Beyond the limits of feed-forward processing in visual object recognition

Albert L. Rothenstein, John K. Tsotsos

York University, Toronto, Canada

The limits of feedforward processing in recognition tasks with images containing multiple objects have been explored theoretically [2], through computational modeling [5], and experimentally [1]. All these diverse methods show that performance degrades when increasing the number of objects.

We will present the results of a computational modeling study of the role visual attention plays in primate object recognition, and in particular in recognition beyond the first feed-forward pass. The main condition for multi-pass recognition is the recovery of spatial information from high level, abstract, and invariant representations. This is equivalent to a solution to the binding problem. We present a particular solution within the context of the Selective Tuning model of visual attention [3, 4]. An object detection and recognition system based on this is implemented, demonstrating the important contribution of top-down object-based attention in improving the ability of the system to eliminate the influence of distractors.

The network consists of two parallel pathways, one detects and categorizes objects in input images, while the second pathway implements an object recognition system. Similar to primate visual performance, detection and categorization are fast and parallel, while detailed recognition is slow and serial. The pathways are trained using standard supervised learning. Image processing starts with the presentation of the stimulus and the feedforward propagation of the information along parallel visual pathways. Of interest in the context of this work are representations for the categorization and recognition of stimuli. The power of the feedforward pass in creating detailed and accurate representations of the input stimuli has been extensively explored. Object based attention is deployed at the level of categoric representations, and triggers a feedback cascade of attentional modulations moving back towards early visual areas and inhibiting distractors. Attentional selection and the inhibition of distracters allows the visual system to reevaluate the stimuli present in the input image, but due to the need to eliminate interference, this process needs to be serial.

References


