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

Problem of learning: a focus for
- math
- computer algorithms
- neuroscience
Message of today

Computational neuroscience tries to understand the brain’s information processing principles.

Just now...computational neuroscience may begin to provide new ideas and approaches to machine learning and AI...
Learning: Math, Engineering, Neuroscience (until recently)

Theorems on foundations of learning
Predictive algorithms

- Bioinformatics
- Computer vision
- Computer graphics, speech synthesis, creating a virtual actor

How visual cortex works

Thursday, July 2, 2009
Learning: Math, Engineering, Neuroscience (now)

Learning Theory + Algorithms

Theorems on foundations of learning
Predictive algorithms

Engineering Applications

- Bioinformatics
- Computer vision
- Computer graphics, speech synthesis, creating a virtual actor

Computational Neuroscience: models+experiments

How visual cortex works

Thursday, July 2, 2009
~15 year old CBCL computer vision research: face detection

- Training Database
- 1000+ Real, 3000+ VIRTUAL
- 50,0000+ Non-Face Pattern

Sung & Poggio 1995
~10 year old CBCL computer vision research: SVM-based pedestrian detection system in Mercedes test car... now becoming a product (MobilEye, Israeli company)

~10 year old CBCL computer vision research: SVM-based pedestrian detection system in Mercedes test car... now becoming a product (MobilEye, Israeli company)

1. Today’s supervised learning algorithms: sample complexity problem and shallow architectures
2. Visual Cortex: hierarchical architecture, from neuroscience to a class of models
3. Physiology, psychophysics, computer vision
4. Models suggest new architectures for learning
5. Extensions and limitations of models
Supervised learning

There is an unknown **probability distribution** on the product space $Z = X \times Y$, written $\mu(z) = \mu(x, y)$. We assume that $X$ is a compact domain in Euclidean space and $Y$ a bounded subset of $\mathbb{R}$. The **training set** $S = \{(x_1, y_1), \ldots, (x_n, y_n)\} = \{z_1, \ldots, z_n\}$ consists of $n$ samples drawn i.i.d. from $\mu$.

$\mathcal{H}$ is the **hypothesis space**, a space of functions $f : X \to Y$.

A **learning algorithm** is a map $L : Z^n \to \mathcal{H}$ that looks at $S$ and selects from $\mathcal{H}$ a function $f_S : x \to y$ such that $f_S(x) \approx y$ in a **predictive way**.
Classical learning algorithms: Kernel Machines (eg Regularization in RKHS)

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

implies

\[
f(x) = \sum_{i}^{n} \alpha_i K(x, x_i)
\]

Equation includes splines, Radial Basis Functions and SVMs (depending on choice of V).

For a review, see Poggio and Smale, The Mathematics of Learning, Notices of the AMS, 2003; see also Schoelkopf and Smola, 2002; Bousquet, O., S. Boucheron and G. Lugosi.
Classical learning theory and Kernel Machines (Regularization in RKHS)

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

implies

\[
f(x) = \sum_{i}^{n} \alpha_i K(x, x_i)
\]

Kernel machines correspond to shallow networks
How do the learning machines described by classical learning theory -- such as kernel machines algorithms -- compare with brains?

One of the most obvious differences is the apparent ability of people and animals to learn from very few examples ("poverty of stimulus" problem).

A comparison with real brains offers another, related, challenge to learning theory. Classical "learning algorithms" correspond to one-layer architectures. The cortex suggests a hierarchical architecture.

Are hierarchical architectures with more layers the answer to the sample complexity issue?

*Present learning algorithms: “high” sample complexity and shallow architectures*

The Mathematics of Learning: Dealing with Data Tomaso Poggio and Steve Smale
1. Today’s supervised learning algorithms: sample complexity problem and shallow architectures
2. Visual Cortex: hierarchical architecture, from neuroscience to a class of models
3. Physiology, psychophysics, computer vision
4. Models suggest new architectures for learning
5. Extensions and limitations of models
The Ventral Stream

Hypothesis: the hierarchical architecture of the ventral stream in monkey visual cortex has a key role in object recognition...of course subcortical pathways may also be important (thalamus, in particular pulvinar...).
The Ventral Stream

visual recognition is a difficult learning problem (e.g., “is there an animal in the image?”)
• Human Brain
  – $10^{10} - 10^{11}$ neurons (~1 million flies 😊)
  – $10^{14} - 10^{15}$ synapses

• Ventral stream in rhesus monkey
  – $10^9$ neurons
  – 15 $10^6$ neurons in AIT (Anterior InferoTemporal) cortex
The ventral stream hierarchy: V1, V2, V4, IT

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

Kobatake & Tanaka, 1994
The ventral stream

Feedforward connections as well as backprojections:

How far we can push the simplest type of feedforward hierarchical models?
The ventral stream

Feedforward connections only?
The ventral stream

Feedforward connections only?
The ventral stream

Feedforward connections only?
The ventral stream

(Thorpe and Fabre-Thorpe, 2001)
Model of Visual Recognition (millions of units) based on neuroscience of cortex

Modified from (Gross, 1998)


[software available online]
Model of Visual Recognition (millions of units) based on neuroscience of cortex

- It is in the family of “Hubel-Wiesel” models (Hubel & Wiesel, 1959; Fukushima, 1980; Oram & Perrett, 1993, Wallis & Rolls, 1997; Riesenhuber & Poggio, 1999; Thorpe, 2002; Ullman et al., 2002; Mel, 1997; Wersing and Koerner, 2003; LeCun et al 1998; Amit & Mascaro 2003; Deco & Rolls 2006…)

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

- Feedforward only: an approximation of the first 100 msec of visual perception.

Thursday, July 2, 2009
Two operations (~OR, ~AND): disjunctions of conjunctions

- Tuning operation (Gaussian-like, AND-like)
  \[ y = \exp(-|x - w|^2) \]
  \[ y = \frac{w \times x}{|x|} \]

- Simple units

- Max-like operation (OR-like)
  \[ y = \max\{x_1, x_2, \ldots\} \]

- Complex units
Two operations (~OR, ~AND): disjunctions of conjunctions

- Tuning operation (Gaussian-like, AND-like)
  \[ y = \exp(-|x - w|^2) \]
  \[ y = \frac{w xx}{|x|} \]

- Simple units

- Max-like operation (OR-like)
  \[ y = \max\{x_1, x_2, \ldots\} \]

- Complex units
Two operations (~OR, ~AND): disjunctions of conjunctions

- Tuning operation (Gaussian-like, AND-like)
  \[ y = \exp(-|x - w|^2) \]
  \[ y = \frac{wx}{|x|} \]

- Simple units

- Max-like operation (OR-like)
  \[ y = \max\{x_1, x_2, \ldots\} \]

- Complex units
Learning: supervised and unsupervised
Overcomplete dictionary of “templates” or image “patches” is learned during an unsupervised learning stage (from ~10,000 natural images) by tuning S units.

see also (Foldiak 1991; Perrett et al 1984; Wallis & Rolls, 1997; Lewicki and Olshausen, 1999; Einhauser et al 2002; Wiskott & Sejnowski 2002; Spratling 2005)
Overcomplete dictionary of “templates” or image “patches” is learned during an unsupervised learning stage (from ~10,000 natural images) by tuning S units.

see also (Foldiak 1991; Perrett et al 1984; Wallis & Rolls, 1997; Lewicki and Olshausen, 1999; Einhauser et al 2002; Wiskott & Sejnowski 2002; Spratling 2005)
Learning: supervised and unsupervised

Task-specific circuits (from IT to PFC)
- **Supervised** learning: ~ classifier

- Overcomplete dictionary of “templates” or image “patches” is learned during an **unsupervised** learning stage (from ~10,000 natural images) by tuning S units.

see also (Foldiak 1991; Perrett et al 1984; Wallis & Rolls, 1997; Lewicki and Olshausen, 1999; Einhauser et al 2002; Wiskott & Sejnowski 2002; Spratling 2005)
Learning: supervised and unsupervised
Learning: supervised and unsupervised

- Preprocessing stages lead to a representation that has lower sampling complexity than the image itself.

Learning: supervised and unsupervised

- Preprocessing stages lead to a representation that has lower sampling complexity than the image itself.

Learning: supervised and unsupervised

We refer to the sample complexity of the preprocessing stage as the # of labeled examples required by the classifier at the top.

- Preprocessing stages lead to a representation that has lower sampling complexity than the image itself.

1. Today’s supervised learning algorithms: sample complexity problem and shallow architectures
2. Visual Cortex: hierarchical architecture, from neuroscience to a class of models
3. Physiology, psychophysics, computer vision
4. Models suggest new architectures for learning
5. Extensions and limitations of models
Hierarchical feedforward models of the ventral stream

Millions of units

CBCL software available on the Web
Hierarchical feedforward models of the ventral stream

Hierarchical Feedforward Models: predict/are consistent with 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)

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)
- 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)

Thursday, July 2, 2009
Hierarchical feedforward models of the ventral stream

Rapid Categorization
Hierarchical feedforward models of the ventral stream

Rapid Categorization
Hierarchical feedforward models of the ventral stream

Rapid Categorization: mask should force visual cortex to operate in

Animal present or not?

Thorpe et al 1996; Van Rullen & Koch 2003; Bacon-Mace et al 2005
Hierarchical feedforward models of the ventral stream

Feedforward Models: “predict” rapid categorization
(82% model vs. 80% humans)
Hierarchical feedforward models of the ventral stream

- Image-by-image correlation:
  - Heads: $\rho=0.71$
  - Close-body: $\rho=0.84$
  - Medium-body: $\rho=0.71$
  - Far-body: $\rho=0.60$
Feedforward Models: perform well compared to engineered computer vision systems (in 2006)
Hierarchical feedforward models of the ventral stream

Feedforward Models: perform well compared to engineered computer vision systems (in 2006)
Hierarchical feedforward models of the ventral stream

Feedforward Models: perform well compared to engineered computer vision systems (in 2006)
1. Today’s supervised learning algorithm: sample complexity problem and shallow architectures
2. Visual Cortex: from neuroscience to a class of models
3. Physiology, psychophysics, computer vision
4. Models suggest new architectures for learning
5. Extensions (video+attention) and limitations of feedforward hierarchical models
Hierarchical feedforward models of visual cortex may be wrong
...but present a challenge for “classical” learning theory:

an unusual, hierarchical architecture with unsupervised and supervised learning working well...

...so... we need theories -- not just models!
GOAL:
Hierarchical architectures to preprocess images/signals in order to reduce the sampling complexity of a classifier trained with labeled examples. The hierarchical architecture is synthesized from a large number of unsupervised examples.

Our ideal goal is to show that more layers are “better” and characterize under which conditions...

Joint work with J. Bouvrie, Lorenzo Rosasco, Steve Smale
Hierarchical/Deep Learning

Neocognitron, from Fukushima et al., 1980

CBCL Model

Hinton’s Deep Autoencoder

Convolutional Neural Networks (LeCun)

Figure: Bengio & LeCun, 2007

Figure: T. Serre

8-class digits problem, 1-nearest neighbor classifier, Euclidean distance vs. 2/3-layer derived distance ($u = 12$, $v = 20$, 500 templates/layer, 3-pixel translations).
The ingredients needed to define the derived kernel consist of:

- A finite *architecture* of nested domains. We’ll call them patches.
- A suitable family of *function spaces* defined on each patch.
- A set of *transformations* defined on patches.
- A set of *templates* which connect the mathematical model to a real world setting.
An Architecture of Patches

We first consider an architecture composed of three layers of patches: $v_1$, $v_2$ and $S_q$ in $\mathbb{R}^2$, with $v_1 \subset v_2 \subset S_q$,
We consider a function space on $S^q$, denoted by

$$\text{Im}(S^q) = \{ f : S^q \rightarrow [0, 1] \},$$

as well as the function spaces $\text{Im}(v_1)$, $\text{Im}(v_2)$ defined on subpatches $v_1$, $v_2$, respectively.

Functions can be interpreted as grey scale images when working with a vision problem for example.
Transformations

Next, we assume a set $H_{v_1}$ of transformations that are maps from the smallest patch to the next larger patch

$$h : v_1 \rightarrow v_2.$$ 

Similarly $H_{v_2}$ with $h : v_2 \rightarrow Sq$.

The sets of transformations are assumed to be finite. Examples of transformations are translations, scalings and rotations... In the vision interpretation, a translation $h$ can be thought of as moving the image over the “receptive field”. $f \circ h$ “restricts $f$ to a smaller patch associated with $h$. 

$\text{ Thursday, July 2, 2009 }$
Template sets are finite, 
\( T_{v_1} \subset \text{Im}(v_1) \) and 
\( T_{v_2} \subset \text{Im}(v_2) \)

- they are image patches sampled from some set of unlabeled images.
- link the mathematical development to real world problems.
Recursive Definition

For a general $n$ layer architecture $v_1 \subset v_2 \subset \cdots \subset v_n = S q$, let $N_n = N_{v_n}$ and $H_n = H_{v_n}$, $T_n = T_{v_n}$.
Recursive Definition

For a general \( n \) layer architecture \( v_1 \subset v_2 \subset \cdots \subset v_n = Sq \), let \( N_n = N_{v_n} \) and \( H_n = H_{v_n} \), \( T_n = T_{v_n} \).

Definition

Let us assume that a non-negative valued, normalized, feature map \( N_1 \) is given. Thus the derived kernel can be defined as \( K_1(f, g) = \langle N_m(f), N_m(g) \rangle_{L^2(T_{m-1})} \). The recursion is

\[
N_m(f)(t) = \max_{h \in H} \hat{K}_{m-1}(f \circ h, t), \quad t \in T_{m-1}
\]

with \( t \in T_{m-1}, H = H_{m-1}, K_m(f, g) = \langle N_m(f), N_m(g) \rangle_{L^2(T_{m-1})} \) and \( \hat{K}_{m-1} \) being the normalized kernel.
Example: Discrimination Results for 1-D strings

Consider an exhaustive architecture: \( v_m = \{1, \ldots, m\} \), \( T_m = \text{Im}(v_m) = S^m \), for \( m = 1, \ldots, n \) and transformations are all possible translations.

**Theorem**

If \( f, g \) are \( n \)-strings and \( \hat{K}_n(f, g) = 1 \) then one of the following statements is true:

- \( f, g \) are the same string
- one is the reversal of the other
- \( f, g \) are the “checkerboard” pattern:
  
  \( f = ababa \cdots, g = babab \cdots \), with \( f \) and \( g \) odd length strings.
Summary

- Compact mathematical description of a feedforward model of the visual cortex
- "Derived" kernel recursively defined
- Initial results on invariance/discrimination properties

- Open problem: Discrimination/Approximation properties
- Open problem: Number of layers and sample complexity (poverty-of-stimulus)
- Open problem: Efficient learning of the templates
- Conjecture: small effective dimensionality at each layer
1. Today’s supervised learning algorithm: sample complexity problem and shallow architectures
2. Visual Cortex: from neuroscience to a class of models
3. Physiology, psychophysics, computer vision
4. Models suggest new architectures for learning
5. Extensions (video+attention) and limitations of feedforward hierarchical models
Computational mechanisms of invariant recognition

Animal vs. non-animal

Complex cells
Tuning
Simple cells
MAX
Main routes
Bypass routes

PG Cortex
PG Cortex
Rostral STS
Prefrontal Cortex
STP
STP
DP
DP
VIP
VIP
LIP
LIP
7a
7a
PP
PP
FST
FST
PO
PO
V3A
V3A
MT
MT
TPO
TPO
PGa
PGa
IPa
IPa
V3
V3
V4
V4
PIT
PIT
TF
TF
TG
TG
36
36
35
35

LIP, VIP, DP, 7a
LIP, VIP, DP, 7a
V2, V3, V4, MT, MST
V2, V3, V4, MT, MST
PIT, AIT
AIT, 36, 35

dorsal stream
'where' pathway
ventral stream
'what' pathway

MST p
MST p
C1
S1
S2
S2b
C2

classification units

Model
layers
RF sizes
Num. units

0.2 - 1.1
0.4 - 1.6
0.6 - 2.4
1.1 - 3.0
0.9 - 4.4
1.2 - 3.2

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

Unsupervised
task-independent learning

Supervised
task-dependent learning

Thursday, July 2, 2009
Computational mechanisms of invariant recognition

From static images...
Computational mechanisms of invariant recognition

... to temporal sequences of images

Thomas Serre, Hueihan Jhuang & Tomaso Poggio collaboration with David Sheinberg at Brown University

Thursday, July 2, 2009
The system
The system
The system

motion-direction selective cells in the primary visual cortex

Shmuel & Grinvald '96

Thursday, July 2, 2009
The system

motion-direction selective cells in the primary visual cortex
The system
Behaviors of interest

- drink
- eat
- groom
- hang
- micro-movement
- rear
- rest
- walk
Dataset

Serre* Jhuang* Garrote Poggio Steele in prep

Thursday, July 2, 2009
Dataset

- Collected ~100 hours of videos to train and test the system
Dataset

- Collected ~100 hours of videos to train and test the system
- Hand-scored by trained students

Serre* Jhuang* Garrote Poggio Steele in prep

Thursday, July 2, 2009
Dataset

- Collected ~100 hours of videos to train and test the system
- Hand-scored by trained students
  - ~30 man-hr/hr of video
Automated analysis
Automated analysis

Circadian Time (min)

hang
rear
walk
mm
groom
eat
drink
rest

Circadian Time (hr)

6  12  18  24

rest
drink
eat
hang
rear
walk
mm
groom

Thursday, July 2, 2009
Automatic recognition of rodent behavior
Automatic recognition of rodent behavior

Performance

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>human agreement</td>
<td>72%</td>
</tr>
<tr>
<td>proposed system</td>
<td>71%</td>
</tr>
<tr>
<td>commercial system</td>
<td>56%</td>
</tr>
<tr>
<td>chance</td>
<td>12%</td>
</tr>
</tbody>
</table>
Quantitative automated phenotyping
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
- Automated quant system to help:
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
- Automated quant system to help:
  - Limit subjectivity of human intervention
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
- Automated quant system to help:
  - Limit subjectivity of human intervention
  - 24/7 home-cage analysis of behavior
Quantitative automated phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs

- Automated quant system to help:
  - Limit subjectivity of human intervention
  - 24/7 home-cage analysis of behavior
  - 24/7 monitoring of animal well-being
Behavioral comparison between 4 strains

- 24 hour monitoring of 4 different strains (n=8):
  - CAST/EiJ (wild-like strain)
  - C57Bl/6J (popular inbred mouse strains)
  - DBA/2J (popular inbred mouse strains)
  - BTBR2 (potential model of autism)
Strain can be predicted from behavior measured by system
Strain can be predicted from behavior measured by system

10 min / 50% accuracy

CAST
C57B
DBA
BTBR2

chance: 25%

Serre* Jhuang* Garrote Poggio Steele in prep

Thursday, July 2, 2009
Strain can be predicted from behavior measured by system

10 min / 50% accuracy

30 min / >80% accuracy

chance: 25%

Serre* Jhuang* Garrote Poggio Steele in prep
Limitations of present feedforward hierarchical models

- Vision is more than categorization or identification: it is image understanding/inference/parsing.
- Our visual system can “answer” almost any kind of question about an image or video (a Turing test for vision…)

Thursday, July 2, 2009
Limitations of present feedforward hierarchical models

- Vision is more than categorization or identification: it is image understanding/inference/parsing
- Our visual system can “answer” almost any kind of question about an image or video (a Turing test for vision…)

Thursday, July 2, 2009
Limitations of present feedforward hierarchical models

- Vision is more than categorization or identification: it is image understanding/inference/parsing
- Our visual system can “answer” almost any kind of question about an image or video (a Turing test for vision…)

Thursday, July 2, 2009
Limitations of present feedforward hierarchical models

- **Vision is more** than categorization or identification: it is image understanding/inference/parsing
- **Our visual system can “answer” almost any kind of question about an image or video (a Turing test for vision…)**
Limitations of present feedforward hierarchical models

- Vision is **more** than categorization or identification: it is image understanding/inference/parsing
- Our visual system can “answer” almost any kind of question about an image or video (a Turing test for vision…)

- Three options: 1) top-down (attentional) control of task-dependent routines 2) probabilistic inference in the ventral stream (~Mumford, Geman) 3) hierarchical probabilistic inference by the brain (Tenenbaum)

Thursday, July 2, 2009
Collaborators in recent work


Extension to attention: dealing with clutter

Zoccolan Kouh Poggio DiCarlo 2007
Reynolds Chelazzi & Desimone 1999
Serre Oliva Poggio 2007

see also Broadbent 1952 1954; Treisman 1960; Treisman & Gelade 1980; Duncan & Desimone 1995; Wolfe, 1997; Tsotsos and many others

Thursday, July 2, 2009
Extension to attention: dealing with clutter

Zoccolan Kouh Poggio DiCarlo 2007
Reynolds Chelazzi & Desimone 1999
Serre Oliva Poggio 2007

Parallel processing (No attention)
Serial processing (With attention)

see also Broadbent 1952 1954; Treisman 1960; Treisman & Gelade 1980; Duncan & Desimone 1995; Wolfe, 1997; Tsotsos and many others

Thursday, July 2, 2009
Attention as Bayesian inference

Chikkerur, Serre, Tan & Poggio '09, CBCL/AI Memo
Attention as Bayesian inference

Chikkerur, Serre, Tan & Poggio '09, CBCL/AI Memo

Thursday, July 2, 2009
Attention as Bayesian inference

\[ m_{O \rightarrow F_i} = P(O) \]
\[ m_{F_i \rightarrow F_i} = \sum_{O} P(F_i | O) P(O) \]
\[ m_{L \rightarrow F_i} = P(L) \]
\[ m_{I \rightarrow F_i} = P(I | F_i) \]
\[ m_{F_i \rightarrow F_i} = \sum_{L} \sum_{F_i} P(F_i | F_i, L) (m_{L \rightarrow F_i}) (m_{I \rightarrow F_i}) \]
\[ m_{F_i \rightarrow L} = \sum_{F_i} \sum_{F_i} P(F_i | F_i, L) (m_{F_i \rightarrow F_i}) (m_{I \rightarrow F_i}) \]

Chikkerur, Serre, Tan & Poggio ‘09, CBCL/AI Memo

Thursday, July 2, 2009
Attention as Bayesian inference

Feature-based attention: Where is object O?

LIP/FEF → IT → V4 → V2 → PFC

Desimone (MIT)

Chikkerur, Serre, Tan & Poggio ‘09, CBCL/AI Memo

Thursday, July 2, 2009
Attention as Bayesian inference

Spatial attention: What is at location L?

Chikkerur, Serre, Tan & Poggio ‘09, CBCL/AI Memo
Model agrees with physiological studies

McAdams and Maunsell ‘99

Model

Chikkerur, Serre, Tan & Poggio ‘09, CBCL/Al Memo
Model agrees with physiological studies

Bichot and Desimone ‘05

Model

Chikkerur, Serre, Tan & Poggio ‘09, CBCL/AI Memo
Model agrees with physiological studies

Trujillo and Treue ‘02

Mc Adams and Maunsell’99

Chikkerur, Serre, Tan & Poggio ‘09, CBCL/AI Memo
Model can predict human eye-movements

Bottom-up attention

<table>
<thead>
<tr>
<th>Method</th>
<th>ROC area (absolute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bruce and Tsotos '06</td>
<td>72.8%</td>
</tr>
<tr>
<td>Itti et al '01</td>
<td>72.7%</td>
</tr>
<tr>
<td>Proposed</td>
<td>77.9%</td>
</tr>
</tbody>
</table>
Model can predict human eye-movements

Top-down spatial and feature attention

<table>
<thead>
<tr>
<th>Method</th>
<th>ROC area (Cars)</th>
<th>ROC area (Pedestrian)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itti et al. '01</td>
<td>42.3%</td>
<td>42.3%</td>
</tr>
<tr>
<td>Torralba et al.</td>
<td>78.9%</td>
<td>77.1%</td>
</tr>
<tr>
<td>Proposed</td>
<td>80.4%</td>
<td>80.1%</td>
</tr>
<tr>
<td>Humans</td>
<td>87.8%</td>
<td>87.4%</td>
</tr>
</tbody>
</table>
Recognition performance improves with attention

Chikkerur, Serre, Tan & Poggio (in prep)
• Just one example…:

Read out data (Hung Kreiman Poggio & DiCarlo 2005)
IT Readout data

77 objects, 8 classes

Chou Hung, Gabriel Kreiman, James DiCarlo, Tomaso Poggio, Science, Nov 4, 2005
Example of one AIT cell
Decoding the neural code ...
population response (using a classifier)

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

Population activity

neuron 1
neuron 2
neuron N

\[ x \]

\[ y \in \{1,\ldots,8\} \]

Categorization
8 groups

\( \text{cat/dog} \)
\( \text{human face} \)
\( \text{toys} \)
\( \text{food} \)
\( \text{monkey face} \)
\( \text{white box contours} \)
\( \text{hand/body} \)
\( \text{vehicles} \)
From neuronal population activity... 

...a classifier can decode and guess what the monkey was seeing...

Categorization

- Toy
- Body
- Human Face
- Monkey Face
- Vehicle
- Food
- Box
- Cat/Dog

Video speed: 1 frame/sec
Actual presentation rate: 5 objects/sec

So...we can decode the brain’s code and read-out from neural activity what the monkey is seeing

We can also read-out with similar results from the model!!!
We can decode from model units as well as from IT.

![Graph showing classification performance vs. number of sites for IT and Model units.](image)
Agreement of model w/ IT Readout data
Reading out category and identity invariant to position and scale

Hung Kreiman Poggio DiCarlo 2005

Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005