Towards a More Complete Understanding of Object Recognition in Cortex

Maximilian Riesenhuber

Artificial Intelligence Laboratory and
The Center for Biological and Computational Learning
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Understanding how biological visual systems recognize objects is one of the ultimate goals in computational neuroscience.

Previous Work:

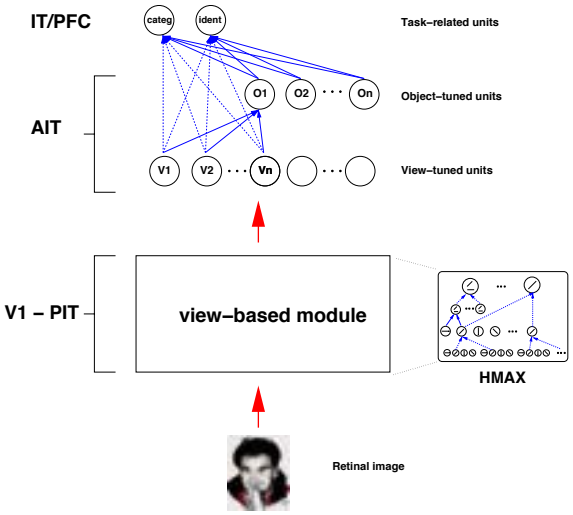

Figure 1: Sketch of our object recognition model. The model builds on our HMAX model of object recognition in cortex [3], which extends up to the layer of view-tuned units. HMAX consists of a hierarchy of layers with linear (template match units, solid lines) and non-linear operations (pooling units, performing a "MAX" operation, dashed lines). These two types of operations, respectively, provide pattern specificity and invariance (to translation, by pooling over afferents tuned to different positions, and scale (not shown), by pooling over afferents tuned to different scales). On top of the view-based module, view-tuned model units (V_n) exhibit tight tuning to rotation in depth (and illumination, and other object-dependent transformations such as facial expression *etc.*), but are tolerant to scaling and translation of their preferred object view. Invariance, for instance to rotation in depth, can then be increased by combining in a learning module several view-tuned units tuned to different views of the same object [2], creating view-invariant units (O_n). These, as well as the view-tuned units, can then serve as input to task modules performing visual tasks such as identification / discrimination or object categorization [4].

Approach: Current areas of interest include the following:

Psychophysics and simulations on the face inversion effect (with A. Folinsky). Here we evaluate how well the model can replicate the existing data on the Face Inversion Effect (a disproportionally large impairment to recognize upside-down vs. upright faces). The model not only replicates the existing data very well, but has also produced an intriguing prediction regarding the relationship of the Inversion Effect to experience with an object class, which we are currently testing in a psychophysical experiment using human subjects.

Feature learning and object detection (with T. Serre). While less crucial for the recognition of isolated objects, the choice of features in HMAX becomes more relevant as background and clutter are introduced. The project compares HMAX performance on a difficult object detection task (faces in arbitrary images) to state-of-the-art machine vision systems developed at CBCL. Proceeding from this baseline, we will investigate how more task-related feature sets can be learned in HMAX and how they affect detection performance.

Configurational vs. feature-based representation of faces (with C. Zrenner, Cambridge). The cognitive literature is rife with abstract theories regarding the representation of faces. One prominent theory posits that faces are "special" compared to other object classes, and in particular, that faces are represented by a scheme that is based on the positional configuration of facial components (eyes, mouth, nose, etc.) relative to a reference face. Using face stimuli generated with an automatic morphing system, we are exploring how well the existing data can be explained within the feature-based HMAX model.

Comparison of neuronal tuning: model and experiment (with E. Brunskill). The nonlinearity of the neuronal response of visual neurons in higher brain areas complicates the determination of a neuron's preferred feature. Exploiting our knowledge of how more complex cell responses are built from simpler ones in HMAX, we will characterize response properties of model neurons on multiple levels, and relate them to experimental data, with the goal of an improved understanding of how feature complexity increases along the visual pathway.

A combined model of object recognition and attention (with D. Walther, Caltech). The human (and macaque) visual system can be coarsely subdivided into two processing streams: the ventral stream, thought to be crucial for object recognition, and the dorsal stream, posited to be key for attentional control. HMAX is a model of the former. Itti and Koch [1] have presented a computational model of visual attention based on a "saliency map" to select regions of interest. We are exploring how to integrate the two systems, and how to include top-down task-related biases into HMAX.

Categorization in HMAX and the macaque (with D. J. Freedman and E.K. Miller). One of the key predictions of our model [4] is that different recognition tasks such as categorization and identification are based on similar neural computations. We are testing this hypothesis in a collaboration with experimental neurophysiologists performing multi-electrode recordings in primate inferotemporal and prefrontal cortices while the animal is performing object recognition tasks.

Exploring the neural basis of the MAX operation (with I. Lampl and D. Ferster, Northwestern). A key element of HMAX is its reliance of a nonlinear pooling mechanism, the MAX function, to achieve invariance and robustness to clutter. In this project we are directly testing whether complex cells actually show MAX-like pooling of stimuli, by performing (I. Lampl and D. Ferster) intracellular recordings of simple and complex cells in cat area 17.

Nice and not-so-nice object classes: implications for recognition. It has been suggested in the literature that recognition performance depends on the "niceness" of an object class [5], *i.e.*, whether objects share a common 3D structure. We are investigating this issue within HMAX, which predicts object-class dependent tuning properties of IT neurons and behavioral performance.

Research Support: Research at CBCL is sponsored by grants from: Office of Naval Research (DARPA) under contract No. N00014-00-1-0907, National Science Foundation (ITR) under contract No. IIS-0085836, National Science Foundation (KDI) under contract No. DMS-9872936, and National Science Foundation under contract No. IIS-9800032. Additional support was provided by: Central Research Institute of Electric Power Industry, Center for e-Business (MIT), Eastman Kodak Company, DaimlerChrysler AG, Compaq, Honda R&D Co., Ltd., Komatsu Ltd., Merrill-Lynch, NEC Fund, Nippon Telegraph & Telephone, Siemens Corporate Research, Inc., Toyota Motor Corporation and The Whitaker Foundation. M.R. is supported by a McDonnell-Pew Award in Cognitive Neuroscience.

References:

[1] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.*, 40:1489–1506, 2000.

[2] T. Poggio and S. Edelman. A network that learns to recognize 3D objects. *Nature*, 343:263–266, 1990.

[3] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, 2:1019–1025, 1999.

[4] M. Riesenhuber and T. Poggio. Models of object recognition. *Nat. Neurosci. Supp.*, 3:1199–1204, 2000.

[5] T. Vetter, A. Hurlbert, and T. Poggio. View-based models of 3D object recognition: invariance to imaging transformations. *Cereb. Cortex*, 3:261–269, 1995.