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

Thomas Serre, Aude Oliva & Tomaso Poggio

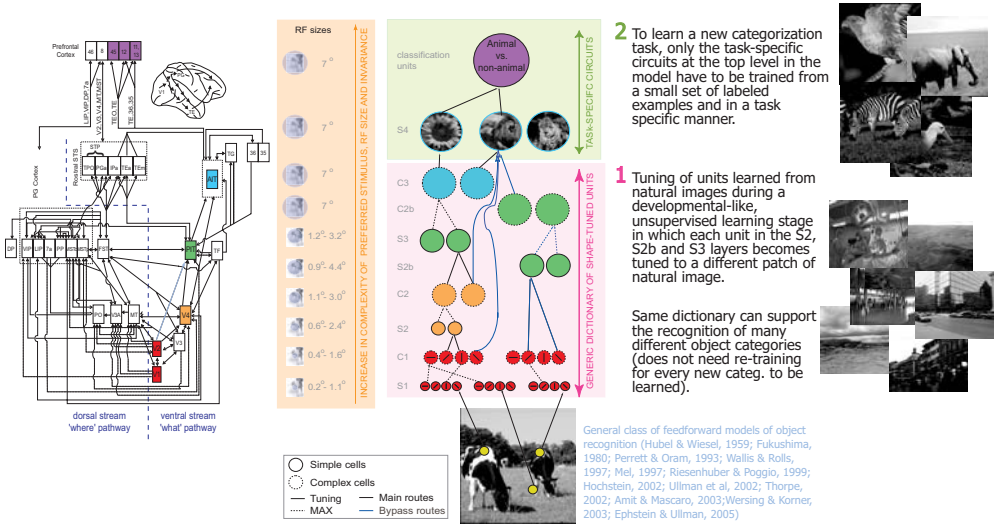
Dept. of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA

722
VSS 2006

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 on a rapid animal vs. non-animal categorization task.

The model



General class of feedforward models of object recognition (Hubel & Wiesel, 1959; Fukushima, 1980; Perrett & Oram, 1993; Wallis & Rolls, 1997; Mel, 1997; Riesenhuber & Poggio, 1999; Hochstein, 2002; Ullman et al, 2002; Thorpe, 2002; Amit & Mascaro, 2003;Wersing & Korner, 2003; Ephstein & Ullman, 2005)

1 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.

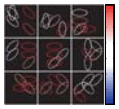
Same dictionary can support the recognition of many different object categories (does not need re-training for every new categ. to be learned).

2 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.

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).



Read-out from C2b units in the model predicts recent read-out experiments in IT (Hung et al, 2005), showing very similar selectivity and invariance for the same set of stimuli.

Robust invariant recognition on natural images

(Serre & Poggio, VSS'05)

Additional Information

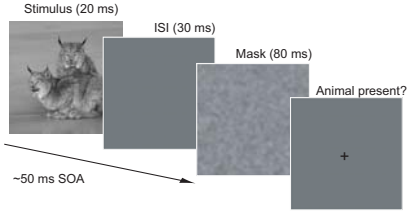
Serre (2006) Learning a dictionary of shape-components: Comparison with neurons, humans and machines, PhD thesis, CBCL Paper #260/MIT-CSAIL-TR-2006-028, MIT, 2006.
Serre et al. (2005) A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex, CBCL Paper #259/AI Memo #2005-036, MIT, 2005.
web: http://web.mit.edu/serre/www/ or **email:** serre@mit.edu

Acknowledgments

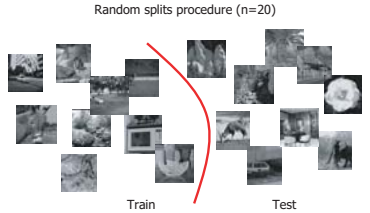
We are grateful to C. Cadieu, B. Desimone, M. Giese, C. Koch, M. Riesenhuber, D. Perrett, U. Knoblich, M. Kouh, G. Kreiman, S. Thorpe, A. Torralba, R. VanRullen and J. Wolfe, for comments and fruitful discussions related to this work. This research was sponsored by grants from NIH, DARPA, ONR and the National Science Foundation. Additional support was provided by Eastman Kodak Company, Daimler Chrysler, Honda Research Institute, NEC Fund, Siemens Corporate Research, Toyota, Sony and the McDermott chair (T.P.).

Animal vs. non-animal categorization task

Assessing human performance

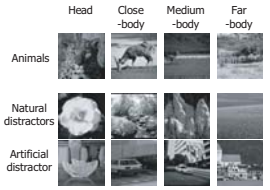


Assessing model performance



The model predicts the level of performance of human observers

The stimulus dataset



(Torralba & Oliva, 2003)

Model vs. human observers

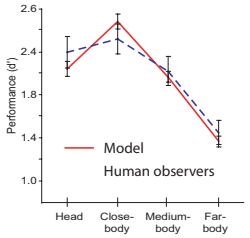
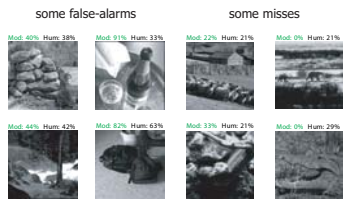


Image-by-image correlation

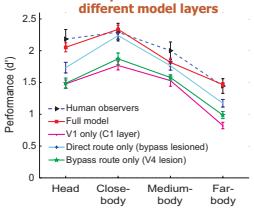


overall corr. = 0.71, 0.84, 0.71 and 0.60 for heads, close-body, medium-body and far-body respectively, $p < 0.01$

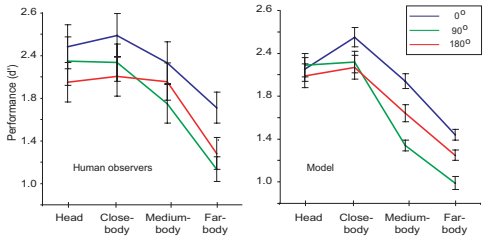
Performance of other systems on the database (d')

	H	C	M	F
Mean luminance	0.28	0.36	0.46	0.34
Gray value SVM	0.23	0.22	0.17	0.13
Textons (Renninger & Malik, 2002)	0.84	0.58	0.69	0.35
Global features (Torralba & Oliva, 2003)	1.43	1.73	1.47	0.74
Model C_1 layer	1.37	1.78	1.53	0.65

Comparison between different model layers



Agreement on (in-plane) rotated images



For longer SOAs... back-projections active?