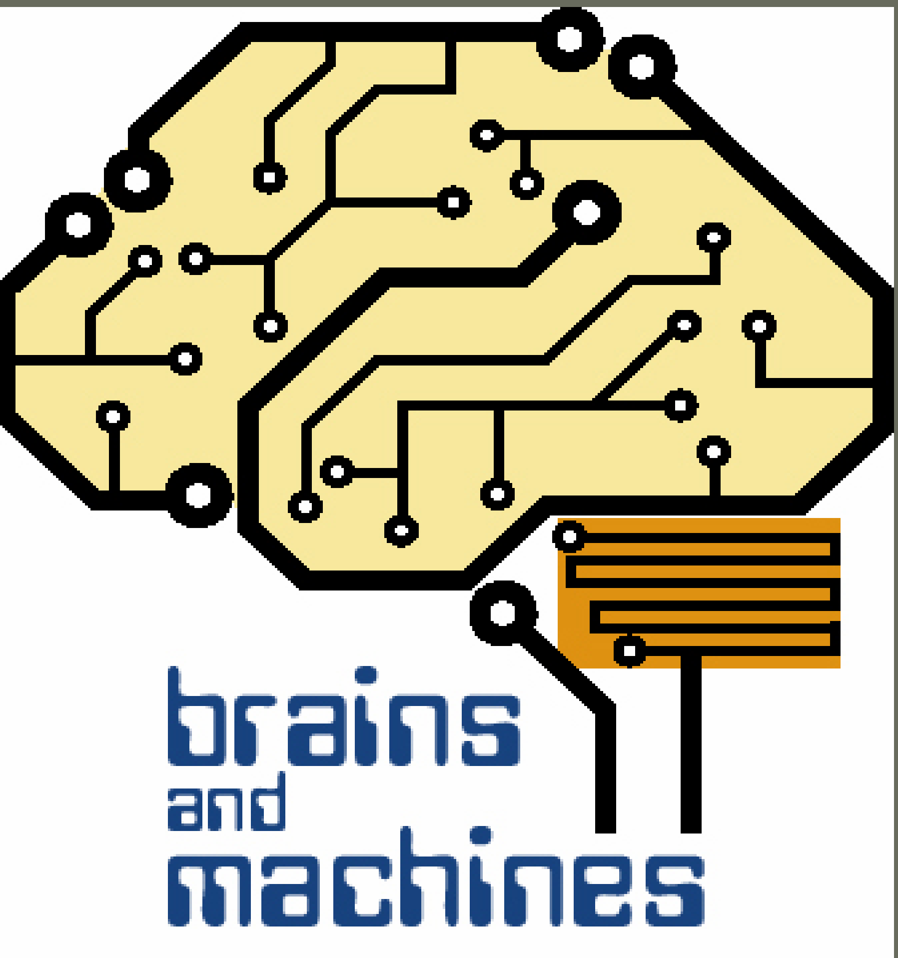
# Learning:
# theory, engineering applications,
## and neuroscience

T. Poggio,
Computer Science and Artificial Intelligence Lab,
McGovern Inst for Brain Research
MIT




Learning is the gateway to understanding the brain and to making intelligent machines

Problem of learning is a focus for

- modern math
- **computer science**
- neuroscience

# Learning from data: today and tomorrow

Two msgs in my talk:

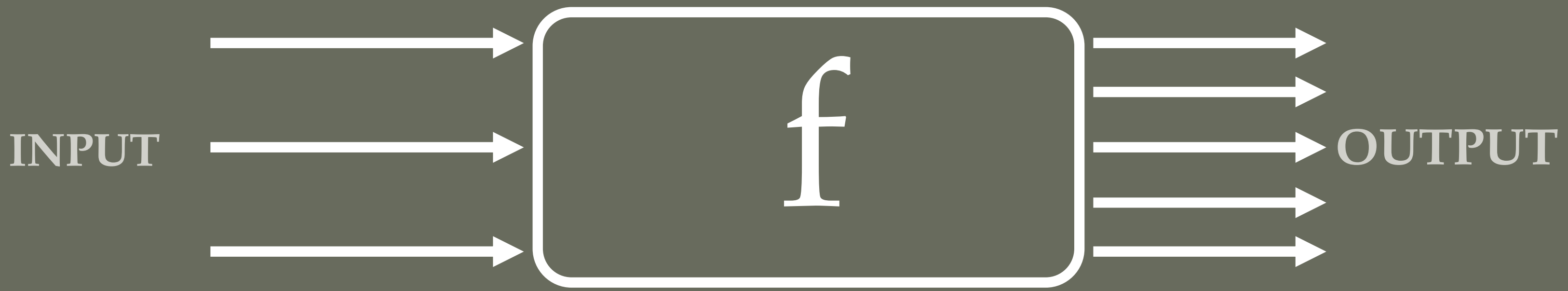
- **Learning theory**: it works (a couple of applications)

- The **brain may teach us how to improve on it**, example of vision

# First message

Supervised learning: a couple of applications

# Learning from examples: goal is not to memorize but to **generalize**, eg *predict*.



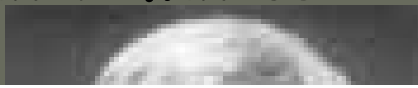

***Given*** a set of ℓ examples (data) $\{(x_1, y_1), (x_2, y_2), \ldots, (x_\ell, y_\ell)\}$

***Question***: find function *f* such that

is a *good predictor* of *y* for a *future* input *x (fitting the data is **not** enough!):*

$$f(x) = \hat{y}$$

# Interesting development: the theoretical foundations of learning are becoming part of mainstream mathematics



BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 39, Number 1, Pages 1–49
S 0273-0979(01)00923-5
Article electronically published on October 5, 2001

## ON THE MATHEMATICAL FOUNDATIONS OF LEARNING

FELIPE CUCKER AND STEVE SMALE

*The problem of learning is arguably at the very core of the problem of intelligence, both biological and artificial.*

### Introduction

(1) A main theme of this report is the relationship of approximation to learning and the primary role of sampling (inductive inference). We try to emphasize relations of the theory of learning to the mainstream of mathematics. In particular, there are large roles for probability theory, for algorithms such as *least squares*, and for tools and ideas from linear algebra and linear analysis. An advantage of doing this is that communication is facilitated and the power of core mathematics is more easily brought to bear.

# A simple algorithm - regularization in RKHS - ensures generalization...

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

implies

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

**Equation includes Regularization Networks (special cases are splines, Radial Basis Functions and Support Vector Machines). Function is nonlinear and general approximator. When V is the square loss, the c are given by**

$$(\mathbf{K} + \lambda \ell I)\mathbf{c} = \mathbf{y}$$

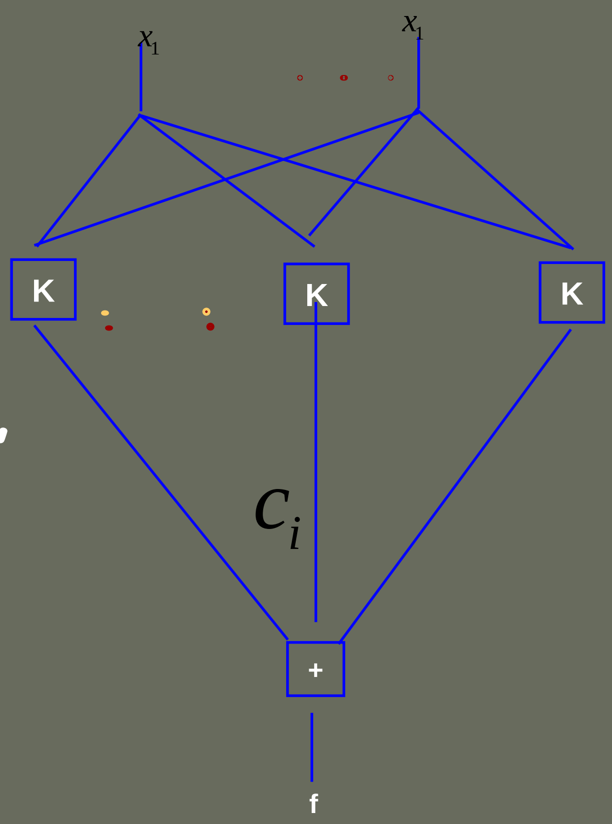
*For a review, see Poggio and Smale, **The Mathematics of Learning**, Notices of the AMS, 2003*

# A remark: equivalence to networks

Many different V lead to the same solution...

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

...and can be "written" as the same type of network...where the value of K corresponds to the "activity" of the "unit" and the $c_i$ correspond to (synaptic) "weights"



# Learning from examples: engineering applications


Bioinformatics
Artificial Markets
Object identification
Image analysis
Graphics
Text Classification
Object categorization
Decoding the neural code

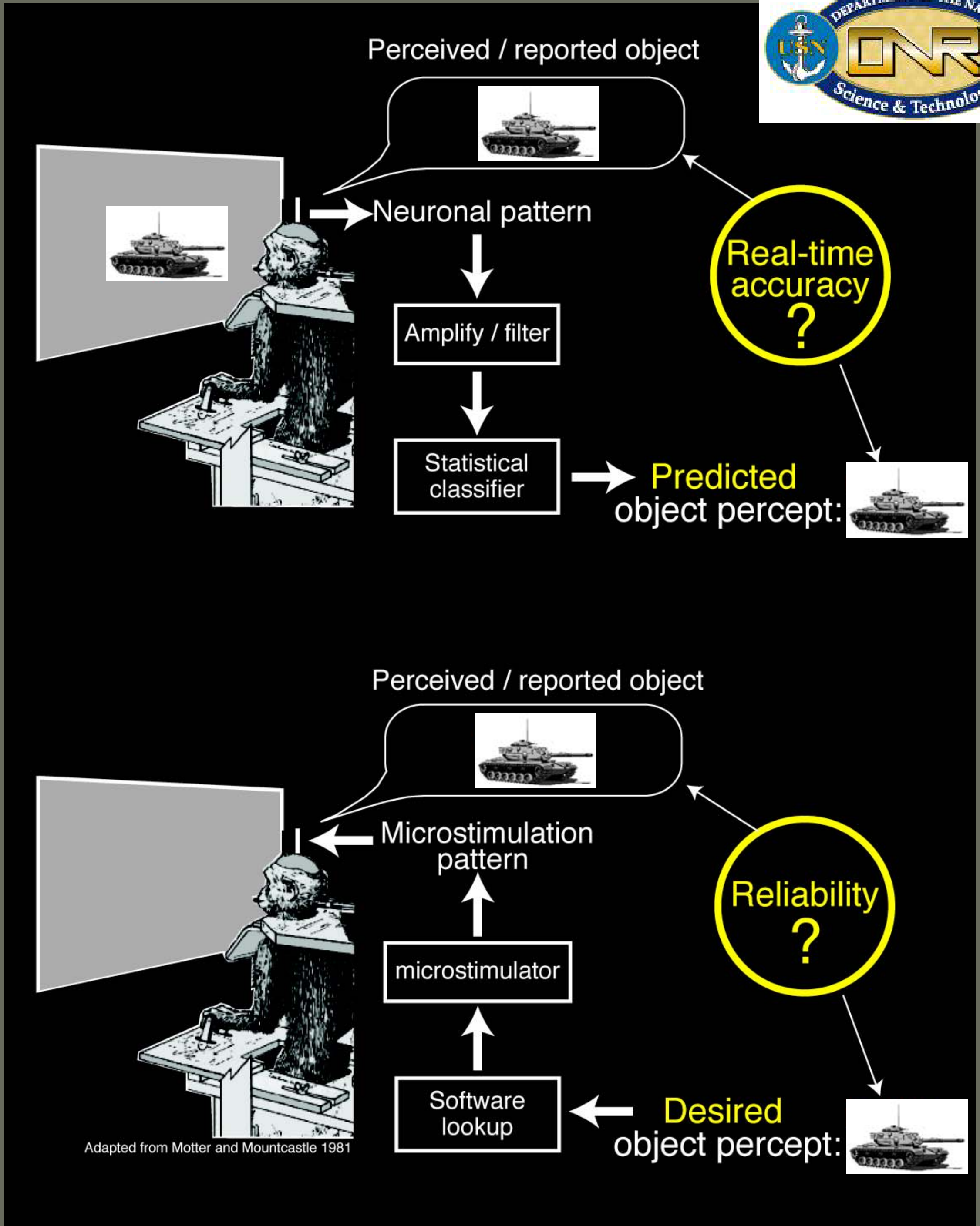
# One recent application of learning: using a classifier to *read-out* the code in IT cortex

Chou Hung, Gabriel Kreiman, James DiCarlo, Tomaso Poggio

Fast Readout of Object Identity from Macaque Inferior temporal Cortex, Science, Nov 4, 2005

(Read-out eg analysis):

Can we “read-out” the subject’s object percept?



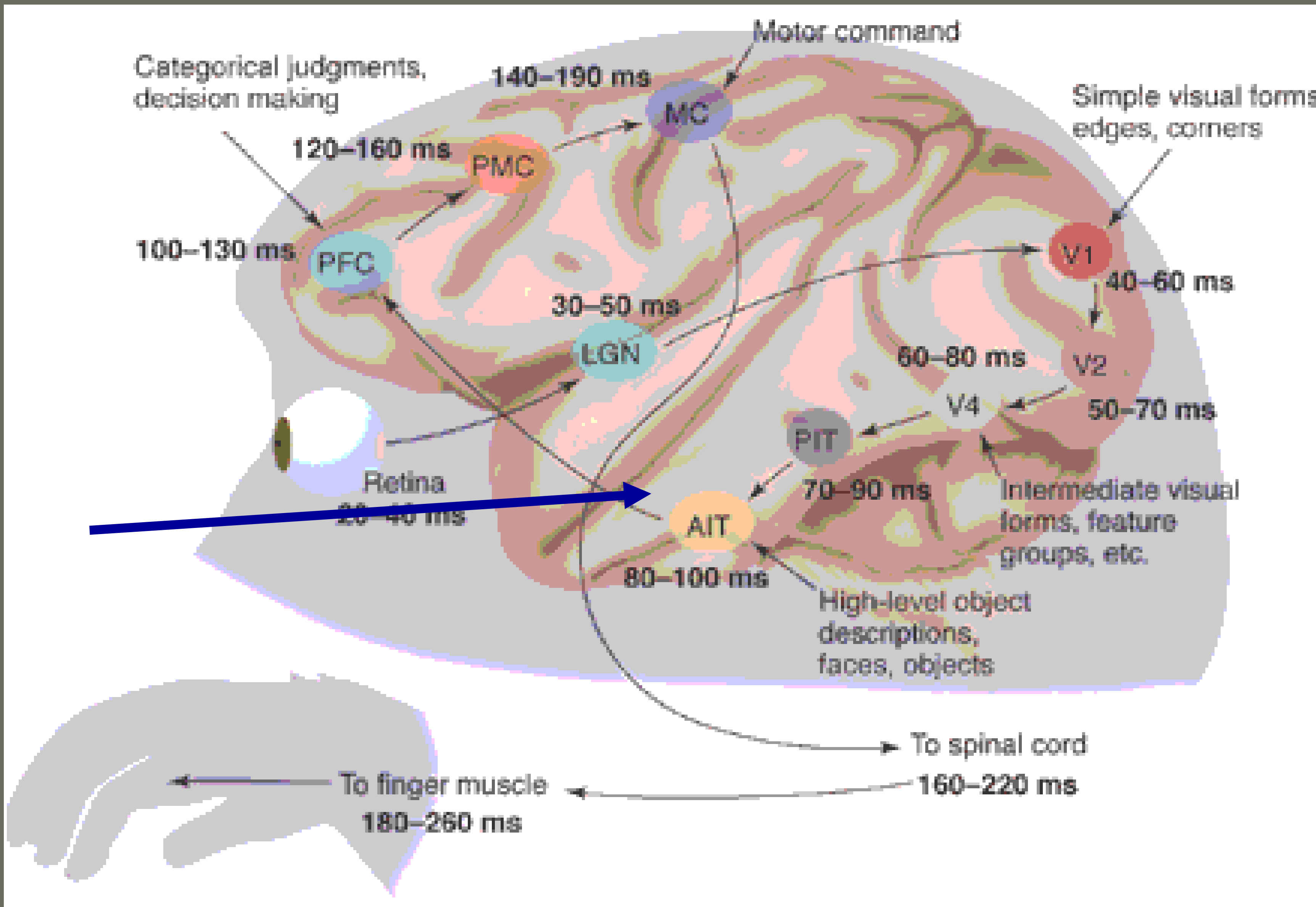
Adapted from Motter and Mountcastle 1981

# Goal 2

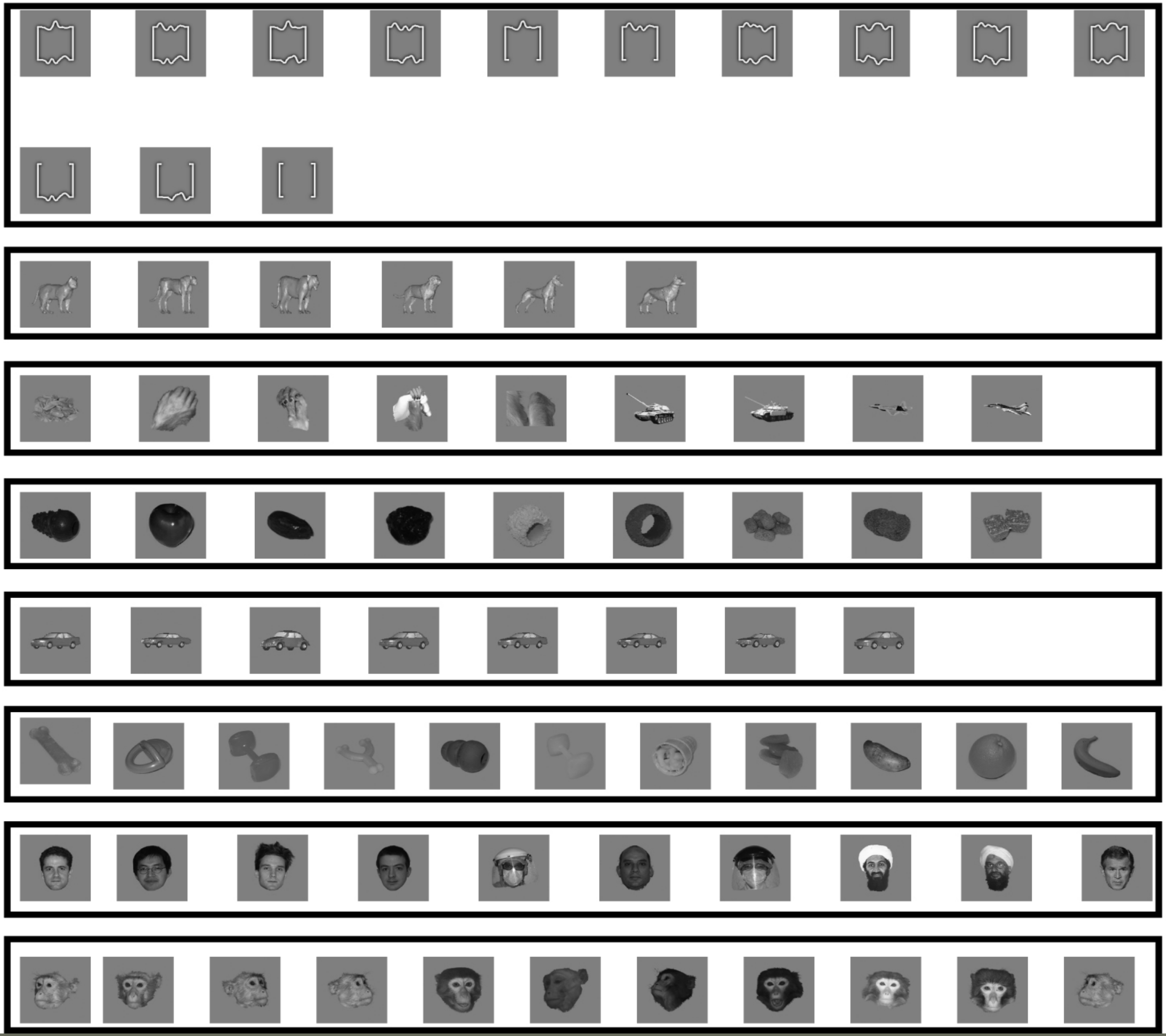
(Write-in eg synthesis):

Can we “write-in” (induce) an object percept?

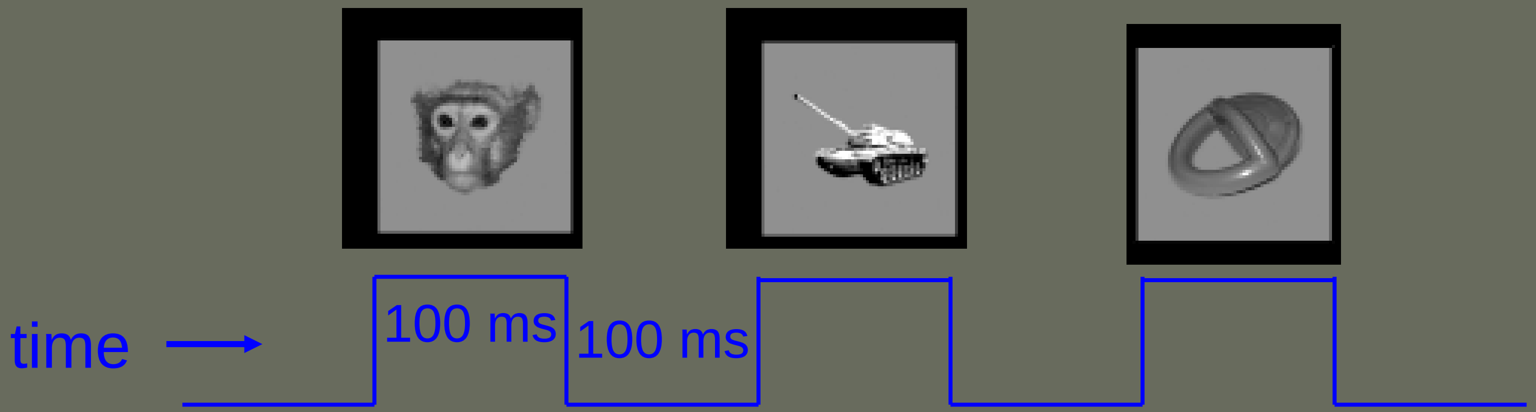
# The end station of the ventral stream in visual cortex is IT



# 77 objects, 8 classes

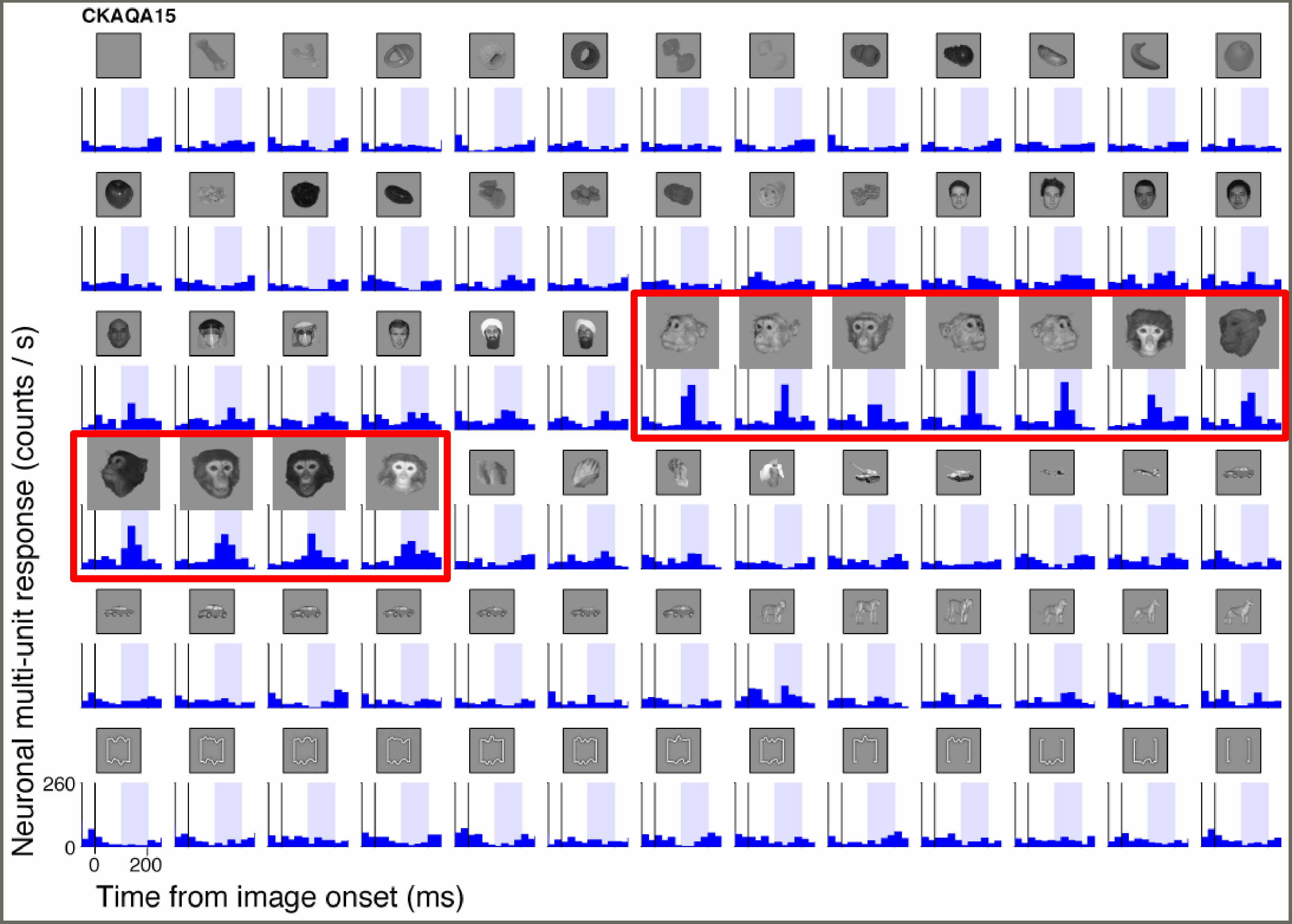


# Recording at each recording site during passive viewing




- 77 visual objects
- 10 presentation repetitions per object
- presentation order randomized and counter-balanced

# Example AIT recording site



# Training a classifier on neuronal activity.

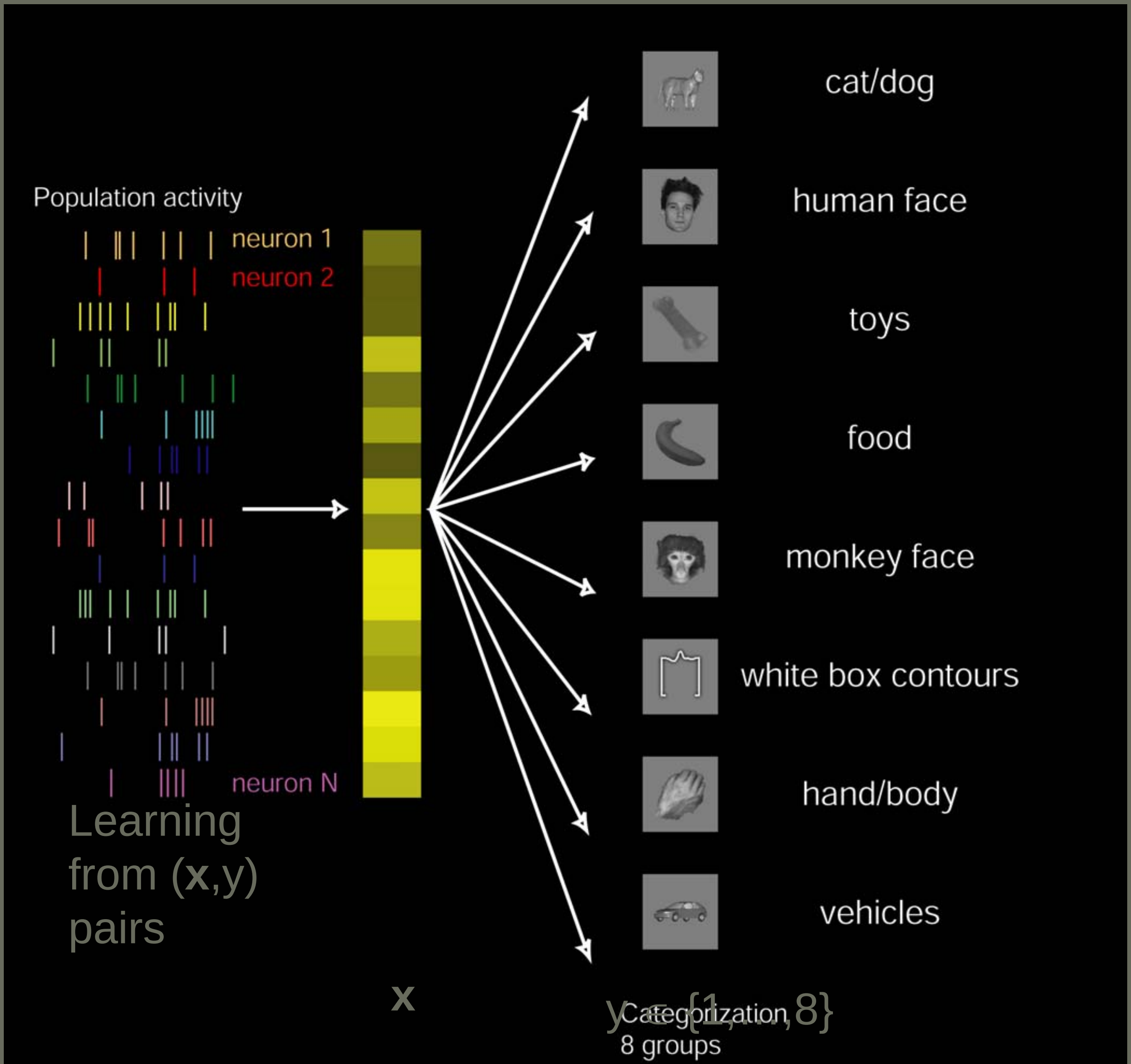

From a set of data (vectors of activity of n neurons (x) and object label (y)

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_\ell, y_\ell)\}$$

Find (by training) a classifier eg a function *f* such that $f(x) = \hat{y}$

is a ***good predictor*** of object label *y* for a ***future*** neuronal activity *x*

# Decoding the population response




Learning from (**x**,y) pairs

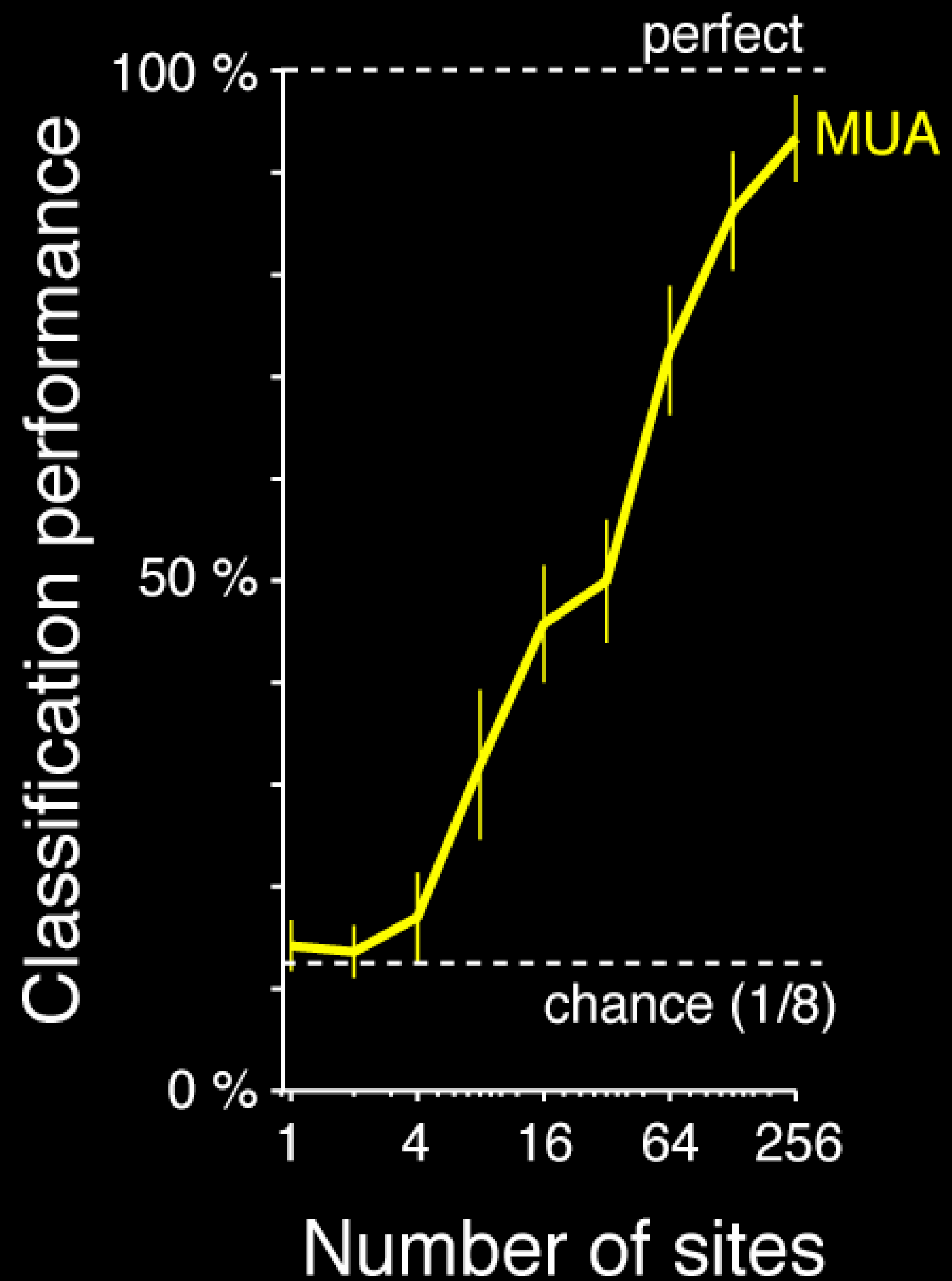
**x**

$y \in \{1,\ldots,8\}$

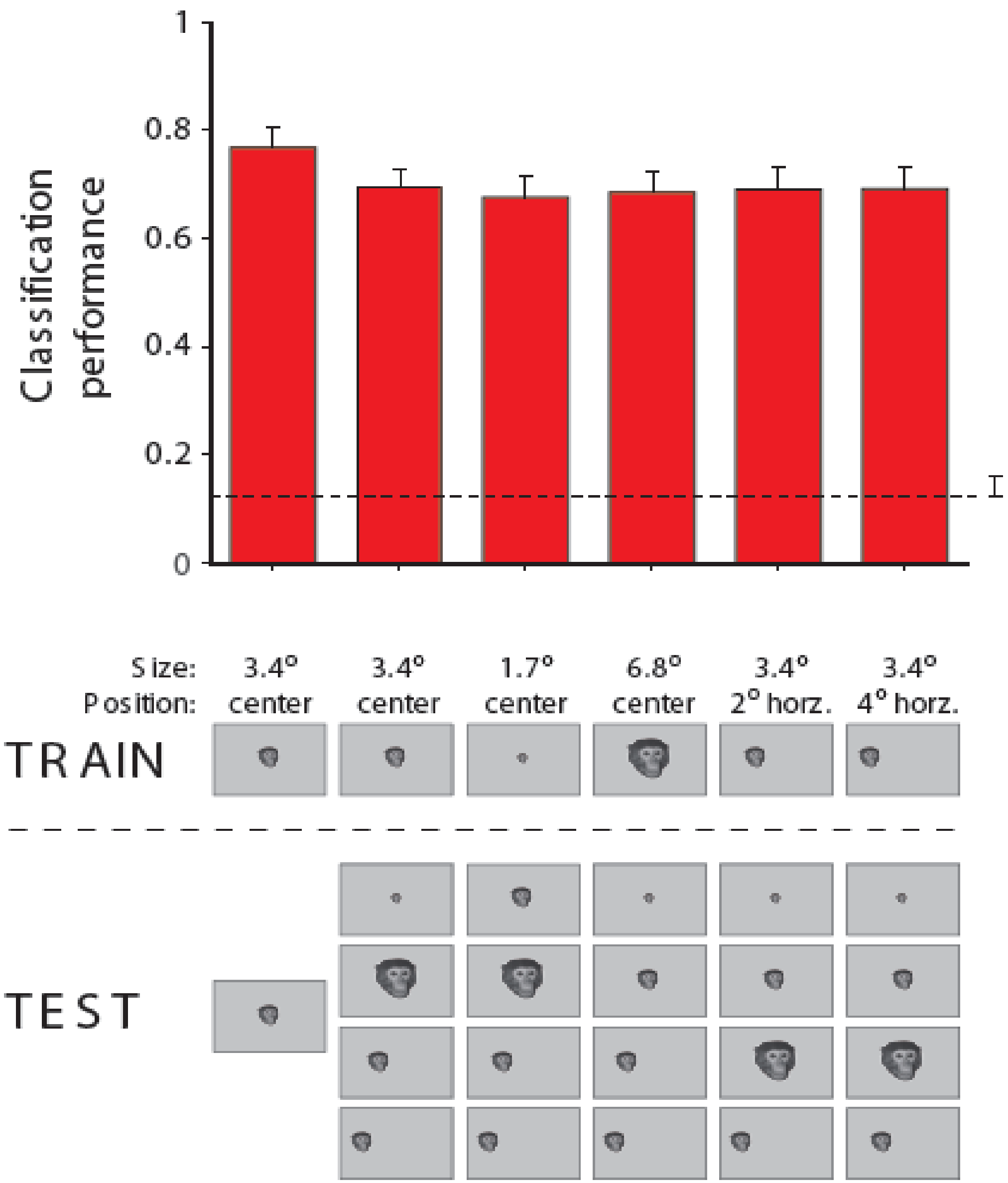
# Results:

**reliable object categorization using ~100 arbitrary AIT sites**

- *[100-300 ms] interval*
- *50 ms bin size*



# IT representation is invariant to changes in position and size



# Learning from examples: engineering applications


Bioinformatics
Artificial Markets
Object identification
Image analysis
Graphics
Text Classification
Object categorization
Decoding neural code

# ~10 years ago: RLSC or SVM works well for image recognition

Scanning in x,y and scale

Preprocessing with overcomplete dictionary of Haar wavelets

SVM Classifier

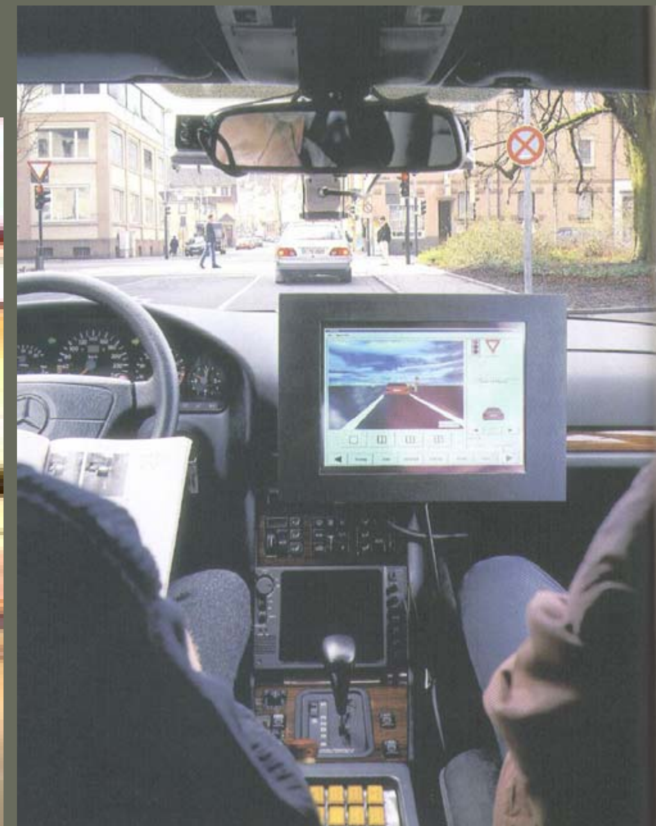
TRAINING

Data Base

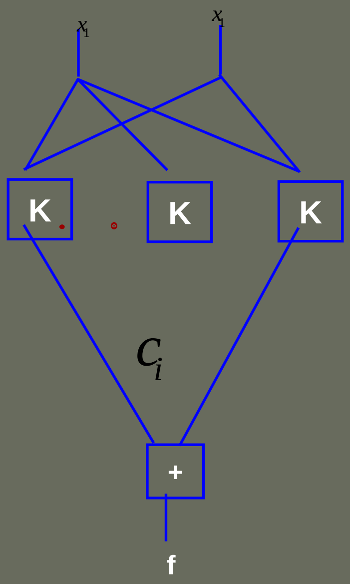
QP Solver

Sung, Poggio 1994; Papageorgiou and Poggio, 1998; also LeCun, Kanade, Schneiderman et al…

# Example: a pedestrian detection system (Mercedes) now about to become a product



# Second msg




More recently…

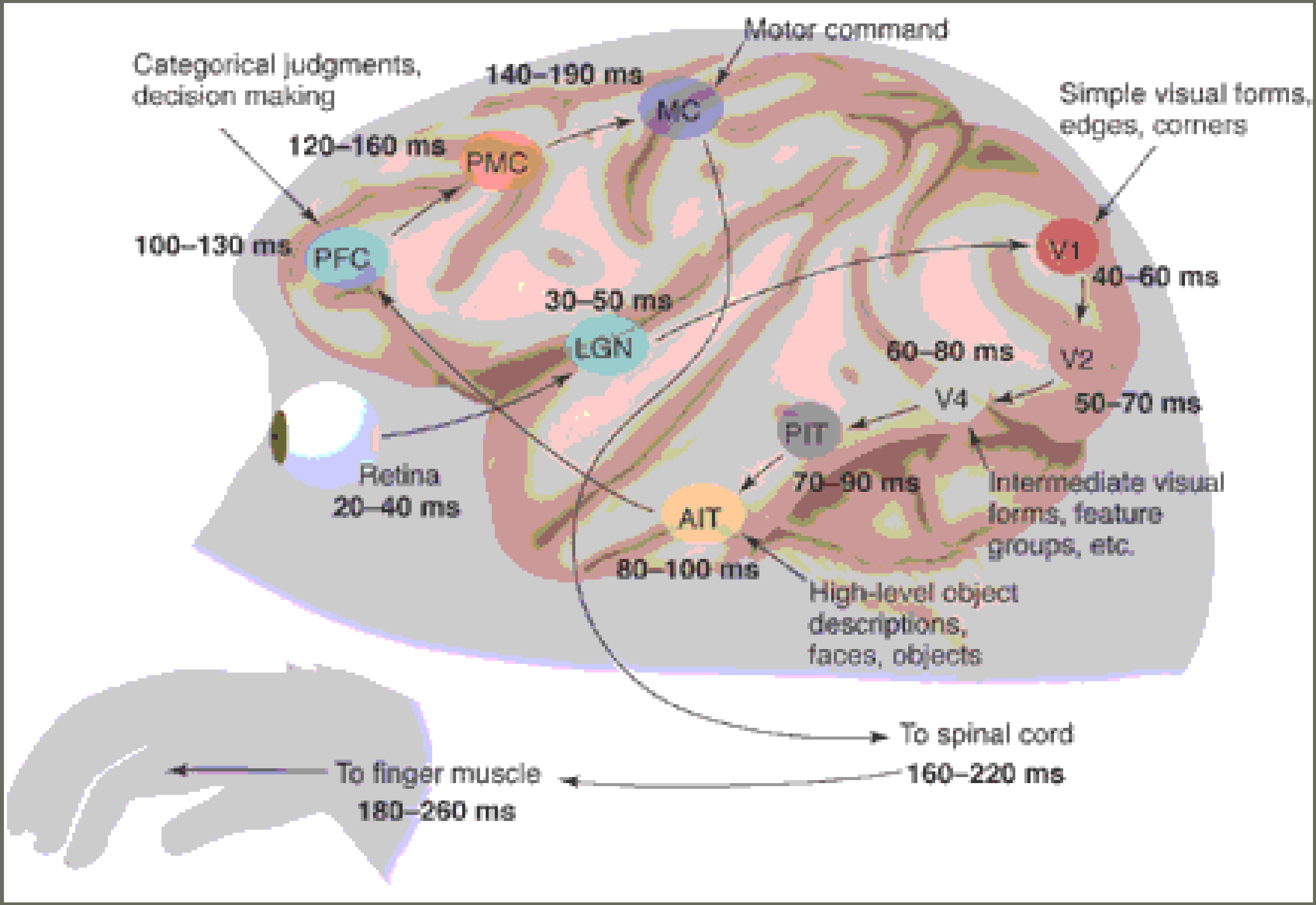
it turns out the brain may teach us

about a better architecture

at least for object recognition

and computer vision…

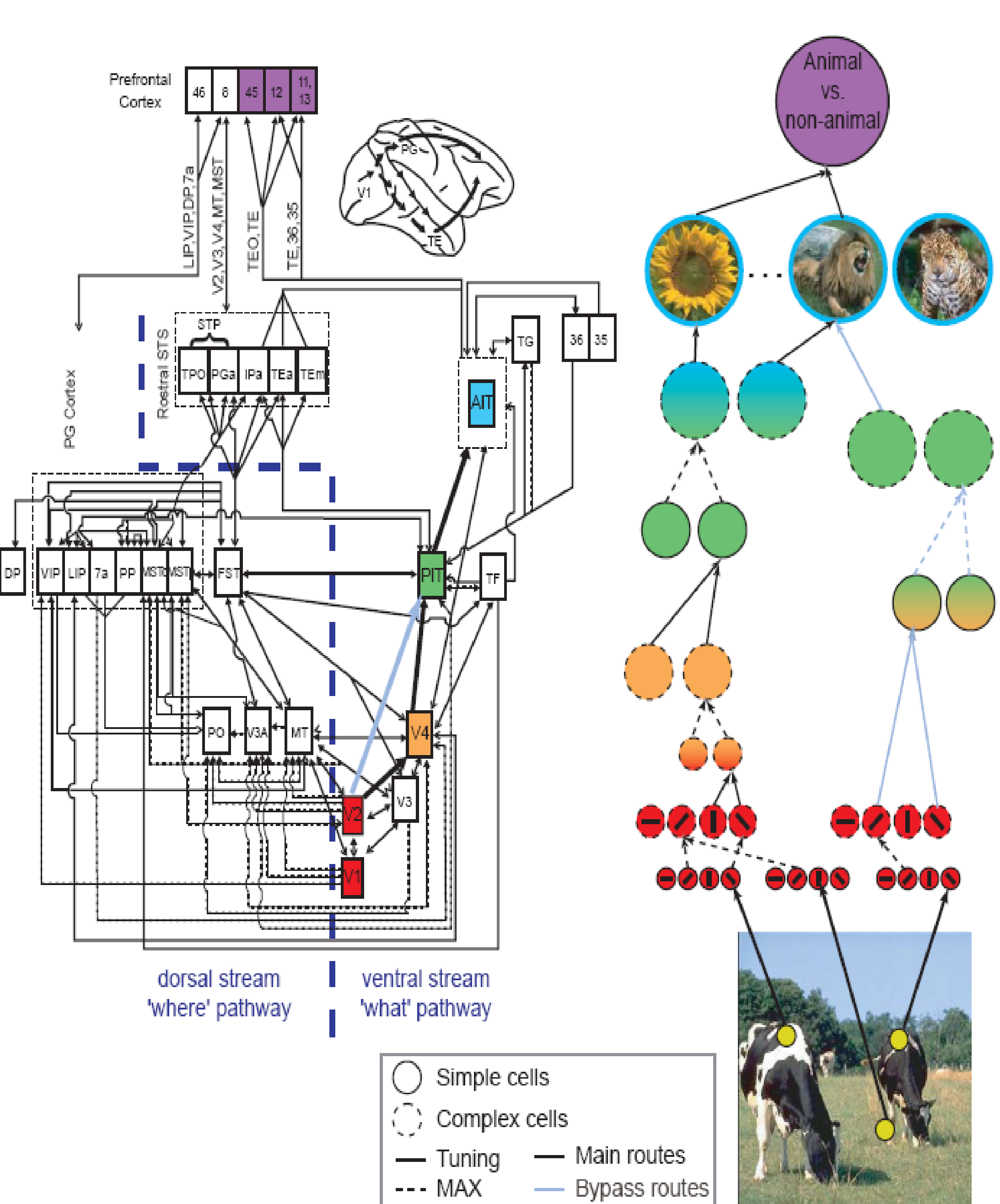
# A theory
# of the ventral stream of visual cortex





Thomas Serre, Minjoon Kouh, Charles Cadieu, Ulf Knoblich
and Tomaso Poggio

# Theory supported by data in V1, V4, IT; works as well as the best computer vision; mimics human performance




| Model layers | Corresponding brain area (tentative) | RF sizes | Number units | |
|---|---|---|---|---|
| classifier | PFC | | $1.0\ 10^{0}$ | |
| S4 | AIT | >4.4° | $1.5\ 10^{2}$ | ~ 5,000 subunits |
| C3 | PIT - AIT | >4.4° | $2.5\ 10^{3}$ | |
| C2b | PIT | >4.4° | $2.5\ 10^{3}$ | |
| S3 | PIT | 1.2°- 3.2° | $7.4\ 10^{4}$ | ~ 100 subunits |
| S2b | V4 - PIT | 0.9°- 4.4° | $1.0\ 10^{7}$ | ~ 100 subunits |
| C2 | V4 | 1.1°- 3.0° | $2.8\ 10^{5}$ | |
| S2 | V2 - V4 | 0.6°- 2.4° | $1.0\ 10^{7}$ | ~ 10 subunits |
| C1 | V1 - V2 | 0.4°- 1.6° | $1.2\ 10^{4}$ | |
| S1 | V1 - V2 | 0.2°- 1.1° | $1.6\ 10^{6}$ | |

Supervised task-dependent learning

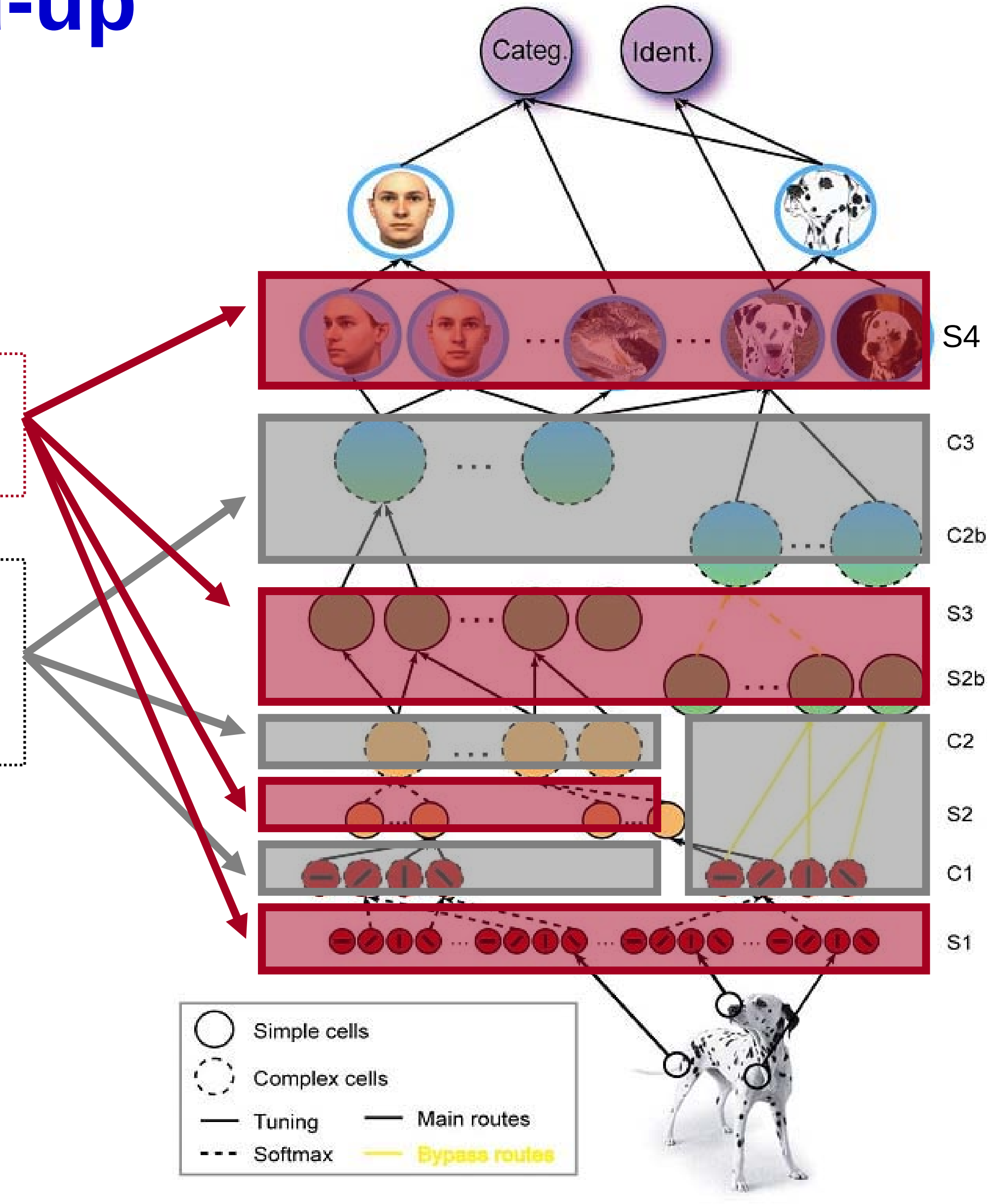
Unsupervised task-independent learning

increase in complexity (number of subunits), RF size and invariance

# Gradual Build-up

(1) **Selectivity (AND-like):** Gaussian-like function for tuning and specificity

(2) **Tolerance (OR-like):** Maximum-like operation for invariance over positions and scales



# 2 operations
# (max for S type units; Gaussian for C type unit) on the neighborhood of afferents

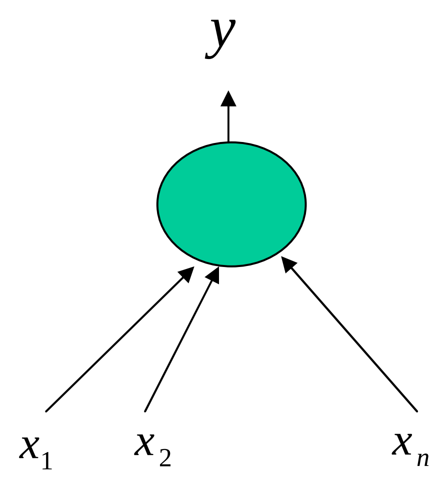
$$y = \max_{i \in N} x_i$$

Sparsification (OR-like)

$$y = e^{-\frac{|\mathbf{x}-\mathbf{w}|^2}{2\sigma^2}}$$

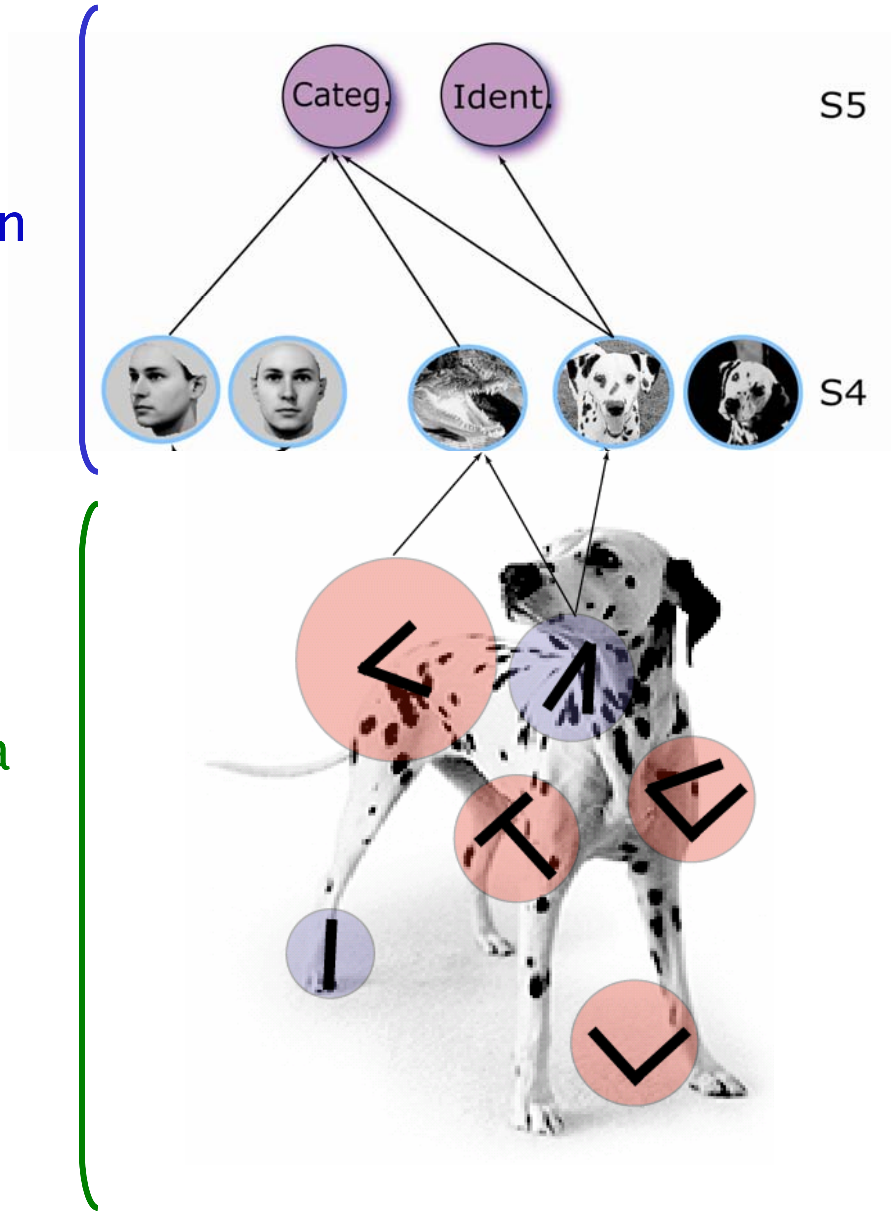
Lifting to higher-dimensional feature space (AND-like)

$$\text{where} \quad \mathbf{x} \in N$$



- Task-specific circuits (from IT to PFC)
  - Supervised learning: ~ Gaussian RBF

- Generic dictionary of shape components (from V1 to IT)
  - Unsupervised learning during a developmental-like stage



# *A list of model predictions which agree with data*

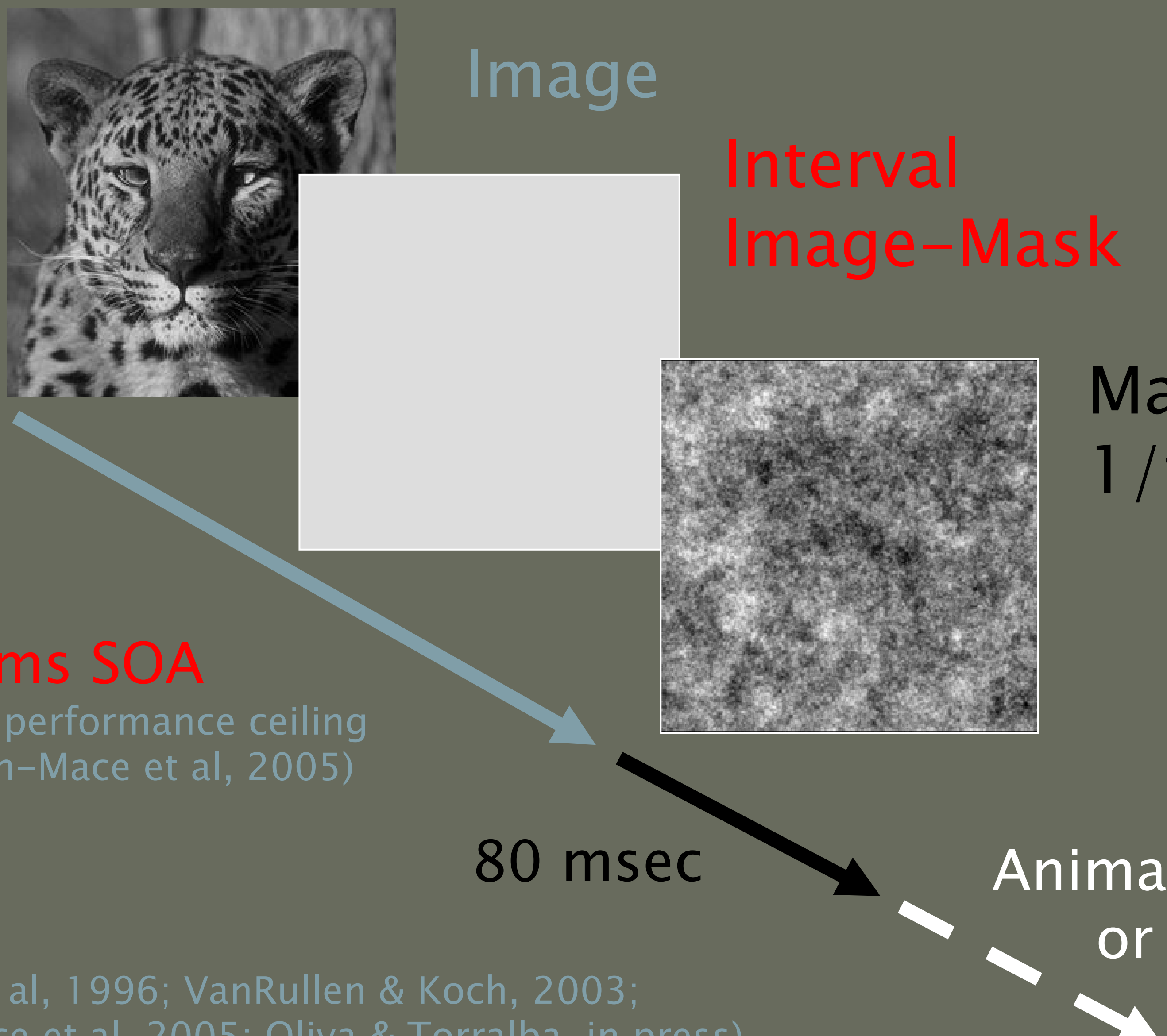
- MAX in V1 (Lampl et al, 2004) and V4 (Gawne et al, 2002)
- Differential role of IT and PFC in categ. (Freedman et al, 2001,2002,2003)
- Face inversion effect (Riesenhuber et al, 2004)
- **IT read out data** (Hung et al, 2005)
- Tuning and invariance properties Of VTUs in AIT (Logothetis et al, 1995)
- Average “average effect” in IT (Zoccolan, Cox & DICarlo, 2005)
- Tow-spot reverse correlation in V1 (Livingstone and Conway, 2003; Serre et al, 2005)
- Tuning for boundary conformation (Pasupathy & Connor, 2001) in V4
- Tuning for Gratings in V4 (Gallant et al, 1996; Serre et al, 2005)
- Tuning for two-bar stimuli in V4 (Reynolds et al, 1999; Serre et al, 2005)
- Tuning to Cartesian and non-Cartesian gratings in V4 (Serre et al, 2005)
- Two-spot interaction in V4 (Freiwald et al, 2005; Cadieu, 2005)

The model fits many physiological data,
predicts several new ones…

recently it provided a surprise (for us)…

…when we compared its performance with
human vision
(Thomas Serre with Aude Oliva)
on rapid categorization of complex natural images
…

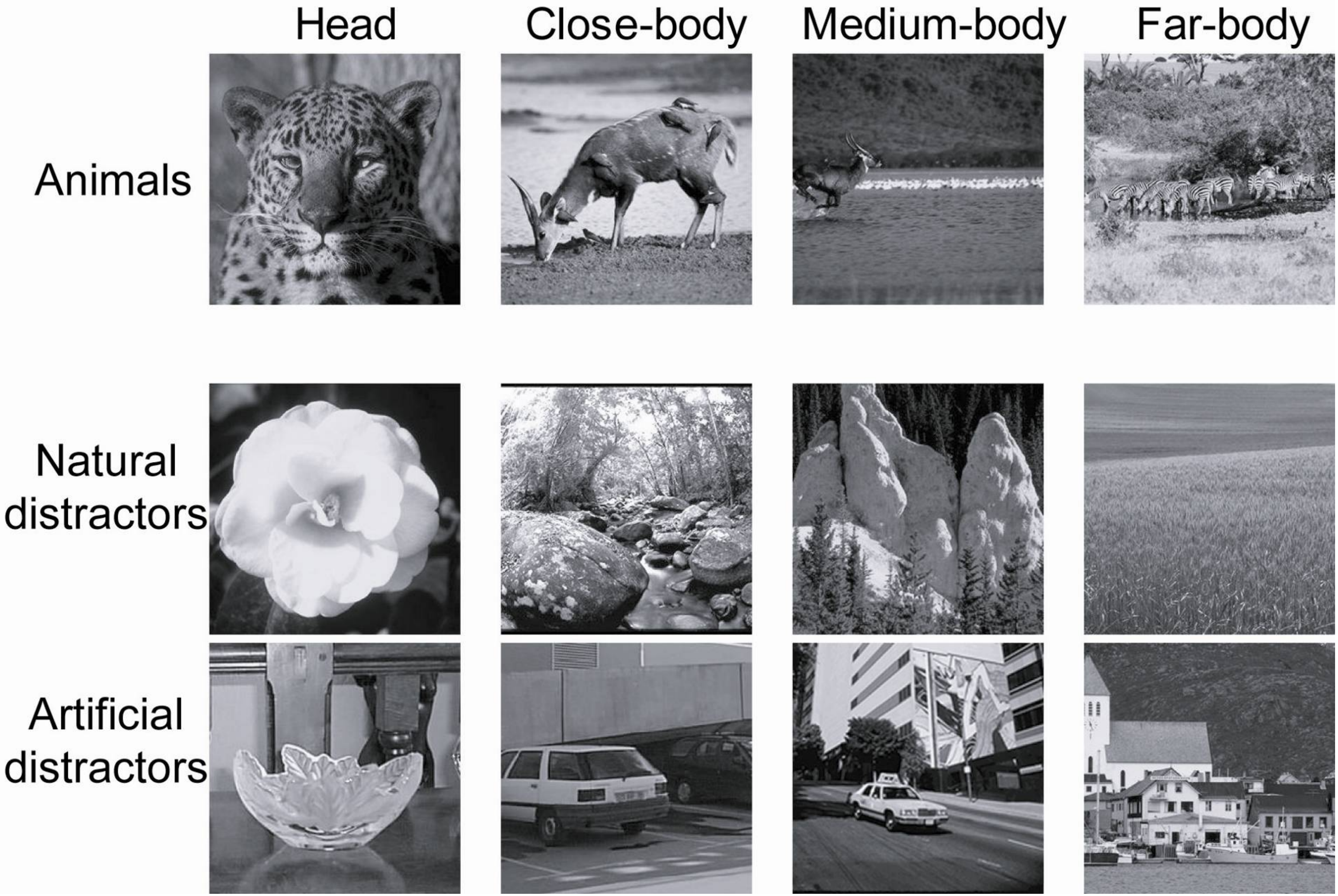
# Rapid categorization task




~ 50 ms SOA

close to performance ceiling
in (Bacon-Mace et al, 2005)

(Thorpe et al, 1996; VanRullen & Koch, 2003;
Bacon-Mace et al, 2005; Oliva & Torralba, in press)



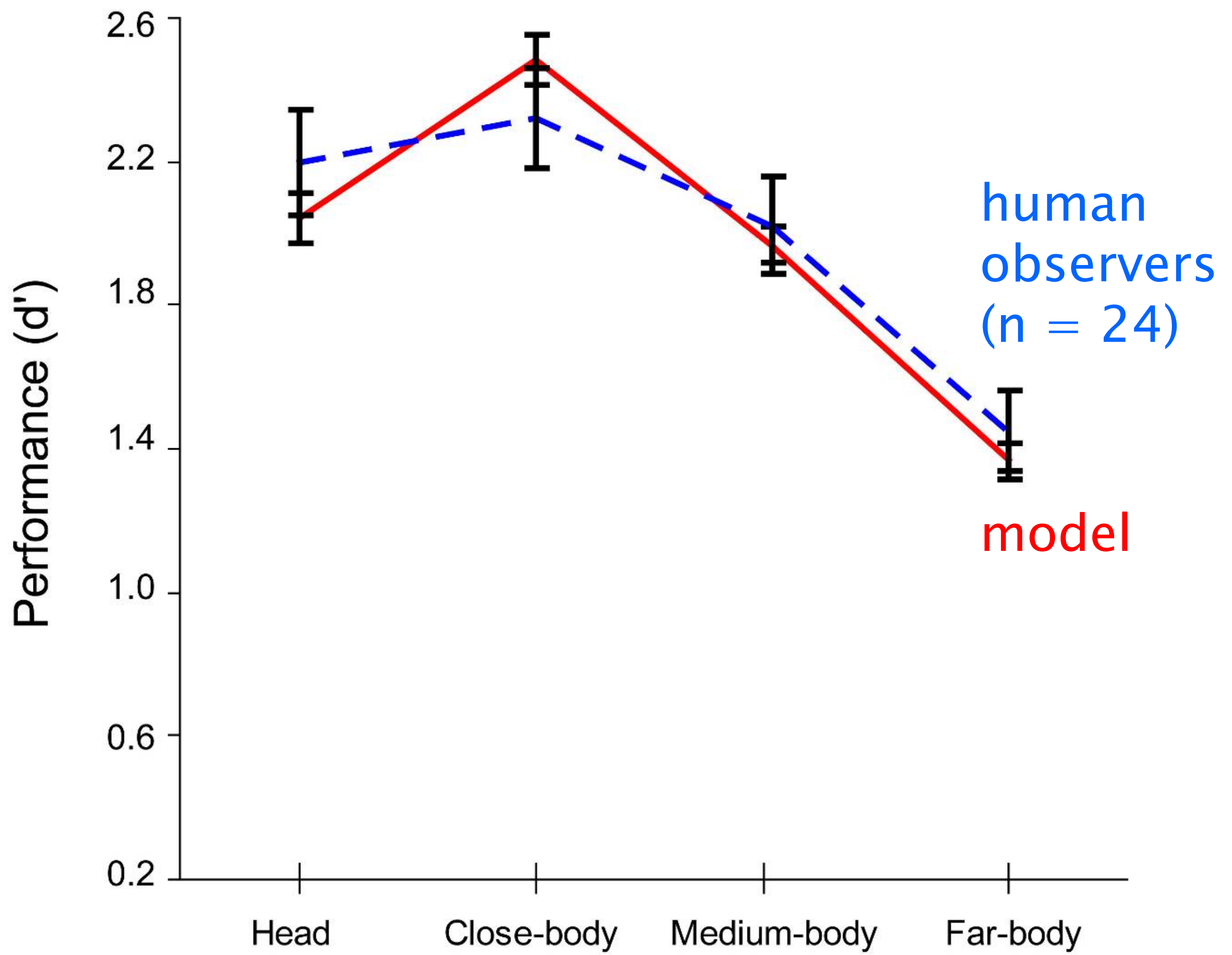

(Torralba & Oliva, 2003; Oliva & Torralba, in press)

# The model predicts human perf.

Model 82%

vs.

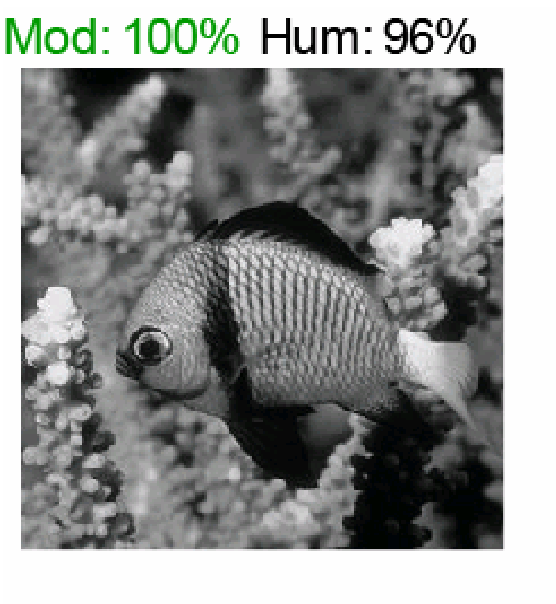
humans 80%




(Serre, Oliva and Poggio, in sub)

# Detailed comparison

- For each individual image
- How many times image classified as animal:
  - For humans: across subjects
  - For model: across 20 runs




- Heads: $\rho=0.71$
- Close-body: $\rho=0.84$
- Medium-body: $\rho=0.71$
- Far-body: $\rho=0.60$

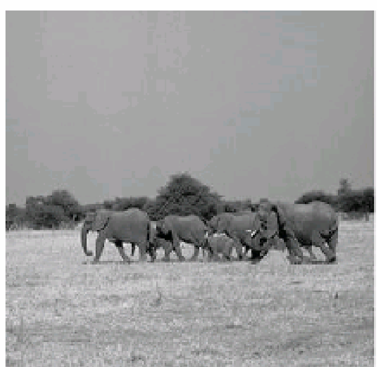
## Some hits

Mod: 100% Hum: 96%

Mod: 91% Hum: 83%

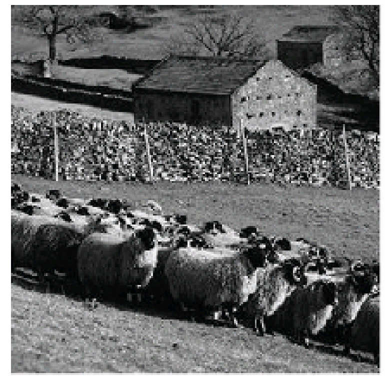
Mod: 100% Hum: 96%

Mod: 100% Hum: 91%



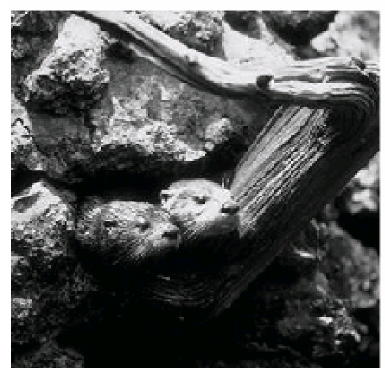
## Some misses

Mod: 22% Hum: 21%



Mod: 0% Hum: 21%

Mod: 33% Hum: 21%



Mod: 0% Hum: 29%

- The model can:
  - predict the tuning of neurons in several cortical areas
  - perform surprisingly well in complex categorization tasks, near human performance

…another surprise…

… was that it works as well as the best machine vision systems…

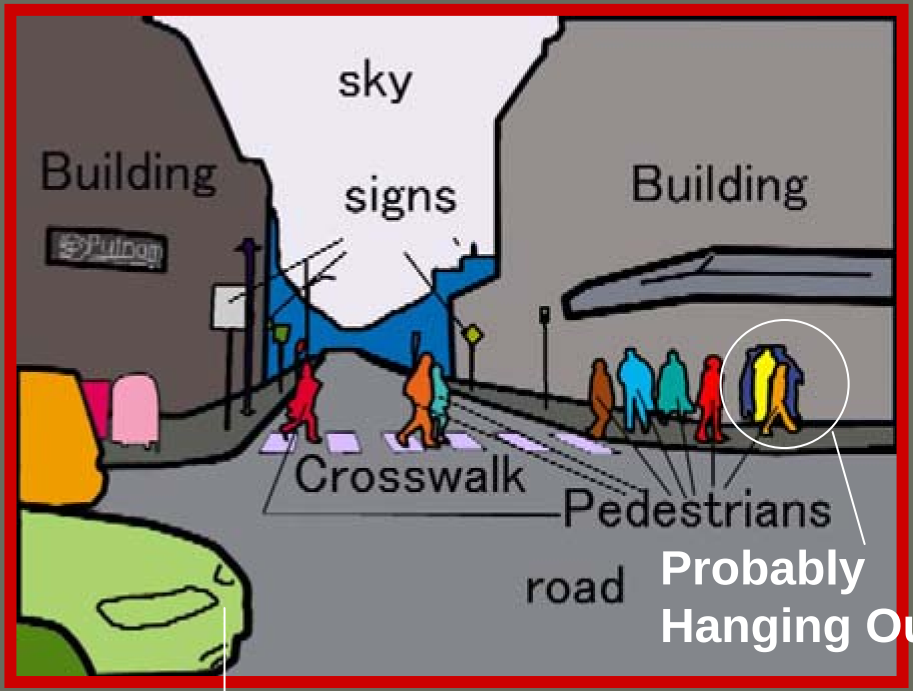
# Comparison with other AI systems

| Datasets | | | AI systems | Model |
|---|---|---|---|---|
| (CalTech) | Leaves | [Weber et al., 2000b] | 84.0 | 97.0 |
| (CalTech) | Cars | [Fergus et al., 2003] | 84.8 | 99.7 |
| (CalTech) | Faces | [Fergus et al., 2003] | 96.4 | 98.2 |
| (CalTech) | Airplanes | [Fergus et al., 2003] | 94.0 | 96.7 |
| (CalTech) | Motorcycles | [Fergus et al., 2003] | 95.0 | 98.0 |
| (MIT-CBCL) | Faces | [Heisele et al., 2002] | 90.4 | 95.9 |
| (MIT-CBCL) | Cars | [Leung, 2004] | 75.4 | 95.1 |

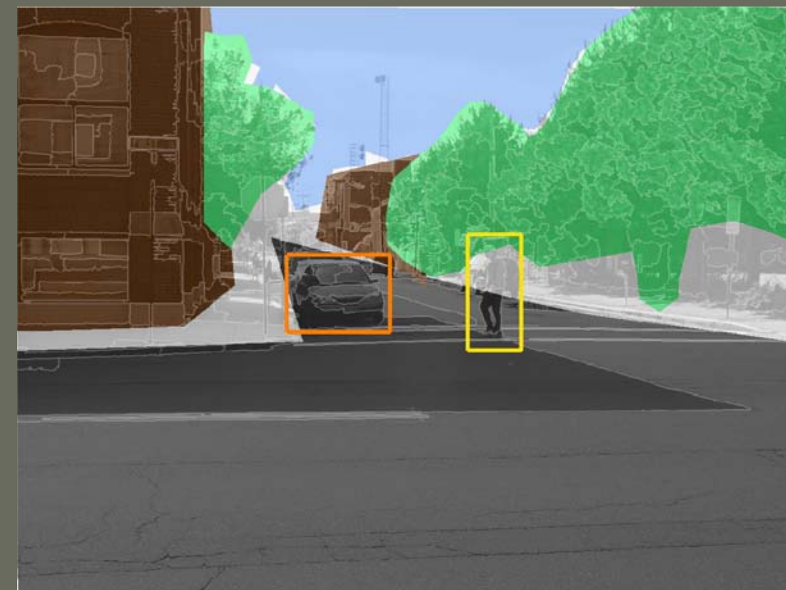

(Serre, Wolf & Poggio, 2005; Serre, Wolf, Bileschi, Riesenhuber & Poggio, to appear)

Since the workshop is on
Massive Datasets…
…here is a more difficult computer vision application
…on which the model of visual cortex does well

# Scene Understanding

# The StreetScenes Database (available on the Web)



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 |

# StreetScenes Database. Subjective Results


Results

# The end…

….with more details on the brain if you want to ask

…