A Neural Model of Visual Attention

Piërre van de Laar
Tom Heskes
Stan Gielen
Department of Medical Physics and Biophysics, University of Nijmegen
RWCP* Novel Functions SNN† Laboratory
Nijmegen, The Netherlands

Abstract

We propose a biologically plausible neural model of selective covert visual attention. We show that this model is able to learn focussing on object-specific features. It has similar learning characteristics as humans in the learning and unlearning paradigm of Shiffrin and Schneider [8].

1 Introduction

In many pattern recognition and navigation tasks humans outperform computers. One explanation for this fact is that the software algorithm has to process all available information at a time, whereas a human quickly focusses its attention on the most important information, without paying too much attention to irrelevant information. The aim of this study is to model this process of information selection. In this article we roughly sketch a model for selective covert visual attention. Inspired by the human visual and attentional systems, we have made a biologically plausible model that might be able to explain some of the results of psychological experiments on attention, and especially on the learning of attention. Our model aims not to "appeal to some intelligent force or agent in explanations of attention phenomena" [3] but rather to be "a mechanism for attention selection" [1].

2 Selective Covert Attention

Johnston and Dark gave the following definition of selective attention: "Selective attention refers to the differential processing of simultaneous sources of information. In nature these sources are internal (memory and knowledge) as well as external (environmental objects and events)" [3]. As Johnston and Dark, we will follow "the vast majority of contemporary investigators and consider only external sources." Covert attention means that no overtly visible action is made to select between the different sources of information.

In our model we will consider both the bottom-up and top-down aspects of attention. In bottom-up attention the stimulus causes neural activity at the

*Real World Computing Partnership
†Foundation for Neural Networks

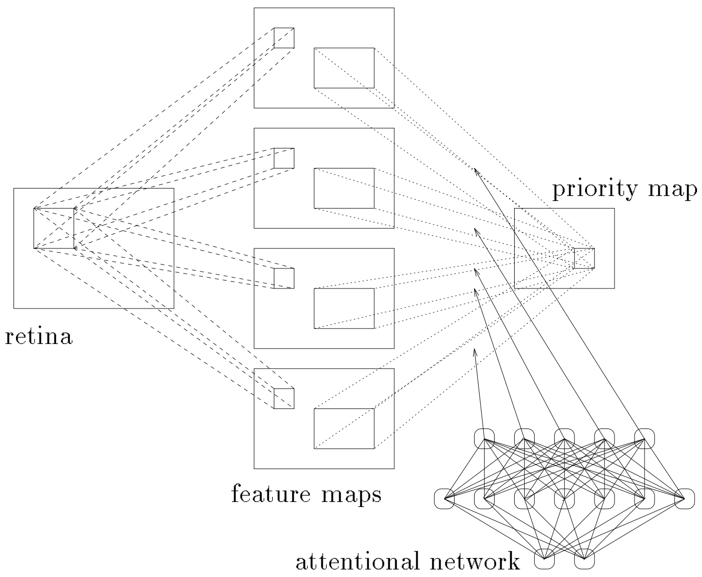

Figure 1: The architecture of the artificial visual system. From left to right: the retina, the feature maps and the priority map. The information going from feature maps to priority map is influenced by the attentional network, which receives a task-depending input.

different processing levels, which may affect the focus of attention. In top-down attention, the individual is internally biased towards particular stimuli, depending on the task or on the interpretation of the task to be performed. For example, an oval shape as the memory-set item in a distractor set of letters has been shown to give rise to a faster response for subjects who think the oval shape is a number (zero) than for those who start searching for a letter (o) [2].

3 A Neural Network Model

Our model, depicted in figure 1, receives only stationary images. These images are viewed by a "retina", which has a homogeneous distribution of receptors. The receptors pass their information on to the feature maps. Neurons in a feature map are topologically ordered, meaning that neighboring neurons have neighboring receptive fields. Translation-invariant pattern recognition is incorporated by applying weight sharing [5]. Each feature map has its own optimal stimulus which could be hardwired or learned, for example a color or a bar of a particular orientation.

The attentional network receives task-dependent input and controls the top-down aspects of attention by affecting the flow of information of the visual system from the lower levels to higher levels. In our model, the attentional network is a multi-layered perceptron, the output of which is used to gate the information flow from the different feature maps to the priority map. The priority map determines which part of the visual field contains the most inter-

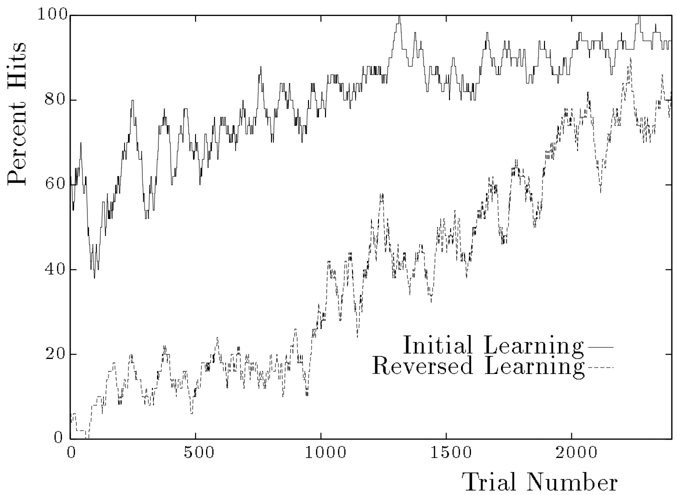

Figure 2: The reversal experiment: performance of detecting a target among distractors as a function of time before and after reversal. A hit occurs when our model focusses its attention on the target. After the initial 2400 learning steps, the target and distractor sets are interchanged and the reversed learning performance is measured over another 2400 learning steps.

esting information. Our attentional network is trained by backpropagation to enhance the relevant and to suppress the irrelevant information for the current task, which is specified by its input (see also [4]).

4 Comparison with Psychological Data

To test the selective attentional properties of the modeled visual system[1] we have compared its performance with the psychological data gathered in the learning and unlearning paradigm of Shiffrin and Schneider [8]. Qualitatively the performance of human subjects can be described as follows. Subjects who have to detect a target selected from a fixed set of targets among distractors selected from a fixed set of distractors, show a performance improving over time. However, when the target and distractor sets are reversed, the performance of these "trained" subjects drops far below the initial untrained level. Of course, continual training again leads to a better performance, but it takes a longer period of time to achieve the same accuracy as before reversal. Figure 2, which should be compared with figure 3 on page 132 of [8], illustrates that our neural-network model yields similar learning characteristics.

[1] Inputs are the characters {GMFP} and {CNHD}, which can be placed in either the left or the right visual field. The feature maps encode closure, bars of particular orientation in particular parts of the visual field and symmetries. The attentional network, which has noisy output neurons, was trained to yield maximal activation in the part of the visual field corresponding to the target position. In both situations, the task is the same: to detect the target among a set of distractors.

5 Discussion

The activity of the priority map can be used as input for control units in so-called shifter models (see e.g. [6]). These shifter models transfer the information inside the locus of attention to the higher levels of the visual system.

Our model only focusses its attention based on shape information. Attention towards a particular spatial position requires an extra module. This module may either influence the flow of information from the feature maps to the priority map or it can directly affect the priority map. Thus "two different attentional systems serve as sources of activation for [...] form and for location, although both might enter and amplify activity within the visual system at the same site" [7].

Our main goal remains to study the learning of attention. There is a vast amount of literature on all kinds of quantitative and qualitative psychological experiments, e.g. [2, 8, 9], that can be simulated with our neural-network model. In this paper we have given an illustrative example. In the future we will continue to test and improve our model in order to comply with the available psychological data without giving in to its biological plausibility.

References

[1] C. Bundesen. A theory of visual attention. *Psychological Review*, 97(4):523–547, 1990.

[2] J. Johnides and H. Gleitman. A conceptual category effect in visual search: O as letter or as digit. *Perception & Psychophysics*, 12(6):457–460, 1972.

[3] W. A. Johnston and V. J. Dark. Selective attention. *Annual Review of Psychology*, 37:43–75, 1986.

[4] J. K. Kruschke. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1):22–44, 1992.

[5] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1990.

[6] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience*, 13(11):4700–4719, 1993.

[7] M. I. Posner and S. Dehaene. Attentional networks. *Trends in Neurosciences*, 17(2):75–79, 1994.

[8] R. M. Shiffrin and W. Schneider. Controlled and automatic human information processing: II. perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2):127–190, 1977.

[9] A. Treisman. Features and objects in visual processing. *Scientific American*, 255(5):106–115, 1986.