

A High-Throughput Screening Approach to Discovering Good Forms of Visual Representation

**David D. Cox^{1,2}, Nicolas Pinto¹, David Doukhan¹, Benoit Corda¹
and James J. DiCarlo¹**

¹. The McGovern Institute for Brain Research, MIT

². The Rowland Institute at Harvard

The experimental study of biological vision and the creation of artificial vision systems are naturally intertwined – while exploration of the neuronal substrates of visual processing provides clues and inspiration for artificial systems, artificial systems can in turn serve as important generators of new ideas and working hypotheses. However, while systems neuroscience has so far provided inspiration and constraints for some of the "broad-stroke" properties of the visual system (e.g. hierarchical organization, synaptic integration of inputs and threshold, normalization, plasticity, etc.), much is still unknown. Even for those qualitative properties that most biological-inspired models hold in common, experimental data currently provide little constraint on their key parameters. Consequently, it is difficult to truly evaluate a set of computational ideas, since the performance of any one model depends strongly on its particular instantiation – e.g. the size of the pooling kernels, the number of units per layer, exponents in normalization operations, etc. Since the number of such parameters (explicit or implicit) is large, and the typical computational cost of evaluating one particular model is high, the space of possible model instantiations usually goes largely unexplored. Compounding the problem, even if a set of computational ideas are on the right track, the instantiated "scale" of those ideas is typically small (e.g. in terms of dimensionality and amount of training experience provided). Thus, when a model fails to approach the abilities of the visual system, we are left uncertain whether this failure is because we are missing a fundamental idea, or because the correct "parts" have not been tuned correctly, assembled at sufficient scale, or provided with sufficient experience.

To explore a possible way forward, we here present a high-throughput approach to more expansively exploring the possible range of biologically-inspired models – including models of larger, more realistic scale – leveraging recent advances in commodity stream processing hardware (high-end graphics cards and the Playstation 3's Cell Broadband Engine processor). In analogy to high-throughput screening approaches in molecular biology and genetics, we generated and trained tens of thousands of potential network architectures and parameter instantiations, and we "screened" the visual representations produced by these models using tasks that engage the core problem of object recognition – tolerance to image variation. From these candidate models, the most promising were selected for further analysis. We show that this approach can yield significant, reproducible gains in performance in a basic object recognition task and that it offers insight into which computational ideas are most important for achieving this performance. Such insights can then be "fed back" into the design of candidate models (constraining the search space and suggesting additional model features), further guiding "evolutionary" progress. As the scale of available computational power continues to expand, we believe that this approach holds great potential both for accelerating progress in artificial vision, and for generating new, experimentally-testable hypotheses for the study of biological vision.