

## Modeling Feature Sharing between Object Detection and Top-down Attention

Dirk Walther<sup>1</sup>, Thomas Serre<sup>2</sup>, Tomaso Poggio<sup>2</sup>, Christof Koch<sup>1</sup>

<sup>1</sup> Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA, corresponding author: D.W. (walther@klab.caltech.edu)

<sup>2</sup> Center for Biological and Computational Learning, Brain and Cognitive Sciences, and McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA

Visual search and other attentionally demanding processes are often guided from the top down when a specific task is given (e.g. Wolfe et al. Vision Research 44, 2004). In the simplified stimuli commonly used in visual search experiments, e.g. red and horizontal bars, the selection of potential features that might be biased for is obvious (by design). In a natural setting with real-world objects, the selection of these features is not obvious, and there is some debate which features can be used for top-down guidance, and how a specific task maps to them (Wolfe and Horowitz, Nat. Rev. Neurosci. 2004).

Learning to detect objects provides the visual system with an effective set of features suitable for the detection task, and with a mapping from these features to an abstract representation of the object. We have developed a model, in which V4-type “S2” features are shared between object detection and top-down attention. As the model familiarizes itself with objects, i.e. it learns to detect them, it acquires a representation for features to solve the detection task. We propose that by cortical feedback connections, top-down processes can re-use these same features to bias attention to locations with higher probability of containing the target object. Our implementation of the model outperforms pure bottom-up attention. The performance of our model, which uses only grayscale information, is comparable to a top-down bias for skin hue.

**Example image (right) with purely bottom-up saliency map (bottom right) and top-down bias map for one of the features that were learned in order to detect faces (bottom left). Observe that while four of the five faces in the image are represented in the top-down bias map, the one on the top right is not. Combinations of various top-down features as well as combination with bottom-up cues are necessary for more robustness.**



