

---

# Learning Object-based Attention Control

---

**Ali Borji, Majid N. Ahmadabadi, Babak N. Araabi**  
School of Cognitive Sciences  
Institute for Studies in Theoretical Physics and Mathematics  
Niavaran Bldg. P.O.Box 19395-5746, Tehran, IRAN  
borji@ipm.ir, {mnili, araabi}@ut.ac.ir

Remarkable efficiency of human vision is the main reason of being the most-studied mode of perception in machine learning. Despite the huge active research in computer vision and robotics, many real-world visumotor tasks that are easily performed by humans are still unsolved. Of special interest is designing efficient learning algorithms, in terms of high accuracy and low computational cost, for enabling autonomous mobile robots to act in visual interactive environments.

Visual attention has been frequently used for reducing the complexity of computationally intensive processes. It solves the problem of information overloading by implementing a bottleneck through which only task-relevant information are allowed to pass. From a massive amount of studies in neuroscience and psychology, it is now known that visual attention is controlled by bottom-up and top-down mechanisms. Several theories for explaining the bottom-up influences of visual attention have been proposed like saliency concept, information theory, game theory, etc. While bottom-up mechanism is well understood, much less is known about the top-down component of visual attention.

There are evidences in AI and biology which motivate learning attention control. In a novel and pragmatic point of view, situated and embodied AI, it is claimed that intelligent behaviors like attention and emotion are the product of relationships between brain of an organism, its body and the environment. Links between attention and decision making as well as learning attention control by past experience has already been shown by experimental studies. Therefore, semi-supervised approaches, especially RL, seem to be the most appropriate tools for interactive and incremental learning of task-driven visual attention control. In this study, top-down attention is learned through ordered selection of objects by maximizing the expected reward of the agent.

Our proposed model consists of three layers. First, in the early visual processing layer, basic layout and gist of a scene are extracted. The most salient location of the scene is simultaneously derived using the biased saliency-based model of visual attention [1]. Then, a cognitive component in the higher visual processing layer performs an application-specific operation such as object recognition and scene classification at the focus of attention. From this information, a state is derived in the decision making layer. Top-down attention in our model is learned by the U-TREE algorithm [2], which successively grows a tree whenever perceptual aliasing occurs. Internal nodes in this tree check the existence of a specific object in the scene and its leaves point to states in the Q-table. Motor actions are associated with the leaves. After performing a motor action, the agent receives a reinforcement signal from the critic which is alternately used for modifying the tree or updating the action selection policy. A long-term memory component holds the bias signals of important task-relevant objects of the environment. Basic saliency-based model of visual attention is devised to consider processing costs of feature channels and image resolutions [3]. To recognize objects, a recent and successful method inspired by the hierarchical organization of the visual ventral stream, is used [4]. Experimental results on visual navigation tasks, lend support to the applicability and usefulness of this approach for robotics.

## References

- [1] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Machine Intell*, 20(1998), 1254-1259.
- [2] A. K. McCallum, Reinforcement learning with selective perception and hidden state, Doctoral dissertation, Department of Computer Science, University of Rochester. 1995.
- [3] A. Borji, M. N. Ahmadabadi, B. N. Araabi, Learning top-down feature-based attention control, Workshop on 'Vision in Action: Efficient strategies for cognitive agents in complex environments', *ECCV 2008*.
- [4] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, Object recognition with cortex like mechanisms, *IEEE Trans. Pattern Anal. Machine Intell*, 29(2007), 3, 411-426.