Elucidating and testing hierarchical visual models using model-optimized synthetic stimuli

Eero Simoncelli
NYU

Primate visual representations may be approximated with hierarchically cascaded models, in which each stage performs a set of canonical operations: convolutions with a set of kernels, followed by nonlinear rectification, and local gain control. The first two operations are now commonly found in deep convolutional neural network (CNN) architectures, but the third is generally not. All three operations are primarily motivated by known functional properties of neural response throughout the visual system, and they can be validated by fitting and testing against physiological or perceptual data. We've recently used such models to capture human abilities to detect/discriminate natural images. But stronger tests emerge when the models are forced to generalize well beyond basic training/testing data. I'll show several examples of this, in which we use models to generate novel stimuli that are designed to test predictions. We find that despite their generality, machine learning models such as CNNs are often outperformed by simpler biologically-constrained models on these tasks.