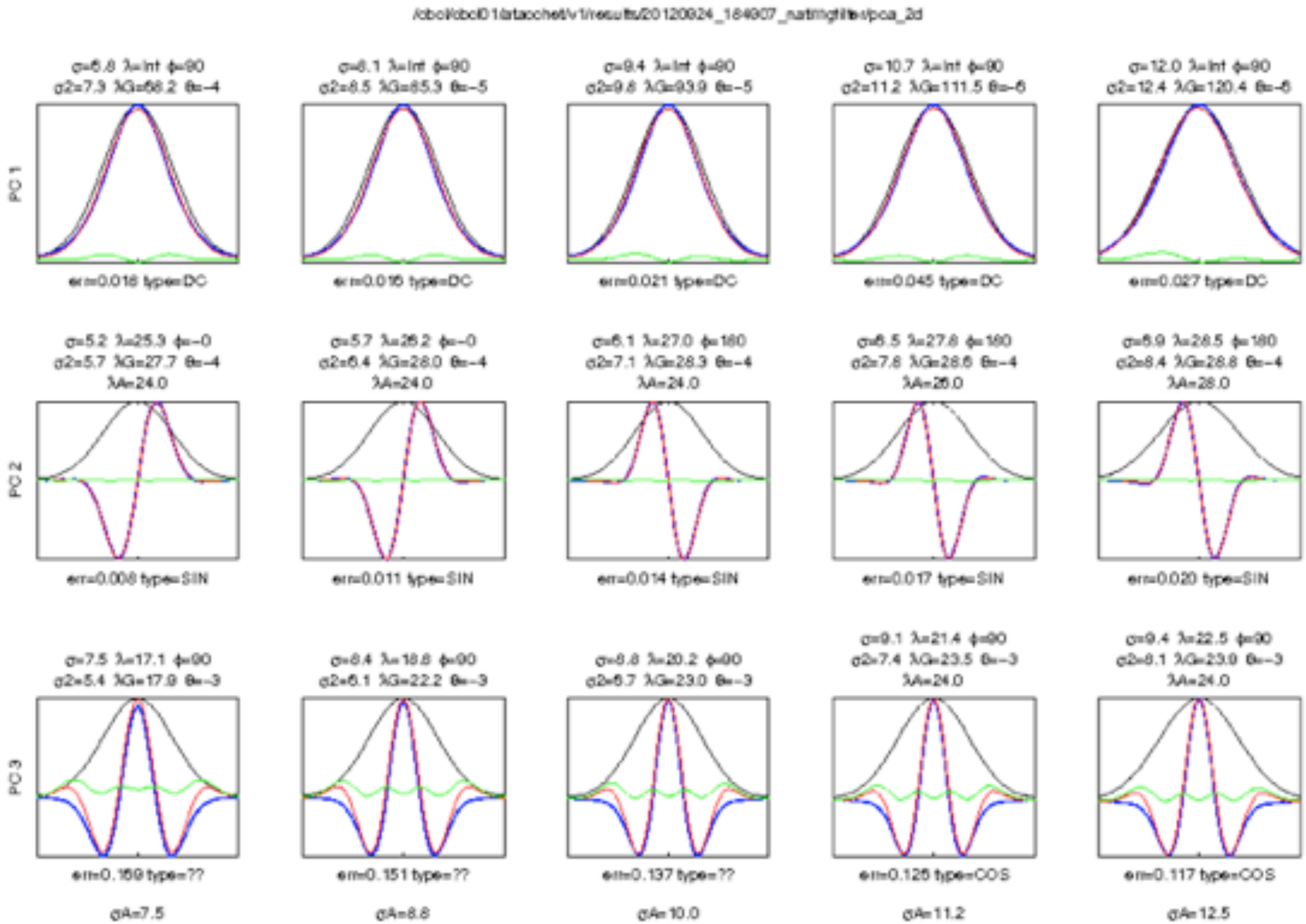
Rust et al. 2005



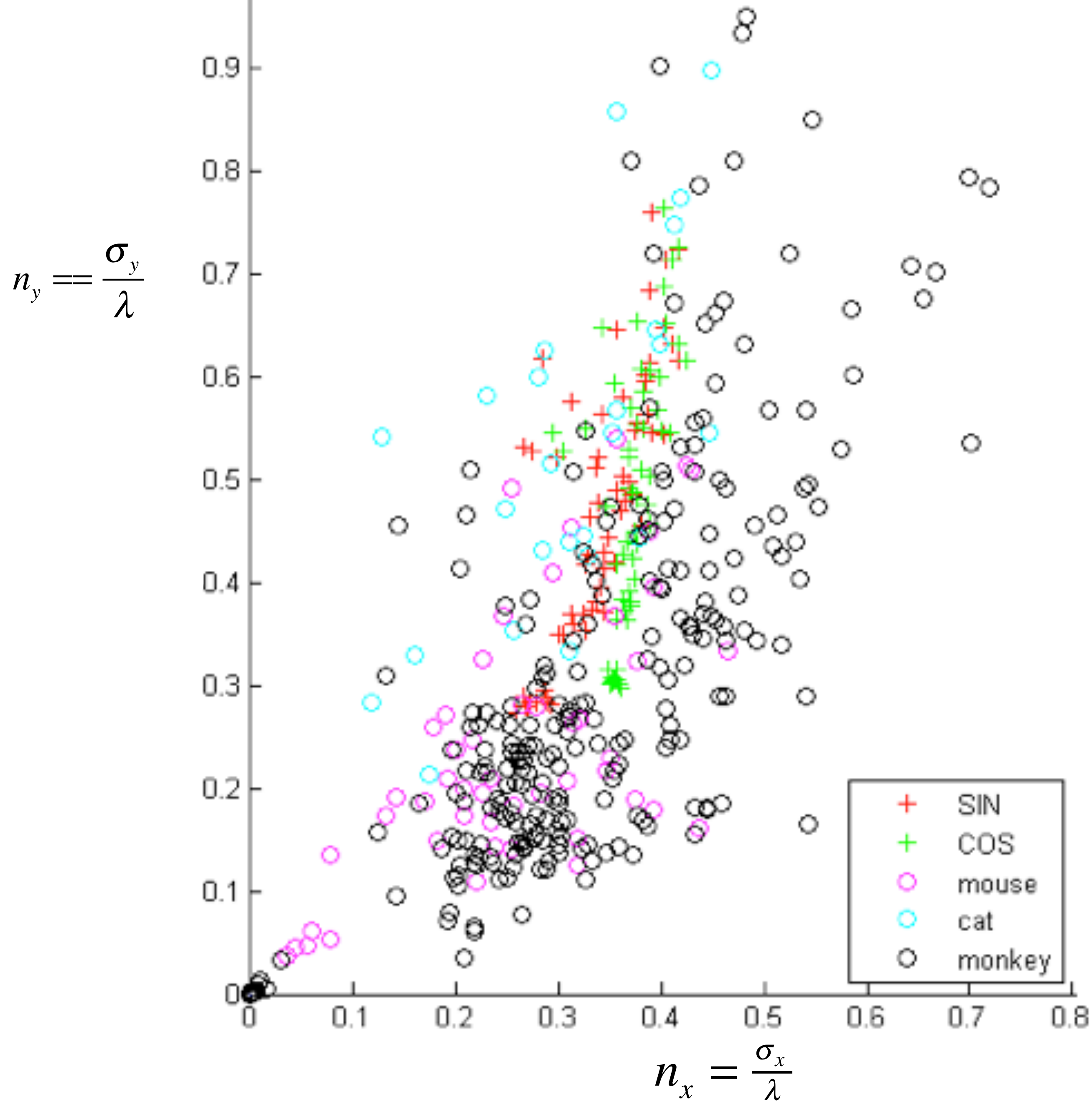
Carandini

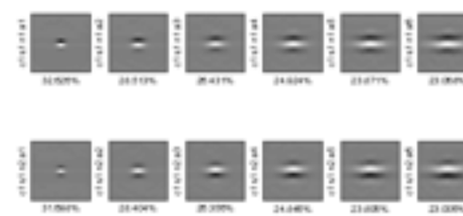


Cortical equation in 2D:
natural images, Gabor-like receptive fields

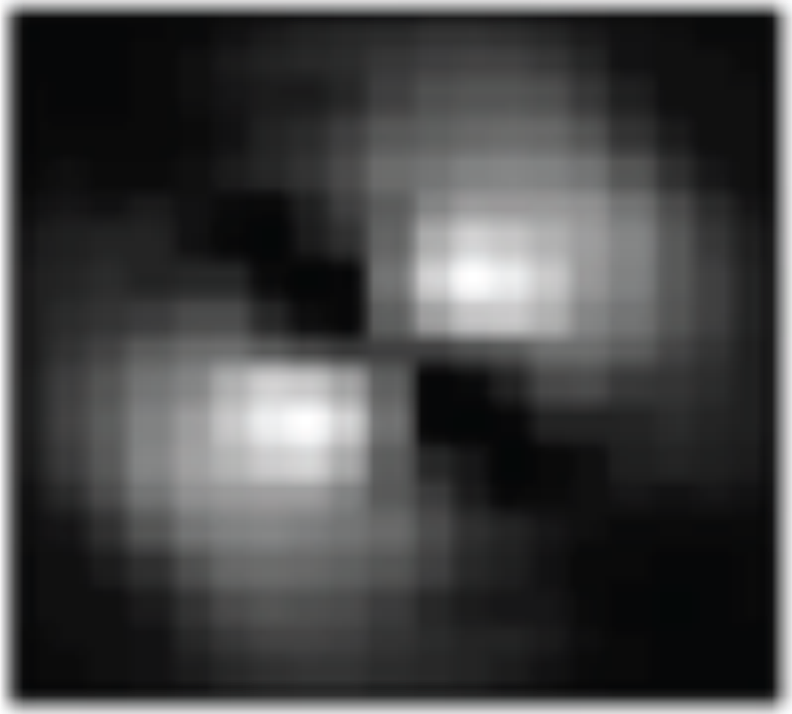


Too nice to be true!





Beyond V1, towards V2 and V4



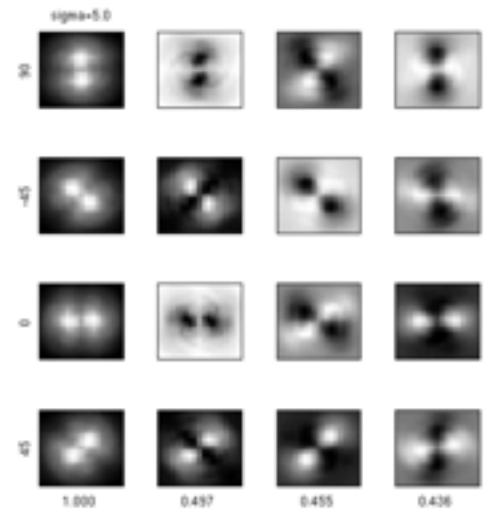
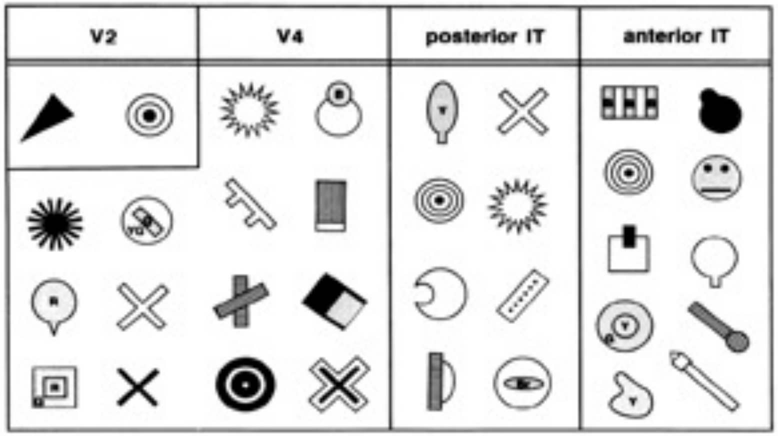
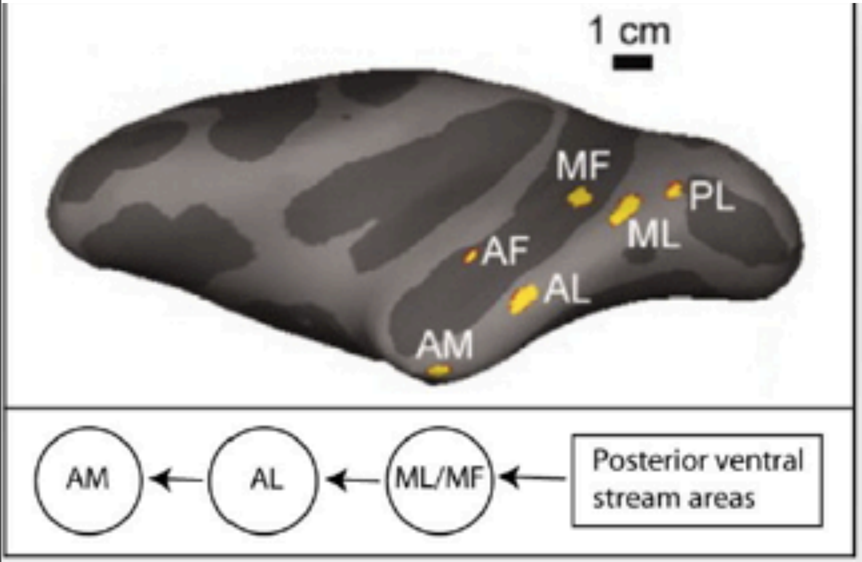


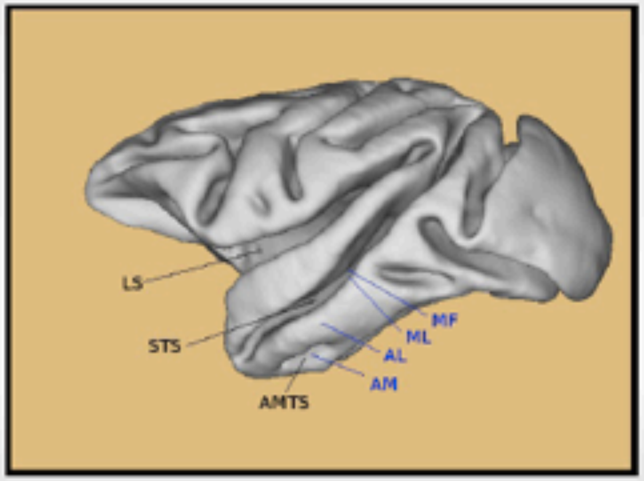
FIGURE 19. *Rotation centered in the RF. 10 images x 360 steps. C1 units of 4 orientations. DOG + DIDT + S1 + C1 + aperture + SVD. 1st four singular values for each aperture.*

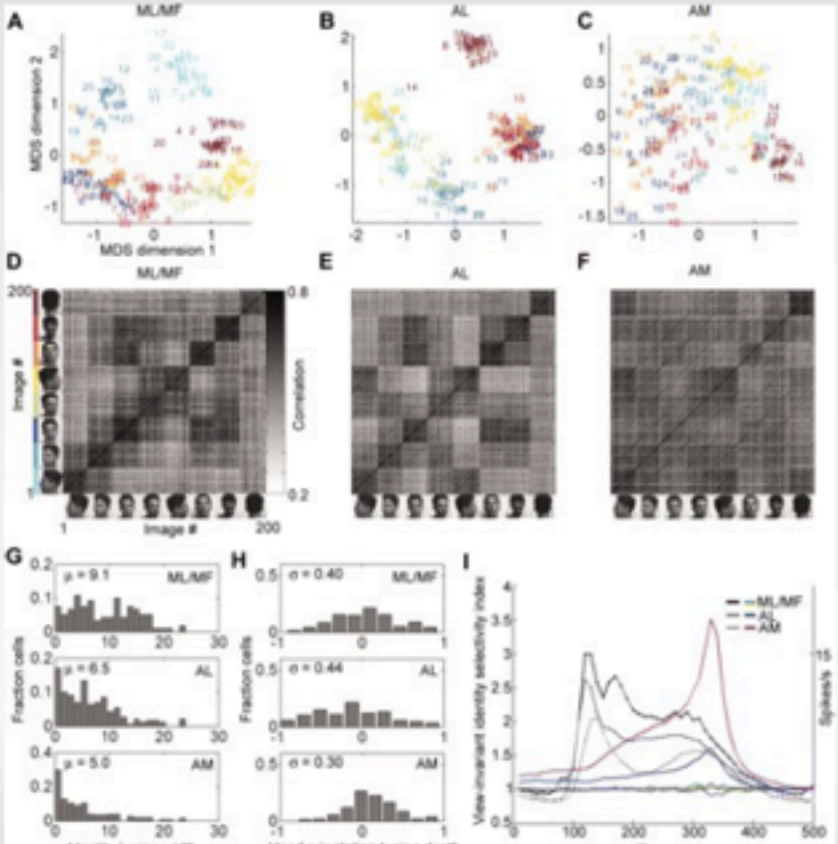


We are working on implementing the full theory
(the corresponding model is an extension of Hmax)
evaluating its performance and comparing it to physiology

Class-specific modules







A theory of biological vision: will it tell us what cortex computes and properties of its neurons?

- The basic equation of physics can be derived from a small number of symmetry properties: invariance wrt space+time, conservation of energy, invariance to measurement units....
- Is it possible that the basic architecture and tuning properties of visual cortex could all be predicted from basic symmetries of geometric transformations of images? Other sensory cortices would be explained in similar ways.
- The brain would be a mirror of the physical world and the tuning of its neurons would reflect symmetry properties of basic physics and geometry.