A cortical theory of super-efficient probabilistic inference based on sparse distributed representations

Rod Rinkus, Neurithmic Systems, 275 Grove St., Suite 2-400, Newton, Mass, 02466
E-mail: rod@neurithmicsystems.com

The remarkable structural homogeneity of isocortex strongly suggests a canonical cortical algorithm that performs the same essential function in all regions [1]. That function is widely construed/modeled as probabilistic inference, i.e., the ability, given an input, to retrieve the best-matching memory, or equivalently, the most likely hypothesis, stored in memory (see [2] for review). In [3], I described a cortical model for which both storage (learning) of new items into memory and probabilistic inference are constant time operations, which is a level of performance not present in any other published information processing system. This efficiency depends critically on: a) representing inputs with *sparse distributed representations* (SDRs), i.e., relatively small sets of binary units chosen from a large pool; and on b) choosing (learning) new SDRs so that more similar inputs are mapped to more highly intersecting SDRs. The cortical macrocolumn ("mac") , specifically, its pool of L2/3 pyramidals, was proposed as the large pool, with its minicolumns ("mincs") acting in winner-take-all (WTA) fashion, ensuring that macrocolumnar codes consist of one winner per minicolumn.

In this work, I will present results of large hierarchical model instances, having many levels and hundreds of macs, performing: a) single-trial learning of sets of sequences derived from natural video; and b) immediate (i.e., no search) retrieval of best-matching stored sequences. Figure 1 shows the major shift in going from the localist coding scheme present in most hierarchical cortical models, e.g., [3], to SDR coding. In the localist hierarchy of Figure 1a (adapted from [4]), the red rounded red rectangles at S1 contain very simple feature sets (lexicons, bases). Each one "sees" a different small region [receptive field (RF), "aperture"] of the visual field (gray hexagons) and corresponds to a mac (i.e., V1 hypercolumn). An edge-filtered image of a dog's head is shown on the input surface. Pink shading emphasizes the three RFs for which we show active macs at S1. The two pink RFs over the dog's nose currently have a vertical and a horizontal edge, respectively. The coactivity of these two features' S1 units activates the "corner" unit in the overlying C1 mac whose RF (dashed circle over the nose) includes these S1 macs. C1 macs contain units representing many possible spatial and spatiotemporal patterns (an overcomplete basis) that might occur in a C1-sized aperture.

In the SDR-based version of the hierarchy of Figure 1b, each 4-unit localist S1 mac in Figure 1a is replaced by a mac (hexagon) consisting of Q=7 winner-take-all (WTA) mincs, each having K=7 binary *representational units* (RUs). Codes for the vertical and horizontal edges forming part of the dog's nose are shown active (red) in two neighboring S1 macs. Round symbols partially overlapping macs indicate the features represented by the depicted codes. As in Figure 1a, when codes are coactive at S1, the resulting bottom-up (U) signals arriving at the C1 mac will cause the code for the "corner" feature to become active. Black lines between RUs at different levels show (a tiny sample of) the synapses (weights) that would be increased in forming the inter-level associations.

This change of representation has a potentially large impact on explaining the storage capacity of cortex, but more importantly on explaining the speed and other characteristics of probabilistic/approximate reasoning possessed by biological brains. The SDR-based model's constant-time inference property means for example, that when the vertical and horizontal edge features for the dog's nose are present in the adjacent apertures, the combined bottom-up signals to the overlying macrocolumn will immediately, i.e., without any search, activate the corner feature, regardless of how many features have been stored in that macrocolumn.

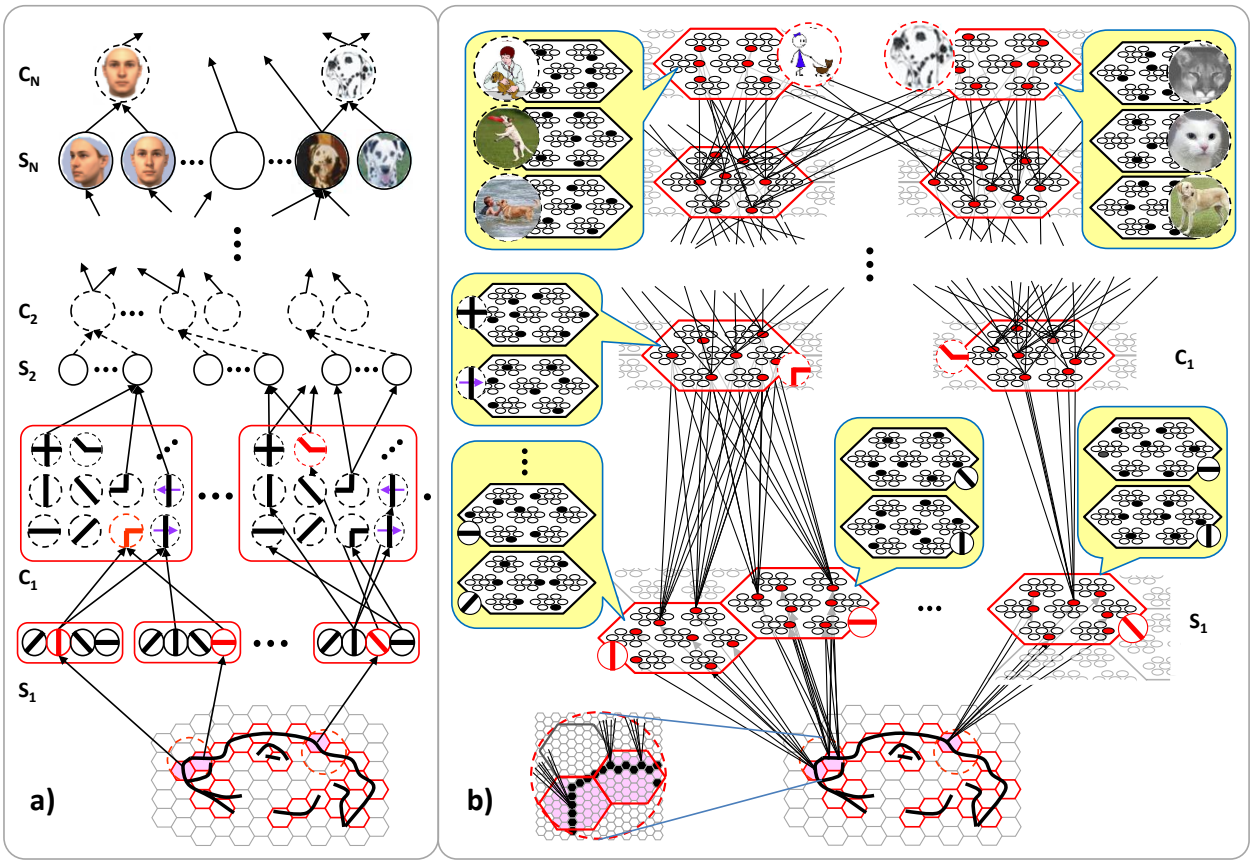

Figure 1: Comparison of localist (a) and SDR-based (b) versions of visual feature hierarchies.

At the left of Figure 2, we show a 40-frame sequence of input (indexed in row-major order) for a 12x12-pixel region (aperture) taken from an edge-filtered, binarized natural video (from the Hollywood 2 database). In the middle of Figure 2, we show the hierarchical model instance to which we presented this sequence. There was only one learning trial. The model has three internal levels, each having one macrocolumn. The L1 macrocolumn had 16 mincs, each with 16 units; the L2 and L3 macs each had 9 mincs, each with 9 units. There is complete bottom-up, top-down, and horizontal connectivity between and within levels, for a total of ~207,000 binary weights (not shown). Another key model property is that codes at higher levels have longer persistence, in this case, L1 codes persist for 1 frame, L2 codes persist for 2 frames, and L3 codes for 4 frames. In view of the model's essentially Hebbian learning rule, this causes higher-level codes to associate with sequences of subjacent codes. This shows the actual codes that were assigned at the three internal levels on the first frame (T=0) of the learning trial. At right, we show the codes that were activated at all three levels at T=0 of the recognition test trial. The two traces are identical except for one error (the red unit in the second L1 minc). Across all 40 frames of the recognition trace, only three errors were made, meaning that with one learning trial, recognition of this sequence was virtually perfect.

Figure 2 provides just a flavor of the kind of results I will describe. Data will be given demonstrating the model's ability to learn, with single trials, and subsequently recognize large sets of sequences where the sequence frames are much larger, e.g., 300x240.

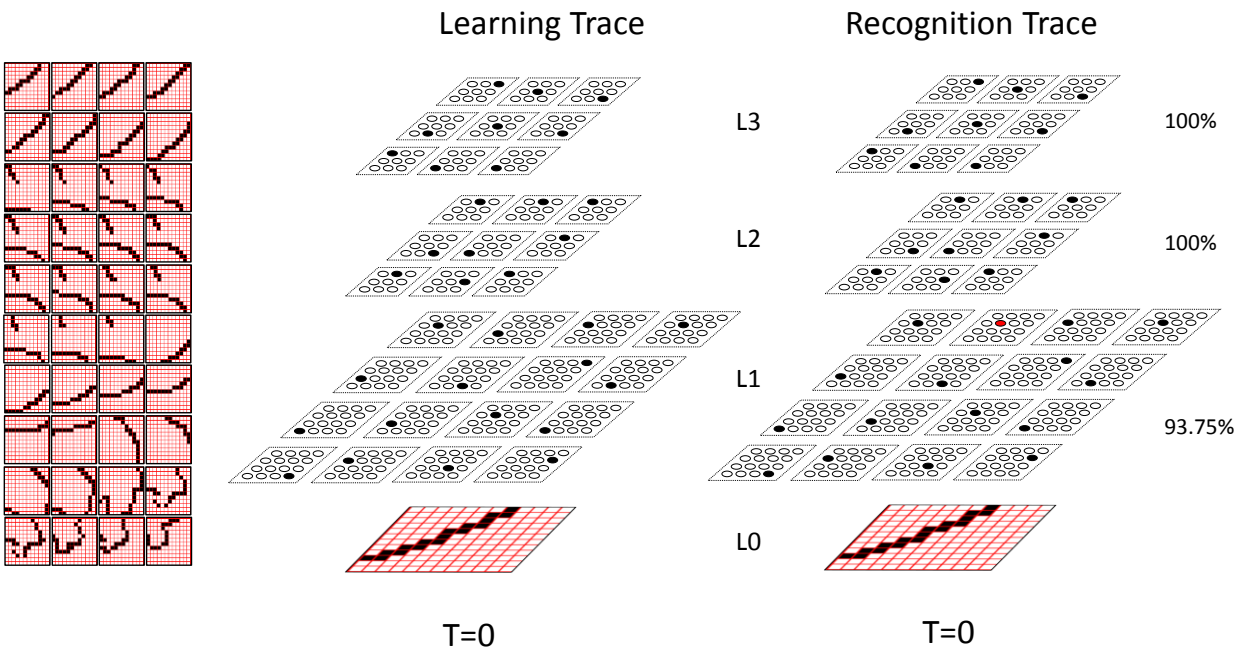

Figure 2: (left) 40-frame sequence of 12x12-pixel video snippet taken from Hollywood 2 video. (middle, right) Side-by-side comparison of the full hierarchical traces of the first (of 40) input frames during the learning presentation and the recognition test presentation. At right are the layer-wise recognition accuracies, i.e., of the macrocolumnar codes. The red unit in the L1 mac during recognition is an error: comparing to the same minicolumn in the learning trial, you can see that a different unit won. This accounts for the 93.75% accuracy at L1, i.e., 15 out of 16 of the code's units were correct. The model achieved virtually perfect accuracy on all 40 frames.

References

1. **Douglas RJ, Martin KA, Witteridge D: A canonical microcircuit for neocortex. *Neural Computation* 1989, 1(4):480-488.**
2. **Bengio Y, Courville A, Vincent P: Representation Learning: A Review and New Perspectives. In.: U. Montreal; 2012.**
3. **Rinkus GJ: A cortical sparse distributed coding model linking mini- and macrocolumn-scale functionality. *Frontiers in Neuroanatomy* 2010, 4.**
4. **Serre T, kouh M, Cadieu C, Knoblich U, Kreiman G, Poggio T: A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex. In: *AI Memo 2005-036.* MIT; 2005.**