

Chapter 1

Neural model of spatiotemporal prediction for object recognition

1.1 Introduction

TODO

1.2 Methods

1.2.1 Model architecture

The model consisted of four layers (Figure 1.1) whose parameters are described in detail here. Two of the layers contained columnar substructure which was necessary for implementing learning using the LeabraTI framework (Chapter ??):

Figure 1.1:

Retina and V1 preprocessing: Input was provided to the model via a 24x24 topographic filter bank that preprocessed images offline from the model proper. This preprocessing step is consistent with a large class of biological models describing object recognition in cortex (e.g.,

Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; O'Reilly, Wyatte, Herd, Mingus, & Jilk, 2013) and in the case of the present model, represents visual processing from the level of the retina through V1 simple cells (Hubel & Wiesel, 1962) Grayscale bitmap images were scaled to 24x24 pixels and convolved with Gabor filters at four orientations (0° , 45° , 90° , and 135°) at two polarities (off-on and on-off) producing a 24x24x4x2 set of inputs. Each Gabor filter was implemented as 6x6 pixel kernel, with a wavelength $\lambda = 6$ and Gaussian width terms of $\sigma_x = 1.8$ and $\sigma_y = 1.2$. A static nonlinearity was applied to the output of the filtering step in the form of a simplified k -Winners-Take-All (k WTA) inhibitory competition function that reduced activation across the 4x2 filter bank to the equivalent of $k = 1$ fully active units (see O'Reilly et al., 2013, Supporting Information).

Early visual columns: 24x24 topographic layer arranged into groups of 4x2 units (4608 total units), decomposed into superficial (Layers 2 and 3) and deep (Layer 6) neuron subtypes. Each superficial unit receives the output of the retina/V1 preprocessing step. k WTA inhibition for superficial units was set to 60% of the average of the top k active units compared to the average of all other superficial units with each 4x2 unit group using a value of $k = 2$. Deep units received from 5x5 columns of superficial units (200 inputs per deep unit) integrated into a single value that was used as the additional context input channel for each superficial unit.

Secondary visual columns: 6x6 topographic layer arranged into groups of 7x7 units (1764 total units), also decomposed into superficial and deep neuron subtypes. Each superficial unit receives from 8x8 topographical neighborhoods of early visual columns (512 afferents per superficial unit) and sends back reciprocal connections with the same topography. k WTA for superficial units was set to 60% of the average of the top k active units compared to the average of all other superficial units with each and 15% activity within each unit group. Deep units received from 3x3 columns of superficial units (441 inputs per deep unit) integrated into a single value that was used as the additional context input channel for each superficial unit.

Output units: 10 x 10 layer without unit group or columnar substructure (100 total units). Each unit receives a full projection from secondary visual columns (1764 afferents per unit) and

fully projects back to all columns. A k WTA value of $k = 1$ was used to enforce a localist representation. The localist representation is a computational simplification that allowed an identity readout of lower-level features similar to that provided by inferior temporal (IT) neurons in cortex albeit using a population code (Hung, Kreiman, Poggio, & DiCarlo, 2005; Li, Cox, Zoccolan, & DiCarlo, 2009).

1.2.2 LeabraTI learning algorithm

LeabraTI was implemented as an extension of the standard Leabra algorithm described in detail in O'Reilly and Munakata (2000) and O'Reilly, Munakata, Frank, Hazy, and Contributors (2012). Standard Leabra learning operates across two phases: a *minus* phase that represents the system's expectation for a given input and a *plus* phase, representing observation of the outcome. The difference between the minus and plus phases, along with various self-organizing mechanisms, is used in computing the synaptic weight update function at the end of each plus phase.

LeabraTI extends

1.2.3 Training and testing environment

1.3 Results

1.4 Discussion

References

- Hubel, D., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*(1), 106–154.
- Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*(5749), 863–866.
- Li, N., Cox, D., Zoccolan, D., & DiCarlo, J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? *Journal of Neurophysiology*, *102*(1), 360–376.
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: The MIT Press.

- O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors (2012). Computational Cognitive Neuroscience. Wiki Book, 1st Edition, URL: <http://ccnbook.colorado.edu>.
- O'Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent processing during object recognition. Frontiers in Psychology, 4, 124.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11), 1019–1025.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Sciences of the United States of America, 104(15), 6424–6429.