V1 Maps from Sparse Neurons with Intrinsic Plasticity

Nicholas Butko (nbutko@cogsci.ucsd.edu) ; Jochen Triesch (triesch@cogsci.ucsd.edu)
UC San Diego Dept. of Cognitive Science

We present a model for learning V1-like maps composed of neurons with Gabor-like (oriented, bandpass, and localized) receptive fields. We only use the neural principles of Hebbian learning, a form of homeostasis called intrinsic plasticity, and Kohonen-style population organization to achieve this result. Each model neuron exhibits an exponential firing distribution in response to natural image patches. An example of a trained population of our model neurons is shown in Figure 1.

Simoncelli and Olshausen reported that image responses to Gabor filters exhibit a sparse, near exponential distribution [1]. Previously, Olshausen and Field had shown that Gabor-like filters are learned as a generative model for natural image patches when a sparse prior is imposed over the coefficients [2], while Bell and Sejnowski showed that Gabor-like filters form the independent components of natural scenes [3]. These results are meant to be complementary to findings such as those by, for example, Baddeley *et al.*, who observed that individual neurons in V1 exhibit exponential firing rate distributions [4].

While [1-3] provide a computational level description of why sparse firing rates are desirable, they do not hypothesize a biologically achievable algorithm for neurons adapting themselves to fire in such regimes. Moreover, they provide no account for the map-like structure in cortex. In our current work, we attempt to provide this intermediate algorithmic layer between the computational description in [1-3], and the physical properties of cortical systems.

In [5], Triesch developed model neurons that have nice computational properties. They receive an internal activation as the inner product of an input vector with a weight vector. They then pass this signal through a non-linearity as output. The non-linearity is a logistic sigmoid with two parameters. The cells adjust these intrinsic parameters based only on instantaneous input and output activations. Meanwhile the cell adapts its synaptic weights using a simple Hebbian learning rule. Triesch showed that, under certain assumptions about the input space, these neurons adapt their weights to directions to which the input projects in an exponentially distributed fashion. One set of weight vectors that has this property for natural images, according to Simoncelli and Olshausen, is Gabor filters. Triesch also showed that in practice these neurons compute a form of non-linear ICA, and we know from Bell and Sejnowski's work that ICA also finds Gabor-like weight vectors. So we can imagine that, using natural image patches as input, these model neurons might learn Gabor filter weight vectors.

The current work extends our previous work by considering *populations* of these neurons. We arrange the neurons in a two dimensional rectangular grid, and then learn using a modified Kohonen learning algorithm. In the Kohonen algorithm, the weight vectors are prototypes. For each input, the closest prototype moves closer to the input and pulls its neighbors along. In our algorithm, the weight vectors are directions, and the most active unit's vector is rotated (along with its neighbors) toward the input by Hebbian correlation-based learning. Our algorithm is further different in that, while Kohonen nets have only a center-pull, we also have a surroundpush. This ensures that neighboring regions of the map cover similar directions, while farther regions are forced to cover different directions. Algorithmically, this is implemented through modulating Hebbian learning with a difference of Gaussians envelope across the map, centered at the maximally activated unit.

Figure 1 shows the result of learning in this framework. We found that a wide variety of oriented, bandpass, localized receptive fields develop. Moreover, we found the learning was extremely robust to changing parameters.

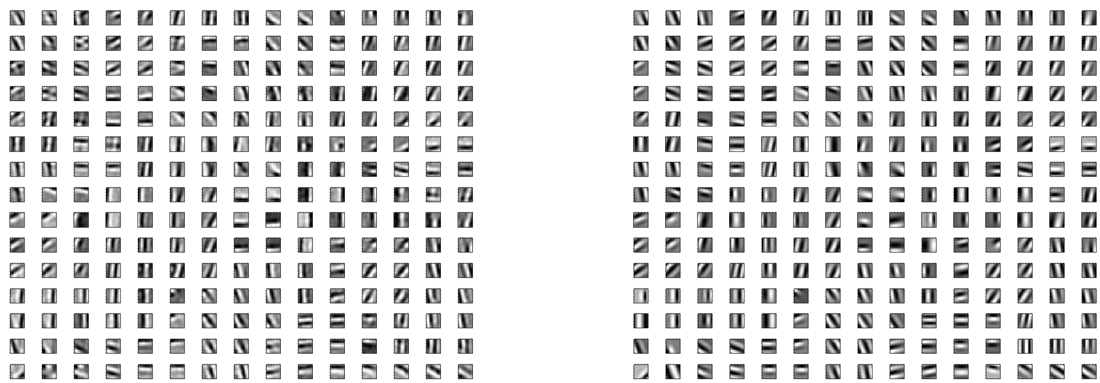

Figure 1: Left: Receptive field map learned by our model when trained on natural images. Right: The best-fit Gabor functions to the receptive fields on the left. Our model learns filters of many different orientations, frequencies, and localizations. Regions of the map exhibit smooth interpolation between these features.

In this model, we have invoked three simple neural principles: correlation-based-learning, homeostasis, and Kohonen-style map formation with surround inhibition. Out of the interactions among these three principles, populations of neurons that exhibit V1-style map organization emerge. The resulting receptive fields are overcomplete, and each neuron exhibits a sparse, exponentially distributed firing rate.

Bibliography:

[1] Simoncelli, E.P., & Olshausen, B.A. (2001). Natural Image Statistics And Neural Representation. *Annual Review of Neuroscience* 24: 1193-1216.

[2] Olshausen B.A., & Field, D.J. (1997). Sparse Coding with an Overcomplete Basis Set: A Strategy employed by V1? *Vision Research 37* 23:3311-3325.

[3] Bell, A.J., & Sejnowski, T.J. (1997). The "Independent Components" of Natural Scenes are Edge Filters. *Vision Research 37* 23:3327-3338

[4] Baddeley, R., Abbott, L. F., Booth, M.C., Sengpiel, F., & Freeman, T. (1998). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. London, Ser. B*, 264:1775.1783.

[5] Triesch, J. (2005). Synergies between Intrinsic and Synaptic Plasticity in Individual Model Neurons. *Advances in Neural Information Processing Systems* 17:1417-1424.