Predictive Coding with Context as a Model of Image Saliency Map

Duzhen Zhang and Chuancai Liu

Abstract Predictive coding/biased competition (PC/BC) is a computational model of primary visual cortex (V1). Recent literature demonstrates that PC/BC model provides an implementation of the V1 bottom-up saliency map hypothesis. In this paper, we propose a novel approach toward natural color images saliency detection via the PC/BC model with top-down cortical feedback as context. We compare our method with the five state-of-the-art models of saliency detectors. Experimental results show that our method performs competitively for visual saliency detection task.

Keywords Saliency map · PC/BC model · Primary visual cortex (V1) · Top-down · Bottom-up · Context

1 Introduction

The visual system pays attention to the salient object. A number of psychophysical experiments suggest that primary visual cortex (V1) may be involved in the computation of visual salience. Spratling introduced the nonlinear predictive coding/biased competition (PC/BC) model [1], a reformulation of predictive

D. Zhang (✉) · C. Liu
School of Computer Science and Engineering, Nanjing University of Science and Technology, NJUST, Nanjing, China
e-mail: zhduzhen@yahoo.cn

C. Liu
e-mail: liu.ccnj@yahoo.com.cn

D. Zhang
School of Computer Science and Technology, Jiangsu Normal University, JSNU, Xuzhou, China

F. Sun et al. (eds.), *Foundations and Practical Applications of Cognitive Systems and Information Processing*, Advances in Intelligent Systems and Computing 215,
DOI: 10.1007/978-3-642-37835-5_2,

coding consistent with the biased competition theory of attention, that can simulate a very wide range of V1 response properties including tuning and suppression [2, 3]. The paper [4] extends his previous work by showing that the PC/BC model of V1 can also simulate a wide range of psychophysical experiments on visual salience, and demonstrates that PC/BC provides a possible implementation of the V1 bottom-up saliency map hypothesis. It proposes that the perceptual saliency of the image is consistent with the relative strength of the prediction error calculated by PC/BC. 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 [4].

Visual saliency plays important roles in natural vision in that saliency can direct eye movements, deploy attention, and facilitate tasks like object detection and scene understanding. Many models have been built to compute saliency map. There are two major categories of factors that drive attention: bottom-up factors and top-down factors [5]. Bottom-up factors are derived solely from the visual scene. Regions of interest that attract our attention are in a bottom-up way and the responsible feature for this reaction must be sufficiently discriminative with respect to surrounding features. Most computational models focused on bottom-up attention, where the subjects are free-viewing a scene and salient objects attract attention. Inspired by the feature-integration theory [6], Itti et al. [7] proposed one of the earliest bottom-up selective attention models by utilizing color, intensity, and orientation of images. Bruce et al. [8] introduced an idea of using Shannon's self-information to measure the perceptual saliency. Saliency using natural image statistics (SUN) is a bottom-up bayesian framework [9]. Recently, Hou et al. [10] proposed a dynamic visual attention approach to calculate the saliency map based on Incremental Coding Length (ICL). Bottom-up attention can be biased toward targets of interest by top-down cues such as object features, scene context and task-demands. Bottom-up and top-down factors should be combined to direct attentional behavior. A recent review of attention models from computational perspective can be found in [11].

Reference [4] uses synthetic stimuli to test the saliency of the PC/BC model. In this paper, inspired by the work of Spratling, we propose an approach toward natural color images saliency detection via the PC/BC model with top-down cortical feedback as context. We compare our method with the five state-of-the-art models of saliency detectors. Experimental results show that our method performs competitively for visual saliency detection task. The rest of this paper is organized as follows. Section 2 introduces and analyzes Spratling's PC/BC model, and based on his work, a novel method combining top-down cortical feedback for measuring image saliency is proposed. Experimental results and comparisons with state-of-the-art models are presented in Sect. 3, and discussions are given in Sect. 4.

2 The Model Description

Figure 1 illustrates the retina/LGN model and the PC/BC model of V1, from left to right, capital characters I, X, E, Y, and A represent input image, image preprocessing stage by the retina/LGN, the error-detecting neurons, the prediction neurons, feedback from higher cortical regions, respectively.

2.1 The Retina/LGN Model

To simulate the effects of circular-symmetric center-surround receptive fields (RFs) in lateral geniculate nucleus (LGN) and retina, input image (I) preprocessed by convolution with a Laplacian-of-Gaussian (LoG) filter (l) and a saturating nonlinearity:

$$X = \tanh\{2\pi(I * l)\}. \tag{1}$$

The positive and rectified negative responses were separated into two images X_{ON} and 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.

2.2 The V1 Model

The PC/BC model of V1 is described by the following equations:

$$E_O = X_O \oslash \left(\varepsilon_2 + \sum_{k=1}^{p} (\hat{\omega}_{ok} * Y_k)\right). \tag{2}$$

$$Y_k \leftarrow (\varepsilon_1 + Y_k) \otimes \sum_{o} (\omega_{ok} \circ E_O). \tag{3}$$

$$Y_k \leftarrow Y_k \otimes (1 + \eta A_k). \tag{4}$$

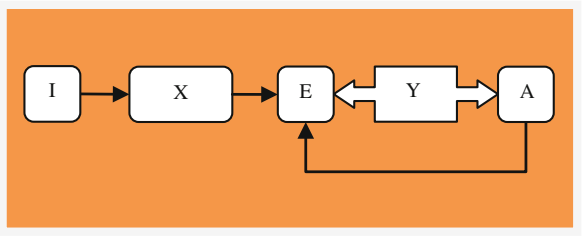

Fig. 1 The retina/LGN model and the PC/BC model of V1

where $o \in$ [ON, OFF]; X_o represents the input to the model of V1, E_o represents the error-detecting neuron responses, Y_k represents the prediction neuron responses, A_k represents the weighted sum of top-down predictions, all of them are two-dimensional array, equal in size to the input image; ω_{ok} is a two-dimensional kernel representing the synaptic weights for a particular class (k) of neuron normalized so that the sum of all the weights is equal to ψ, $\hat{\omega}_{ok}$ is a two-dimensional kernel representing the same synaptic weights as ω_{ok} but normalized so that the maximum value is equal to ψ, the Gabor function is used to define the weights of each kernel ω_{ok} and $\hat{\omega}_{ok}$ (a family of 32 Gabor functions with eight orientation (0°–157.5° in steps of 22.5°) and four phases (0°, 90°, 180°, and 270°) were used); p is the total number of kernels; $\varepsilon_1, \varepsilon_2, \eta$ and ψ are parameters; $\varnothing$ and $\otimes$ indicate element-wise division and multiplication, respectively; o represents cross-correlation (which is equivalent to convolution without the kernel being rotated 180°); and * represents convolution (which is equivalent to cross-correlation with a kernel rotated 180°). Parameter values $\psi = 5000, \varepsilon_1 = 0.0001, \varepsilon_2 = 250$, and $\eta = 1$ were used in the experiments.

Equation (2) describes the calculation of the neural activity for each population of error-detecting neurons. The activation of the error-detecting neurons can be interpreted as representing the residual error between the input and the reconstruction of the input generated by the prediction neurons. The values of E are related to the image saliency, with high error values corresponding to high saliency.

Equation (3) describes the updating of the prediction neuron activations. The values of Y_k represent predictions of the causes underlying the inputs to the model of V1. If the input remains constant, the values of Y_k will converge to steady-state values that reconstruct the input with minimum error.

2.3 Modeling the Top-Down Effects

Equation (4) describes the effects on the V1 prediction neuron activations of top-down inputs from prediction neurons at later processing stages (i.e., in extra-striate cortical regions). In Eq. (4), the effects of cortical feedback are modeled by using an array of inputs (A) to the V1 model which represents the weighted sum of top-down predictions. In the simulations of Ref. [4], feedback was either simple orientation preferences, the elements of A were set to values of 0.25 and zero, or assumed to be negligible, the elements of A were given a value of zero, in which cases Eq. (4) had no effect. We add the following equation between Eq. (3) and Eq. (4) to model the top-down effects:

$$A_k \leftarrow \sum_{k=1}^{p} (\hat{\omega}_{ok} * Y_k). \tag{5}$$

This top-down feedback will have two effects on the PC/BC model of V1. (1) Increasing the response of the prediction neurons that represent information consistent with the top-down expectation [see Eq. (4)]. This will result in these prediction neurons sending stronger feed-forward activation, and hence, make this information more conspicuous for cortical regions at subsequent stages along the processing hierarchy. (2) The enhanced activity in the prediction neurons consistent with top-down expectations will in turn decrease the response of the error-detecting neurons from which these prediction neurons receive their input [see Eq. (2)] [4]. Since the strength of the responses of the error-detecting neurons is assumed to be related to saliency, in this way, top-down feedback modulates bottom-up saliency.

3 Experimental Comparisons

3.1 Saliency Results Comparison

We evaluated our method on human visual fixation data from natural images. The dataset we used was collected by Bruce and Tsotsos [8] as the benchmark dataset for comparing human eye predictions between methods. The dataset contains eye fixation data from 20 subjects for a total of 120 natural images.

Figure 2 affords a qualitative comparison of the output of the proposed models (without/with context) for a variety of images. Visually, top-down effects increase the performance of salient object detection, i.e., top-down signals modulate bottom-up saliency. This is in line with preceding analysis. Figure 2d is fixation density map based on experimental human eye tracking data as the "ground truth" saliency map of each image.

3.2 Comparing Our Saliency Results with Other Methods

We compare our saliency method with context against other five state-of-the-art methods using the database from the publicly available database used by Achanta et al. [12]. Each of the 1,000 images in the database contains a salient object or a distinctive foreground object, so we can compare the performance of different algorithms.

The five saliency detectors are Itti et al. [7], Harel et al. [13], Hou and Zhang [14], Achanta [12], and Goferman et al. [15], hereby referred to as IT, GB, SR, IG, and CA. We refer to our proposed method as PC. The choice of these algorithms is motivated by the following reasons: citation in literature (the classic approach of IT is widely cited), recency (IG, and CA are recent), and variety (IT is biologically

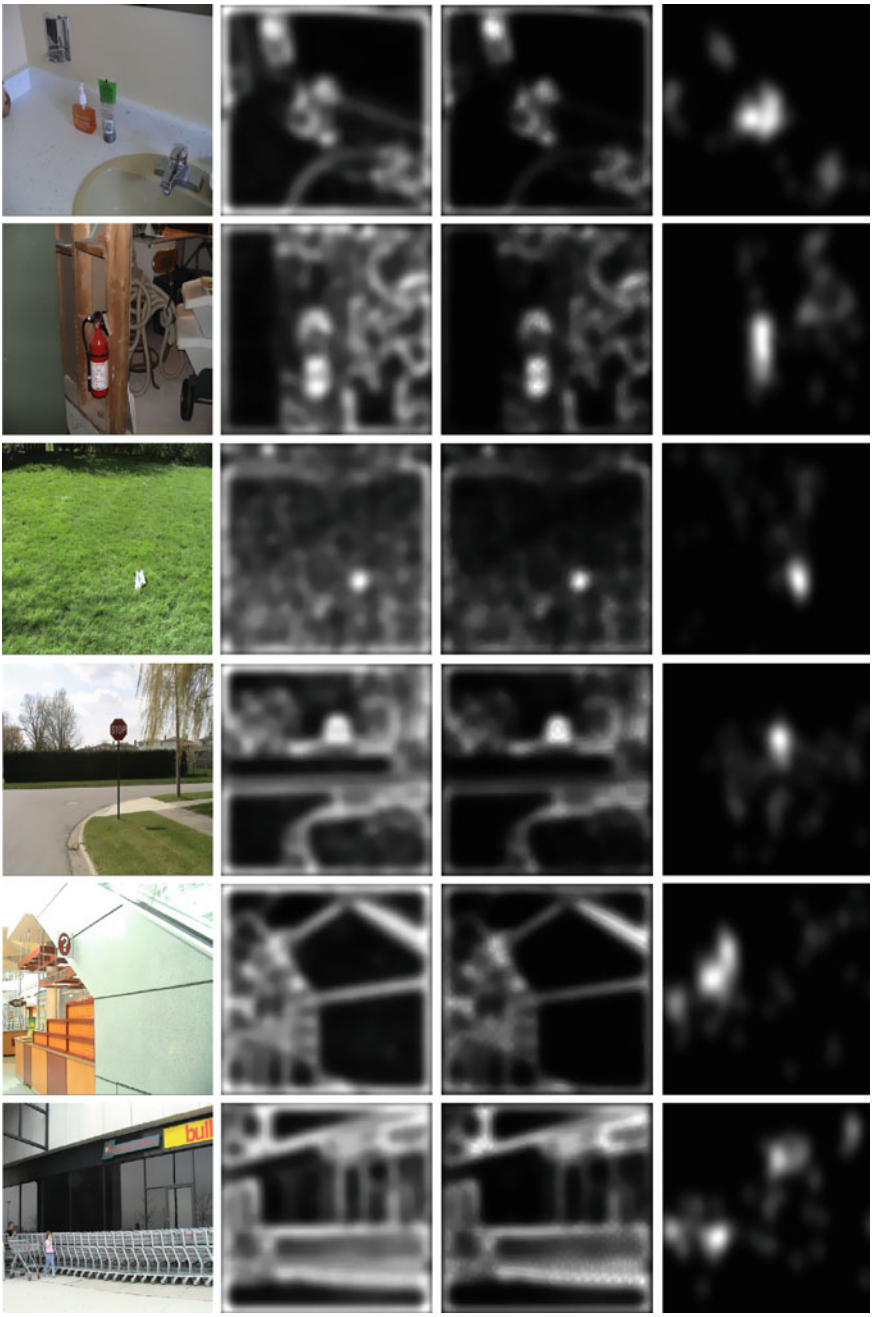

Fig. 2 Results for qualitative comparison: **a** Original image; **b** Saliency map without context; **c** Saliency map with context; **d** Fixation density map based on experimental human eye tracking data

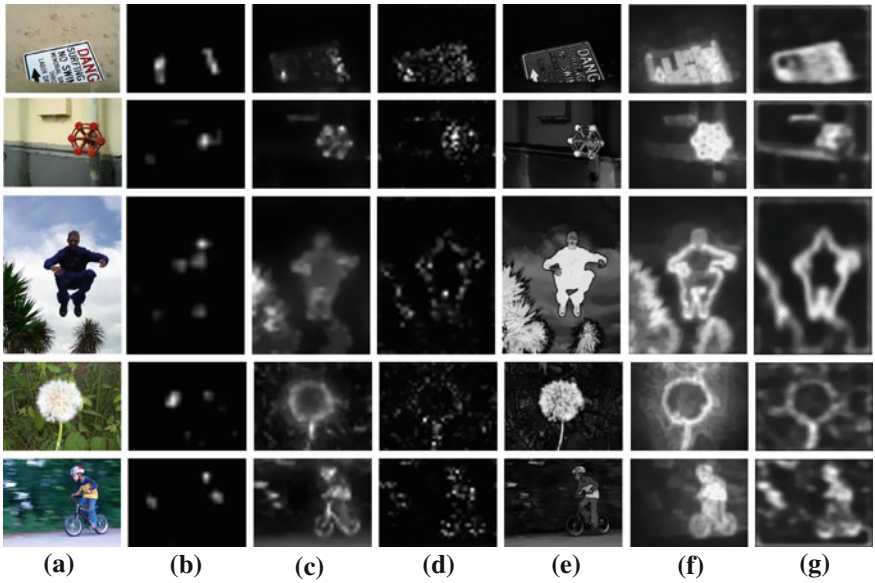

Fig. 3 Visual comparison of saliency map. **a** Original, **b** IT [7], **c** GB [13], **d** SR [14], **e** IG[12], **f** CA [15], **g** PC

motivated, CA is purely computational, GB is a hybrid approach, SR and IG estimates saliency in the frequency domain).

We randomly choose some images from the database. Figure 3 is the output of the five state-of-the-art methods and our method for comparison. Our method is a competitive, promising algorithm.

3.3 Quantitative Evaluation

To obtain a quantitative evaluation we compare ROC curves and Area Under Curve (AUC) on the database presented in [8]. Figure 4 is the result of our method and other three methods.

4 Discussions

PC/BC is a computational model of primary visual cortex (V1) which provides an implementation of the V1 bottom-up saliency map. In this paper, we propose a novel approach to natural color image saliency detection method with top-down cortical feedback as context. Our experimental result is consistent with recent literature conclusion: top-down signals modulate (override) bottom-up saliency (in

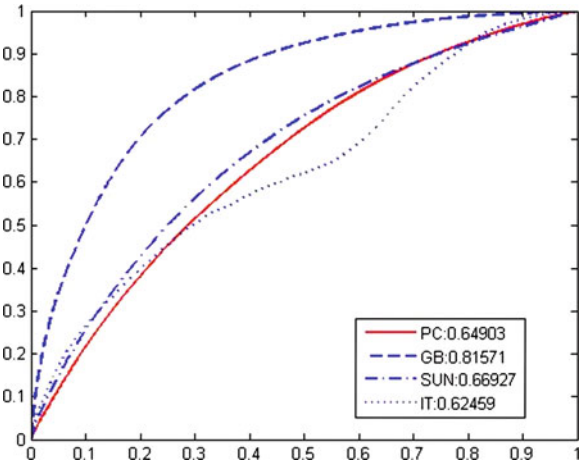

Fig. 4 ROC curves for the database of [8]

a feature-specific way) [16]. We compare our method with the five state-of-the-art models of saliency detectors. Experimental results show that our method performs competitively for visual saliency detection task.

When the organism is not actively searching for a particular target (the free-viewing condition), the organism's attention should be directed to the most salient points which potential targets in the visual field. Bottom-up attention mechanisms have been more thoroughly investigated than top-down mechanisms. One reason is that data-driven stimuli are easier to control than cognitive factors such as task-demands, knowledge, and expectations. Even less is known on the interaction between the two processes [17].

In future work, we will incorporate color feature and other task-demands features as context to detect saliency, "Combining such features-specific top-down signals with (learnt) contextual priors on target location therefore may provide a promising approach to searching for real-world objects in their natural context [16]", and develop applications of our model.

Acknowledgments This work is supported by the National Natural Science Fund of China (Grant Nos. 60632050, 9082004) and by the basic key technology project of Ministry of Industry and Information Technology of China (Grant No. E0310/1112/JC01).

References

1. Spratling MW (2008) Predictive coding as a model of biased competition in visual attention. Vis Res 48:1391–1408
2. Spratling MW (2010) Predictive coding as a model of response properties in cortical area V1. J Neurosci 30:3531–3543
3. Spratling MW (2011) A single functional model accounts for the distinct properties of suppression in cortical area V1. Vis Res 51:563–576

4. Spratling MW (2011) Predictive coding as a model of the V1 saliency map hypothesis. Neural Networks 20:1–22
5. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Ann Rev Neurosci 18:193–222
6. Treisman A, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97–136
7. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259
8. Bruce N, Tsotsos J (2005) Saliency based on information maximization. In: Proceedings in advances in neural information processing systems, vol 18, pp 155–162
9. Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW (2008) SUN: a Bayesian framework for saliency using natural statistics. J Vis 8:1–20
10. Hou X, Zhang L (2008) Dynamic visual attention: searching for coding length increments. In: Proceedings of advances in neural information processing systems, pp 681–688
11. Borji A, Itti L (2012) State-of-the-art in visual attention modeling. http://doi.ieeecomputersociety.org/10.1109/TPAMI.2012.89
12. Achanta R, Hemami S, Estrada F, Süsstrunk S (2009) Frequency-tuned salient region detection. In: Proceedings of IEEE international conference on computer vision and pattern recognition, pp 1597–1604
13. Harel J, Koch C, Perona P (2007) Graph-based visual saliency. Adv Neural Inf Proc Syst 19:545–552
14. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 18–23
15. Goferman S, Zelnik-Manor L, Tal A (2010) Context-aware saliency detection. In: Proceedings of IEEE international conference on computer vision and pattern recognition, pp 2376–2383
16. Einhäuser W, Rutishauser U, Koch C (2008) Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. J Vis 8:1–19
17. Frintrop S, Rome E, Christensen HI (2010) Computational visual attention systems and their cognitive foundations: a survey. ACM Trans Appl Percept 7:1–46