Feedforward theories of visual cortex predict human performance in a rapid categorization task

Thomas Serre, Aude Oliva & Tomaso Poggio
Dept. of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA

Motivation
1. High accuracy of primates in ultra-rapid object categorization (Thorpe et al., 1996) and rapid serial visual processing (Potter, 1975) unmatched by best machine vision systems.
2. Evidences suggest feedforward processing for "immediate recognition", yet so far no biologically plausible feedforward model of visual cortex shown to perform at human level. Underlying computational mechanisms still debated.
3. We show that a specific implementation (Riesenhuber & Poggio, 1999; Serre et al., 2005) of a class of feedforward theories of object recognition can predict the level and the pattern of performance achieved by humans in a rapid animal vs. non-animal categorization task.

The model
1. To learn a new categorization task, only the task-specific circuits at the top level in the model have to be trained from a small set of labeled examples and in a task-specific manner.
2. Tuning of units learned from natural images during a developmental-like, unsupervised learning stage in which each unit in the S2, S2b and S3 layers becomes tuned to a different patch of natural image.
3. Some dictionary can support the recognition of many different object categories (does not need re-training for every new catag, to be learned).

Additional Information
web: http://www.mit.edu/~serre/ or email: serre@mit.edu

Acknowledgments
We are grateful to C. Girkin, J. Desimone, M. Giese, C. Koch, M. Riesenhuber, D. Ferre, U. Kasch, M. Koff, G. Kogman, & Thorpe, A. Giovio, R. Nayar, and S. Wolfe for comments and fruitful discussions related to this work. This research was sponsored by grants from NIH, DARPA, MIT, and the National Science Foundation. Additional support was provided by Eastman Kodak Company, Dieter van Oorschot, Chrysal, Horst Research Institute, INET Fund, Siemens Corporate Research, and the McDonnell chair (T.P.).

Animal vs. non-animal categorization task

Assessing human performance
- Stimulus (20 ms)
- ISI (30 ms)
- Mask (80 ms)
- Animal present?

Model predicts the level of performance of human observers
- Performance (hit)
- Performance (false alarm)

Assessing model performance
- Random splits procedure (n=20)
- Train
- Test

Agreement with data from V1, V4, IT, PFC
Model predicts, at the C1 and C2 levels respectively, the max-like behavior of a subclass of complex cells in V1 (Lampl et al., 2004) and V4 (Gawne & Martin, 2004).
Model agrees with other data in V4 (Reynolds et al., 1999) about the response of neurons to combinations of simple two-bar stimuli (within the receptive field of the S2 units) and some of the C2 units show a tuning for boundary conformations consistent with recordings from V4 (Pasupathy & Connor, 1999).

Robust invariant recognition on natural images

Agreement on (in-plane) rotated images

Performance of other systems on the database (d')

Mod: 22% Hum: 21%
Mod: 0% Hum: 21%
Mod: 91% Hum: 33%
Mod: 82% Hum: 63%
Mod: 100% Hum: 29%
Mod: 100% Hum: 96%
Mod: 91% Hum: 83%

For longer SOAs... back-projections active?