Global Scene Features and Contextual Guidance of Attention

Antonio Torralba\(^1\), Aude Oliva\(^2\), Monica S. Castelhano\(^3\) and John M. Henderson\(^3\)

\(^1\) Computer Science and Artificial Intelligence, MIT
\(^2\) Department of Brain and Cognitive Sciences, MIT
\(^3\) Department of Psychology and Cognitive Science Program, Michigan State University

Questions about this research and request for additional information should be addressed to A.T (torralba@csail.mit.edu) or A.O. (oliva@mit.edu)

Behavioral experiments have shown that the human visual system makes extensive use of contextual information for facilitating object search in natural scenes. However, the question of how to formally model such contextual influences is still largely open. Based on a Bayesian framework, we present an original approach of attentional guidance by scene context. From the probabilistic model, two parallel pathways, one for local features (salience) and one for global (or scene-centered) features emerge naturally. A top down control enhances task-relevant information by emphasizing regions which the model has learned are likely to contain target category objects. Results show that the model accurately predicts where human eye movements are directed in a scene image. The Contextual Guidance model of attention provides the first theoretical and formal demonstration of how bottom-up saliency, scene context and top-down mechanisms can be integrated at an early stage of visual processing, much as human observers do.

Fig. 2. Schematic illustration of the Contextual Guidance Model that integrates bottom-up saliency, contextual scene information and top-down scene priors. In the local pathway, a vector of features that describe local image properties represents each location. It could be a collection of templates (e.g., object detection) or a vector composed by the output of wavelets at different orientations and scales (e.g., saliency models of attention). In the global pathway, the entire image is represented by a unique set of features that summarizes the structure of the whole scene without encoding specific objects or regions. The Top down control enhances task-relevant information in the global scene pathway by emphasizing particular regions in a given scene, which the model has learned are likely to contain target category objects. The green and red circles represent the receptive fields of a local and a global feature respectively.