Available online at www.sciencedirect.com
SCIENCE DIRECT®
Neurocomputing 58–60 (2004) 613–618
NEUROCOMPUTING
www.elsevier.com/locate/neucom

A normative model of attention: receptive field modulation

Santiago Jaramillo*, Barak A. Pearlmutter

Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland

Abstract

When sensory stimuli are encoded in a lossy fashion for efficient transmission, there are necessarily tradeoffs between the represented fidelity of various aspects of the stimuli. In the model of attention presented here, a top-down signal informs the encoder of these tradeoffs. Given the stimulus ensemble and tradeoff requirements, our system learns an optimal encoder. This general model is instantiated in a simple network: an autoencoder with a bottleneck, innervated by a top-down attentional signal, and trained using backpropagation. The modulation of neural activity learned by this model qualitatively matches that measured in animals during visual attention tasks.

Keywords: Attention; Neural coding; Learning; Receptive field

1. Introduction

Normative models explain structure as being optimized to perform some function well, and have recently found fruitful application in the study of low-level sensory processing (e.g. [1]). Optimized representation models of visual receptive fields have to date used fixed and homogeneous fidelity requirements. We introduce a normative model of top-down attention in which an attentional signal (originating outside the model) breaks this symmetry by modulating the tradeoffs in transmission fidelity of the features of an input pattern.

The limited capacity for processing information and the ability to filter out unwanted information are the basic phenomena that define attention [4]. Physiological studies have shown that when two stimuli are presented simultaneously inside a cell's receptive field,

* Corresponding author. Tel.: +353-1-708-6100; fax: +353-1-708-6269.
E-mail addresses: sjara@ieee.org (S. Jaramillo), barak@cs.may.ie (B.A. Pearlmutter).

0925-2312/$ - see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/j.neucom.2004.01.103

the cell's response is strongly influenced by which of the two stimuli was attended [7,8,11]. But the fashion in which this modulation occurs remains a subject of debate.

Previous computational models of top-down attention used gating mechanisms or synaptic modulation to implement selective attention [3,5,6,9,12], and the particulars of the modulation is thus built into those models. This paper introduces a new class of model, in which the attentional signal is presented to the processing layers in the same fashion as is the sensory input, and the system learns to assign resources to different parts of the stimulus, and to modulate this assignment according to the attentional signal, without any special structure or architectural bias. The only information concerning the semantics of the attentional signal comes from the error measure, which for the specific visual phenomena being modeled here penalizes the system more heavily for errors in an attentional spotlight.

2. Methods

Network architecture: The system consists of an auto-associative network with five layers (Fig. 1). The bottom layers encode the input signal, while the top layers decode it. The central bottleneck layer represents the input pattern using fewer units[1] than the input layer. Each layer is fully connected to the next, and they all receive an additional *top-down attentional signal* input. The layer sizes are 256–20–10–20–256, where the input and output layers are treated as 16×16 grids for display purposes. The attentional signal consists of a two-element vector representing the center of the attentional spotlight in Cartesian coordinates scaled into the range ± 1. The hyperbolic tangent activation function was used throughout.

Training: The encoder and decoder were jointly optimized to minimize $E(\mathbf{p}) = \sum_i c_i(\mathbf{p})(y_i(\mathbf{p}) - d_i(\mathbf{p}))^2$, where i indexes locations in the 16×16 grids holding the stimulus and its reconstruction, $c_i(\mathbf{p})$ is the intensity of the attentional spotlight, $y_i(\mathbf{p})$ is the output of the network, $d_i(\mathbf{p})$ is the desired output, which is in our case the same as the input, and $\mathbf{p}$ represents the complete pattern of information coming into the system at one point in time, i.e. the input pattern as well as the top-down attentional signal. The gradient was calculated using backpropagation. Optimization used online gradient descent with a weight decay term of 10^{-6} and a learning rate $\eta = 0.005$. All weights were plastic during learning, and the attention coefficients in the penalty function formed a simple soft mask $c_i(\mathbf{p}) = 1/(1 + k^2\|i - a(\mathbf{p})\|^2)$, with $a(\mathbf{p})$ being the attentional input (a two-dimensional vector in our case) and i being a location in the plane. The width of the attentional spotlight was set by k, which was held constant at $k = 12$ in our simulations.

Training set: The 2000-element training set consisted of 16×16 pixel images, with the pixels being zero mean and having standard deviation $\sigma = \frac{1}{3}$. The images were created by convolving (filtering) white Gaussian noise images with a rotationally symmetric 2D Gaussian with $\sigma_{\text{filter}} = 2$. Edge effects were avoided by extracting only the

[1] Fewer units is not a strict requirement, as other means, such as injected noise, can serve to limit the capacity of the bottleneck.

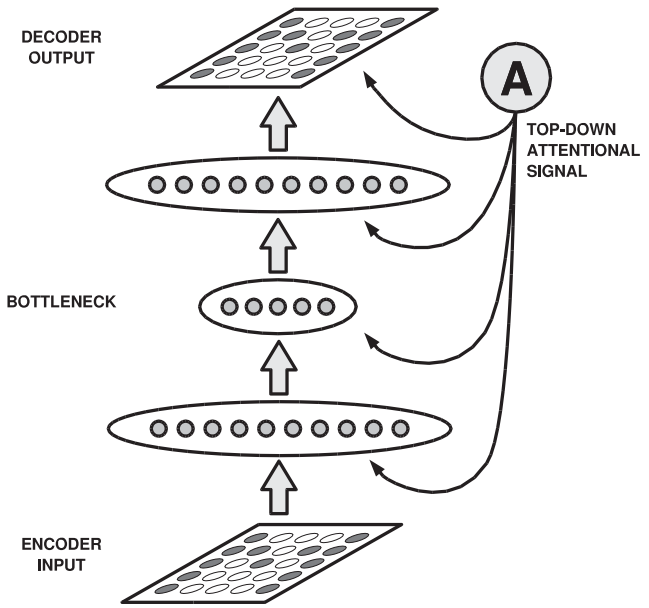

Fig. 1. Network architecture. The network contains five layers forming an encoder/decoder system with a bottleneck. Each layer receives the attentional signal.

16×16 center of the resulting image. These images were later scaled to have the desired variance. The center of the attentional mask was drawn independently of the input image, and uniformly distributed within the input image.

Controls: To limit the capacity of the system, which is theoretically unbounded for real-valued units, zero-mean Gaussian noise with standard deviation 0.1 was added to each bottleneck unit's total input during training. Moreover, we confirmed that the system is in fact reassigning resources appropriately, and not just degrading perfomance for unattended location, by comparing results using a flat attentional mask to those exhibited with the peaked mask described above.

3. Results

Encoding/decoding: An example of the encoding/decoding results for a testing pattern (i.e. a pattern not included in the training set) is shown in Fig. 2. This figure presents the output of the system when the center of the spotlight of attention is located in different corners of the image while the image itself is held constant. The dashed circles indicate the location of the attentional spotlight, but should not be interpreted as hard-edged masks. Obtaining a lower error inside the dashed circles is consistent with the hypothesis that attention assigns more resources to attended locations, thus giving better reconstruction of some features of the input stimulus.

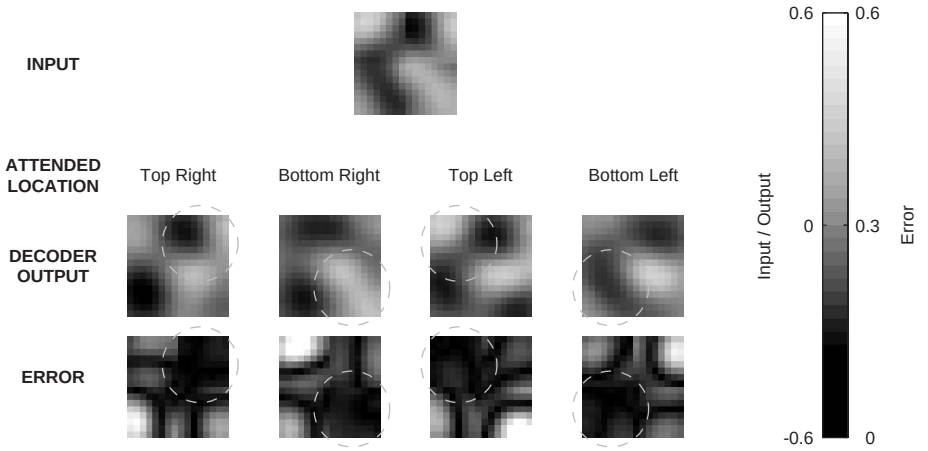

Fig. 2. Reconstruction example for a single input pattern and four different attentional states as indicated by the dashed circles. The error is calculated as the absolute intensity difference between input and output.

Modulation of unit activation: A linear approximation of the response of each neuron in the bottleneck was found using the reverse correlation method [2,10]. These values were used to define "excitatory" and "inhibitory" stimuli for each bottleneck unit, with respect to its activation. Images containing combinations of excitatory and inhibitory regions were created and presented to the network. Fig. 3a shows the activation of one bottleneck unit for two attentional states (right or left, as indicated by the dashed ellipse) and four different input images. The $+/-$ symbols indicate which part of the image contains excitatory/inhibitory input. The difference of activation of the bottleneck units as attention is shifted from right to left is presented in Fig. 3b. The height of each bar represents the average over all bottleneck units (10 in our case). Standard errors are also shown. These figures show a clear modulation of the activation of a unit: when the same stimulus is presented (one side excitatory, and the other inhibitory)

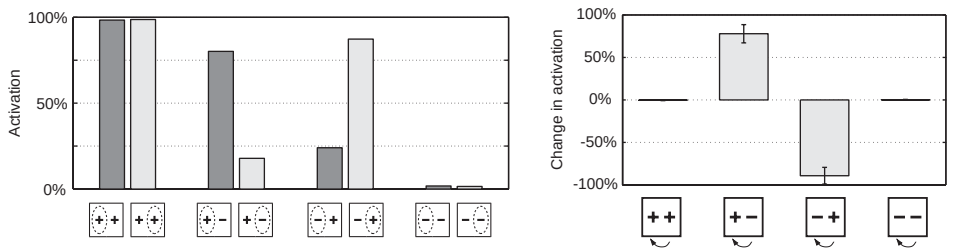

Fig. 3. Attentional modulation of activation. (a) Activation of one bottleneck unit for two attentional states. Squares indicate stimuli, created by combining left and right halves of excitatory $(+)$ and inhibitory $(-)$ inputs. Dashed ellipse indicates attentional spotlight. (b) Average change in activation over all units in the bottleneck with right/left attentional shift, $\pm$ standard error. Squares at bottom represent stimuli.

the activation changes dramatically depending on the top-down attentional signal. This result is common for all units in the bottleneck, as indicated by the averages and small standard errors in Fig. 3b. These results also show that stimulus changes in the unattended location produce smaller changes in activation than changes in the attended location. This can be seen by comparing the second and third dark bars (attention to the left, with one half excitatory) with the first dark bar (attention to the left with both halves excitatory) in Fig. 3a.

These results qualitatively match neuronal response changes found in animals during selective visual attention tasks [8,11].

4. Conclusion

We presented an unstructured model that learns a covert top-down attentional mechanism. Top-down attentional signals innervate the entire network, and each unit treats them no differently than bottom-up sensory signals. The only information received by the network about the semantics of the attentional input comes from the objective optimized during learning. This model accounts for attentional modulation of neural response in a unified framework that includes both attention and receptive field formation, and as a consequence of an underlying normative principle, rather than by tuning a complex special-purpose architecture.

One general prediction of this class of models is that a system with a narrower bottleneck (or richer input) will have stronger attentional modulation than a system with sufficient capacity to represent its input with high fidelity. This might be tested by raising animals in visually rich vs. impoverished environments and measuring differences in the magnitude of attentional modulation of receptive fields.

The model reproduces neuronal modulation observed in physiological experiments, and can be naturally applied across tasks and across sensory modalities. The model has the potential of being extended to attentional goals where modulation would be less intuitive, such as acoustic source segregation, or feature-driven attentional goals such as priming. It may also be possible to build a hierarchical system whose modules not only receive top-down attention signals, but generate such signals for lower-level modules.

Acknowledgements

We thank Heather L. Read for helpful comments. Supported by US NSF PCASE 97-02-311, the MIND Institute, an equipment Grant from Intel, a gift from the NEC Research Institute, and Science Foundation Ireland Grant 00/PI.1/C067.

References

[1] J. Atick, J. Redlich, A.N. Fall, Towards a theory of early visual processing, Neural Comput. 2 (3) (1990) 308–320.

[2] R. de Boer, P. Kuyper, Triggered correlation, IEEE Trans. Biomed. Eng. 15 (3) (1968) 169–179.
[3] G. Deco, J. Zihl, A neurodynamical model of visual attention: feedback enhancement of spatial resolution in a hierarchical system, Comput. Neurosci. 10 (3) (2001) 231–253.
[4] R. Desimone, J. Duncan, Neural mechanisms of selective visual attention, Ann. Rev. Neurosci. 18 (1995) 193–222.
[5] D. Heinke, G.W. Humphreys, Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (SAIM), Psychol. Rev. 110 (1) (2003) 29–87.
[6] G.E. Hinton, K.J. Lang, Shape recognition and illusory conjunctions, in: the Ninth International Joint Conference on Artificial Intelligence, Vol. 1, Morgan Kaufmann, Los Angeles, 1985, pp. 252–259.
[7] S.J. Luck, L. Chelazzi, S.A. Hillyard, R. Desimone, Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex, J. Neurophysiol. 77 (1) (1997) 24–42.
[8] J. Moran, R. Desimone, Selective attention gates visual processing in the extrastriate cortex, Science 229 (4715) (1985) 782–784.
[9] B.A. Olshausen, C.H. Anderson, D.C. Van Essen, A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, J. Neurosci. 13 (11) (1993) 4700–4719.
[10] F. Rieke, D. Warland, R. de Ruyter van Steveninck, W. Bialek, Spikes: Exploring the Neural Code, a Bradford Book, MIT Press, Cambridge, MA, 1996.
[11] S. Treue, J.H. Maunsell, Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas, J. Neurosci. 19 (17) (1996) 7591–7602.
[12] P. van de Laar, T. Heskes, S. Gielen, Task-dependent learning of attention, Neural Networks 10 (6) (1997) 981–992.