Measuring the attentional effect of the bottom-up saliency map of natural images

Cheng Chen[1,3], Xilin Zhang[2,3], Yizhou Wang[1,3], and Fang Fang[2,3,4,5]

[1]National Engineering Lab for Video Technology and School of Electronics
Engineering and Computer Science, [2]Department of Psychology
[3]Key Laboratory of Machine Perception (Ministry of Education)
[4]Peking-Tsinghua Center for Life Sciences
[5]IDG/McGovern Institute for Brain Research
Peking University, Beijing 10081, P.R. China
`chencheng880829@gmail.com`
`{zhangxilin,Yizhou.Wang,ffang}@pku.edu.cn`

Abstract. A saliency map is the bottom-up contribution to the deployment of exogenous attention. It, as well as its underlying neural mechanism, is hard to identify because of the existence of top-down signals. In order to exclude the contamination of top-down signals, invisible natural images were used as our stimuli to guide attention. The saliency map of natural images was calculated according to the model developed by Itti *et al.* [1]. We found a salient region in natural images could attract attention to improve subjects' orientation discrimination performance at the salient region. Furthermore, the attraction of attention increased with the degree of saliency. Our findings suggest that the bottom-up saliency map of a natural image could be generated at a very early stage of visual processing.

Keywords: Bottom-up saliency map, Natural image, Visual Attention, Unconsciousness

1 Introduction

Because of the limited resources of the visual system, visual attention is essential for us to select the most valuable information from extremely complex natural scenes, and thus plays an important role in understanding the world. The information selection process can be achieved by directing visual attention to a target under a top-down goal, or be triggered by a salient stimulus. The former process is executed voluntarily, while the latter process is automatic and guided by the saliency map. Relative to extensive studies on the neural basis of top-down selection, the neural basis of bottom-up saliency map is controversial because of the possible contamination by top-down signals in higher brain areas.

In this paper, we measured the attentional effect of the bottom-up saliency map of natural images. Low luminance natural images were presented very briefly, which rendered them invisible to subjects and also excluded the contamination by top-down signals. Natural images were used here instead of simple textures because of their rich naturalistic low-level features that the human visual system is tuned to. Although these natural images were invisible, the difference of visual saliency (calculated by a famous computational saliency model [1]) between inside and outside a local region could attract attention to improve the performance of an orientation discrimination performance at the salient region. We use the degree of saliency to refer to the saliency difference in the paper. Then we could measure the attentional effect generated by different degrees of saliency. Investigating this topic not only provide evidence for the bottom-up saliency map in our brain, but also is helpful to many important applications, such as object detection.

1.1 Related work

The representation of the strength of the bottom-up attention attraction from our visual input [2] is a saliency map. It is constructed in our brain and can direct our attention along with top-down signals. Several studies had tried to measure the effect of visual saliency, and also found brain regions that realize the saliency map. For example, Geng and Mangun found that anterior intraparietal sulcus could realize the saliency map [3]. Mazer and Gallent found a goal-related activity in V4, which provided evidence that V4 could realize the saliency map [4]. However, these studies can not rule out the top-down attentional control, which makes it hard to identify the neural basis of the bottom-up saliency map. So it's important to probe the bottom-up attention attraction free from the top-down influence.

Several methods can be used to reduce the top-down signals influence, such as backward masking, binocular rivalry and continuous flash suppress (CFS). Zhang *et al*. had adopted the backward masking method to investigate the neural substrate of the bottom-up saliency map [5]. In their study, stimuli were presented so briefly and followed by a high contrast mask so that subjects could not perceive the stimuli. Similarly, we also used backward masking to make low-luminance stimuli invisible. Stimuli in our experiment were natural images collected from the Internet instead of simple pattern or texture, for natural images contain multiple and naturalistic low-level features [6]. Consider that some studies had used checkerboard to mask objects [7], random checkerboard was used as mask in our experiment.

In our experiment, we adopt a revised version of the cueing effect paradigm proposed by Posner *et al*. [8]. In this paradigm, a target appears in one of two locations randomly, and subjects need to finish a discrimination task about this target. Prior to this target, a cue indicates the location of the following target. Trials with a correct cue are called valid cue trials, while trials with an incorrect cue are called invalid cue trials. A classical result demonstrates that performance (response time or accuracy) in the valid cue trials is significantly higher than that in the invalid cue trials [9]. The salient region of a natural image was used as a cue in our experiment.

Many studies had also proposed a computational model to generate the saliency map of an image. An example can be seen in Fig. 1, the value of each

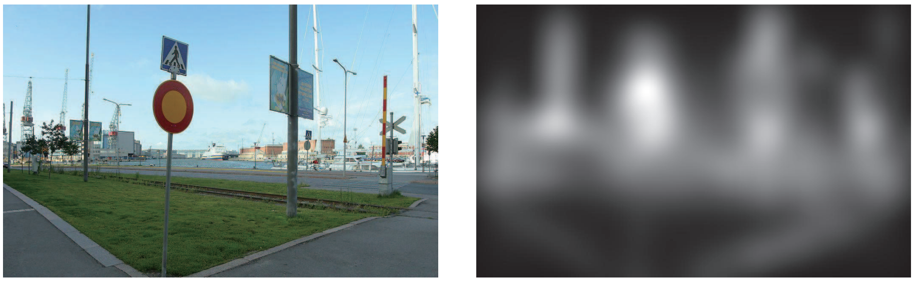


Fig. 1. An example of a color image (left) and its saliency map (right). White region in the right image indicate its salient region.

pixel in the saliency map ranges from 0 to 1, higher value correlated with more saliency. Itti *et al.* proposed a biologically-plausible saliency model based on a center-surround mechanism, by combining information from three channels: color, intensity and orientation [1]. According to the spectrum of natural images, Hou *et al.* compute the spectral residual of an input image and transform the spectral residual to spatial domain to obtain its saliency map [10]. By simulating the information transmitting between neurons, Wang *et al.* proposed a saliency model based on information maximization [11]. These saliency models can provide a prediction about the attentional effect of a bottom-up saliency map.

Moreover, the underlying neural mechanism of the bottom-up saliency map has been subject to debate. A dominant view assumes that saliency results from pooling different visual features (e.g. [2], [12]), thus could be realized by higher cortical areas such as parietal cortex. However, Li proposed the V1 theory which claimed the saliency map was created by V1 (e.g. [13], [14]). It was completed via intra-cortical interactions that are manifest in contextual influences [15]. By combing psychophysical data and brain imaging results, Zhang *et al.* found that neural activities in V1 could create a bottom-up saliency map of simple texture [5], which supported the V1 theory. But evidence on natural images is still lack.

The rest of this paper is organized as follows. In Section 2 we introduce the details of our approach, including the information of subjects, the stimuli and the procedure of psychophysical experiment. The results of our experiment are given in Section 3. Finally, we conclude and discuss our work in Section 4.

2 Our Approach

2.1 Subjects

16 human subjects (7 male and 9 female) participated in the psychophysical experiments. All subjects were right-handed, reported normal vision or corrected-to-normal vision, and had no known neurological or visual disorders. Ages ranged from 19 to 26. All of them were naive to the purpose of our study except for

one subject who was one of the authors. They were given written, informed consent in accordance with the procedures and protocols approved by the human subjects review committee of Peking University.

2.2 Stimuli

We collected a large number of grayscale images about natural scenes from the Internet, resized them into the same size (384×1024 pixels), and decreased the luminance of these images to a low level (about 2.9 cd/m^2), Fig. 2 (a) shows a sample image. To quantitatively measure the attentional effect, we adopt a visual saliency model proposed by Itti *et al.* [1] and calculate the saliency map of each image. After that we selected 50 images, and each of them had a round salient region centered at about 7.2° eccentricity in the lower left quadrant(called left-salient images). The diameter of the salient region was about 4°. By flipping each image across its vertical midline, we can generate 50 new images, each of them had a local salient region in the lower right quadrant(called right-salient images). Notice that the content between the two groups of images were totally the same, the only difference between the two groups was the location of the salient region. The average saliency map of the 50 left-salient images can be seen in Fig. 2 (b).

Based on the bottom-up saliency, we classified all images into two groups: high salient images and low salient images. We proposed a salient index to measure the degree of saliency based on the following formulation:

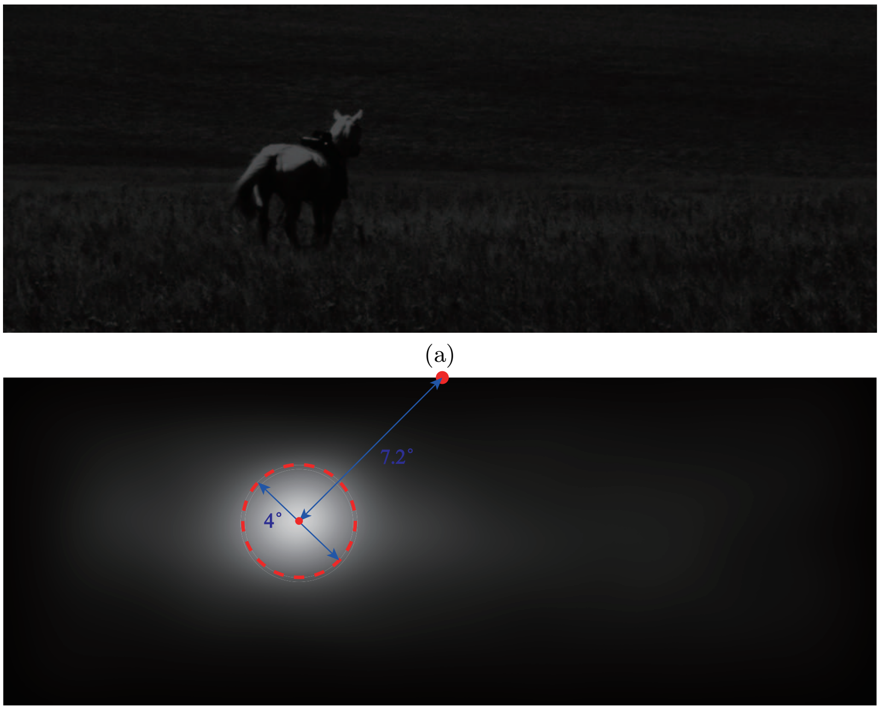
$$Index(n) = \frac{S_I(n) - S_O(n)}{S_O(n)} \ . \tag{1}$$

In the above formulation, n denoted the index of an image. For left-salient images, S_I denoted the averaged saliency value of the round region in Fig. 2 (b), and S_O denoted the averaged saliency value of the residual region. The higher $Index$ value indicated the higher saliency. We selected half of images with a higher $Index$ in left-salient images as the high salient images, and selected the other half as the low salient images. The same manipulation was adopted on right-salient images. Thus, stimuli used for psychophysical experiment had two groups: high salient and low salient groups. Each group contained 50 images, half of them were left-salient and the other were right-salient.

Mask stimuli were high contrast checkerboards that randomly arranged (see Fig. 3), the size of each checker was about 0.25° × 0.25°. The luminance of a black checker was 1.8 cd/m^2, while the luminance of a white checker was 79 cd/m^2.

2.3 Psychophysical experiment

In the psychophysical experiment, all stimuli were displayed on a Gamma-corrected Iiyama HM204DT 22 inches monitor, with a spatial resolution of 1024 × 768 and a refresh rate of 60Hz. The viewing distance was 83 cm, and



(a)

(b)

Fig. 2. (a) A sample of a low-luminance image used as our stimulus. (b) The averaged saliency map of left-salient images, a circle-like local salient region can be seen on this map.

their head position was stabilized using a chin rest and a head rest. A white fixation cross was always present at the center of the monitor, and subjects were asked to fixate the cross throughout the experiment.

We adopt a modified version of the cueing effect paradigm proposed by Posner to measure the attentional effect of the visual saliency of invisible natural images. Each trial started with a fixation. A low-luminance (2.9 cd/m^2) image was presented on the lower half of the screen for 50 ms, followed by a 100ms mask at the same position, and another 50ms fixation interval. The bottom-up saliency map of the image served as a cue to attract spatial attention, and the mask could ensure that the image was invisible to subjects. Then a grating orientated at about ± 1.5° which centered at about 7.2° eccentricity from the fixation was presented randomly at either the lower left quadrant or lower right quadrant with equal probability for 50 ms. The location of the grating was either at or symmetric with the salient region of the previous image, thus indicated the valid cue condition or the invalid cue condition. The grating had a spatial frequency of 5.5 cpd (cycle per degree) and its diameter was 2.5 with full contrast. Subjects

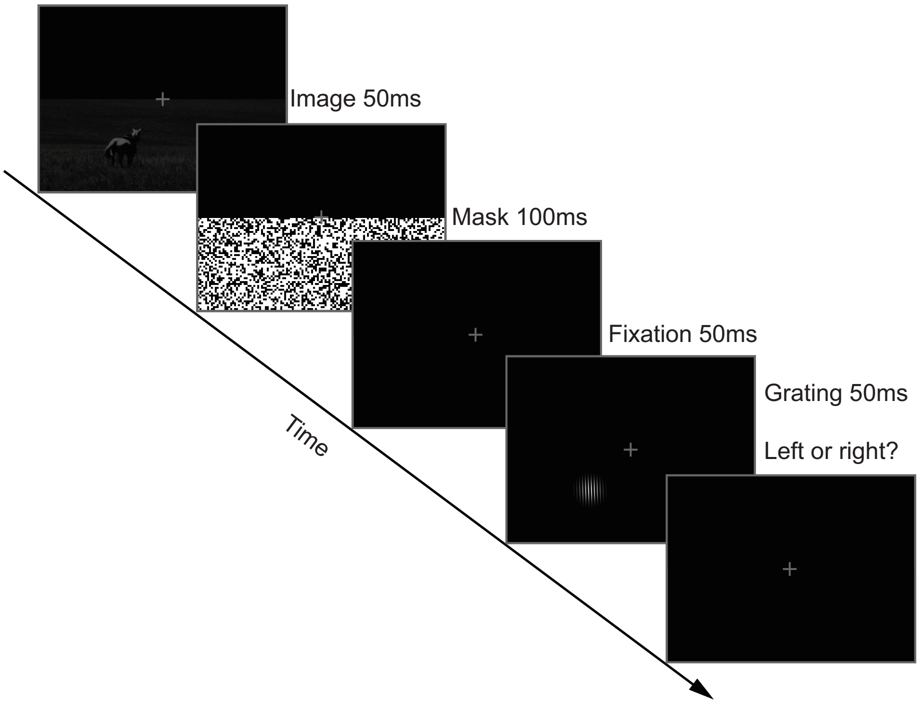


Fig. 3. The procedure of our experiment.

were asked to press one of the two keys to indicate the orientation of the grating. The duration of each trial was 2s, Fig. 3 shows the procedure of our experiment.

The experiment consisted of 10 blocks. Each block contained 100 trials with two conditions: high salient condition and low salient condition. Images for the first condition were selected randomly from the high salient group, and images for the second condition were selected randomly from the low salient group. The attentional effect of bottom-up saliency maps of invisible images for each condition was measured by the difference between the performance of the valid cue condition and invalid cue condition in the grating orientation discrimination task (see Section 3.2 for details).

Moreover, in order to determine whether the image was indeed invisible, subjects were asked to complete a two-alternative forced choice (2AFC) experiment in a criterion-free way before the attentional effect experiment. Each trial began with either a low-luminance image or a blank, followed by a mask. Subjects were asked to make a forced choice response to judge whether there was an image presented before the mask. The performance at chance level in this experiment could provide an objective confirmation that the masked images were indeed invisible.

3 Experimental Result

3.1 Images Invisibility

The purpose of the 2AFC experiment was to evaluate whether those natural images used as the cue in the attentional experiment were indeed invisible. High salient images and low salient images were counterbalanced in this task. Subjects had to report whether they can see an image before the mask (details can be found in Section 2.3.

We found that percentages of correct detection (mean ± std) were 48.6 ± 6.0% and 50.9 ± 5.7% for high salient and low salient images respectively. Paired t-test results showed that the percentages of correct detection were statistically indistinguishable from the chance level for both high salient and low salient images(paired t-test: high salient images: $t_{15} = -0.934,\ p = 0.365$; low salient images: $t_{15} = 0.6324,\ p = 0.537$; significant level $\alpha = 0.5$), indicated that natural images in both groups were indeed invisible for subjects in our experiment.

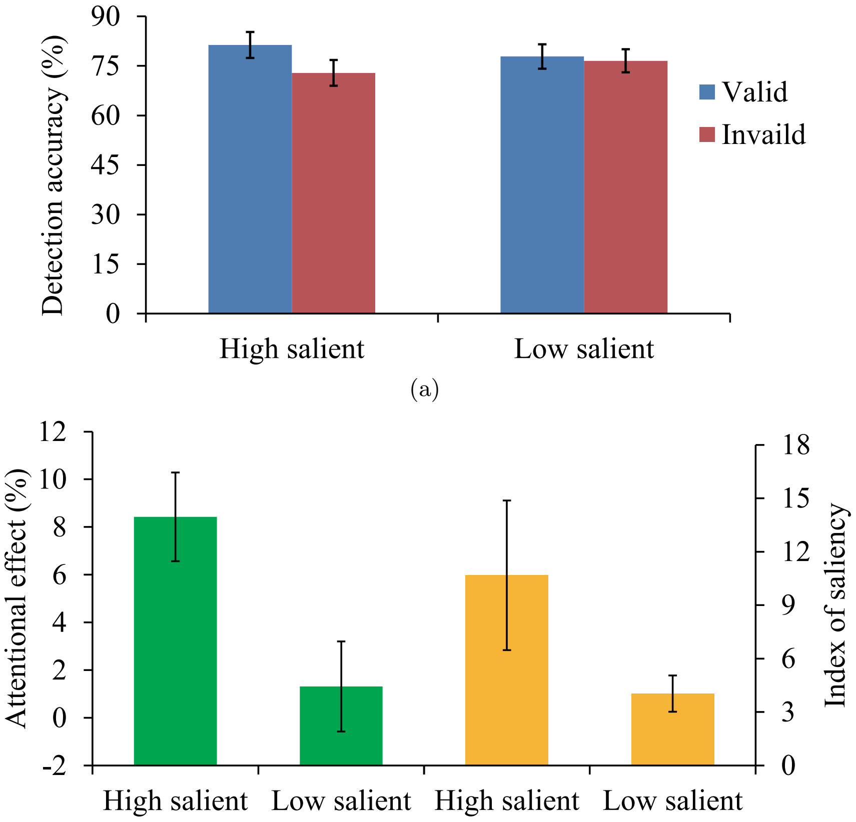
3.2 Attentional Effect

The attentional effect of bottom-up saliency maps of invisible images was measured by the difference between the accuracy of grating orientation discrimination performance in the valid cue condition, and that in the invalid cue condition. The grating appeared at randomly either the same location with the salient region of an image (valid cue condition) or its contralateral counterpart (invalid cue condition) with equal probability.

We found that the discrimination accuracy was higher in the valid cue condition than that in the invalid cue condition (see Fig. 4 (a)), for both high salient images (Valid: 81.31 ± 3.93%; Invalid 72.88 ± 3.92%) and low salient images (Valid: 77.86 ± 3.7%; Invalid 76.54 ± 3.52%). The results indicated that the bottom-up saliency map exhibited a positive cueing effect even when the image was invisible, which suggested that subjects' attention was attracted to the salient region of an invisible image, so that they performed better in the valid cue condition than in the invalid cue condition.

Moreover, we measured the attentional effect of bottom-up saliency maps for both high salient and low salient images(see Fig. 4 (b), the left two green bars), the results suggested that the attentional effect of high salient images (8.43 ± 1.32%) and that of low salient images (1.486 ± 1.89%) were both significantly higher than zero(high salient:$t_{15} = 18.126,\ p < 0.001$; low salient: $t_{15} = 2.782,\ p = 0.014$; significant level $\alpha = 0.05$). The attentional effect of high salient images was significantly higher than that of low salient images($t_{15} = 9.665,\ p < 0.001$).

We also calculated the proposed index of the high salient and the low salient images(high salient: 10.67 ± 4.20%; low salient: 4.03 ± 1.02%), the index predicted the degree of the attention attraction of a bottom-up saliency map(see Fig. 4 (b), the right two yellow bars). Psychophysical data were consistent with the prediction from the computational model.



(a)

(b)

Fig. 4. Results of our experiment. (a) The performance of the grating orientation discrimination task for high salient images and low salient images. (b) The left two green bars indicate the attentional effect of bottom-up saliency maps in high salient and low salient groups. The right two yellow bars indicate the predication of the attentional effect in two groups.

4 Conclusion and Discussion

In this paper, we proposed a method to measure the attentional effect of bottom-up saliency maps. By using backward masking, we could eliminate the contamination of top-down signals. We selected natural images which had a local round salient region and found that even those natural images were invisible, the salient region could attract attention to improve the orientation discrimination performance on a grating in the cueing effect paradigm. Furthermore, we found that the attraction of attention increased with the degree of saliency.

In our experiment, we assume that the absence of awareness to the whole image could maximally reduced top-down signals, even if it did not completely abolish them [5]. These top-down signals may include feature and object perception, as well as subjects' intentions [16]. Compared to previous studies, such manipulation could help us observe the attentional effect based on a relatively pure bottom-up saliency signal. Our findings may suggest that the bottom-up saliency map of a natural image could be generated at a very early stage of visual processing.

In the future, we will extend our study to find the neural substrate of bottom-up saliency maps of natural images. Moreover, consider it's difficult to modulate the degree of saliency on the same content, we will also extend our work on synthesized textures so that we could quantitatively change the degree of saliency on one image.

5 Acknowledgement

This work was supported by the Ministry of Science and Technology of China (2011CBA00400 and 2009CB320904) and the National Natural Science Foundation of China (Project 30925014, 31230029 and 90920012).

References

1. Itti, L., Koch, C., Niebur, E.: A model of saliency based visual attention for rapid scene analysis. IEEE Trans. Patt. Anal. Mach. Intell. 20, 1254–1259 (1998)
2. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. 4, 219–227 (1985)
3. Geng, J.J.,Mangun, G.R. Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. J. Cogn. Neurosci. 21, 1584C-1601 (2008)
4. Mazer, J.A., Gallent, J.L. Goal-related activity in V4 during free viewing visual search: evidence for a ventral stream visual salience map. Neuron. 40, 1241–1250 (2003)
5. Zhang, X., Zhaoping, L., Zhou, T., Fang, F.: Neural activities in V1 create a bottom-up saliency map. Neuron. San Francisco. 73, 183–192 (2012)
6. Bogler, C., Bode, S., Haynes, J. D. Decoding Successive Computational Stages of Saliency Processing. Curr. Biol. 21, 1667-1671 (2011)
7. Fang, F., He, S. Cortical responses to invisible objects in human dorsal and ventral pathways. Nature Neuroscience. 8, 1380–1385 (2005)

8. Posner, M.I., Snyder, C.R.R., Davidson, B.J. Attention and the detection of signals. J. Exp. Psychol. 109, 160C-174 (1980)
9. Eckstein, M.P., Schimozaki, S.S., Abbey, C.K. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. J. Vis. 2, 25–45 (2002)
10. Hou, X., Zhang, L. Saliency detection: A spectral residual approach. IEEE Computer Vision and Pattern Recognition. (2007)
11. Wang W., Wang, Y., Huang, Q., Gao, W. Measuing visual saliency by site entropy rate. IEEE Computer Vision and Pattern Recognition. (2010)
12. Itti, L., Koch, C., Niebur, E.: Computational Modelling of Visual Attention. Nat. Rev. Neurosci. 2, 194–203 (2001)
13. Li, Z.: Contextual influences in V1 as a basis for pop out and asymmetry in visual search. Proc. Natl. Acad. Sci. USA. 96, 10530–10535 (1999)
14. Li, Z.: A saliency map in primary visual cortex. Morgan Kaufmann. Trends Cogn. Sci. 6, 9–16 (2002)
15. Allman, J., Miezin, F., McGuinness, E.: Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for localCglobal comparisons in visual neurons. Annu. Rev. Neurosci. 8, 407–430 (1985)
16. Jiang, Y., Costello, P., Fang, F., Huang, M., He, S. A gender-and sexual orientation-dependent spatial attentional effect of invisible images. Proc. Natl. Acad. Sci. USA. 103, 17048C-17052 (2006)