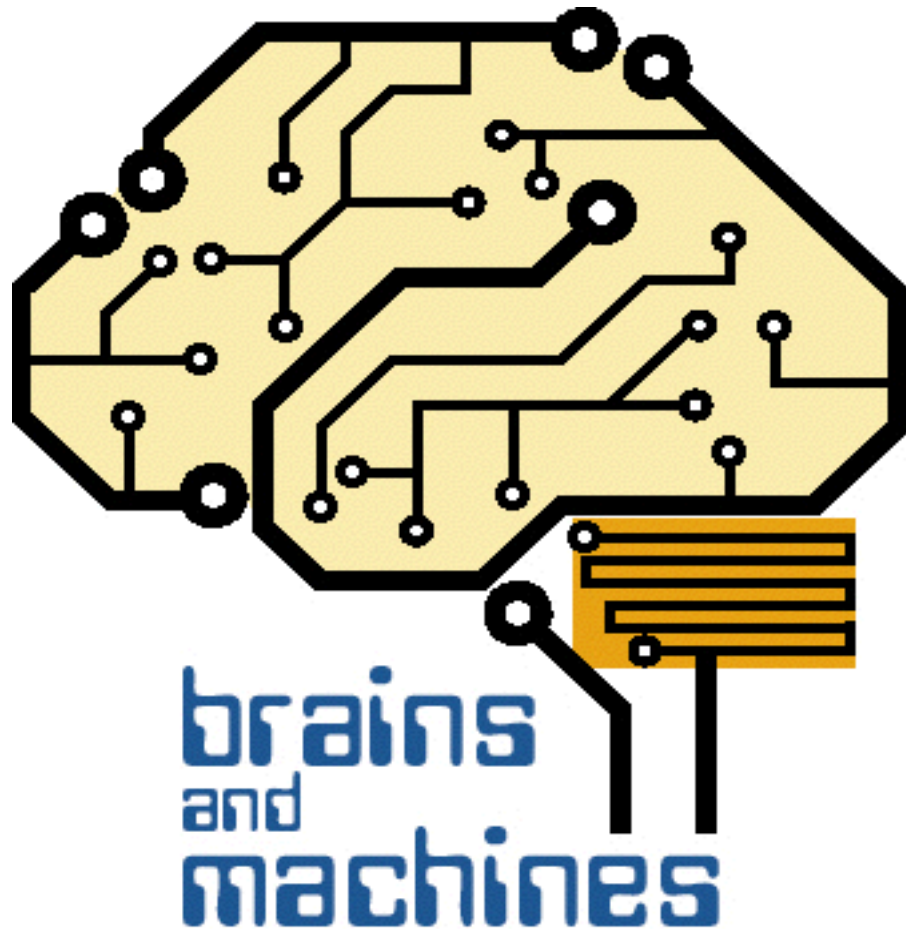


Invariant recognition: a theory of the visual system



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

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 continu

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

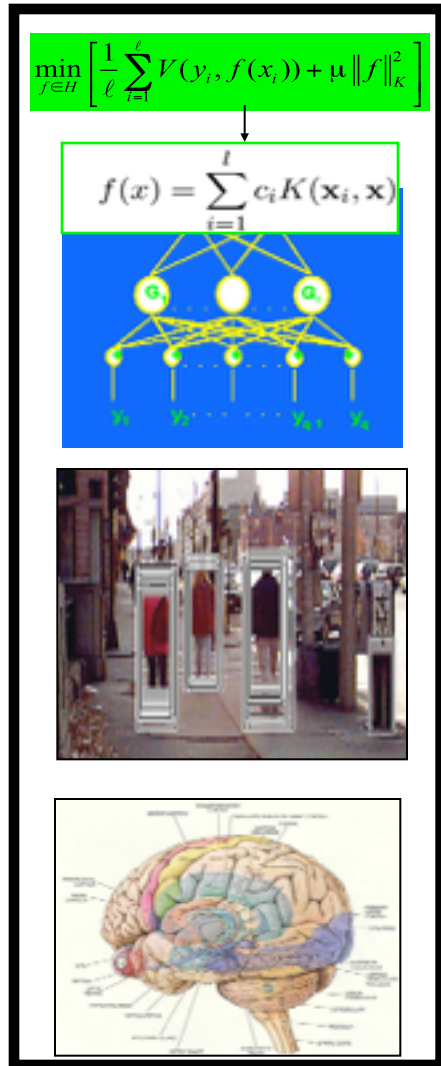
These advances will be critical to of our society's

- future prosperity
- education, health, security





Vision @CBCL, ~20 years ago



LEARNING THEORY + ALGORITHMS

Theorems on foundations of learning
Predictive algorithms

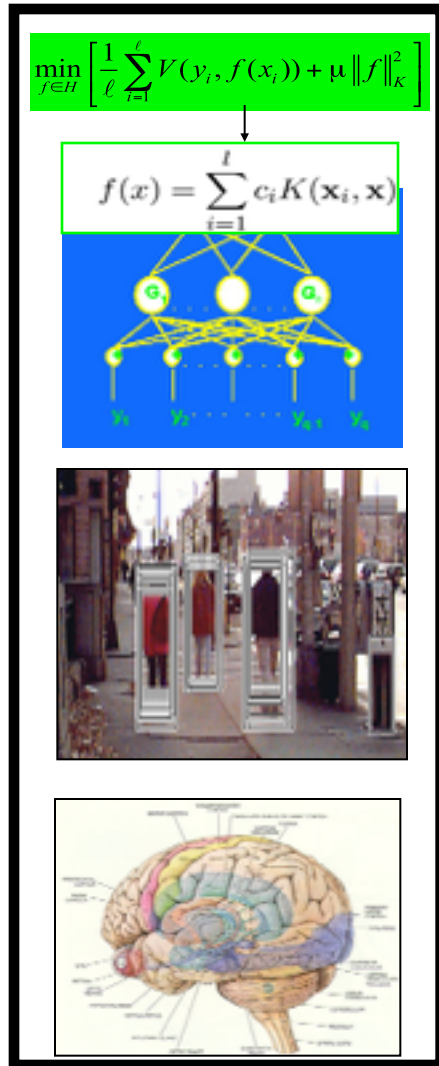


Sung & Poggio 1995, also Kanade & Baluja....

COMPUTATIONAL NEUROSCIENCE: models+experiments

How visual cortex works

Vision @CBCL, ~20 years ago



LEARNING THEORY + ALGORITHMS

Theorems on foundations of learning

Predictive algorithms

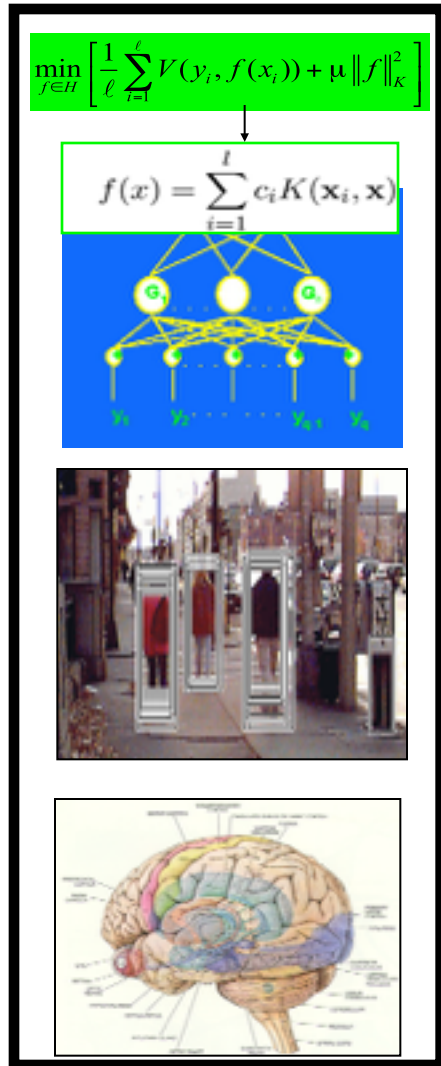


Face detection is now available
in digital cameras (commercial
systems)

COMPUTATIONAL NEUROSCIENCE: models+experiments

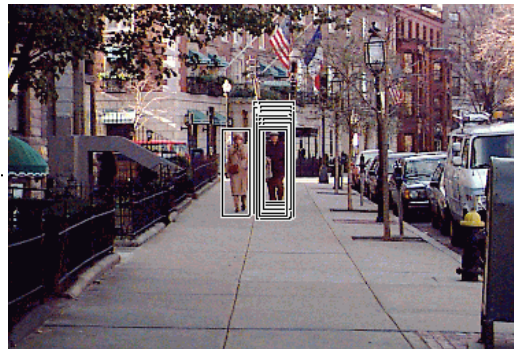
How visual cortex works

Vision @CBCL, ~18 years ago



LEARNING THEORY + ALGORITHMS

Theorems on foundations of learning
Predictive algorithms

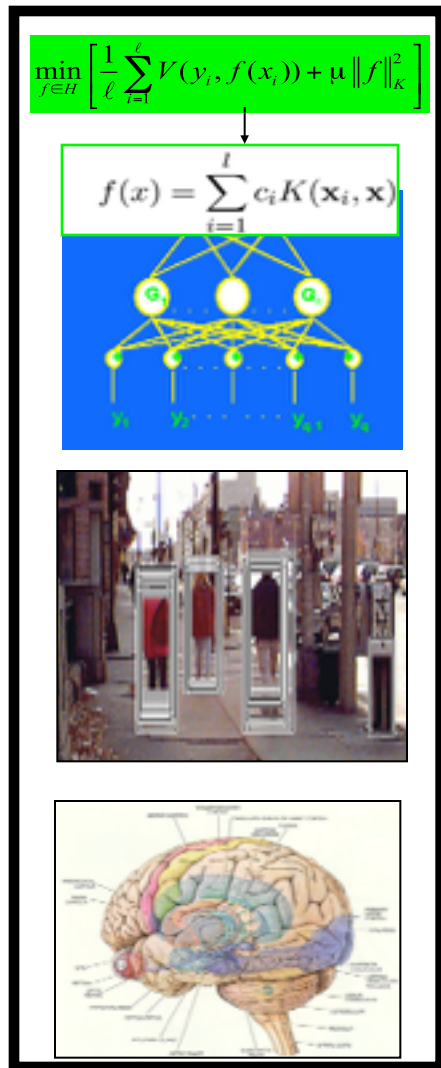


Papageorgiou&Poggio, 1997, 2000
also Kanade&Scheiderman

COMPUTATIONAL NEUROSCIENCE: models+experiments

How visual cortex works

Vision @CBCL, ~18 years ago



LEARNING THEORY + ALGORITHMS

Theorems on foundations of learning
Predictive algorithms



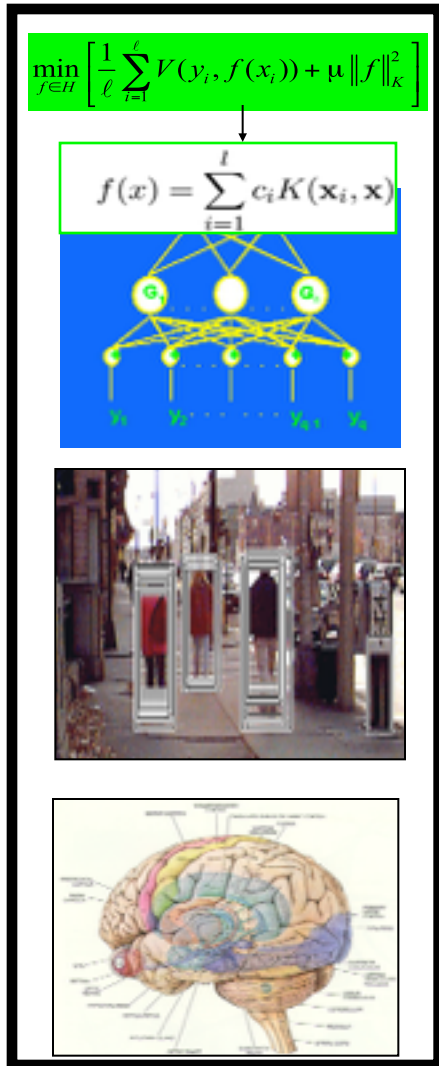
Papageorgiou&Poggio, 1997, 2000
also Kanade&Scheiderman

COMPUTATIONAL NEUROSCIENCE: models+experiments

How visual cortex works



Vision, ~ now



LEARNING THEORY + ALGORITHMS

Theorems on foundations of learning
Predictive algorithms



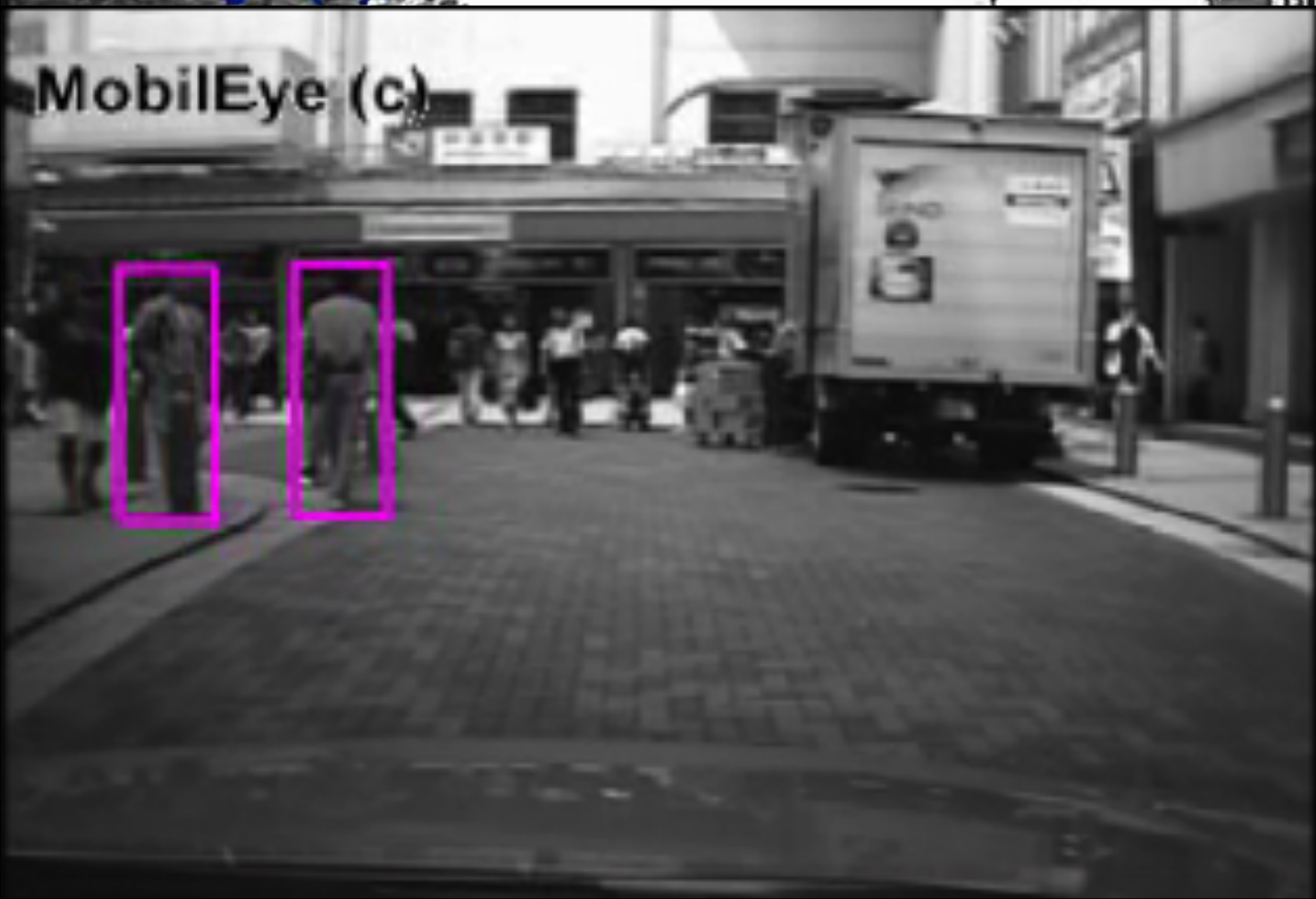
Pedestrian and car detection
are also “solved” (commercial
systems, *MobilEye*,
Jerusalem)

COMPUTATIONAL NEUROSCIENCE: models+experiments

How visual cortex works

Mobileye (c) 2004

MobilEye (c)



Mobileye (c) 2004



Pedestrian accidents occur every day
in our increasingly intensive traffic environment.



Golden age for the technology of
narrowly intelligent machines
but...

┐ ...even the MobilEye vision system is
not intelligent

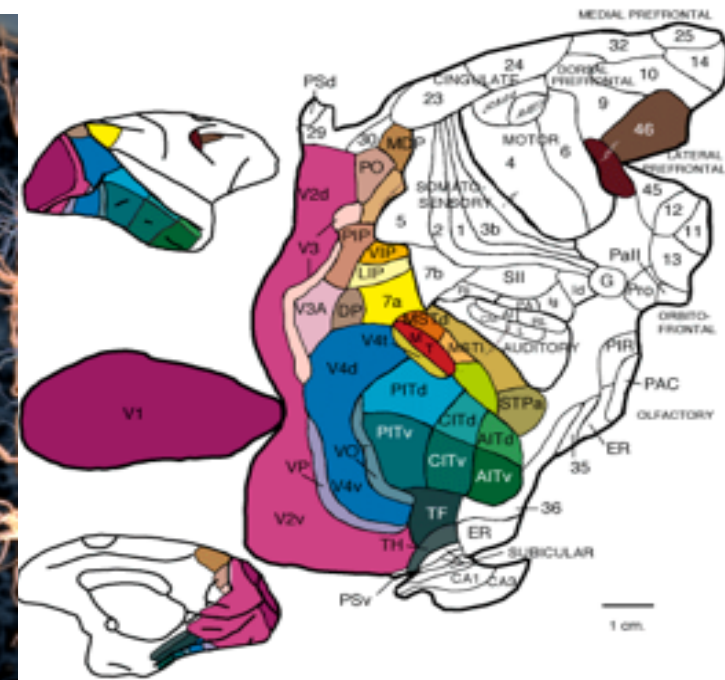
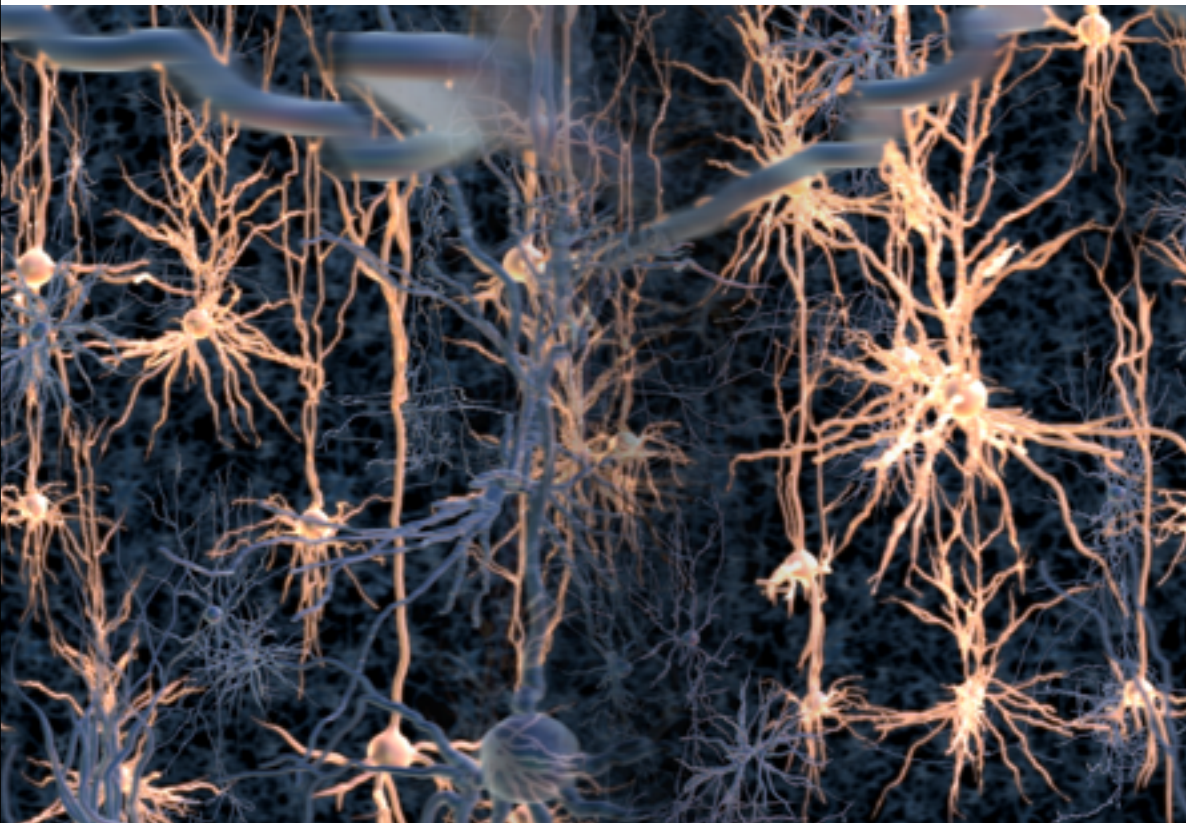
It cannot deal with a *Turing* test of vision:
understanding a scene.

A “Turing” 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)

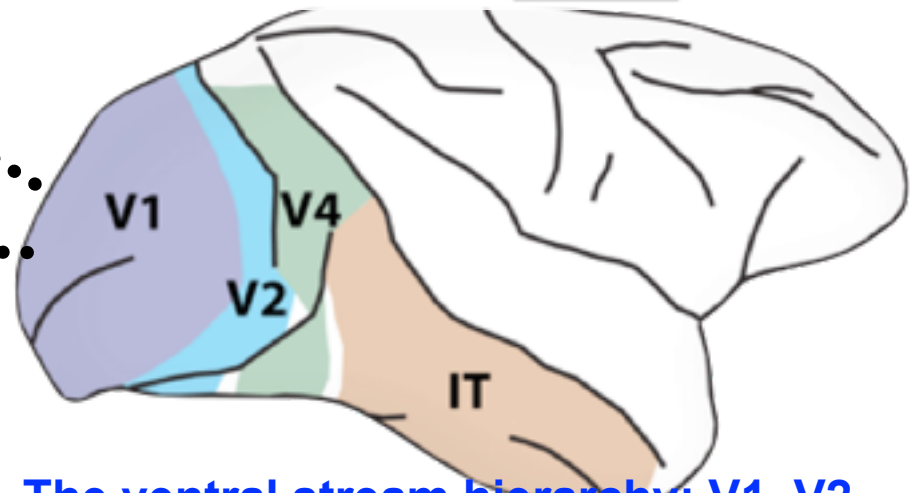
Visual Object Recognition: the ventral stream (macaque)

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









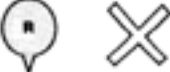



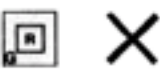




A gradual increase in the receptive field size, in the “**complexity**” of the preferred stimulus, in “**invariance**” to position and scale changes

Kobatake & Tanaka, 1994

V2	V4	posterior IT	anterior IT

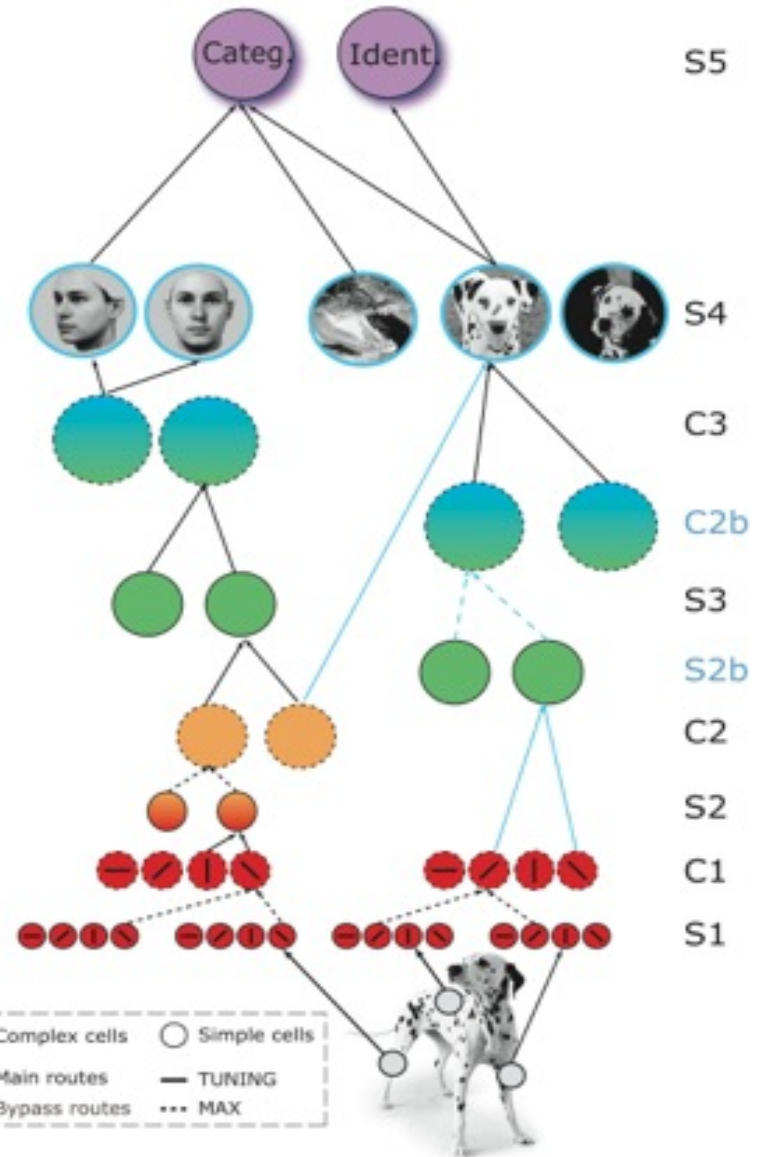
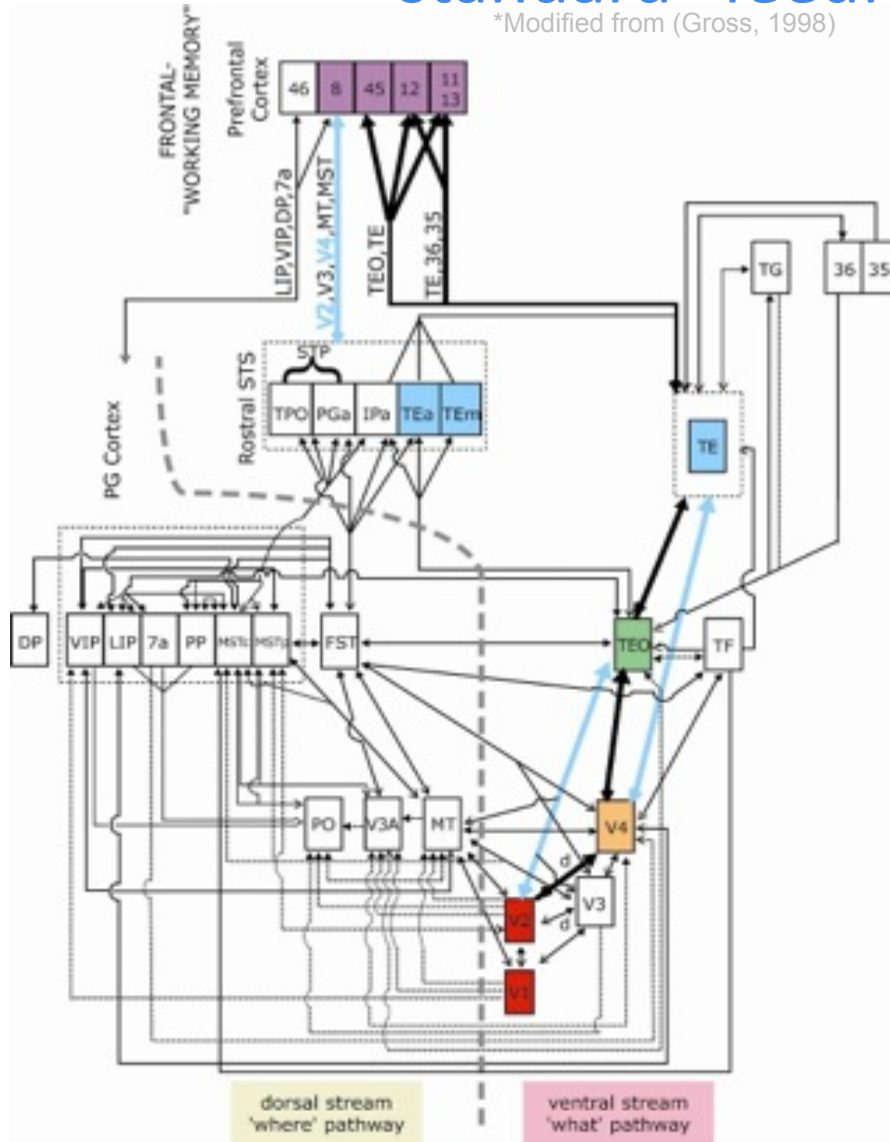


A gradual increase in the receptive field size, in the “**complexity**” of the preferred stimulus, in “**invariance**” to position and scale changes

V2	V4	posterior IT	anterior IT
			
			
			
			
			

Recognition in the Ventral Stream: “standard” feedforward model

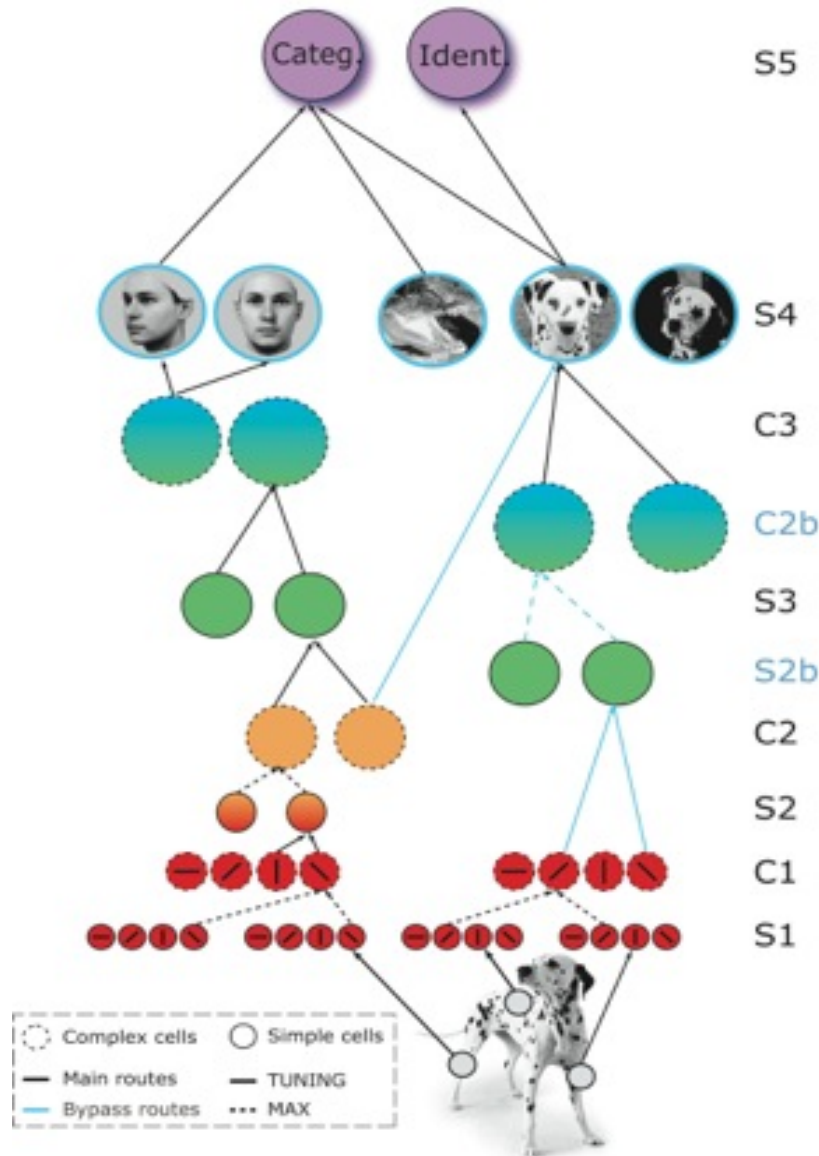
*Modified from (Gross, 1998)



[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



- 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

[software available online]

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)

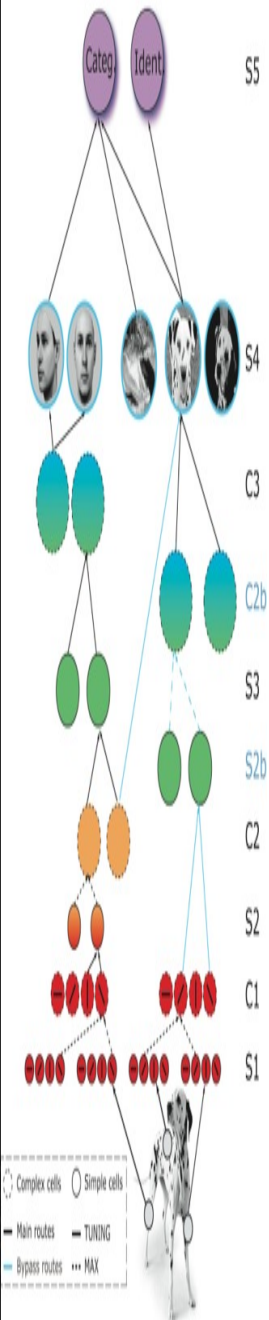
Read out results (Hung Kreiman Poggio & DiCarlo 2005)

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

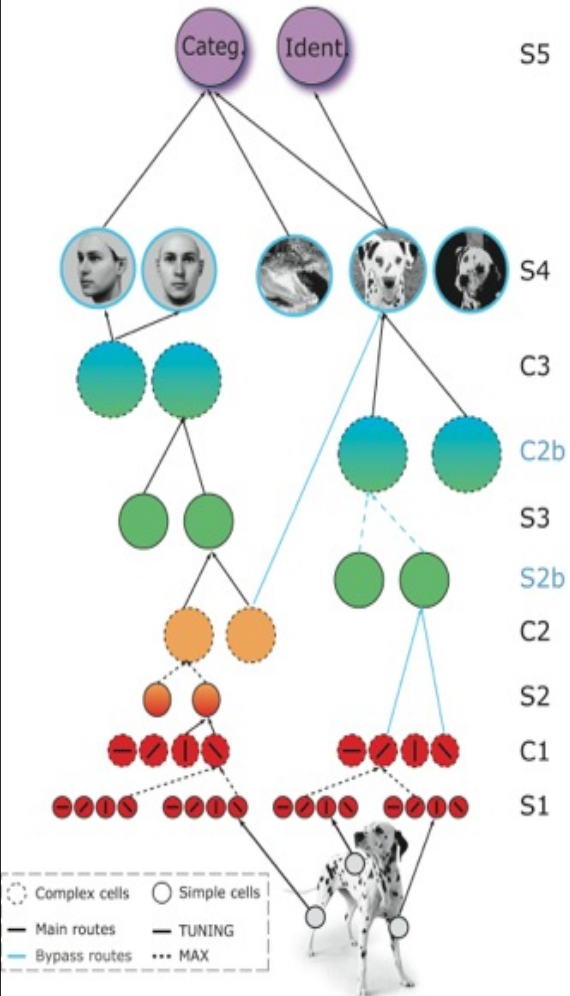
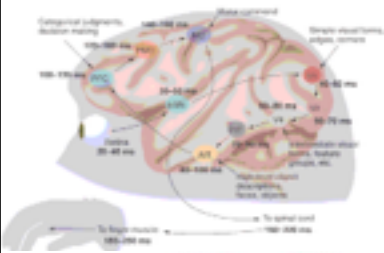
Human:

Rapid categorization (Serre Oliva Poggio 2007)

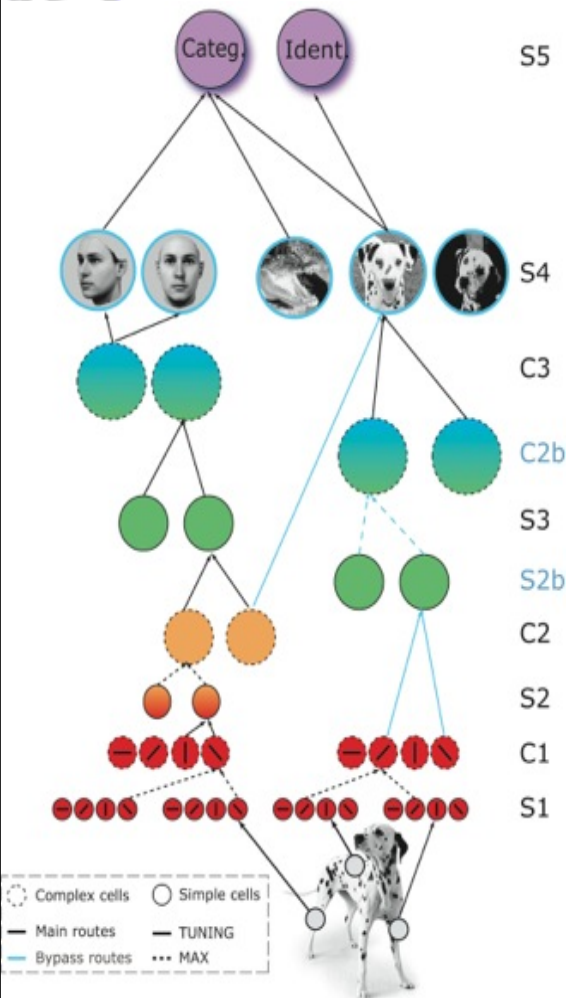
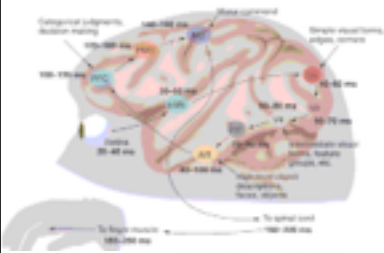
Face processing (fMRI + psychophysics) (Riesenhuber et al 2004; Jiang et al 2006)



Model “works”: it accounts for psychophysics



Model “works”: it accounts for psychophysics



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

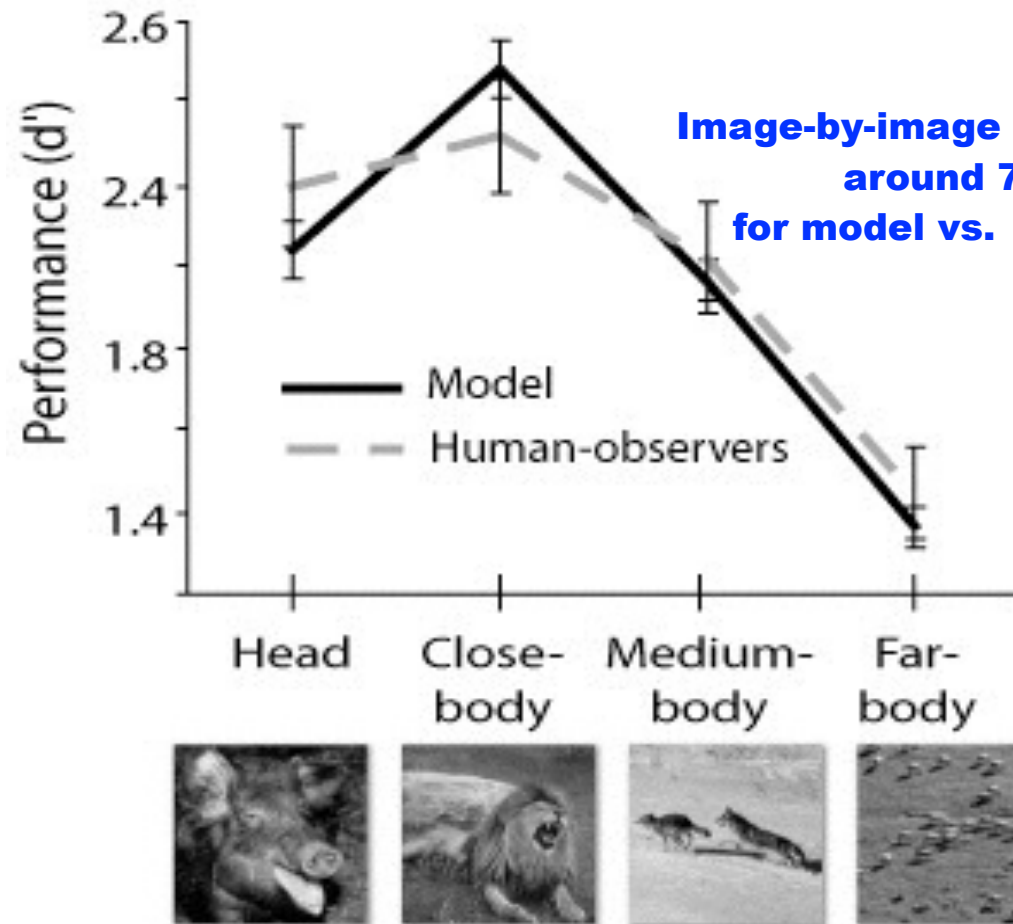
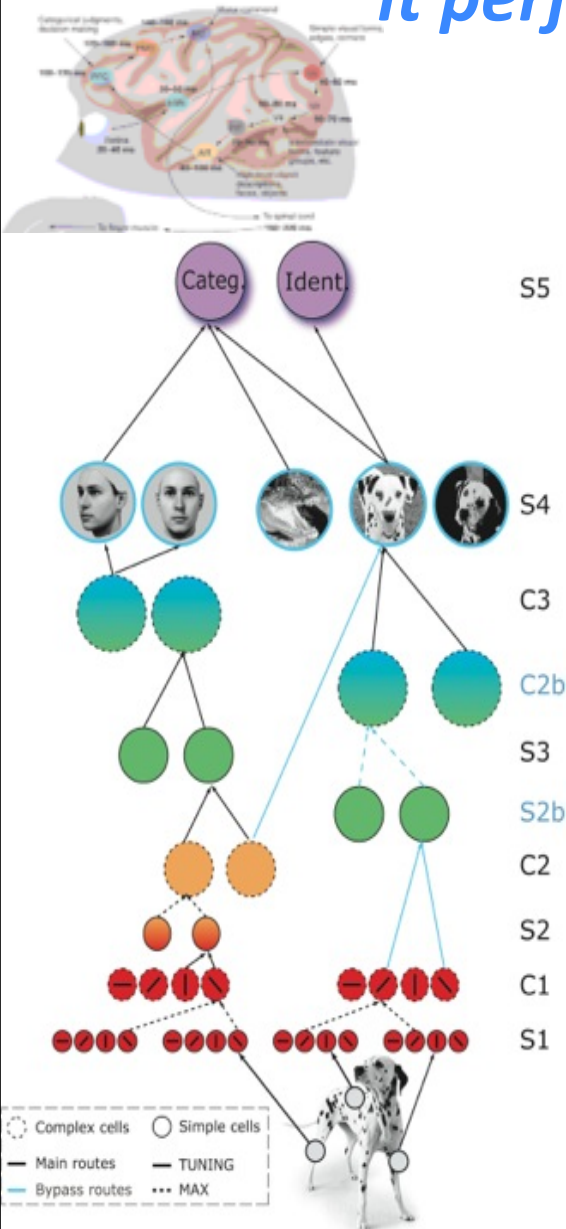


Image-by-image correlation:
around 73%
for model vs. humans)

Model “works”: it performs well at computational level

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



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

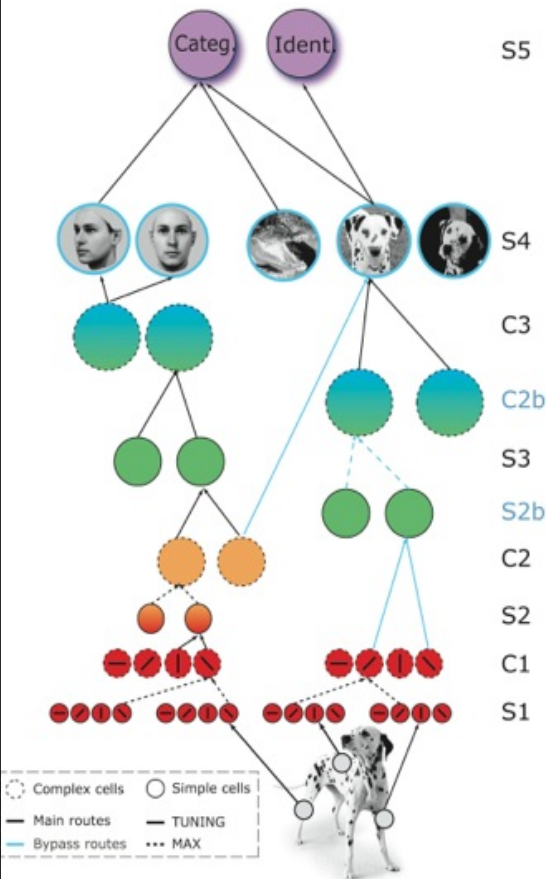


Visual Cortex: models and theories

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

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

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



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[†])

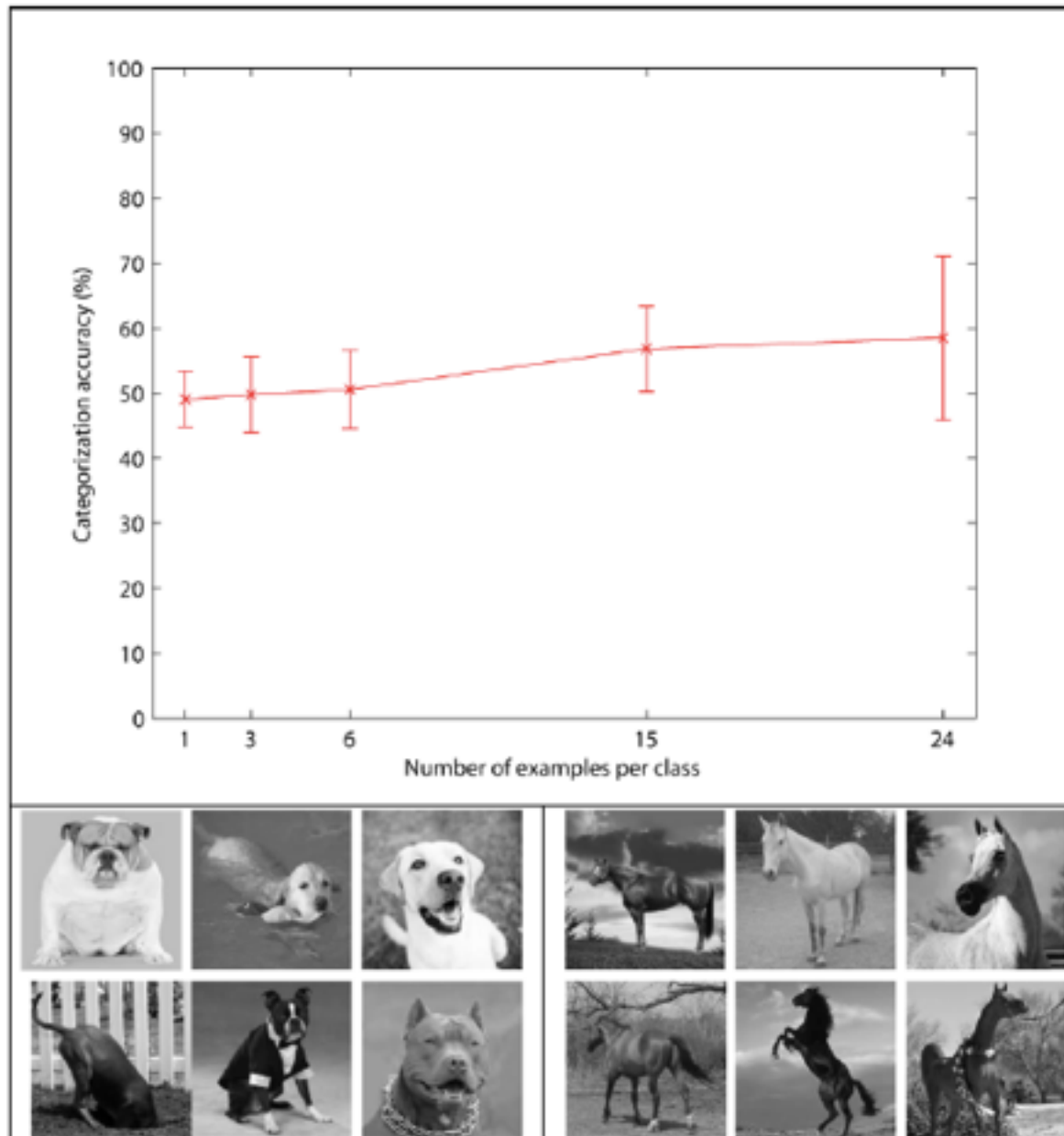
^{*} CBCL, McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA, USA

[†] Istituto Italiano di Tecnologia, Genova, Italy

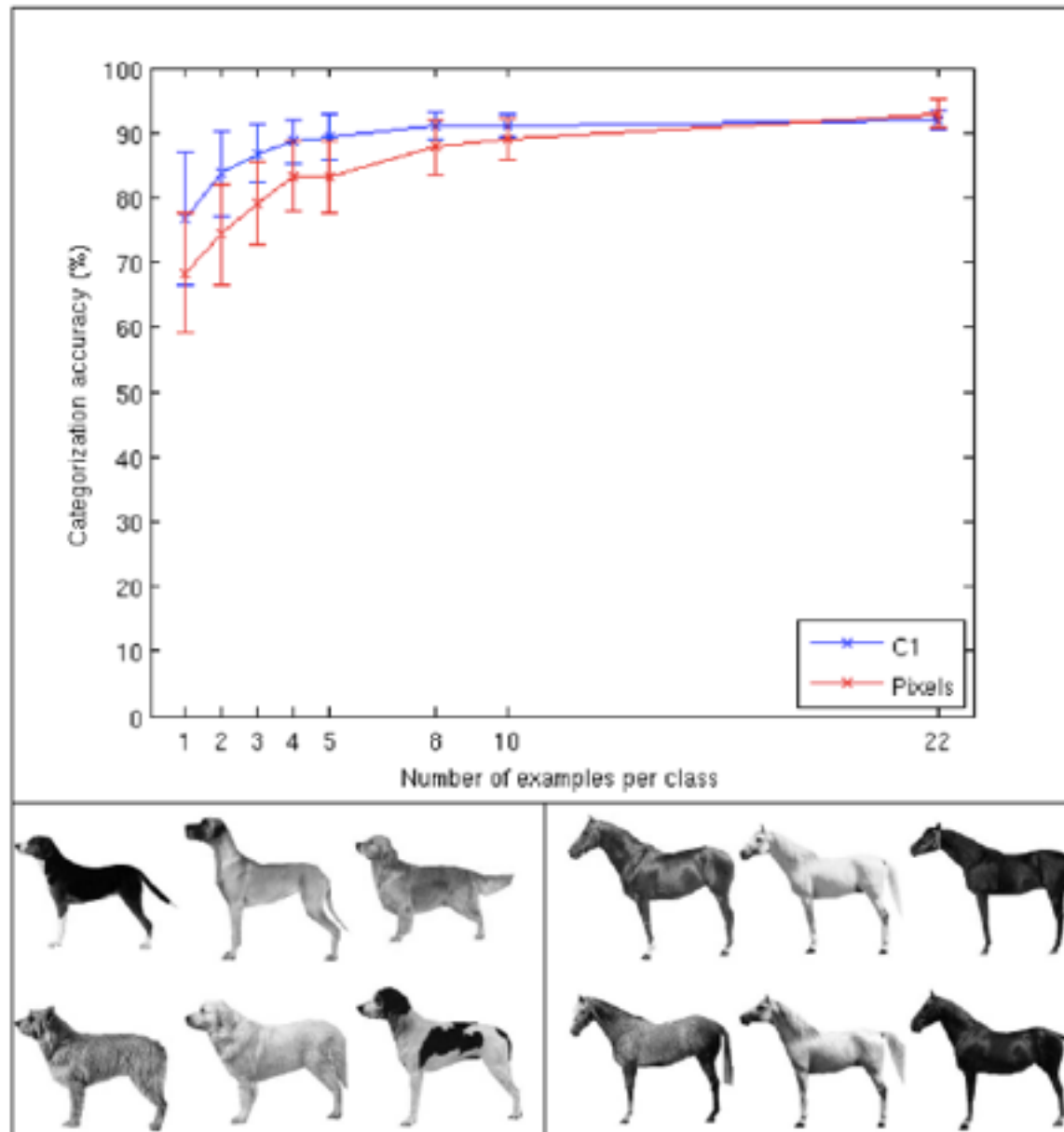
Nature Precedings, [doi:10.1038/npre.2011.6117.1](https://doi.org/10.1038/npre.2011.6117.1) July 16, 2011: outdated version;

new ones will be posted in the future.

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



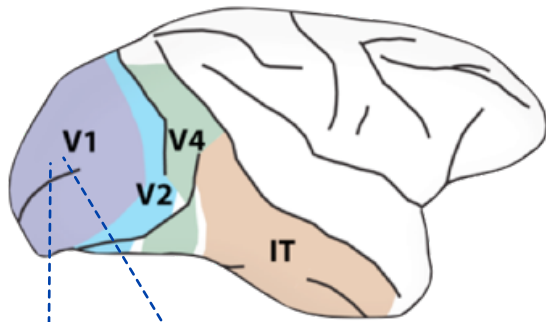
Motivation: transformations may be the main difficulty for (biological) object recognition



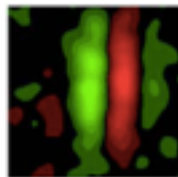
Some of the questions answered by the theory

- What is the main computational task of the ventral stream?
- Why do simple cells in V1 have Gabor tuning curves?
- What are V2, V4, IT computing?
- Why do cells in the AL *face* patch show mirror symmetric tuning curves?

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

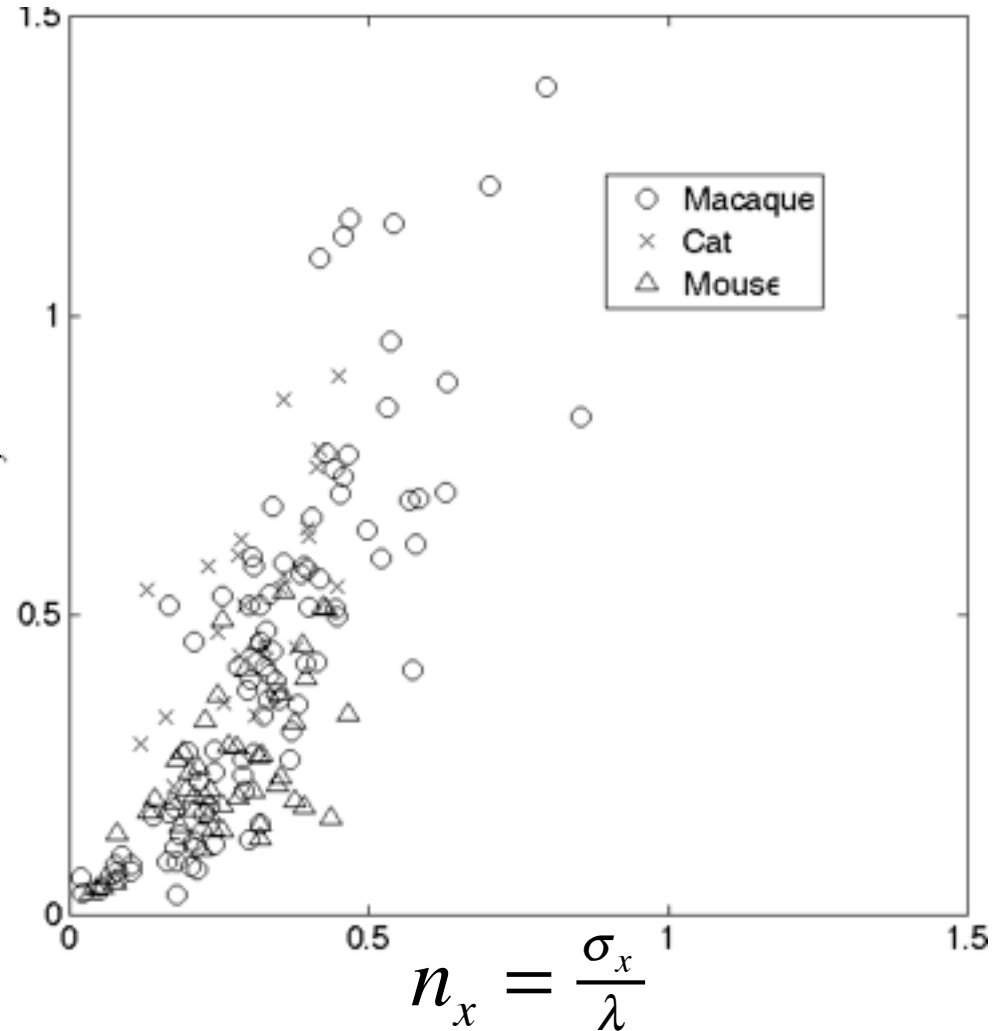


STC - E



Carandini

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



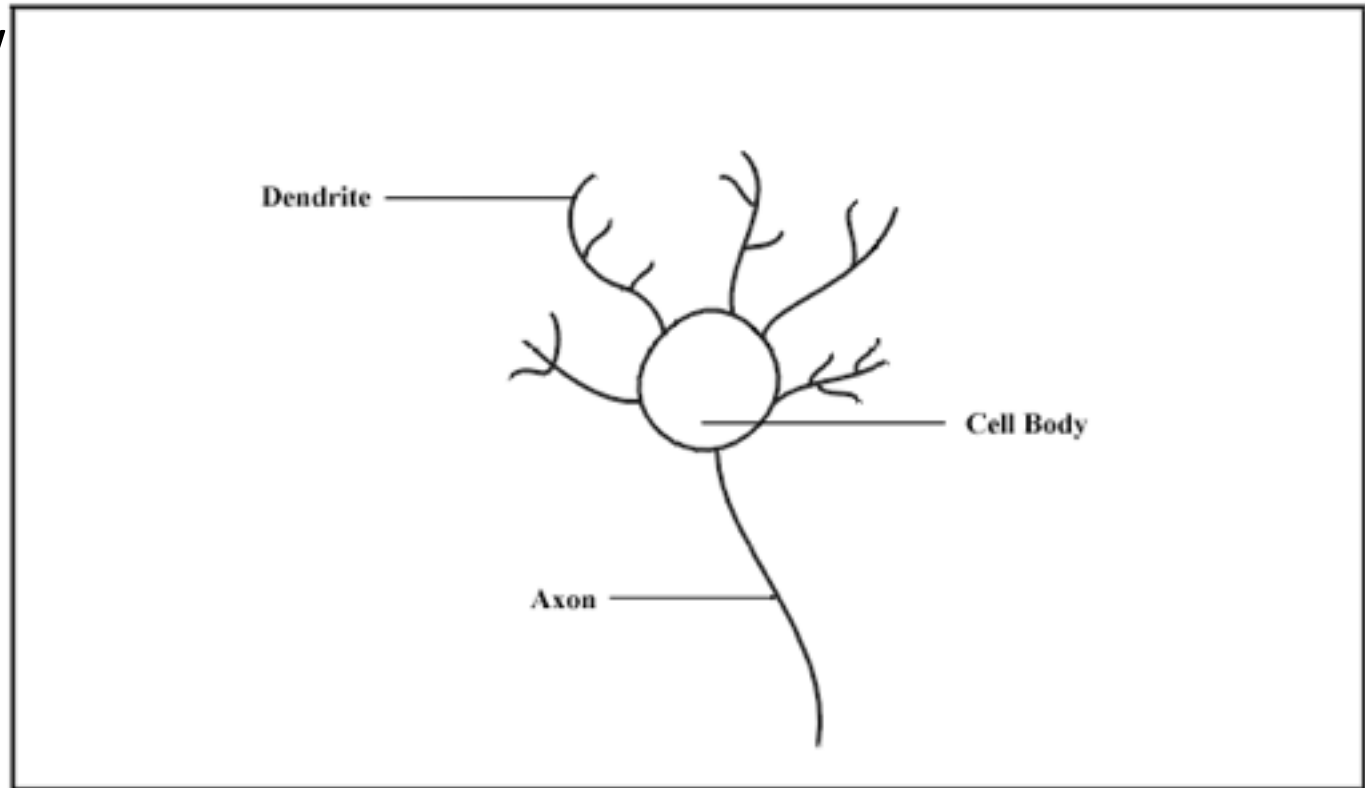
2 Different stages in the theory

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, or 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.

$$x \cdot t$$



Templates and signature

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

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

Thus an image I is represented by a set of neurons as a *signature vector* of I defined with respect to the template set \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 $P : \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that for all $u, v \in V$

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

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

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

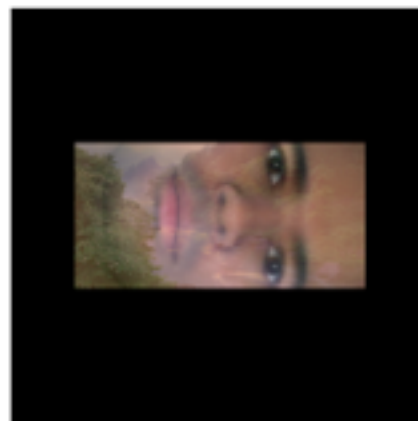
JL suggests that good image representations for classification and discrimination of n objects can be provided by k dot products with *random* templates!



(a) *Input*



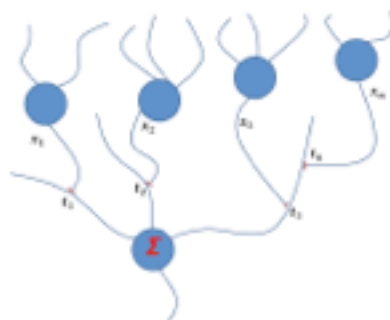
(b) *Template*



(c) *Transformed Input and Template*



(d) *Input and Transformed Template*



(e) *A neuron's dendritic tree with inputs at its synapses*

Figure 2: The dot product between a transformed image and a template (c) is equivalent to the dot product between the image with the inversely transformed template (d). Neurons can easily perform high-dimensional dot products between inputs on their dendritic tree and stored synapses weights (indicated in (d)).

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

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

Initial observation: learning to be invariant for any new object

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

$$g_0 t, g_1 t \dots g_n t$$

Then if the group is compact

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

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

Projections of Probabilities

As argued later, simple operations for neurons are (high-dimensional) dot products between inputs and stored “templates” which are images. It turns out that classical results (such as the Cramer-Wold theorem) ensure that lower dimensional projections of a probability distribution on the unit ball uniquely characterize it.

Theorem *Let P and Q two probability distributions on \mathbb{R}^d . Let $\Gamma = \{t \in \mathbb{S}(\mathbb{R}^d), \text{ s.t. } P_t = \langle P, t \rangle = \langle Q, t \rangle = Q_t\}$, where $\mathbb{S}(\mathbb{R}^d)$ is the unit ball in \mathbb{R}^d . Let $\lambda(\Gamma)$ its normalized measure. We have that if $\lambda(\Gamma) > 0$ then $P = Q$. This implies that the probability of choosing t such that $P_t = Q_t$ is equal to 1 if and only if $P = Q$ and the probability of choosing t such that $P_t = Q_t$ is equal to 0 if and only if $P \neq Q$.*

Analog of JohnsonLindenstrauss for probabilities

Fineteness of the number of templates in practical cases is ensured by

Theorem (*Heppes et al., 1956*) *Let P be a discrete probability distribution on \mathbb{R}^d with a support made with exactly k distinct atoms. Assume that V_1, \dots, V_{k+1} are subspaces of \mathbb{R}^d of respective dimensions d_1, \dots, d_{k+1} such that no couple of them is contained in a hyperplane (i.e. no straight line is perpendicular to more than one of them). Suppose, e.g. $d_1 = 1 = \dots = d_{k+1}$ and call the subspaces t_i , $i = 1, \dots, k + 1$. Then, for any probability distribution Q in \mathbb{R}^d , we have $P = Q$ if and only if $t_i \in \Gamma$, for every $1 \leq i \leq k + 1$.*

In particular, for a probability distribution made with k atoms in \mathbb{R}^d , we see that at most $k + 1$ directions are enough to characterize the distribution. Thus a finite – albeit large – number of one-dimensional projections is equivalent to the full distribution.

Group Invariance

The estimation of $P(gl \cdot t^k)$ seems to require the observation of the image *and* “all” its transforms. Ideally we would like to compute an invariant signature for a new object seen only once (we can recognize a face at a different distances after just one observation). The key here is the simple observation that $gl \cdot t^k = l \cdot g^{-1}t^k$. Thus it is possible for the system to store for each template t^k all its transformations gt^k and thus later obtain an invariant signature for new images.

Group Invariance

- ▶ The full $P(gI)$ is a probability density induced by “all” $g \in G$; not surprisingly it is a full and invariant characterization of I and all its transforms.
- ▶ The Cramer Wold-like theorems say that a proxy for $P(gI)$ is a set of K one dimensional $P(gI \cdot t^k)$. This still requires observation of all the transformations of I induced by the group.
- ▶ Since $gI \cdot t^k = I \cdot g^{-1}t^k$ it is however possible possible to obtain an invariant signature from a single image I by storing for each template t^k all its transformations gt^k .

Group Invariance

The following holds since the distributions $P_g(gI \cdot t^k)$ and $P_g(I \cdot g^{-1}t^k)$ are equivalent (the inverse g^{-1} is an element of the group):

Theorem *Empirical estimates of the probability distribution $P_g(I \cdot g^{-1}t^k)$ for $k = 1, \dots, K$ represent a ϵ -unique (empirical) invariant associated with the orbit of I under the group G .*

Neurons ways to compute invariance

During development of the visual system a group of $|G|$ (simple) cells store in their synapses an image patch t^k and its transformations $g_1 t^k, \dots, g_{|G|} t^k$. This is done for several image patches (templates). Later when an image is presented, the simple cells compute $I \cdot g_i t^k$ for $i = 1, \dots, |G|$. Complex cells pool the outputs of the simple cells and compute $\mu_n^k = \sum_{i=1}^{|G|} \sigma(I \cdot g_i t^k + n\Delta)$ where σ is a smooth step function ($\sigma(x) = 0$ for $x \leq 0$, $\sigma(x) = 1$ for $x > 0$) and $n = 1, \dots, N$.

Neural signature: invariance and *uniqueness*

Linear combinations of the μ_n^k for various n could provide an effective binning of $P(I \cdot gt^k)$ and thus an estimate of the empirical distribution at resolution Δ . Of course we are not interested in reconstructing the full probabilities from the empirical estimate; we do not even need the empirical estimate of $P(I \cdot gt^k)$; what is important is that the μ_n^k determine uniquely the probabilities and the associated orbits. Following this argument it can be proved that *a vector with KN components μ_n^k represents a unique and invariant signature for image I .*

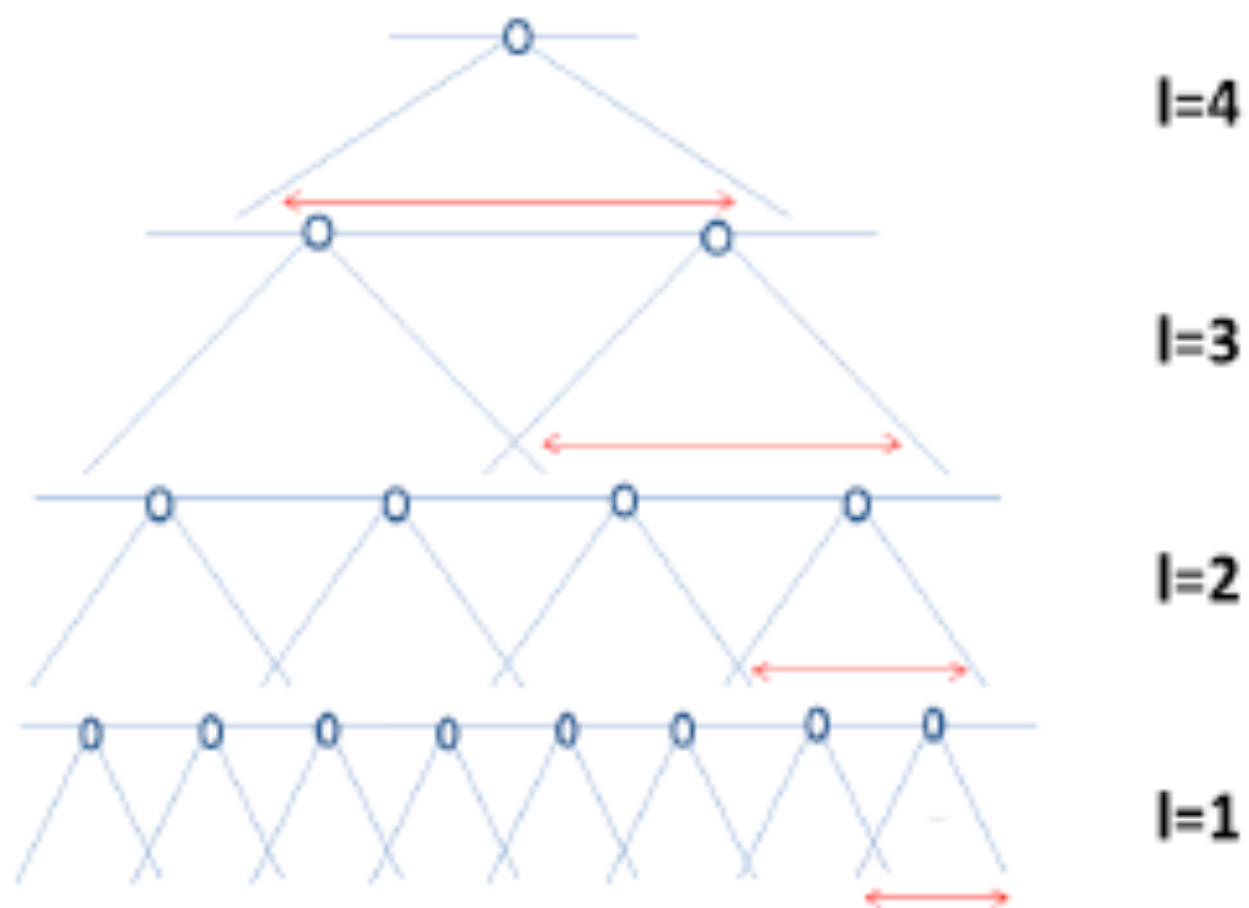
Neural signature: energy model

An invariant signature can be computed in other, equivalent ways at the level of complex cells. Instead of the μ_n^k components, the moments $m_n^k = \int_G (I \cdot g_i t^k)^n dg$ can be computed (they characterize the projections of the probability distributions and can be regarded as group averages. Under some rather weak conditions, they characterize uniquely the distribution $P(I \cdot t)$. For $n = 2$ this corresponds to an energy model of complex cells; for very large n it corresponds to a *max* operation by complex cells. Other nonlinearities are also possible. The available evidence suggests that simple/complex cells in V1 and cells in AL may be described better in terms of energy models than in terms of the sigmoidal nonlinearity.

A theory of hierarchical architectures

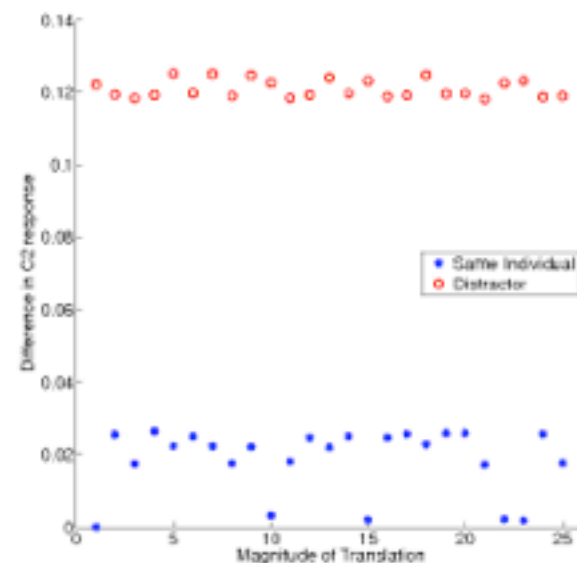
1. Hierarchical architectures divide and conquer.
2. In a hierarchical architecture different types of transformations can be factorized in different layers. This property ensures significant advantages in terms of sample complexity of learning.
3. In multilayer architectures with modules of the simple-complex type the following covariance-invariance property holds: *For a given transformation of an image or part of it, the signature from complex cells at a certain level is either invariant or covariant w.r.t. the group of transformations; if it is covariant there will be a higher layer in the network at which it is invariant.*
4. Invariant hierarchical architectures reflect the hierarchy of *wholes and parts*—of objects and components—in the visual world as described by a special metric defined by a *derived kernel* that is iteratively obtained from the initial similarity defined at the first layer.

The last two properties are related to the problem of clutter and context in object recognition.





(a) Reference input and distractor.

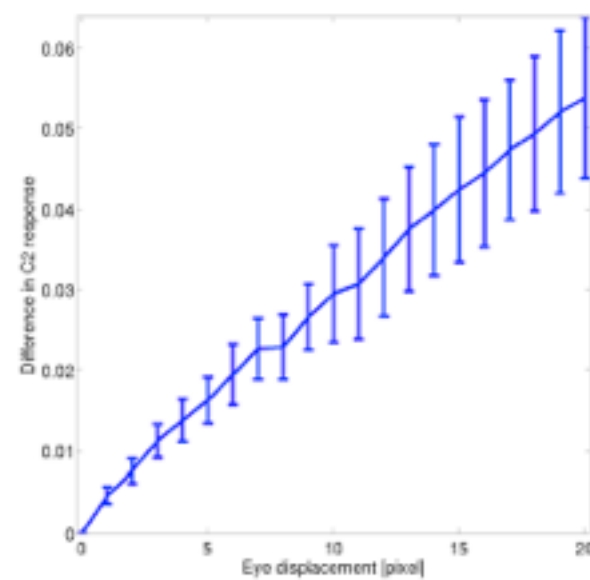
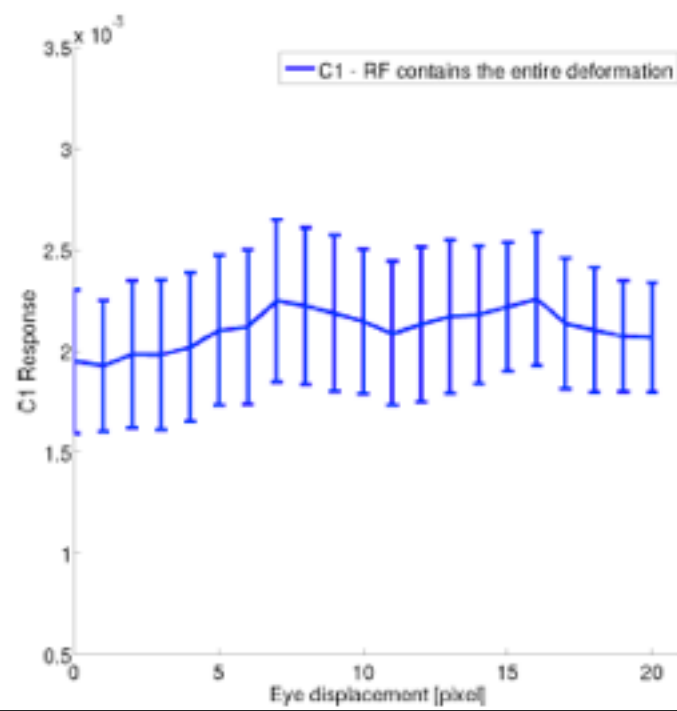


(b)

Figure 3: Two distinct stimuli (left) are presented at various location in the visual field. The Euclidean distance between C2 response vectors in HMAX is reported (right). It can be seen how the response are invariant to global translation and discriminative. The C2 units represent the top of a hierarchical, convolutional architecture.



(a)



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.

Unsupervised tuning (during development) and eigenvectors of covariance matrix

Hebb synapses imply 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 follows the Oja flow

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

Different cells are exposed (during development) to 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.

2D eigenvectors

In 1D the eigenvectors are Gabor like functions. In 2D the solutions are also Gabor with an orientation orthogonal to the direction of motion. Motion, together with high-pass filtering in the retina induces *symmetry breaking* that allows non-symmetric solution to emerge. Note that for motion at constant speed

$$\frac{d}{dt} = v \frac{d}{dx}$$

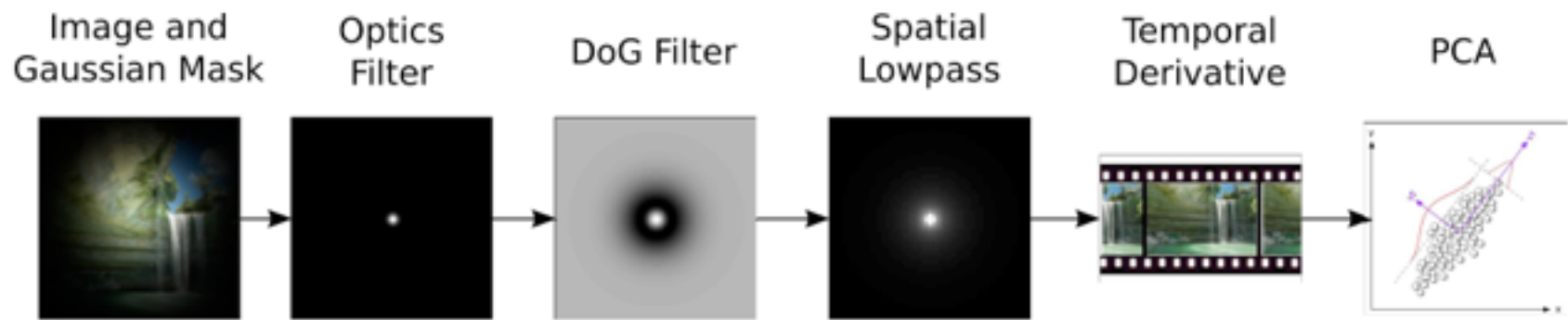
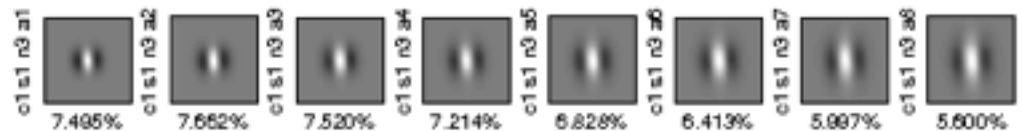
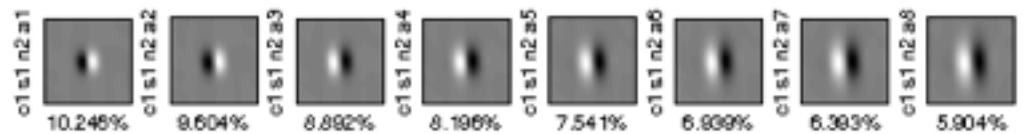
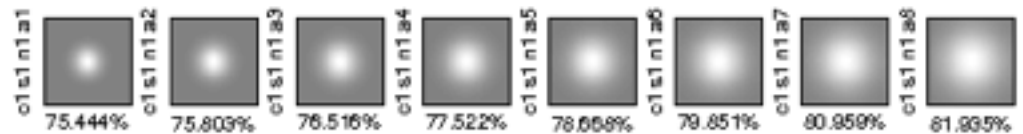
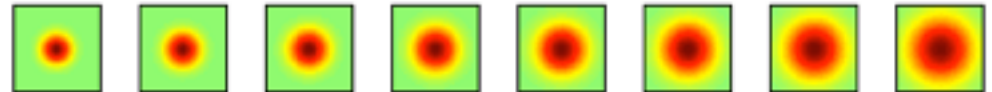
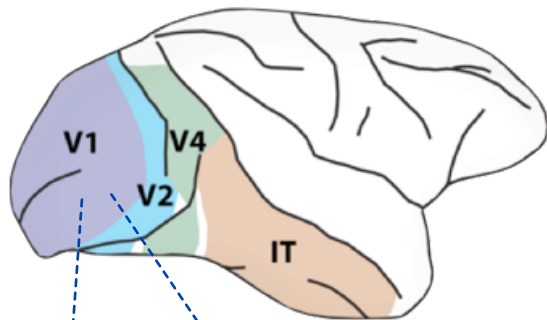
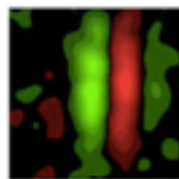


Figure 4: *Retinal processing pipeline used for V1 simulations.*

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



STC - E

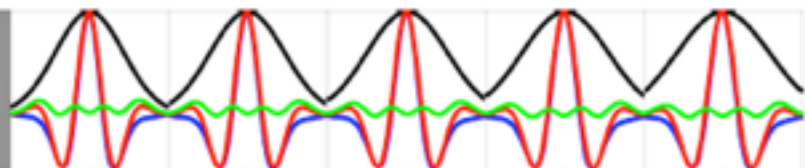
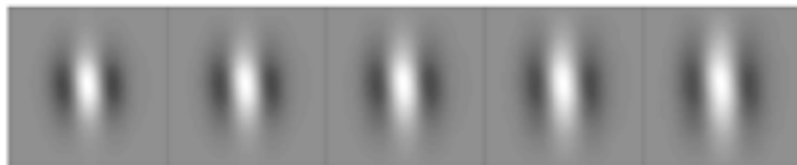
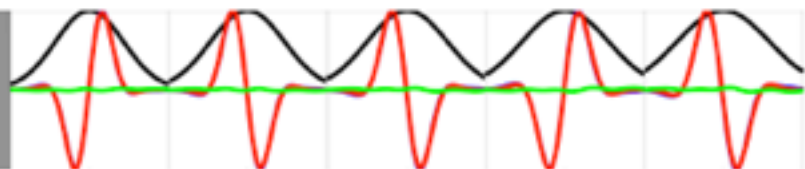
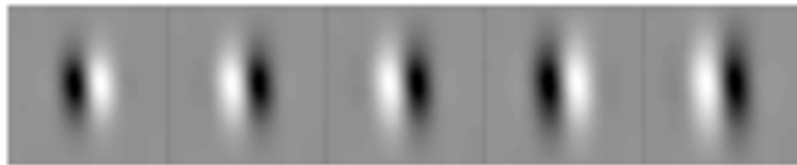
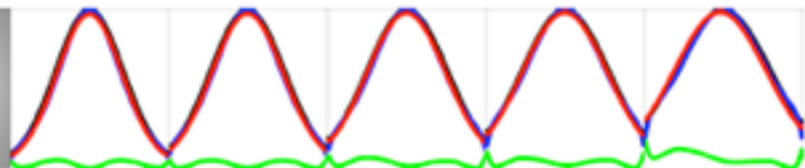
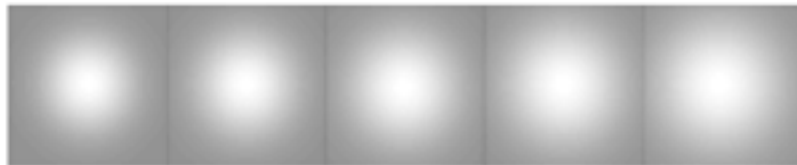
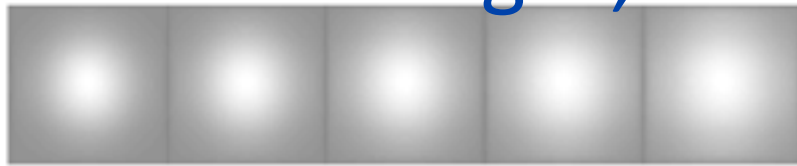


Carandini



Rust et al. 2005

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

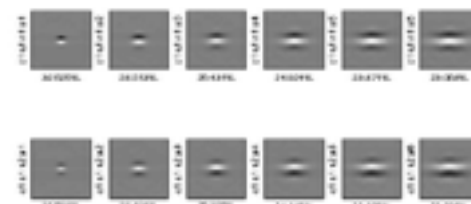
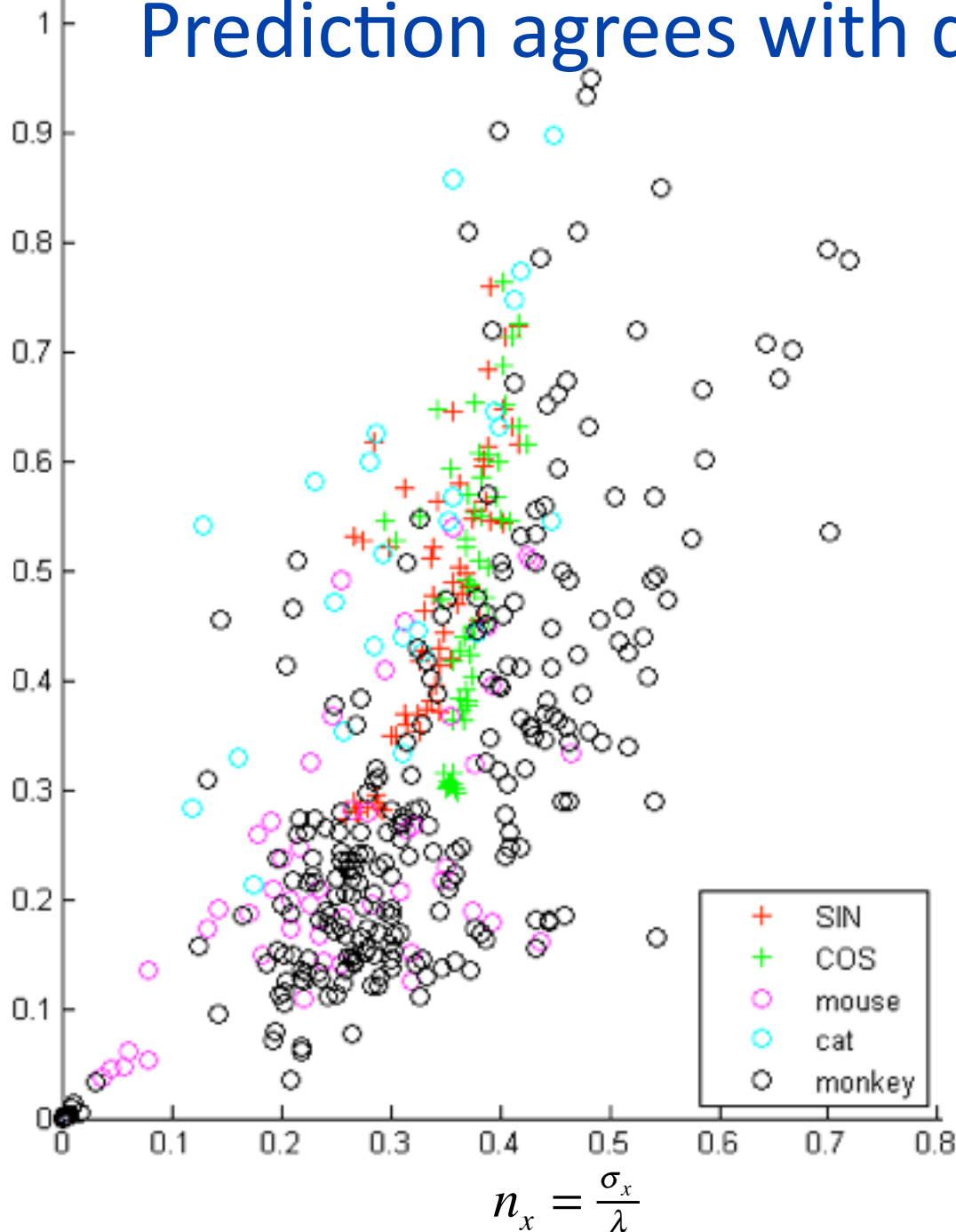


$\sigma = 8.3$ $\sigma = 9.0$ $\sigma = 9.7$ $\sigma = 10.5$ $\sigma = 11.3$

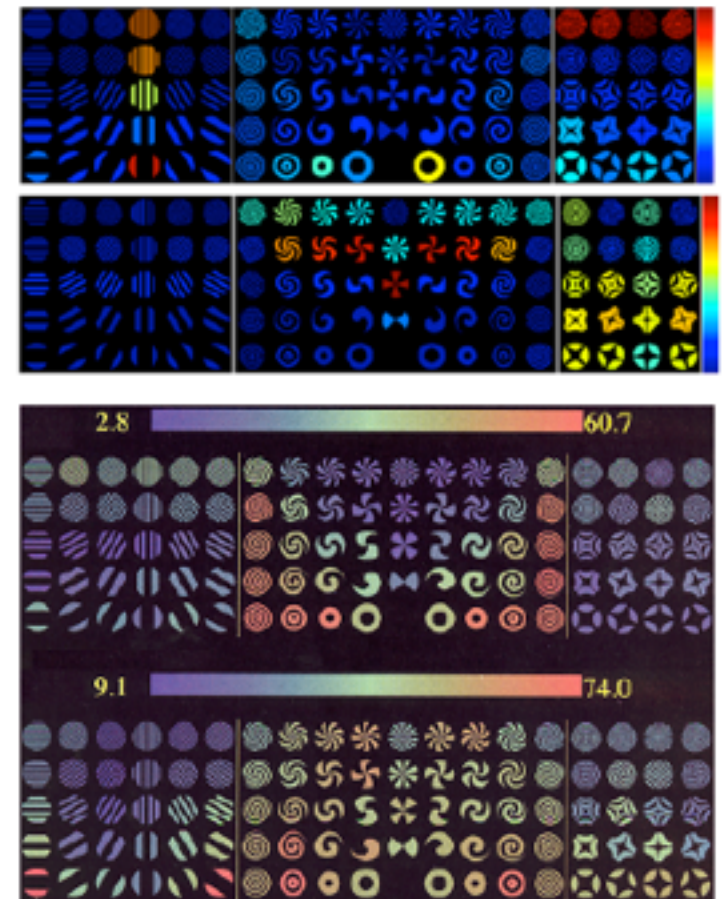
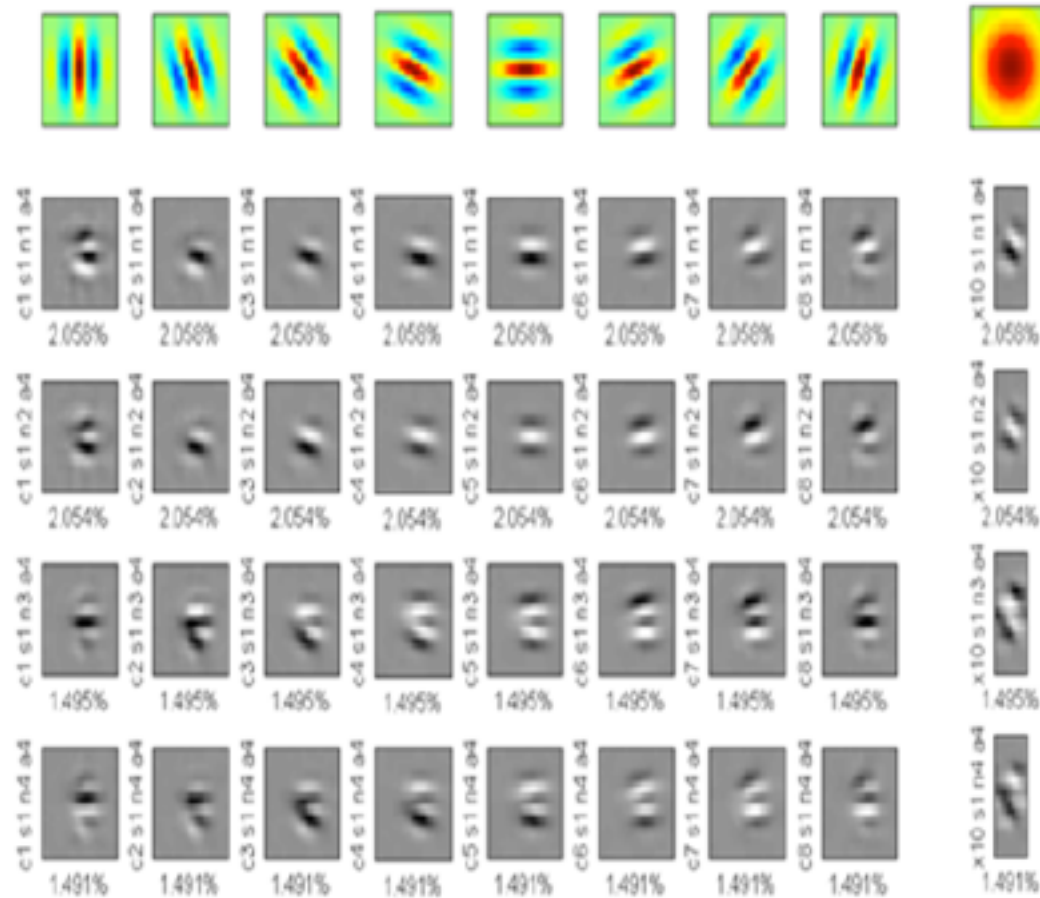
$\sigma = 8.3$ $\sigma = 9.0$ $\sigma = 9.7$ $\sigma = 10.5$ $\sigma = 11.3$

Prediction agrees with data!

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

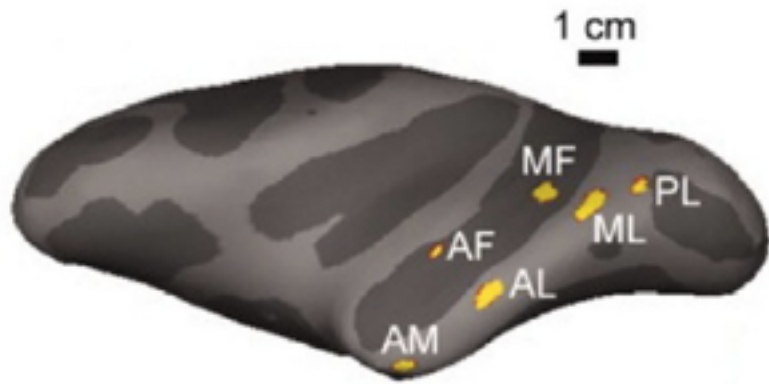


Beyond V1, towards V2 and V4: wavelets of wavelets



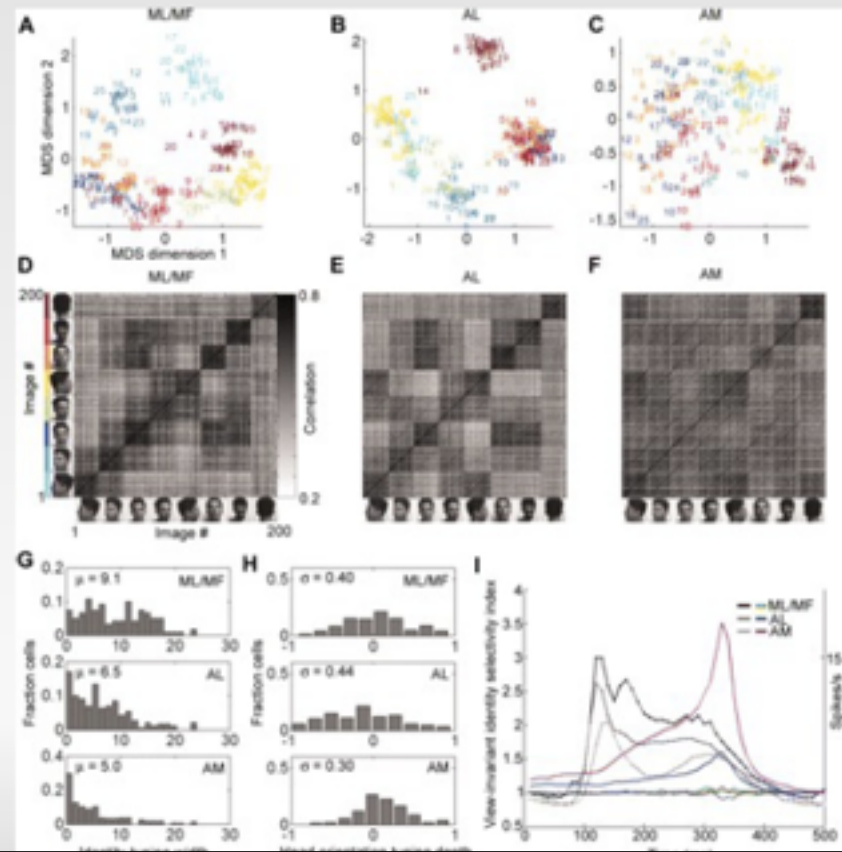
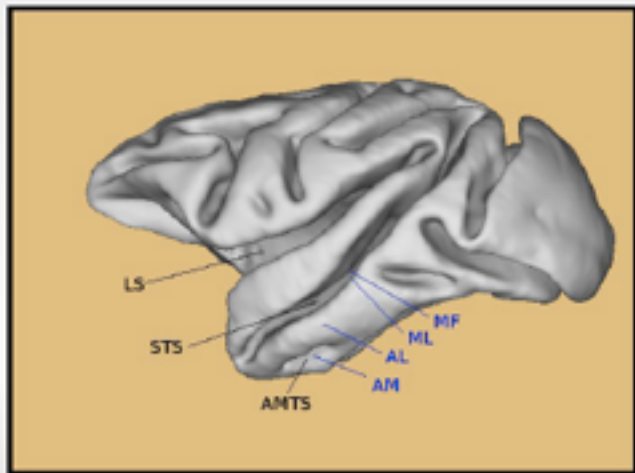
We are working on implementing
the full theory
(the corresponding model is an
extension of Hmax and
convolutional networks)

Class-specific modules



View-invariant

View-specific



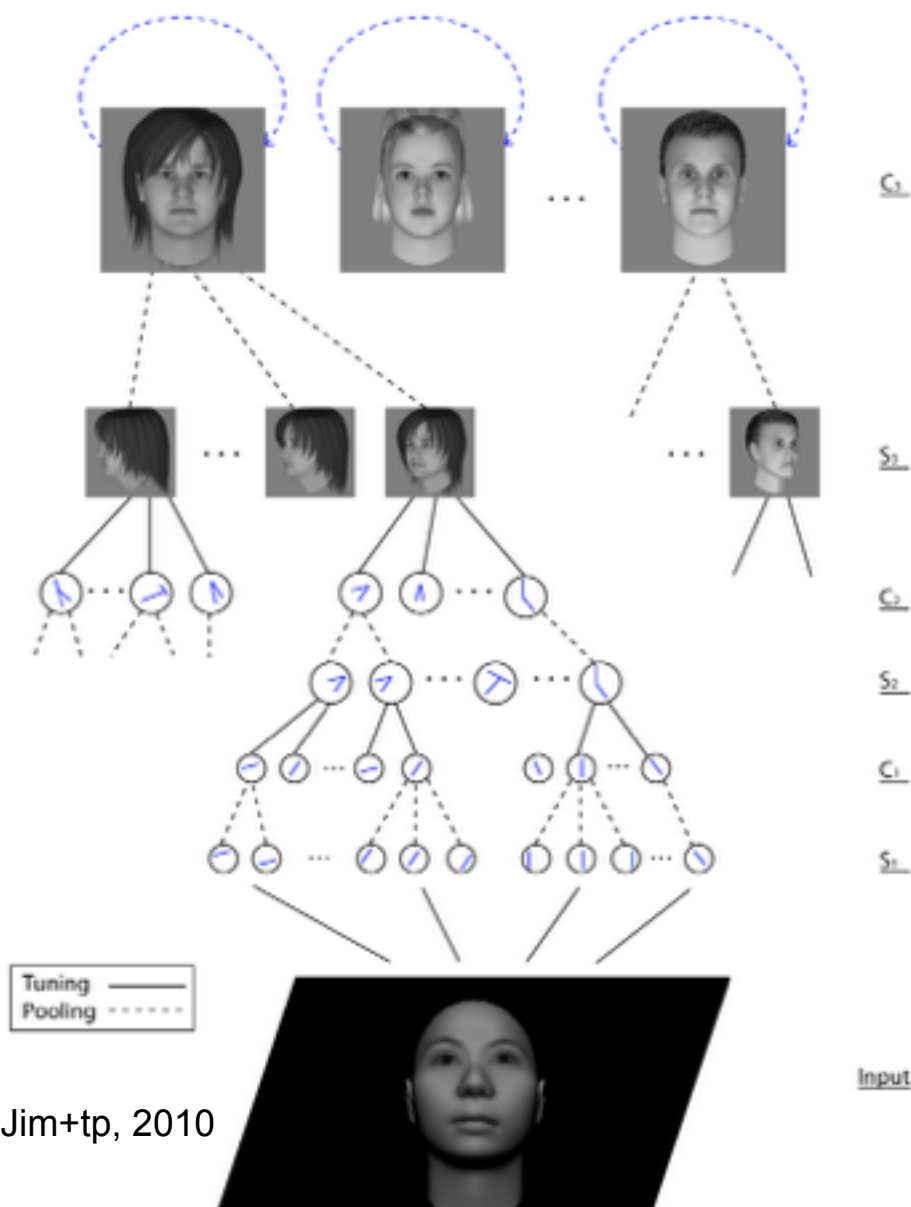
Class-specific modules

In general, global transformations – such as rotation in 3D of an object – can be represented only approximately. For specific classes of objects... good approximations of global non-affine transformations are possible, using the dot-products-and-templates approach.

We consider faces. An additional layer storing a set of face-specific templates for different rotations of a face can provide the required class-specific approximate invariance.

The transformations here are *class-specific* and not *generic*.

Recognizing a face from a different viewpoint

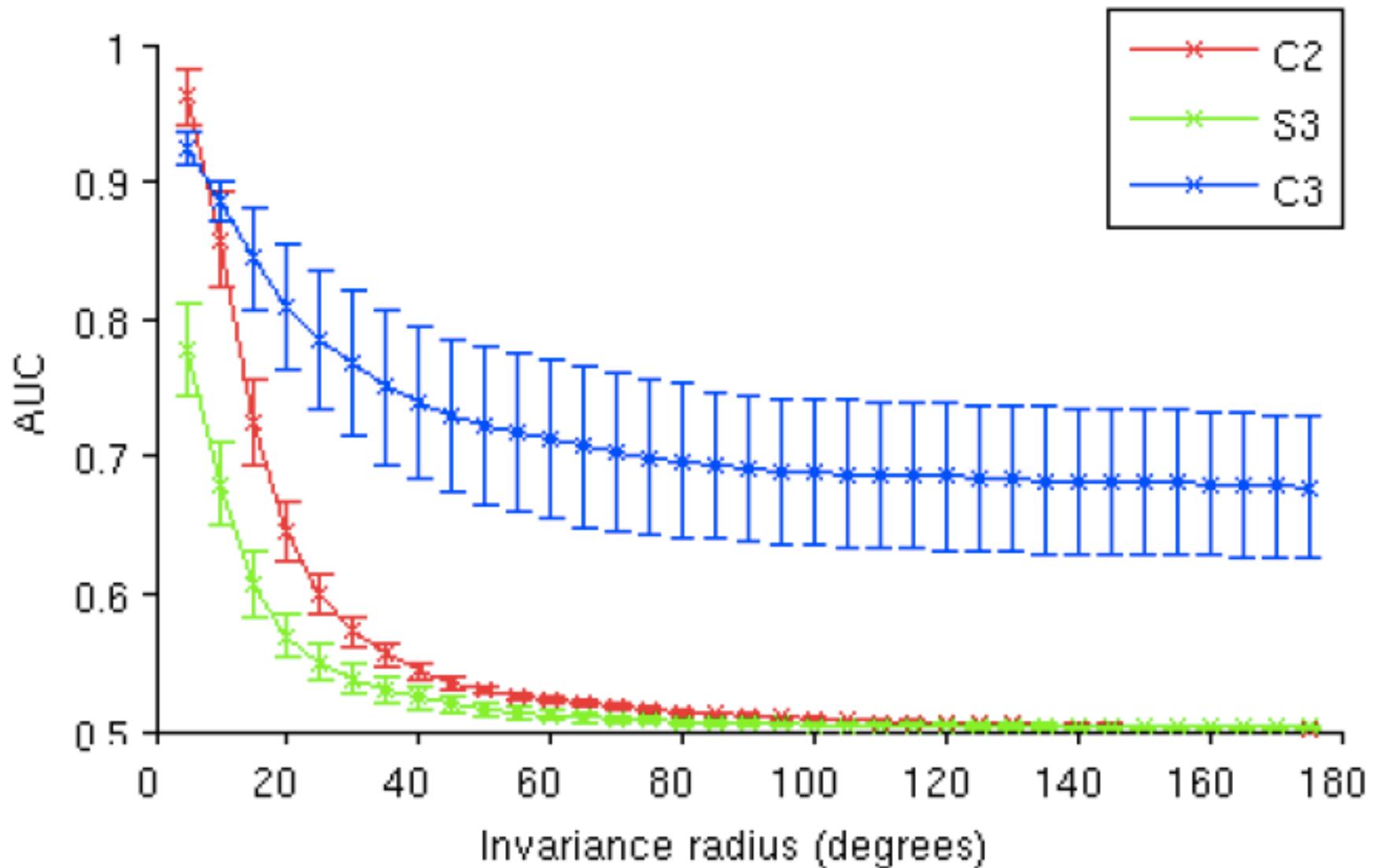


Viewpoint tolerant units
(complex units)

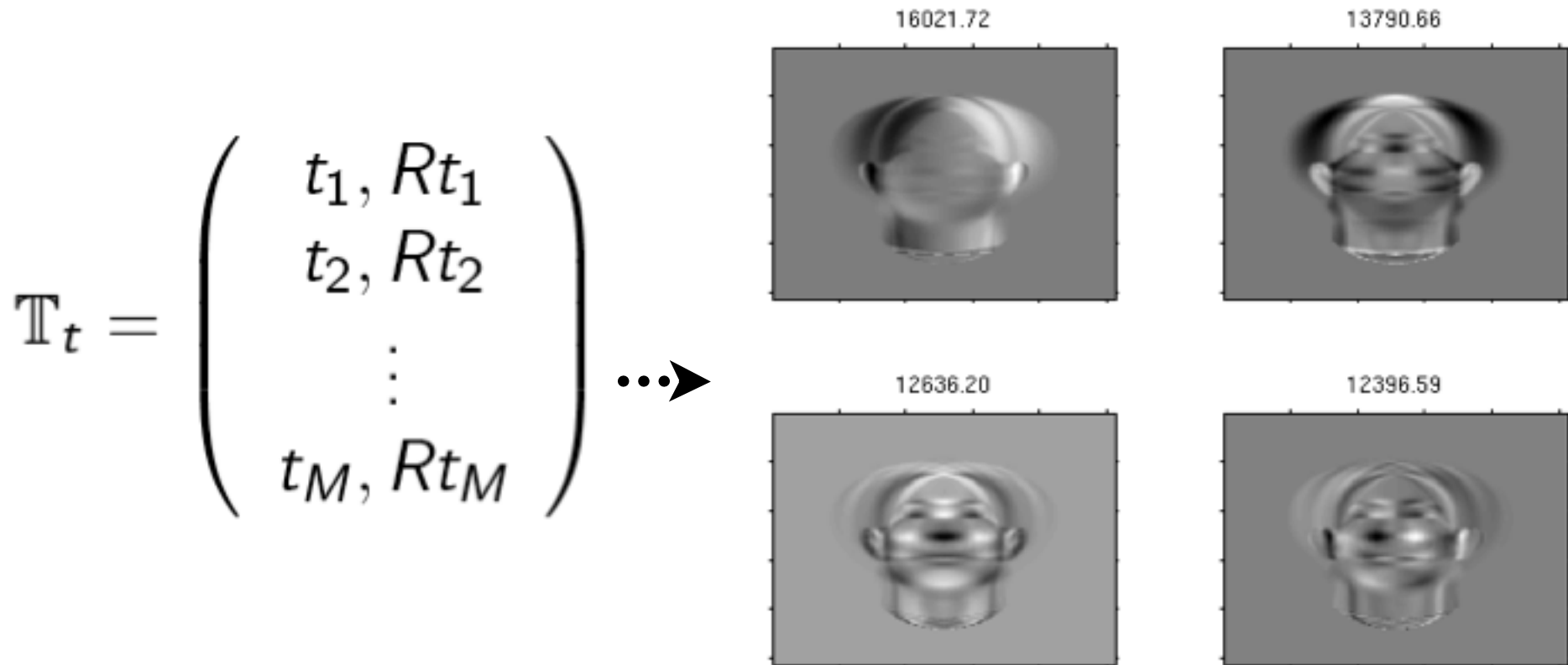
View-tuned units, tuned to full-face
templates for different view angles

*Tolerance to a transformation
may be learned from examples
of a class*

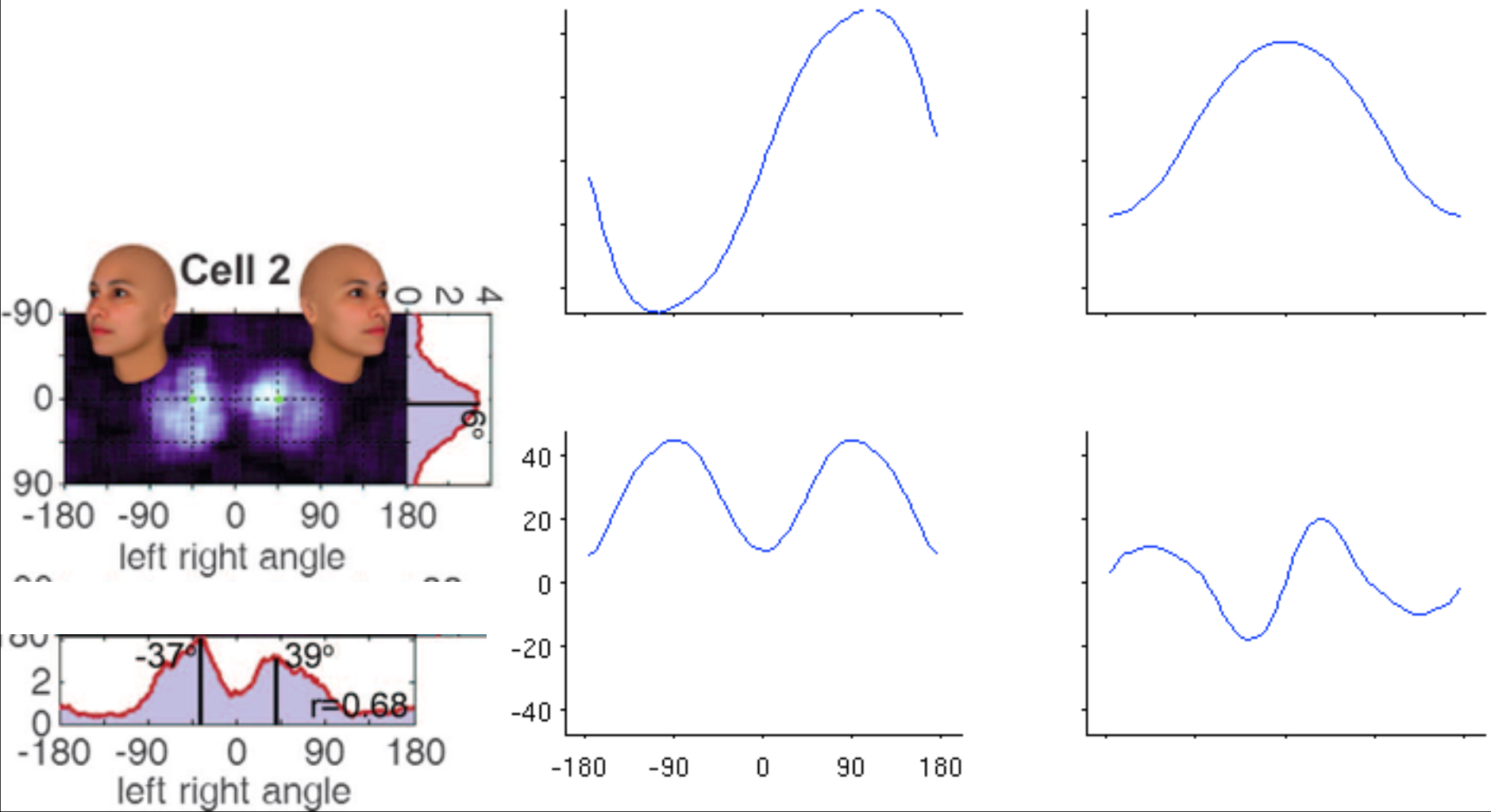
Learning class specific transformations: quasi-invariance to pose for faces



PCA of face views are tuning of AL neurons: what are they?
Lemma: PCAs here are odd or even functions, and so the complex cells always even (because of square)!



Response of simple AL “model” cells to different views of a face



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 the architecture and tuning properties of visual cortex predicted from basic symmetries of geometric transformations of images?
- 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.

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