Sampling l_0 Sparse Codes Leades to Improved Classification Performance

William B. Shainin · Dylan M. Paiton · Garrett T. Kenyon · Bruno Olshausen

Abstract Weights learned from natural scenes via the l_0 sparse coding energy function have been shown to match well to classical and non-classical biological receptive field properties of pyramidal cells in primary visual cortex. However, the l_0 energy function is non-convex and thus without a known global minima. Therefore, a network that minimizes an l_0 cost function could fall into any number of minima that lie along a manifold representing the space of possible representations of the input data. It is not clear how the amount of information differs between these minima, or how the energy level varies between minima. We propose the hypothesis that these local minima have varying support for whole-scene object classification. We test the hypothesis by stochastically sampling multiple fixed points of the Hopfield-like LCA sparse solver network for a basis set learned with an l_0 sparse coding energy function. We show that sampling multiple local minima in the energy landscape enables improved classification performance.

$$\underset{a}{\operatorname{argmin}}\left(E = \overbrace{\tfrac{1}{2}\|S-R\|_2^2}^{\text{Preserve Information}} + \overbrace{\lambda\sum_m C(a_m)}^{\text{Limit Activations}}\right) \tag{1}$$

$$R = a\phi = \sum_m^M a_m\phi_m \tag{2}$$

$$E(t) = \tfrac{1}{2}\sum_p^P\left[S_p - \sum_m^M a_m(t)\phi_{m,p}\right]^2 + \lambda\sum_m^M C(a_m(t)) \tag{3}$$

$$\dot{u_i}(t) \propto -\frac{1}{\tau}\frac{\partial E}{\partial a_i(t)} \tag{4}$$

WB Shainin
The New Mexico Consortium
Los Alamos, New Mexico
E-mail: wshainin@gmail.com

$$\begin{aligned} \dot{u_i}(t) &\propto -\frac{\partial E}{\partial a_i(t)} \\ &= \frac{1}{\tau}\left[b_i - u_i(t) - \sum_{m \neq i}^{M} G_{i,m} a_m(t)\right] \end{aligned} \tag{5}$$

$$\lambda \frac{\partial C(a_i(t))}{\partial a_i(t)} = u_i(t) - a_i(t) = u_i(t) - T_\lambda(u_i(t)) \tag{6}$$

$$T_\lambda(u_m(t)) = \begin{cases} 0, & u_m(t) \ \leq \ \lambda \\ u_m(t), & u_m(t) \ > \ \lambda \end{cases} \tag{7}$$