

Explicit Object Representation by Sparse Neural Codes

Stephen Waydo¹ and Christof Koch²

¹Control & Dynamical Systems, ²Computation & Neural Systems
California Institute of Technology

Highly sparse representations of objects in the visual environment in which individual neurons display a strong selectivity for only one or a few stimuli (such as familiar individuals or landmark buildings) out of on the order of 100 presented to a test subject have been observed in the human medial temporal lobe (MTL), a brain area believed to be crucial to the formation of new semantic memories [1, 2]. The process by which more distributed representations earlier in the visual pathway are transformed to produce such highly selective and invariant units results in information represented only implicitly by the pattern of light impinging on the retina and in the firing of neurons in early visual areas being made explicit at the level of MTL. This “sparsification” may be an important design principle underlying the structure of this brain region.

We apply a modified version of the model of Olshausen and Field [3], in which a network of nonlinear neurons generates a sparse representation of its inputs through an unsupervised learning process, to the outputs of a biologically plausible model of the human ventral visual pathway [4]. We train this system on real-world images from multiple categories taken from the Caltech-256 dataset. This training is carried out in an entirely unsupervised manner, without specifying image categories or even the number of categories present. Although the underlying constraint in the model is merely to produce a sparse representation of its input set, units emerge that respond selectively to specific image categories such as faces and airplanes. The sparseness constraint thus facilitates the formation of explicit representations of image categories, despite the category information being represented only implicitly in the input images.

Acknowledgments

We thank Thomas Serre and Minjoon Kouh of MIT for providing the visual system model used here as well as assistance with its operation, and Richard Murray, Jerry Marsden, and Pietro Perona at Caltech for valuable feedback. This work was funded by a Fannie and John Hertz Foundation Fellowship (to S.W.), as well as by grants from the ONR, NIMH, and NGA.

References

- [1] Invariant Visual Representation by Single Neurons in the Human Brain. R. Quian Quiroga et al, *Nature* 435:1102-1107, 2005.
- [2] Sparse Representation in the Human Medial Temporal Lobe. S. Waydo et al, *Journal of Neuroscience* 26(40):10232-10234, 2006.
- [3] Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? B.A. Olshausen and D.J. Field, *Vision Research* 37(23):3311-3325, 1997.
- [4] Object Recognition with Features Inspired by Visual Cortex. T. Serre, L. Wolf, and T. Poggio, *Proc. CVPR*, 2005.