

# Saliency Detection Based on Background and Foreground Modeling

Zhengbing Wang, Guili Xu<sup>(✉)</sup>, Yuehua Cheng, and Zhengsheng Wang

College of Automation Engineering,  
Nanjing University of Aeronautics and Astronautics, Nanjing, China  
guilixu@nuaa.edu.cn

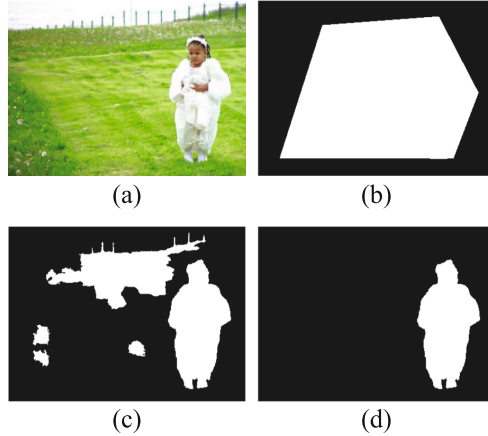
**Abstract.** In this paper, a novel saliency detection algorithm is proposed to fuse both the background and foreground information while detecting salient objects in complex scenes. Firstly, we extract background seeds as well as their spatial information from image borders to construct a background-based saliency map. Then, an optimal contour closure is selected as the foreground region according to the first-stage saliency map. The optimal contour closure can provide a preferable description for salient object. We compute a foreground-based saliency map using the selected foreground region and integrate it with the background-based one. Finally, the unified saliency map is further refined to obtain a more accurate result. Experimental results show that the proposed algorithm can achieve favorable performance compared to the state-of-the-art ones.

**Keywords:** Saliency detection · Background · Foreground  
Optimal contour closure

## 1 Introduction

Saliency detection aims at highlighting the most attractive regions in a scene. It has been further studied in recent years, and numerous computational models have been presented. As a preprocessing operation, saliency detection can benefit many other tasks, including image segmentation [9, 14], image compression [4], object localization and recognition [3].

Saliency detection algorithms can be roughly divided into two categories from the perspective of information processing. The top-down approaches [13, 22] driven by specific tasks need to learn the visual information of specific objects to form the saliency maps. In contrast, the bottom-up methods [5, 12, 19, 20] usually exploit low-level cues such as color, lamination and texture to highlight salient objects. Early researches address saliency detection via heuristic principles [12], including contrast prior, center prior and background prior. Most works based on these principles exploit low-level features directly extracted from images [5, 19]. They perform well in many cases, but are still unfavorable in complex scenes. Due to the shortcomings of low-level features, many algorithms are presented



**Fig. 1.** Examples of foreground regions. (a) Input image, Foreground region used in (b) XIE [20], (c) BFS [17], (d) Our method.

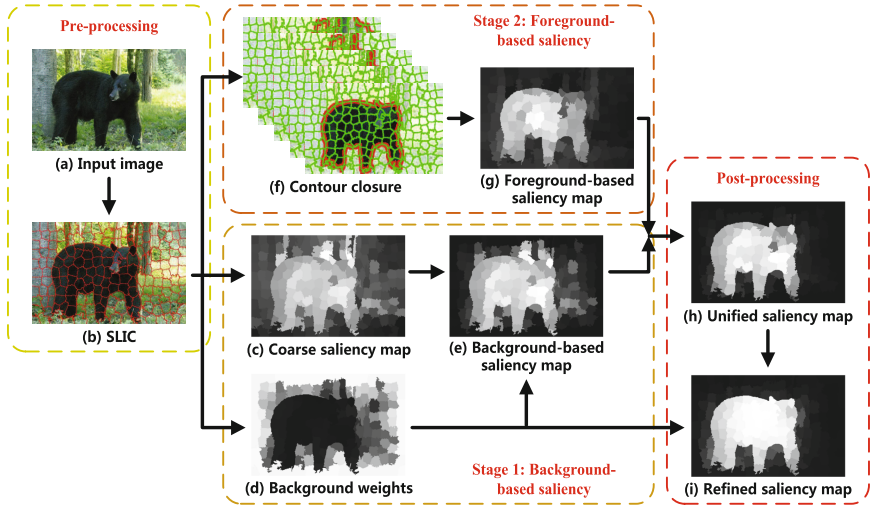
to incorporate high-level features in saliency detection. Xie et al. [20] propose a bottom-up approach which integrates both low and mid level cues using the Bayesian framework. Some learning methods [6, 16] are also presented to integrate both low and high level features to compute saliency based on parameters trained from sample images.

Recently, to achieve better performance, some object-level cues are introduced as hints of the foreground. Some examples are shown in Fig. 1. Xie et al. [20] detect the salient points in the image and a convex hull is computed to denote the approximate location of the salient object. Wang et al. [17] binarize the coarse saliency map using an adaptive threshold and select the super-pixels whose saliency values are larger than the threshold as foreground seeds. While the extracted foreground information can be used to improve performance of saliency detection, the false foreground region may have unfavorable influence.

In this paper, we propose an effective method to incorporate foreground information in saliency detection. First, we extract background seeds and their spatial information to construct a background-based saliency map. Then, several compact regions are generated using the contour information. We select the optimal one as the foreground region and calculate the foreground-based saliency map accordingly. To achieve better performance, two saliency maps are finally integrated and further refined.

## 2 Saliency Detection Algorithm

This section explains the details of the proposed saliency detection algorithm. In order to preserve the structural information, we over-segment the input image to generate  $N$  super-pixels [2] and use them as the minimum units. After that, a background-based saliency map is firstly constructed using the background



**Fig. 2.** Pipeline of our method, including input image pre-processing, background-based saliency, foreground-based saliency and post-processing.

information (Subsect. 2.1). We then select the optimal contour closure as the foreground region according to the first-stage saliency map and compute the foreground-based saliency map (Subsect. 2.2). Finally, these two saliency maps are integrated and further refined to form a more accurate result (Subsect. 2.3). The pipeline of our saliency detection method is illustrated in Fig. 2.

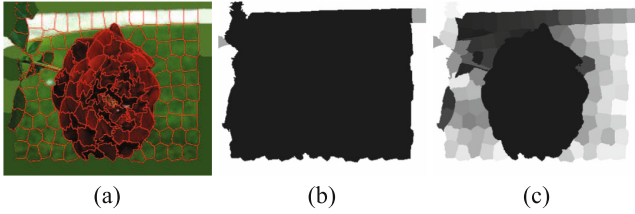
## 2.1 Saliency Detection via Background Information

Border regions of the image have been proved to be good visual cues for background priors in saliency detection [19]. Observing that background areas are usually connected to the image borders, we select the super-pixels along the image borders as background seeds and define the coarse saliency of super-pixels as their color contrast to the background ones. Denote the background seeds set as  $BG$ , and the coarse saliency value of super-pixel  $s_i$  is computed as

$$S_i^c = \sum_{s_j \in BG} d_c(s_i, s_j) * w_l(s_i, s_j) \quad (1)$$

where  $d_c(s_i, s_j)$  is the Euclidean color distance between two super-pixels and  $w_l(s_i, s_j)$  denotes the spatial weight.

As shown in Fig. 2(c), the coarse saliency map may include a large amount of background noises and is visually unfavorable. Therefore, we further consider the spatial information of the selected background seeds to define the background weight for each super-pixel, which can be used to restrain undesirable noises. The process of computing background weight is given as follows: First, we cluster the super-pixels in  $BG$  into  $K$  clusters using K-means clustering algorithm.



**Fig. 3.** Definition of background weights. (a) Background seeds clustering; (b) Background weights of the selected background seeds; (c) Background weights of the other super-pixels.

The number of clusters  $K$  is set to 3 in our experiments as shown in Fig. 3(a). For each cluster  $k$ , determine the shortest continuous super-pixel link  $SL_k$ , which contains all the super-pixels belonging to cluster  $k$ . Denote the length of this super-pixel link as  $L_s$ , and the background weight for cluster  $k$  can be calculated as

$$P_k = 1 - \exp(-\alpha(L_s + L_o)) \quad (k = 1, 2, \dots, K) \quad (2)$$

where  $L_o$  is the number of super-pixels belonging to the other clusters in  $SL_k$ . As shown in Fig. 3(b), for each super-pixel  $s_j$  in cluster  $k$ , we assign the same value  $P_k$  to its background weight  $p_{s_j}$ . The background weights of the remainder super-pixels are determined as

$$p_{s_i} = \frac{p_{s_j^*}}{d_{geo}^*} \quad (s_i \notin BG) \quad (3)$$

where  $d_{geo}^*$  is the shortest geodesic distance from super-pixel  $s_i$  to the background seeds and  $s_j^*$  is the corresponding seed in  $BG$ .

The background-based saliency value of super-pixel  $s_i$  is finally calculated as

$$S_i^b = S_i^c * (1 - p_{s_i}) \quad (4)$$

As shown in Fig. 2(e), the background-based saliency map can be substantially improved by considering the spatial information of background seeds. However, some background regions with discriminative appearance are still incorrectly highlighted. The foreground information is therefore incorporated to suppress the background noises.

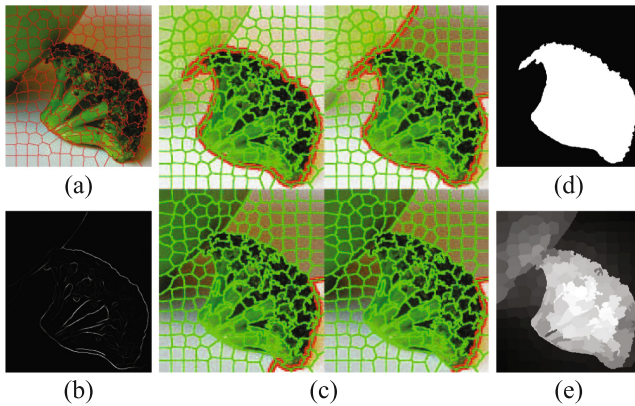
## 2.2 Saliency Detection via Optimal Contour Closure

The background-based saliency map can highlighted all the regions with high contrast to the background seeds but may be invalid for background noises. Some recent works [17, 18, 20] incorporate foreground information to restrain noises. However, the false foreground information may have unfavorable influence on saliency detection. According to the research of visual psychology [15], compact regions grouped by contour information can provide important cues for selective

attention. We adopt Levinshtein et al.'s mechanism [10] to generate foreground regions. Given the contour image and the assumption that the salient contours that define the boundary of the object will align well with super-pixel boundaries, we obtain several contour closures by solving a parametric maxflow problem as shown in Fig. 4(c). We select the optimal contour closure as

$$\mathbf{x}^* = \arg \min_{\mathbf{x}^m} \sum_{i=1}^N |\mathbf{x}_i^m - S_i^b| + V(\mathbf{x}^m) \quad (m \leq M) \quad (5)$$

where  $\mathbf{x}^m$  is a binary mask, which denotes the  $m$ -th foreground region (contour closure) and  $M$  is the number of previously obtained contour closures.  $V(\mathbf{x}^m)$  denotes the spatial variance of a foreground region.



**Fig. 4.** Foreground-based saliency detection. (a) Super-pixels of input image; (b) Salient contours; (c) Examples of obtained contour closures; (d) Optimal contour closure; (e) Foreground-based saliency map.

The selected optimal contour closure is shown in Fig. 4(d), and we collect all the super-pixels in this contour closure to compose the foreground seeds set  $FG$ . The foreground-based saliency value of each super-pixel is computed as

$$S_i^f = \sum_{s_j \in FG} \frac{1}{d_c(s_i, s_j) + \beta d_l(s_i, s_j)} \quad (6)$$

where  $d_l(s_i, s_j)$  is the spatial distance between two super-pixels.

### 2.3 Integration and Refinement Operation

Referring to [17], the background-based saliency map can uniformly highlight the salient object while the foreground-based one can well restrain the background

noises. In order to take advantage of both the background-based saliency and the foreground-based one, we integrate two saliency maps as

$$S_i^u = S_i^b * (1 - \exp(-\theta * S_i^f)) \tag{7}$$

where  $\theta$  is set to 4 in our experiments.

To obtain a better result, we further refine the unified saliency map by the energy function presented in [23]. The used energy function can not only assign large saliency value to foreground region but promote the smoothness of refined saliency map. The energy function is given as

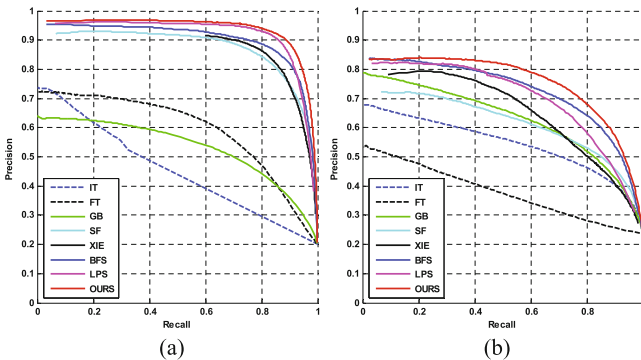
$$S^r = \arg \min_{\mathbf{S}} \left( \sum_{i,j=1}^N w_c(s_i, s_j) (S_i - S_j)^2 + \sum_{i=1}^N p_{s_i} S_i^2 + \sum_{i=1}^N S_i^u (S_i - 1)^2 \right) \tag{8}$$

where  $w_c(s_i, s_j)$  denotes the color similarity between two adjacent super-pixels and  $p_{s_i}$  is the background weight of super-pixel  $s_i$  obtained in Subsect. 2.1.  $\mathbf{S}^r = [S_1^r, S_2^r, \dots, S_N^r]^T$  denotes the refined saliency value vector.

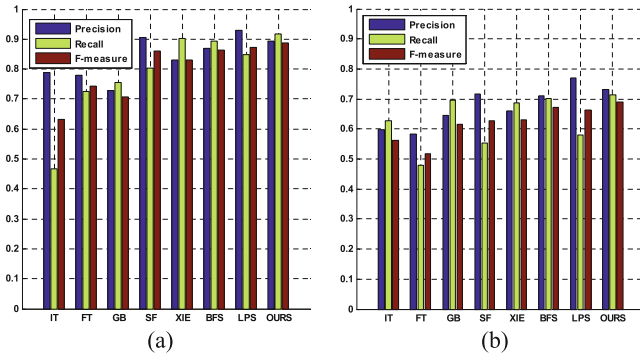
### 3 Experiments

In this section, we evaluate our algorithm on two public datasets: ASD [1] and ECSSD [21]. Both of them consist of 1000 images with pixel-wise labeled ground truth, while the ECSSD dataset is more challenging as many images contain more complex scenes. We compare our algorithm with 7 state-of-the-art methods, including IT [5], FT [1], GB [7], SF [8], XIE [20], BFS [17], and LPS [11].

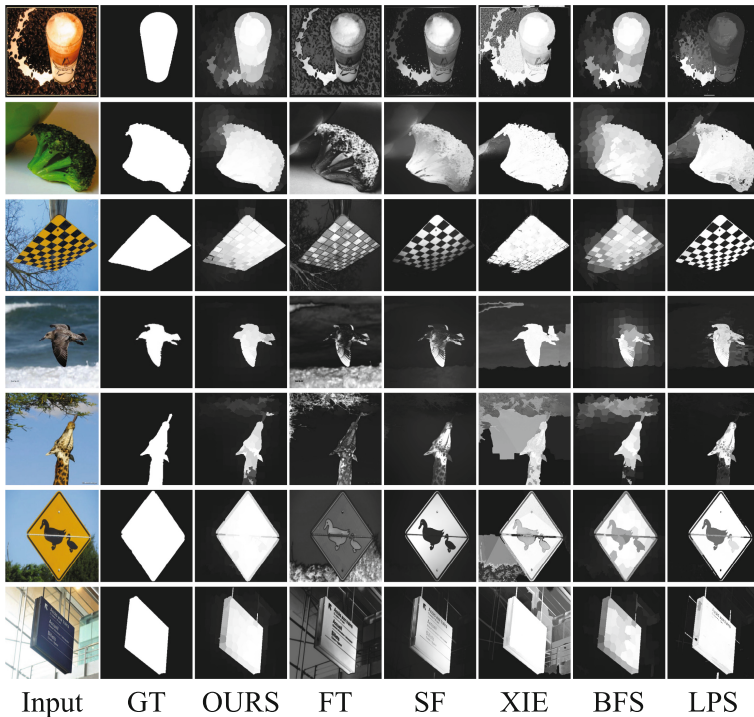
To make a fair comparison, the precision-recall curve and F-measure are used for quantitative analysis. Given a saliency map, we segment it with the



**Fig. 5.** Precision-recall curves of compared methods on (a) MSRA dataset, (b) ECSSD dataset.



**Fig. 6.** Average precision, recall and F-measure of compared methods on (a) MSRA dataset, (b) ECSSD dataset.

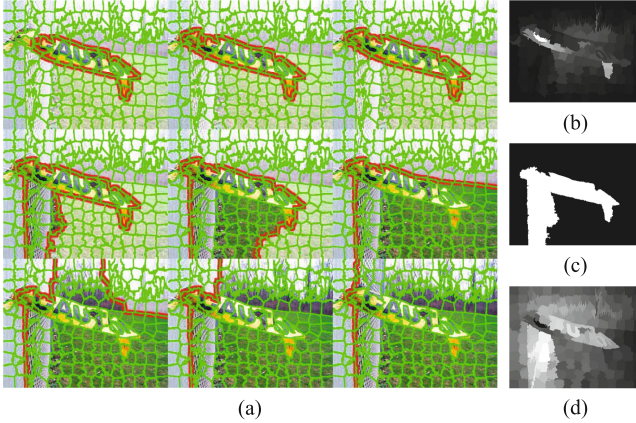


**Fig. 7.** Visual comparisons of our algorithm and 5 state-of-the-art methods.

thresholds ranging from 0 to 255, and compare each result with ground truth to generate the precision-recall curve. The precision-recall curves of compared methods are shown in Fig. 5, which demonstrates that our result performs better than others. To compute the F-measure, we first over-segment the original image using the mean-shift algorithm. A binary map can be obtained by a threshold,

**Table 1.** Average values of precision and recall for ASD and ECSSD

Methods	BFS		OURS	
Datasets	ASD	ECSSD	ASD	ECSSD
$p_F$	0.8197	0.6197	0.8715	0.7136
$r_F$	0.8726	0.7010	0.8944	0.6241

**Fig. 8.** Failure case of foreground region selection. (a) All the generated contour closures; (b) Background-based saliency map; (c) Selected foreground region; (d) Foreground-based saliency map.

which is set to twice the mean saliency value. For each binary map, we compute the F-measure as

$$F - measure = \frac{(1 + \gamma^2) Precision \times Recall}{\gamma^2 Precision + Recall} \quad (9)$$

where  $\gamma^2$  is set to 0.3 according to [1]. As shown in Fig. 6, our result achieves the highest recall and F-measure, although the precision is not always the best.

Figure 7 shows some visual comparison results. We note that our method can not only highlight the salient object uniformly, but well restrain the background noises. The presented algorithm achieves good performance against other state-of-the-art methods, especially in complex scenes.

The effectiveness of the proposed algorithm is partially due to the more accurate foreground information compared to the previous methods [17, 18, 20]. To evaluate the foreground information incorporated in the presented algorithm, we compute the precision  $p_F$  and recall  $r_F$  for our foreground regions and compare them to the Otsu segmentations used in the BFS [17]. The precision  $p_F$  and recall  $r_F$  for each foreground region are calculated as

$$\begin{cases} p_F = \frac{|R_F \cap R_{GT}|}{|R_F|} \\ r_F = \frac{|R_F \cap R_{GT}|}{|R_{GT}|} \end{cases} \quad (10)$$



where  $R_F$  denotes the estimated foreground region and  $R_{GT}$  is the ground truth foreground region. The average values of precision and recall for each dataset is shown in Table 1. It indicates that the selected foreground regions are usually more favourable than the Otsu segmentations, since the high-level cue is incorporated.

Note that, the Levinshtein et al.'s mechanism [10] usually generates a dozen of contour closures and we select an optimal one using Eq. (5), which may not always obtain the best region. Figure 8 illustrates a failure case. Figure 8(a) presents all the contour closures generated by [10] and Fig. 8(c) is the selected contour closure. It is clear that the presented method selects an acceptable foreground region instead of the best one.

## 4 Conclusions

In this paper, we propose an effective method to fuse both the background and foreground information in saliency detection. To efficiently suppress the background noises, we employ two techniques: (1) the background weights defined by the spatial information of background seeds. (2) a foreground-based saliency map constructed from the optimal contour closure. The experimental results show that the presented algorithm can achieve favorable performance compared to the state-of-the-art methods.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China (61473148) and the Funding of Jiangsu Innovation Program for Graduate Education (KYLX16-0337).

## References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: CVPR, pp. 1597–1604 (2009)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI* **34**(11), 2274 (2012)
3. Gao, D., Han, S., Vasconcelos, N.: Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. *IEEE TPAMI* **31**(6), 989 (2009)
4. Guo, C., Zhang, L.: A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *Oncogene* **3**(5), 523–529 (2010)
5. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI* **20**(11), 1254–1259 (1998)
6. Jiang, H., Wang, J., Yuan, Z., Wu, Y.: Salient object detection: a discriminative regional feature integration approach. *Int. J. Comput. Vis.* **9**(4), 1–18 (2014)
7. Jonathan, H., Christof, K., Pietro, P.: Graph-based visual saliency. In: *Advances in Neural Information Processing Systems*, pp. 545–552 (2006)
8. Krahenbuhl, P.: Saliency filters: contrast based filtering for salient region detection. In: CVPR, pp. 733–740 (2012)

9. Lempitsky, V., Kohli, P., Rother, C., Sharp, T.: Image segmentation with a bounding box prior. In: ICCV, pp. 277–284 (2009)
10. Levinshtein, A., Sminchisescu, C., Dickinson, S.: Optimal contour closure by superpixel grouping. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6312, pp. 480–493. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15552-9\\_35](https://doi.org/10.1007/978-3-642-15552-9_35)
11. Li, H., Lu, H., Lin, Z., Shen, X., Price, B.: Inner and inter label propagation: salient object detection in the wild. IEEE TIP **24**(10), 3176–86 (2015)
12. Liu, H., Tao, S., Li, Z.: Saliency detection via global-object-seed-guided cellular automata. In: ICIP, pp. 2772–2776 (2016)
13. Liu, T., Sun, J., Zheng, N.N., Tang, X.: Learning to detect a salient object. In: CVPR, pp. 1–8 (2007)
14. Qin, C., Zhang, G., Zhou, Y., Tao, W., Cao, Z.: Integration of the saliency-based seed extraction and random walks for image segmentation. Neurocomputing **129**(4), 378–391 (2014)
15. Qiu, F., Sugihara, T., Von Der Heydt, R.: Figure-ground mechanisms provide structure for selective attention. Nat. Neurosci. **10**(11), 1492–1499 (2007)
16. Siva, P., Russell, C., Xiang, T., Agapito, L.: Looking beyond the image: unsupervised learning for object saliency and detection. In: CVPR, pp. 3238–3245 (2013)
17. Wang, J., Lu, H., Li, X., Tong, N., Liu, W.: Saliency detection via background and foreground seed selection. Neurocomputing **152**(C), 359–368 (2015)
18. Wang, Z., Xu, G., Wang, Z., Zhu, C.: Saliency detection integrating both background and foreground information. Neurocomputing **216**, 468–477 (2016)
19. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-33712-3\\_3](https://doi.org/10.1007/978-3-642-33712-3_3)
20. Xie, Y., Lu, H., Yang, M.H.: Bayesian saliency via low and mid level cues. IEEE TIP **22**(5), 1689–1698 (2013)
21. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: CVPR, pp. 1155–1162 (2013)
22. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: Sun: a bayesian framework for saliency using natural statistics. J. Vis. **8**(7), 1–20 (2008)
23. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: CVPR, pp. 2814–2821 (2014)