Why is Real-World Object Recognition Hard?: Establishing Honest Benchmarks and Baselines for Object Recognition

Nicolas Pinto¹, David D. Cox¹,², Benoit Corda¹, David Doukhan¹ and James J. DiCarlo¹

¹. The McGovern Institute for Brain Research, MIT
². The Rowland Institute at Harvard

Progress in understanding the brain mechanisms underlying vision requires the construction of computational models that not only emulate the brain's anatomy and physiology, but ultimately match its performance on visual tasks. In recent years, large databases of ostensibly "natural" images have become popular in the study of vision and have been used to show apparently impressive progress in building such models. However, while these test sets have grown increasingly large, it remains difficult to know whether they properly capture the essence of the problem domain that they are intended to represent. Here, we challenge this assumption for a range of standard test sets in the domains of object and face recognition.

Specifically, we show that a simple V1-like model – a neuroscientist's "null" model, which should perform poorly at real-world visual object recognition tasks – matches or outperforms existing state-of-the-art recognition systems (biologically-inspired, and otherwise) on a variety of standard object and face recognition tests. As a counterpoint, we designed a series of "simpler" synthetic recognition tests that better span the real world variation in object pose, position, and scale, and we show that these tests correctly expose the inadequacy of the V1-like model. Taken together, these results demonstrate that tests based on uncontrolled "natural" images can be seriously misleading, potentially guiding progress in the wrong direction.

Going forward, we examine classes of image test sets that might be more valuable for guiding object recognition efforts. Similarly, we explore what measures could represent fair baselines for tests of object and face recognition, and show how these can be extended to the related problem of object tracking, where standardized tests are even less common. At the center of these issues is the need to clearly define what the problem is, why it is difficult, and what results would constitute success. The path forward will not be easy, but we argue that it is time for the field to give this problem much more central attention.