A High-Throughput Screening Approach to Biologically-Inspired Object Recognition

Nicolas Pinto[1], James J. DiCarlo[1] & David D. Cox[2]
1. McGovern Inst/Dept of Brain & Cog Sci, MIT
2. Rowland Institute, Harvard

While many models of biological object recognition share a common set of "broad-stroke" properties, the performance of any one model depends strongly on the choice of parameters in a particular instantiation of that model – e.g. the number of units per layer, the size of pooling kernels, exponents in normalization operations, etc. Since the number of such parameters (explicit or implicit) is typically large, and the computational cost of evaluating one particular parameter set is high, the space of possible model instantiations goes largely unexplored. Thus, when a model fails to approach the abilities of biological visual systems, 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 enough training. Here, we present a high-throughput approach to the exploration of such parameter sets, leveraging recent advances in stream processing hardware (high-end NVIDIA graphics cards and the PlayStation 3's IBM Cell Processor). In analogy to high-throughput screening approaches in molecular biology and genetics, we explored thousands of potential network architectures and parameter instantiations, screening those that show promising object recognition performance for further analysis.

In this first foray into "high-throughput" model discovery, we constructed a basic, but inclusive "super-family" of models encompassing a range of possible models built upon principles gleaned from systems neurophysiology research (hierarchical organization, linear-nonlinear operations, simple unsupervised learning rules, etc.). We show that this high-throughput search approach can yield significant, reproducible gains in performance across an array of basic object and face recognition tasks, consistently outperforming a variety of state-of-the-art purpose-built vision systems from the literature. We further show that even with unsupervised learning rules disabled, a pure brute-force "learning-free" search approach can be used to find feature representations that achieve excellent performance (in combination with "off-the-shelf" supervised learning techniques) across a variety of test sets, including the popular "Labeled Faces in the Wild" unconstrained face recognition test set.

As the scale of available computational power continues to expand, we argue that this approach has the potential to greatly accelerate progress in both artificial vision and our understanding of the computational underpinning of biological vision.