

Object-specific Features, Biological Vision and Real World Object Recognition

Thomas Serre, Jennifer Louie, Maximilian Riesenhuber [CBCL, with Prof. T. Poggio]

The Problem: We propose a biologically plausible model of object recognition in cortex that handles a real-world face detection task at the level of state-of-the-art machine vision systems.

Motivation: Models of object recognition in cortex have been mostly applied to tasks involving the recognition of isolated objects presented on blank backgrounds. Ultimately models of the visual system have to prove themselves in real world object recognition tasks, such as face detection in cluttered scenes, a standard computer vision benchmark task.

For such tasks, recent advances in machine vision have shown the benefit of image representations based on target object class-specific features [1]. We here wish to explore the learning of object class-specific features in intermediate stages of a model of object recognition in cortex recently presented [3], and to test its performance on a face detection task.

Previous Work: We propose an extension of the HMAX model of object recognition in cortex [3] that characterizes the ventral visual pathway in cortex, extending from primary visual cortex, V1, to inferotemporal cortex, IT, a brain area thought to be crucial for object recognition. The model consists of a hierarchy of layers with two different types of pooling mechanisms (linear operations “S”, to build more complex features from simple ones and nonlinear MAX pooling operations, “C”, to increase the invariance of units to stimulus scaling and translation, see figure). The model explains how so called view-tuned neurons in IT can exhibit highly specific tuning to views of complex objects while showing invariance to changes in stimulus position and scale.

In the model, object-specific learning so far only occurs in the higher levels. We found that the model performed rather poorly on a face detection task, due to the low specificity of the hardwired feature set of C2 units in the model (corresponding to neurons in intermediate visual area V4) that do not show any particular tuning for faces vs. background. We extended the previous model and showed how visual features of intermediate complexity can be learned in HMAX using a simple learning rule [4].

Approach: Input images are first filtered through a continuous layer S1 of overlapping simple cell-like receptive fields (first derivative of gaussians) at different scales and orientations. Neighboring S1 cells in turn are pooled by C1 cells through a MAX operation. The difference to standard HMAX lies in the C1→S2 connectivity: While in standard HMAX these connections are hardwired to produce $256 \times 2 \times 2$ combinations of C1 input, they are now learned from the data. S2 units are RBF-like units centered on features \mathbf{u} that were obtained by performing vector quantization (VQ, using the k-means algorithm) over randomly chosen pattern of C1 activation \mathbf{w} extracted at random position over face images. Given a certain *patch size* p , a feature corresponds to a $p \times p \times 4$ pattern of C1 activation \mathbf{w} , where the last 4 comes from the four different preferred orientations of C1 units. On top of the system, C2 cells perform a MAX operation over the whole visual field and provide the final encoding of the stimulus, constituting the input to an SVM classifier.

Extensive comparisons between computer vision systems [1, 5] have shown that HMAX with feature learning handles a face detection task at the level of state of the art classifier [4]. Also, we showed that a simple feature selection technique that is biologically plausible (maximally activated S2 units) would allow unsupervised feature learning from both faces and non-faces parts while maintaining high level performances [2].

Impact: Feature learning in a hierarchy is a difficult computational problem, and so is face detection in natural images. Using a simple rule to learn object-specific features, HMAX performs at the level of classical machine vision face detection systems presented in the literature. This suggests an important role for the set of features in intermediate visual areas in object recognition. Moreover features are not chosen according to their discriminative power for any classification task (between-class discrimination) but rather for their within-class representativeness. We expect the same features to be used for other recognition tasks, however, their weight in the decision task might vary from one to another.

Future Work: We plan to look at model performances with respect to non-affine transformations such as rotation in depth and illumination changes. Future work will also include a comparison with humans on a face detection task and extension to other object classes.

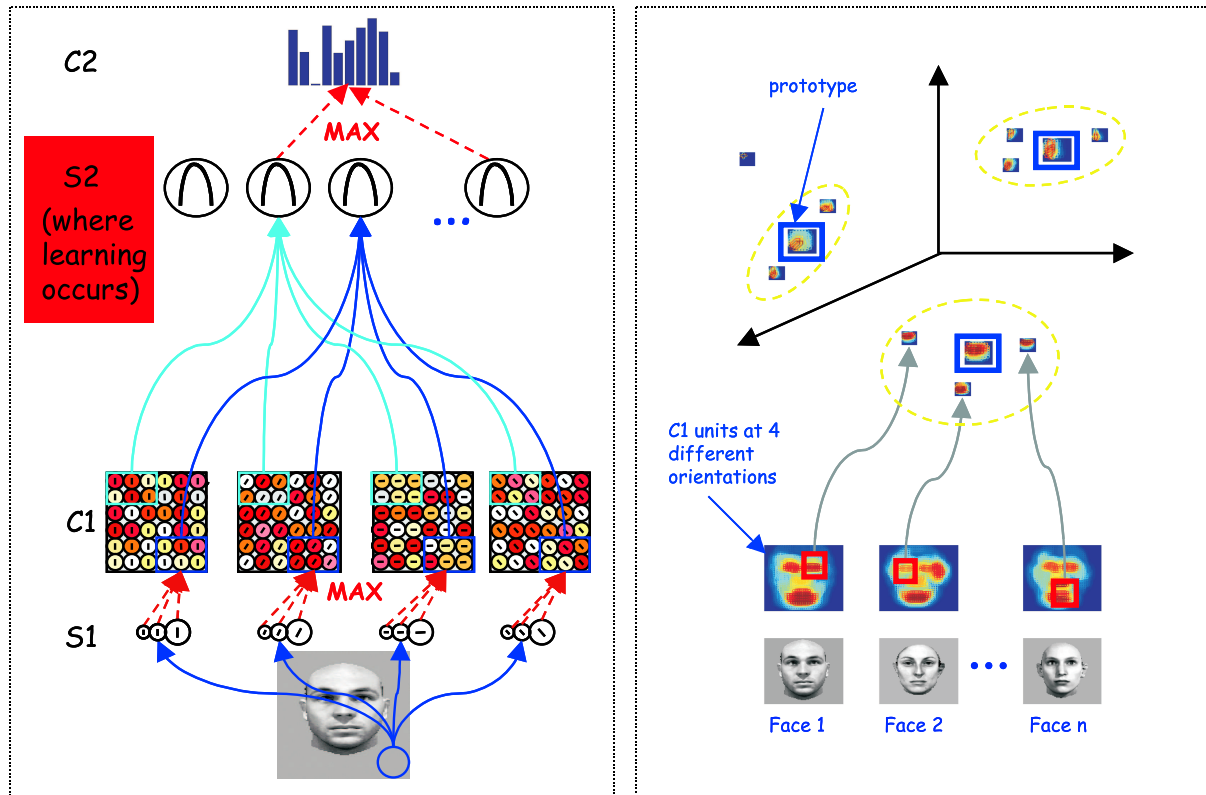


Figure 1: Left: Sketch of the model. Right: Learning object specific features at the S2 level.

Research Support: Research at CBCL is sponsored by grants from: Office of Naval Research (DARPA) Contract No. N00014-00-1-0907, National Science Foundation (ITR/IM) Contract No. IIS-0085836, National Science Foundation (ITR) Contract No. IIS-0112991, National Science Foundation (KDI) Contract No. DMS-9872936, and National Science Foundation Contract No. IIS-9800032.

Additional support was provided by: AT&T, Central Research Institute of Electric Power Industry, Center for e-Business (MIT), DaimlerChrysler AG, Compaq/Digital Equipment Corporation, Eastman Kodak Company, Honda R&D Co., Ltd., ITRI, Komatsu Ltd., Merrill-Lynch, Mitsubishi Corporation, NEC Fund, Nippon Telegraph & Telephone, Oxygen, Siemens Corporate Research, Inc., Sumitomo Metal Industries, Toyota Motor Corporation, WatchVision Co., Ltd., and The Whitaker Foundation.

References:

- [1] B. Heisele, T. Serre, M. Pontil, T. Vetter, and T. Poggio. Categorization by learning and combining object parts. In *NIPS 14*, 2002.
- [2] J. Louie. A biological model of object recognition with feature learning. Master's thesis, MIT, 2003.
- [3] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, 2(11):1019–25, 1999.
- [4] T. Serre, J. Louie, M. Riesenhuber, and T. Poggio. On the role of object-specific features for real world recognition in biological vision. In *BMCV*, 2002.
- [5] K.-K. Sung. *Learning and Example Selection for Object and Pattern Recognition*. PhD thesis, MIT, Cambridge, MA, 1996.