Learning Complex Cell Receptive Field Properties using a Generative Model

Jayanta K. Dutta, Bonny Banerjee

Institute for Intelligent Systems, and Dept. of Electrical & Computer Engineering, University of Memphis
206 Engineering Science Bldg., 3815 Central Ave, Memphis, TN 38152, USA
jkdutta@memphis.edu, BonnyBanerjee@yahoo.com

Learning invariances to transformations within a complex neuron's receptive field plays a crucial role in learning feature hierarchies by facilitating abstraction of more stable spatiotemporal features as we ascend up the hierarchy. Real-world transformations can be arbitrarily complex. For example, the transformation of a balloon from deflated to inflated is much more complex than mere scaling or translation. In hierarchical neural models of vision, such as HMAX (Serre et al., 2007), transformation invariance is typically achieved in the complex layer by pooling from a subset of lower layer simple neurons. Deciding the set to pool from is the crux of the invariance learning problem.

We hypothesize that, any arbitrary transformation may be learned from time-varying data by pooling from the subset of simple neurons that are repeatedly used in close temporal proximity for explaining (or reconstructing) the data. To verify this hypothesis we propose a generative neural model consisting of a simple and complex layer with exhaustive bottom-up (simple to complex), top-down (complex to simple) and lateral (in each layer) connections. The complex neurons have a temporal receptive field, and reconstruct the activations of the simple neurons occurring within that receptive field. The bottom-up connections to a complex neuron encode the subset of simple neurons to pool from. The top-down connections encode correlations between the activations of simple and complex neurons while lateral connections in the complex layer encode correlations between the activations of different complex neurons. The top-down and lateral connections may be conceived as short-term memory while the bottom-up is longer-term.

Our model was learned from natural videos recorded with a camera mounted on a cat's head (Betsch et al., 2004). These videos provide a continuous stream of stimuli similar to that received naturally by the cat's visual system, preserving its temporal structure, and has been used for evaluating models on learning complex cell receptive field properties (e.g., Einhauser et al., 2002; Masquelier et al., 2007). Upon learning, each complex neuron in our model becomes invariant to bars or edges moving in different directions. The lateral connections in the simple layer were learned to encode transition probabilities. Using these, the complex neurons learn sequences of varying lengths (bounded by their temporal receptive field size) and can predict the stimuli for the remaining part of the sequence.

References:

B. Y. Betsch, W. Einhauser, K. P. Kording, and P. Konig. The world from a cat's perspective – statistics of natural videos. Biological Cybernetics, 90(1):41–50, 2004.

W. Einhauser, C. Kayser, P. Konig, and K. P. Kording. Learning the invariance properties of complex cells from their responses to natural stimuli. European Journal of Neuroscience, 15(3):475–486, 2002.

T. Masquelier, T. Serre, S. Thorpe, and T. Poggio. Learning complex cell invariance from natural videos: A plausibility proof. Technical Report 60, MIT, Cambridge, MA, December 2007.

T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci., 104(15):6424–6429, 2007.