i-theory:
visual cortex and deep networks

The Center for Brains, Minds and Machines

tomaso poggio,
CBMM, BCS, CSAIL, McGovern
MIT
Theoretical/conceptual framework for vision

- The first 100ms of vision: feedforward and invariant: what, who, where
- Top-down needed for verification step and more complex questions: generative models, probabilistic inference, top-down visual routines.

Following this conceptual framework we are working on:

1. *theory of invariance* in feedforward networks (visual cortex)
2. *a generative approach*, probabilistic in nature
3. *visual routines*, and of how they may be learned.
Object recognition

Figure removed due to copyright restrictions. Please see the video.

© Elsevier. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Vision: what is where

- **Human Brain**
  - $10^{10}$-$10^{11}$ neurons (~1 million flies)
  - $10^{14}$-$10^{15}$ synapses

- **Ventral stream in rhesus monkey**
  - ~$10^9$ neurons in the ventral stream
    (350 $10^6$ in each hemisphere)
  - ~$15\ 10^6$ neurons in AIT (Anterior InferoTemporal) cortex

- ~200M in V1, ~200M in V2, 50M in V4

Figure removed due to copyright restrictions. Please see the video. Source: Figure 2 from Felleman, Daniel J., and David C. Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." Cerebral cortex 1, no. 1 (1991): 1-47.

• 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

Hierarchical feedforward models of the ventral stream

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

Why do these networks including DLCNs work so well?

Models are not enough… we need a theory!
Plan

• i-theory (main results)

• equivalence to DCLNs, theory notes on DCLNs

• Some predictions + perspectives in i-theory

• Details and ML remarks
Learning of invariant & selective Representations in Sensory Cortex


© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

What i-theory can answer for you

- Why some hierarchical nets work well
- What is visual cortex computing?
- Function and circuits of simple-complex cells
- Why Gabor-like tuning in simple cells?
- Why generic, Gabor-like tuning in early areas and specific selective tuning higher up?
- What is the computational reason for the eccentricity-dependent size of RFs in V1, V2, V4?
- What are the roles of back projections?
A main computational goal of the feedforward ventral stream hierarchy — and of vision — is to compute a representation for each incoming image which is invariant to transformations previously experienced in the visual environment.
Empirical demonstration: invariant representation leads to lower sample complexity for a supervised classifier

**Theorem (translation case)**

Consider a space of images of dimensions $d \times d$ pixels which may appear in any position within a window of size $rd \times rd$ pixels. The usual image representation yields a sample complexity (of a linear classifier) of order $m = O(r^2d^2)$; the oracle representation (invariant) yields (because of much smaller covering numbers) a sample complexity of order

$$m_{\text{oracle}} = O(d^2) = \frac{m_{\text{image}}}{r^2}$$

An algorithm that learns in an unsupervised way to compute invariant representations

\[ P(\nu) \]

\[ \mu^k_n(I) = \frac{1}{|G|} \sum_{i=1}^{|G|} \sigma(I \cdot g_i t^k + n \Delta) \]
Invariant signature from a single image of a new object
We need only a finite number of projections, $K$, to distinguish among $n$ images. Similar in spirit to Johnson-Lindestrauss

$$d(I, I') \text{ distance using all templates}$$

$$\hat{d}_K(I, I') \text{ distance using } K \text{ templates}$$

Suppose we have $n$ images

$$\left\| d(I, I') - \hat{d}_K(I, I') \right\| \leq \epsilon \text{ with probability } 1 - \delta^2 \text{ if}$$

$$K \geq \frac{2}{c\varepsilon^2} \log\left(\frac{n}{\delta}\right)$$
I-Theory

So far: compact groups in $R^2$

I-theory extend proves invariance+uniqueness theorems for

- partially observable groups
- non-group transformations
- hierarchies of magic HW modules (multilayer)

Courtesy of NIPS. Used with permission.
Invariance, sparsity, wavelets

**Theorem:** Sparsity is *necessary and sufficient* condition for translation and scale invariance. Sparsity for translation (respectively scale) invariance is equivalent to the support of the template being small in space (respectively frequency).

**Theorem:** Maximum simultaneous invariance to translation and scale is achieved by Gabor templates:

\[
t(x) = e^{-\frac{x^2}{2\sigma^2}} e^{i\omega_0 x}
\]
Non-group transformations: approximate invariance in class-specific regime

$\mu_n^k(I)$ is locally invariant if:

- $I$ is sparse in the dictionary of $t^k$
- $I$ transforms in the same way (belong to the same class) as $t^k$
- the transformation is sufficiently smooth
Hierarchies of magic HW modules: key property is covariance

Courtesy of The Center for Brains, Minds and Machines, MIT.
Local and global invariance: whole-parts theorem

For any signal (image) there is a layer in the hierarchy such that the response is invariant w.r.t. the signal transformation.

biophysics: prediction on simple-complex cell
Basic machine: a HW module
(dot products and histograms/moments for image seen through RF)

- The cumulative histogram (empirical cdf) can be be computed as

\[
\mu^k_n(I) = \frac{1}{|G|} \sum_{i=1}^{|G|} \sigma(\langle I, g_i t^k \rangle + n\Delta)
\]

- This maps directly into a set of simple cells with threshold \( n\Delta \)

- \( \ldots \)and a complex cell indexed by \( n \) and \( k \) summatting the simple cells

The nonlinearity can be rather arbitrary for invariance provided it is stationary in time
Robust and bio plausible

- nonlinearity can be almost anything
- pooling is average but softmax is OK
- low bit precision
- Details and ML remarks
Dendrites of a complex cells as simple cells…

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.