Neural Networks 26 (2012) 7–28

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Predictive coding as a model of the V1 saliency map hypothesis

M.W. Spratling*

King's College London, Department of Informatics and Division of Engineering, London, UK

ARTICLE INFO

Article history:
Received 3 June 2010
Received in revised form 15 July 2011
Accepted 10 October 2011

Keywords:
Saliency map
Texture segmentation
Visual search
Contour integration
Neural network model
Attention

ABSTRACT

The predictive coding/biased competition (PC/BC) model is a specific implementation of the predictive coding theory that has previously been shown to provide a detailed account of the response properties of orientation tuned cells in primary visual cortex (V1). Here it is shown that the same model can successfully simulate psychophysical data relating to the saliency of unique items in search arrays, of contours embedded in random texture, and of borders between textured regions. This model thus provides a possible implementation of the hypothesis that V1 generates a bottom-up saliency map. However, PC/BC is very different from previous models of visual salience, in that it proposes that saliency results from the failure of an internal model of simple elementary image components to accurately predict the visual input. Saliency can therefore be interpreted as a mechanism by which prediction errors attract attention in an attempt to improve the accuracy of the brain's internal representation of the world.

1. Introduction

A number of psychophysical experiments suggest that primary visual cortex (V1) may be involved in the computation of visual salience (Koene & Zhaoping, 2007; Zhaoping, 2008; Zhaoping, Guyader, & Lewis, 2009a; Zhaoping & May, 2007; Zhaoping, May, & Koene, 2009b). These experiments thus support the hypothesis that V1 operates as a bottom-up, pre-attentive, saliency map (Li, 2002). Previous work (Spratling, 2010, 2011) has demonstrated that a simple functional model (PC/BC), derived from the predictive coding and biased-competition theories of cortical function (Spratling, 2008a, 2008b), can simulate a very wide range of V1 response properties including orientation tuning, size tuning, spatial frequency tuning, temporal frequency tuning, cross-orientation suppression, and surround suppression. This article extends that work by showing that the PC/BC model of V1 can also simulate a wide range of psychophysical experiments on visual salience, and hence, demonstrates that PC/BC provides a possible implementation of the V1 saliency map hypothesis.

Predictive coding is a scheme for combining bottom-up evidence with prior knowledge to infer the most likely causes of a sensory stimulus (Bubic, von Cramon & Schubotz, 2010; Rao & Ballard, 1999). This is achieved through an iterative process in which a prediction about the underlying causes of the sensory data (*e.g.*, an internal representation of the world), is used to reconstruct the expected sensory input. These predicted inputs are compared with the actual stimulus-driven activity in order to calculate the residual error between the predicted data and the sensory evidence. This error is then used to modify the predicted causes to form a more accurate internal model of the world, which will in turn reduce the residual error. Predictive coding is a specific example of more general theories of efficient encoding or redundancy reduction (Attneave, 1954; Barlow, 2001; Olshausen & Field, 1996b, 1997), of generative models of inference and learning (Hinton, 2002; Hinton, Dayan, Frey & Neal, 1995; Hoyer, 2004; Hoyer & Hyvärinen, 2000; Lee & Seung, 1999; Olshausen & Field, 1996a), and of theories of hierarchical perceptual inference or analysis-by-synthesis (Barlow, 1994; Friston, 2005; Lee & Mumford, 2003; Mumford, 1992; Yuille & Kersten, 2006). See Spratling (in press) for a more in-depth discussion of the relationship between predictive coding and other theories of cortical function.

Predictive coding is also a particular instantiation of the free-energy principle (Friston, 2010, 2009). Free-energy suggests that sensory prediction errors give rise to action that will reduce this error (Friston, 2010; Friston, Daunizeau, Kilner & Kiebel, 2010). Hence, if perceptual salience has a role in the control of action (*e.g.*, in directing eye movements or in the allocation of endogenous attention) then saliency should be correlated with the prediction errors generated in a predictive coding model. In order to test this hypothesis, measurements were made of the residual errors generated by the PC/BC model. For a very wide range of images, the relative strength of the error calculated by PC/BC at different locations in the image was found to be consistent with the perceptual saliency of those different parts of the image. The model was tested by comparing the saliency values calculated from the residual error generated by the PC/BC model with

* Correspondence to: Division of Engineering, King's College London, Strand, London WC2R 2LS, UK. Tel.: +44 20 7848 2027.
E-mail address: michael.spratling@kcl.ac.uk.

0893-6080/$ – see front matter
doi:10.1016/j.neunet.2011.10.002

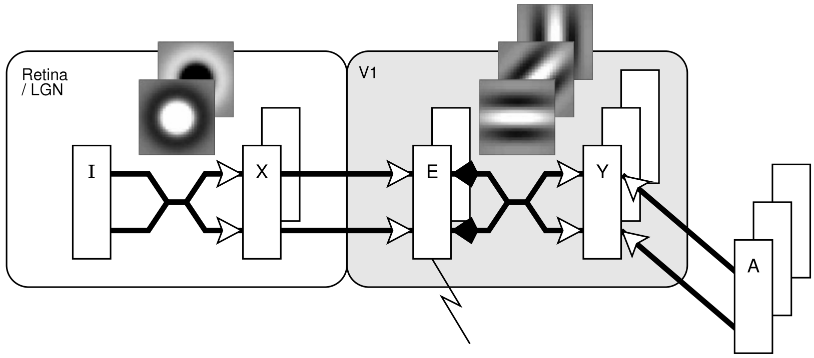

Fig. 1. The PC/BC model of V1 (right) and the model of retina/LGN (left). The input image I was preprocessed by convolution with a circular-symmetric on-center/off-surround kernel to generate the input to the ON channel of the V1 model, and a circular-symmetric off-center/on-surround kernel to generate the input to the OFF channel of the V1 model. The prediction neurons, labeled **Y**, represent V1 simple cells. The activity of these neurons was simulated by convolving the outputs of the ON and OFF channels of the error-detecting neurons, labeled **E**, with the ON and OFF channels of a number of weight kernels (defined by Gabor functions) representing V1 RFs. This convolution process effectively reproduces the same RFs at every pixel location in the image. The prediction neuron responses could also be modulated by feedback from higher cortical regions which were not explicitly modeled, rather these effects were simulated by additional inputs to the V1 model, labeled **A**. The responses of the error-detecting neurons were influenced by divisive feedback from the prediction neurons, which was also calculated by convolving the prediction neuron outputs with the weight kernels. Responses of the error-detecting neurons were recorded during experiments and the strength of response at each location was assumed to be related to the saliency of that location in the image.

psychophysical measures of saliency (*i.e.*, reaction times and response accuracy) recorded in experiments on texture segmentation, visual search, and contour integration. The model was found to provide an accurate account of perceptual saliency in all of these domains.

Specifically, in tasks evaluating the saliency of the border between textured regions, the model is shown to account for: the effects of orientation contrast (Section 3.1); the effects of element spacing (Section 3.2); and of superimposed irrelevant texture elements (Section 3.3). In tasks evaluating the saliency of unique elements in search arrays, the model is shown to account for the range of search efficiencies ('pop-out', 'serial', and 'parallel' search) found in psychophysical experiments (Section 3.4); asymmetries in visual search (Section 3.5); the effects of element spacing (Section 3.6); the effects of superimposed irrelevant texture elements (Section 3.7); the effects of abrupt element onsets (Section 3.8); to account for the preview effect (Section 3.9); and to account for 'flicker' induced change blindness (Section 3.10). The model is also used to explore the possible effects on saliency of cortical feedback generated by expectation, attention, or object familiarity. In these experiments, the model is shown to account for the saliency of a contour embedded in random texture elements (Section 4.1); the effects of element novelty and familiarity in feature search (Section 4.2); the effects of prior exposure to features of either the target or distractors in a subsequent search array, *i.e.*, "priming of pop-out" and the "distractor previewing effect" (Section 4.3); the preview effect (Section 4.4); the increase in target saliency when distractor elements form a familiar contour (Section 4.5); the effects of contextual guidance in feature search (Section 4.6); and the saliency of objects that are incongruous with the scene (Section 4.7). The results of these experiments also suggest that predictive coding can provide a natural explanation for the faster recognition of objects congruent with a visual scene despite the fact that incongruent objects are more likely to attract attention (Section 4.7), and inhibition-of-return (Section 4.8).

For all these seemingly diverse experiments on visual saliency, the model provides a single computational explanation of the results. The model proposes that V1 encodes visual information using an over-complete set of Gabor basis functions, which provided a means of accurately and efficiently representing natural images (Field, 1987, 1994; Olshausen & Field, 1997, 2005). Salient locations in an image are those locations where this representation is least accurate.

2. Model description

2.1. The retina/LGN model

To simulate the effects of circular-symmetric center–surround receptive fields (RFs) in LGN and retina, the input to the PC/BC model of V1, described below, was an input image (I) pre-processed by convolution with a Laplacian-of-Gaussian (LoG) filter (l) with a standard deviation equal to 1.5 pixels, The output from this filter was subject to a multiplicative gain (the strength of which was determined by parameter κ) followed by a saturating non-linearity, such that:

$$\mathbf{X} = \tanh\{\kappa(I * l)\}. \quad (1)$$

A value of $\kappa = 2\pi$ was used in all experiments reported here.

The positive and rectified negative responses were separated into two images $\mathbf{X}_{ON}$ and $\mathbf{X}_{OFF}$ simulating the outputs of cells in retina and LGN with on-center/off-surround and off-center/on-surround RFs respectively. These ON- and OFF-channels provided the input to the PC/BC model of V1. This pre-processing stage described above is illustrated on the left of Fig. 1.

2.2. The V1 model

The PC/BC model of V1 is illustrated on the right of Fig. 1 and described by the following equations:

$$\mathbf{E}_o = \mathbf{X}_o \oslash \left(\epsilon_2 + \sum_{k=1}^{p}\left(\hat{w}_{ok} * \mathbf{Y}_k\right)\right) \quad (2)$$

$$\mathbf{Y}_k \leftarrow (\epsilon_1 + \mathbf{Y}_k) \otimes \sum_{o}\left(w_{ok} \star \mathbf{E}_o\right) \quad (3)$$

$$\mathbf{Y}_k \leftarrow \mathbf{Y}_k \otimes (1 + \eta\mathbf{A}_k) \quad (4)$$

where $o \in [ON, OFF]$; $\mathbf{X}_o$ is a two-dimensional array, equal in size to the input image, that represents the input to the model of V1; $\mathbf{E}_o$ is a two-dimensional array, equal in size to the input image, that represents the error-detecting neuron responses; $\mathbf{Y}_k$ is a two-dimensional array, equal in size to the input image, that represents the prediction neuron responses; $\mathbf{A}_k$ is a two-dimensional array, equal in size to the input image, that represents the weighted sum of top-down predictions arising from extrastriate cortical regions not explicitly modeled here; w_{ok}