Learning Transformational Invariants from Time-Varying Natural Images

Charles F. Cadieu and Bruno A. Olshausen

Redwood Center for Theoretical Neuroscience, University of California, Berkeley

How does the brain represent and learn the structure contained in dynamic visual scenes? Previous work using unsupervised learning in the time domain has shown that sparse coding can uncover direction-selective components that are tuned to specific spatial and temporal frequency bands. However, these models do not capture more complex motion selectivity such as speed tuning or pattern motion, which are response characteristics found in higher visual area MT.

We have developed a hierarchical, probabilistic generative model that learns the higher-order temporal structure in natural movies and produces representations that are similar to the known properties of visual area MT. The model consists of three layers: the first layer represents the pixel values of dynamic visual stimuli (natural movies); the second layer decomposes visual information into two sets of variables, one that is sparse and temporally stable and another that is quickly changing; the top layer learns the sparse, higher-order structure of the quickly changing variables. After learning on time-varying natural images, the second layer represents the structural visual content (local orientation) separately from the dynamical visual content (motion). This allows the top layer to learn the sparse causes of the dynamical visual content. After learning, the top layer units code transformational invariants: they are selective for the speed and direction of a moving pattern, but are invariant to the appearance and spatial frequency content in the moving pattern.

The diversity of units in the middle and top layer provides a set of testable predictions for representations that might be found in V1 and MT. The top layer represents a variety of transformational invariants and contains a population that is selective for pattern motion, a characteristic of visual area MT. Interestingly, because the model is selective to complex motions beyond translation, it predicts a set of MT responses that extends beyond the component-pattern classification. These results suggest that the response properties of neurons in both primary and extra-striate cortex may be accounted for in terms of an efficient coding strategy adapted to time-varying natural images.

Acknowledgments

This work was supported by NGA grant MCA 015894-UCB and NSF grant IIS-06-25223.