



Saliency Detection via Bidirectional Absorbing Markov Chain

Fengling Jiang^{1,2,3(✉)}, Bin Kong^{1,4}, Ahsan Adeel⁵, Yun Xiao⁶,
and Amir Hussain⁵

¹ Institute of Intelligent Machines, Chinese Academy of Sciences,
Hefei 230031, China
bkong@iim.ac.cn

² University of Science and Technology of China, Hefei 230026, China
fljiang@mail.ustc.edu.cn

³ Hefei Normal University, Hefei 230061, China

⁴ Anhui Engineering Laboratory for Intelligent Driving Technology and Application,
Hefei 230088, China

⁵ University of Stirling, Stirling FK9 4LA, UK
{aad,ahu}@cs.stir.ac.uk

⁶ Anhui University, Hefei 230601, China
xiaoyun@ahu.edu.cn

Abstract. Traditional saliency detection via Markov chain only consider boundaries nodes. However, in addition to boundaries cues, background prior and foreground prior cues play a complementary role to enhance saliency detection. In this paper, we propose an absorbing Markov chain based saliency detection method considering both boundary information and foreground prior cues. The proposed approach combines both boundaries and foreground prior cues through bidirectional Markov chain. Specifically, the image is first segmented into superpixels and four boundaries nodes (duplicated as virtual nodes) are selected. Subsequently, the absorption time upon transition node's random walk to the absorbing state is calculated to obtain foreground possibility. Simultaneously, foreground prior as the virtual absorbing nodes is used to calculate the absorption time and obtain the background possibility. Finally, two obtained results are fused to obtain the combined saliency map using cost function for further optimization at multi-scale. Experimental results demonstrate the outperformance of our proposed model on 4 benchmark datasets as compared to 17 state-of-the-art methods.

Keywords: Saliency detection · Markov chain
Bidirectional absorbing

1 Introduction

Saliency detection aims to effectively highlight the most important pixels in an image. It helps to reduce computing costs and has widely been used in

various computer vision applications, such as image segmentation [6], image retrieval [36], object detection [8], object recognition [21], image adaptation [23], and video segmentation [26]. Saliency detection could be summarized in three methods: bottom-up methods [22,33,37], top-down methods [10,15,35] and mixed methods [7,27,31]. The top-down methods are driven by tasks and could be used in object detection tasks. The authors in [34] proposed a top-down method that jointly learns a conditional random field and a discriminative dictionary. Top-down methods could be applied to address complex and special tasks but they lack versatility. The bottom-up methods are driven by data, such as color, light, texture and other basic features. Itti et al. [13] proposed a saliency method by using these basic features. It could be effectively used for real-time systems. The mixed methods are considered both bottom-up and top-down methods.

In this paper, we focus on the bottom-up methods, the proposed method is based on the properties of Markov model, there are many works based on Markov model, such as [3,4]. Traditional saliency detection via Markov chain [14] is based on Markov model as well, but it only consider boundaries nodes. However, in addition to boundaries cues, background prior and foreground prior cues play a complementary role to enhance saliency detection. We consider four boundaries information and the foreground prior saliency object, using absorbing Markov chain, namely, both boundary absorbing and foreground prior are considered to get background and foreground possibility. In addition, we further optimize our model by fusing these two possibilities, and exploit multi-scale processing. Figure 1 demonstrates and compares the results of our proposed method with the traditional saliency detection absorbing Markov chain (MC) method [14], where the outperformance of our method is evident.

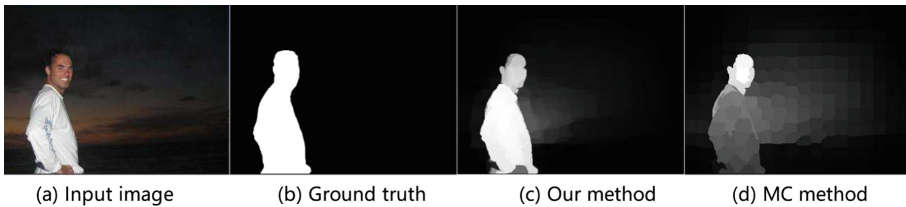


Fig. 1. Comparison of the proposed method with the ground truth and MC method.

2 Principle of Absorbing Markov Chain

In absorbing Markov chain, the transition matrix P is primitive [9], by definition, state i is absorbing when $P(i, i) = 1$, and $P(i, j) = 0$ for all $i \neq j$. If the Markov chain satisfies the following two conditions, it means there is at least one or more absorbing states in the Markov chain. In every state, it is possible to go to an absorbing state in a finite number of steps (not necessarily in one step), then we

call it absorbing Markov chain. In an absorbing Markov chain, if a state is not a absorbing state, it is called transient state.

An absorbing chain has m absorbing states and n transient states, the transfer matrix P can be written as:

$$P \rightarrow \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}, \quad (1)$$

where Q is a n -by- n matrix, giving transient probabilities between any transient states, R is a nonzero n -by- m matrix giving these probabilities from transient state to any absorbing state, 0 is a m -by- n zero matrix and I is the m -by- m identity matrix.

For an absorbing chain P , all the transient states can achieve absorbing states in one or more steps, we can write the expected number of times $N(i, j)$ (which means the transient state moves from i state to the j state), its standard form is written as:

$$N = (I - Q)^{-1}, \quad (2)$$

namely, the matrix N with invertible matrix, where n_{ij} denotes the average transfer times between transient state i to transient state j . Supposing $c = [1, 1, \dots, 1]_{1 \times n}^N$, the absorbed time for each transient state can be expressed as:

$$z = N \times c. \quad (3)$$

3 Bidirectional Absorbing Markov Chain Model

To obtain more robust and accurate saliency maps, we propose a method via bidirectional absorbing Markov chain. This section explains the procedure to find the saliency area in an image in two orientations. Simple linear iterative clustering (SLIC) algorithm [2] has been used to get the superpixels. The pipeline is explained below (Fig. 2):

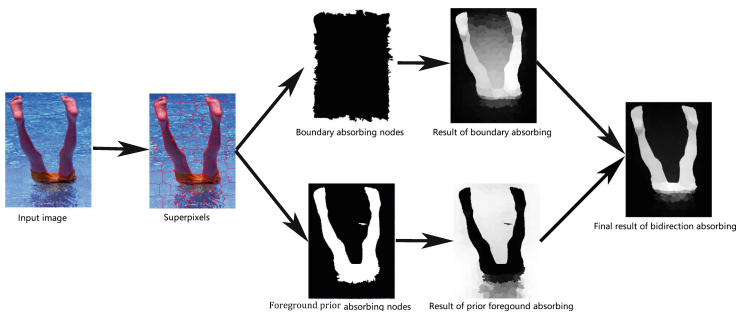


Fig. 2. The processing of our proposed method

3.1 Graph Construction

The SLIC algorithm is used to split the image into different pitches of superpixels. Afterwards, two kinds of graphs $G^1(V^1, E^1)$ and $G^2(V^2, E^2)$ are constructed, where G^1 represents the graph of boundary absorbing process and G^2 represents the graph of foreground prior absorbing process. In each of the graphs, V^1, V^2 represent the graph nodes and E^1, E^2 represent the edges between any nodes in the graphs. For the process of boundary absorbing, superpixels around the four boundaries as the virtual nodes are duplicated. For the process of foreground prior absorbing, superpixels from the regions (calculated by the foreground prior) are duplicated. There are two kinds of nodes in both graphs, transient nodes (superpixels) and absorbing nodes (duplicated nodes). The nodes in these two graphs constitute following three properties: (1) The nodes (including transient or absorbing) are associated with each other when superpixels in the image are adjacent nodes or have the same neighbors. And also boundary nodes (superpixels on the boundary of image) are fully connected with each other to reduce the geodesic distance between similar superpixels. (2) Any pair of absorbing nodes (which are duplicated from the boundaries or foreground nodes) are not connected (3) The nodes, which are duplicated from the four boundaries or foreground prior nodes, are also connected with original duplicated nodes. In this paper, the weight w_{ij} of the edges is defined as

$$w_{ij} = e^{-\frac{\|x_i - x_j\|}{\sigma^2}}, i, j \in V^1 \text{ or } i, j \in V^2 \quad (4)$$

where σ is the constant parameter to adjust the strength of the weights in CIELAB color space. Then we can get the affinity matrix A

$$a_{ij} = \begin{cases} w_{ij}, & \text{if } j \in M(i) \quad 1 \leq i \leq j \\ 1, & \text{if } i = j \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where $M(i)$ is a nodes set, in which the nodes are all connected to nodes i . The diagonal matrix is given as: $D = \text{diag}(\sum_j a_{ij})$, and the obtained transient matrix is calculated as: $P = D^{-1} \times A$.

3.2 Saliency Detection Model

Following the aforementioned procedures, the initial image is transformed into superpixels, now two kinds of absorbing nodes for saliency detection are required. Firstly, we choose boundary nodes and foreground prior nodes to duplicate as absorbing nodes and obtain the absorbed times of transient nodes as foreground possibility and background possibility. Secondly, we use a cost function to optimize two possibility results together and obtain saliency results of all transient nodes.

Absorb Markov Chain via Boundary Nodes. In normal conditions, four boundaries of an image rarely have salient objects. Therefore, boundary nodes are assumed as background, and four boundaries nodes set H^1 are duplicated as absorbing nodes set $D^1, H^1, D^1 \subset V^1$. The graph G^1 is constructed and absorbed time z is calculated via Eq. 3. Finally, foreground possibility of transient nodes $z^f = \bar{z}(i) \quad i = 1, 2, \dots, n$, is obtained, and \bar{z} denotes the normalizing the absorbed time vector.

Absorb Markov Chain via Foreground Prior Nodes. We use boundary connectivity to get the foreground prior f_i without using the down-top method [38].

$$f_i = \sum_{j=1}^N (1 - \exp(-\frac{BC_j^2}{2\sigma_b^2})) d_a(i, j) \exp(-\frac{d_s^2(i, j)}{2\sigma_s^2}) \quad (6)$$

where $d_a(i, j)$ and $d_s(i, j)$ denote the CIELAB color feature distance and spatial distance respectively between superpixel i and j , the boundary connectivity (BC) of superpixel i is defined as $BC_i = \frac{\sum_{j \in \mathcal{H}} w_{ij}}{\sqrt{\sum_{j=1}^N w_{ij}}}$ in Fig. 3, $\sigma_b = 1, \sigma_s = 0.25$. \mathcal{H} denotes the boundary area of image and w_{ij} is the similarity between nodes i and j . N is the number of superpixels. Afterwards, nodes ($\{i | f_i > avg(f)\}$) with high level values are selected to get a set H^2 , which are duplicated as absorbing nodes set $D^2, H^2, D^2 \subset V^2$. The graph G^2 is constructed and absorbed time z is calculated using Eq. 3. Finally, the background possibility of transient nodes $z^b = \bar{z}(i) \quad i = 1, 2, \dots, n$, is obtained, where \bar{z} denotes the absorbed time vector normalization.

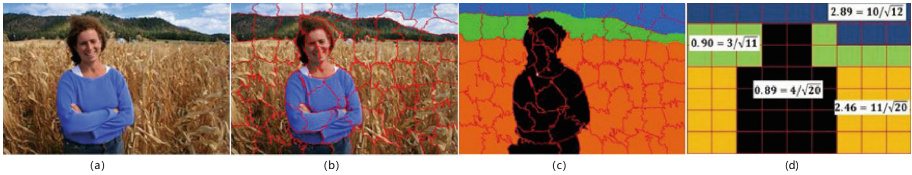


Fig. 3. An illustrative example of boundary connectivity. (a) input image (b) the superpixels of input image (c) the superpixels of similarity in each pitches (d) an illustrative example of boundary connectivity.

3.3 Saliency Optimization

In order to combine different cues, this paper has used the optimization model presented in [38], which fused background possibility and foreground possibility for final saliency map. It is defined as

$$\sum_{i=1}^N z_i^b s_i^2 + \sum_{i=1}^N z_i^f (s_i - 1)^2 + \sum_{i,j} w_{ij} (s_i - s_j)^2 \quad (7)$$

where the first term defines superpixel i with large background probability z^b to obtain a small value s_i (close to 0). The second term encourages a superpixel i with large foreground probability z^f to obtain a large value s_i (close to 1). The third term defines the smoothness to acquire continuous saliency values.

In this work, the used super-pixel numbers N are 200, 250, 300 in the super-pixel element, and the final saliency map is given as: $\mathbf{S} = \sum_h S^h$ at each scale, where $h = 1, 2, 3$.

4 Simulation Results

The proposed method is evaluated on four benchmark datasets ASD [1], CSSD [30], ECSSD [30] and SED [5]. ASD dataset is a subset of the MSRA dataset, which contains 1000 images with accurate human-labeled ground truth. CSSD dataset, namely complex scene saliency detection contains 200 complex images. ECSSD dataset, an extension of CSSD dataset contains 1000 images and has accurate human-labeled ground truth. SED dataset has two parts, SED1 and SED2, images in SED1 contains one object, and images in SED2 contains two objects, in total they contain 200 images. We compare our model with 17 different state-of-the-art saliency detection algorithms: CA [12], FT [1], SEG [20], BM [28], SWD [11], SF [19], GCHC [32], LMLC [29], HS [30], PCA [18], DSR [17], MC [14], MR [33], MS [24], RBD [38], RR [16], MST [25]. The tuning parameters in the proposed algorithm is the edge weight $\sigma^2 = 0.1$ that controls the strength of weight between a pair of nodes.

The precision-recall curves and F-measure are used as performance metrics. The precision is defined as the ratio of salient pixels correctly assigned to all the pixels of extracted regions. The recall is defined as the ratio of detected salient pixels to the ground-truth number. A PR curve is obtained by the threshold sliding from 0 to 255 to get the difference between the saliency map (which is calculated) and ground truth (which is labeled manually). F-measure is calculated by the weighted average between the precision values and recall values, which can be regarded as overall performance measurement, given as:

$$F_\beta = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 Precision + Recall}, \quad (8)$$

we set $\beta^2 = 0.3$ to stress precision more than recall. PR-curves and the F-measure curves are shown in Figs. 4–7, where the outperformance of our proposed method as compared to 17 state-of-the-art methods is evident. Figure 8 presets visual comparisons selected from four datasets. It can be seen that the proposed method achieved best saliency results as compared to the state-of-the-art methods.

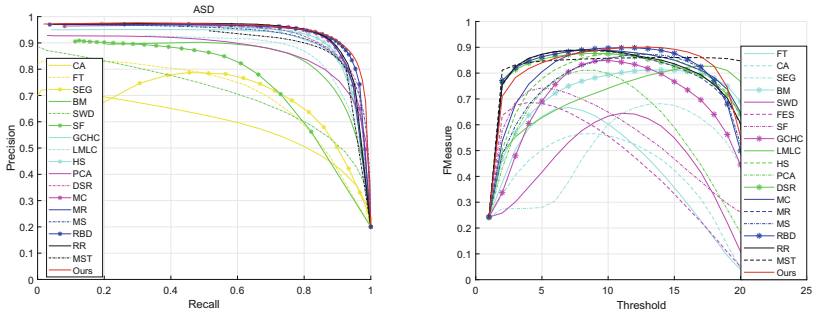


Fig. 4. PR-curves and F-measure curves comparing with different methods on ASD dataset.

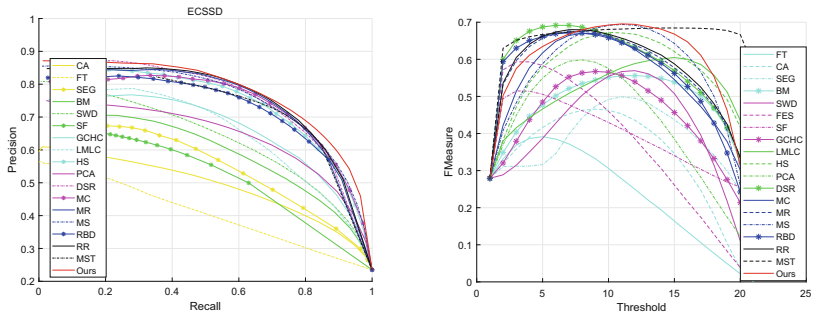


Fig. 5. PR-curves and F-value curves comparing with different methods on ECSSD dataset.

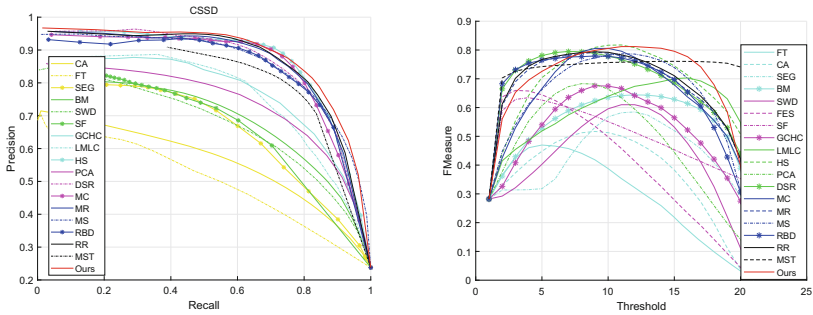


Fig. 6. PR-curves and F-measure curves comparing with different methods on CSSD dataset.

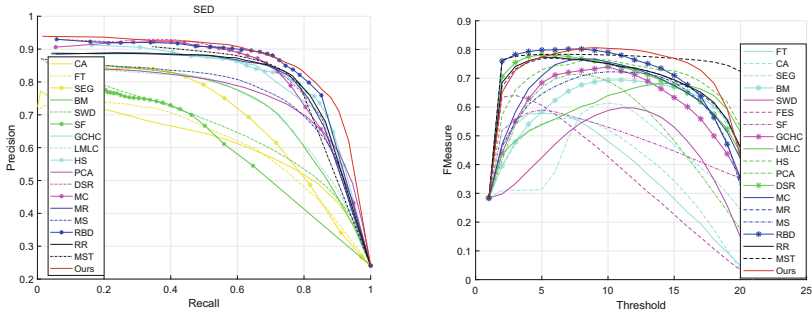


Fig. 7. PR-curves and F-measure curves comparing with different methods on SED dataset.

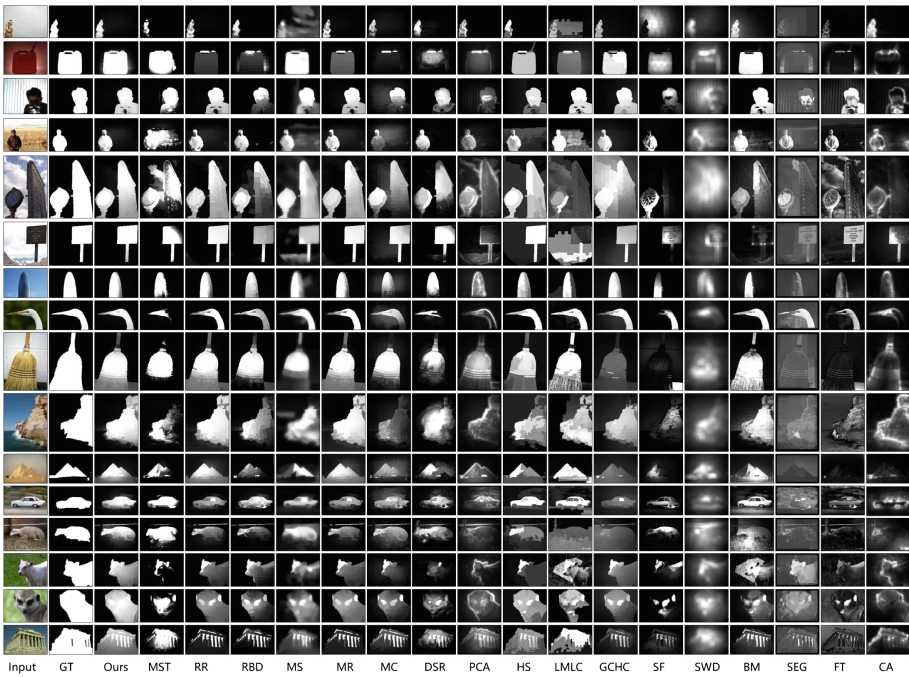


Fig. 8. Examples of output saliency maps results using different algorithms on the ASD, CSSD, ECSSD and SED datasets

5 Conclusion

In this paper, a bidirectional absorbing Markov chain based saliency detection method is proposed considering both boundary information and foreground prior cues. A novel optimization model is developed to combine both background and foreground possibilities, acquired through bidirectional absorbing Markov chain. The proposed approach outperformed 17 different state-of-the-art

methods over four benchmark datasets, which demonstrate the superiority of our proposed approach. In future, we intend to apply our proposed saliency detection algorithm to problems such as multi-pose lipreading and audio-visual speech recognition.

Acknowledgement. This work was supported by China Scholarship Council, the National Natural Science Foundation of China (No. 913203002), the Pilot Project of Chinese Academy of Sciences (No. XDA08040109). Prof. Amir Hussain and Dr. Ahsan Adeel were supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant No. EP/M026981/1.

References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1597–1604. IEEE (2009)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **34**(11), 2274–2282 (2012)
3. AlKhateeb, J.H., Pauplin, O., Ren, J., Jiang, J.: Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition. *Knowl.-Based Syst.* **24**(5), 680–688 (2011)
4. AlKhateeb, J.H., Ren, J., Jiang, J., Al-Muhtaseb, H.: Offline handwritten Arabic cursive text recognition using hidden Markov models and re-ranking. *Pattern Recogn. Lett.* **32**(8), 1081–1088 (2011)
5. Alpert, S., Galun, M., Brandt, A., Basri, R.: Image segmentation by probabilistic bottom-up aggregation and cue integration. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **34**(2), 315–327 (2012)
6. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **33**(5), 898–916 (2011)
7. Borji, A., Sihite, D.N., Itti, L.: Probabilistic learning of task-specific visual attention. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 470–477. IEEE (2012)
8. Chang, K.Y., Liu, T.L., Chen, H.T., Lai, S.H.: Fusing generic objectness and visual saliency for salient object detection. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 914–921. IEEE (2011)
9. Charles, M., Grinstead, J., Snell, L.: *Introduction to Probability*. American Mathematical Society, Providence (1997)
10. Cholakkal, H., Johnson, J., Rajan, D.: Backtracking SCSPM image classifier for weakly supervised top-down saliency. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5278–5287. IEEE (2016)
11. Duan, L., Wu, C., Miao, J., Qing, L., Fu, Y.: Visual saliency detection by spatially weighted dissimilarity. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 473–480. IEEE (2011)
12. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **34**(10), 1915–1926 (2012)
13. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **20**(11), 1254–1259 (1998)

14. Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M.H.: Saliency detection via absorbing Markov chain. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 1665–1672. IEEE (2013)
15. Kocak, A., Cizmeciler, K., Erdem, A., Erdem, E.: Top down saliency estimation via superpixel-based discriminative dictionaries. BMVA Press (2014). <https://doi.org/10.5244/C.28.73>
16. Li, C., Yuan, Y., Cai, W., Xia, Y., Feng, D.D., et al.: Robust saliency detection via regularized random walks ranking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2710–2717. IEEE (2015)
17. Li, X., Lu, H., Zhang, L., Ruan, X., Yang, M.H.: Saliency detection via dense and sparse reconstruction. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 2976–2983. IEEE (2013)
18. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1139–1146. IEEE (2013)
19. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: contrast based filtering for salient region detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 733–740. IEEE (2012)
20. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6315, pp. 366–379. Springer, Heidelberg (2010). <https://doi.org/10.1007/978-3-642-15555-0-27>
21. Ren, Z., Gao, S., Chia, L.T., Tsang, I.W.H.: Region-based saliency detection and its application in object recognition. *IEEE Trans. Circuits Syst. Video Technol.* **24**(5), 769–779 (2014)
22. Riche, N., Mancas, M., Gosselin, B., Dutoit, T.: Rare: a new bottom-up saliency model. In: Proceedings of the 19th IEEE International Conference on Image Processing (ICIP), pp. 641–644. IEEE (2012)
23. Sun, J., Xie, J., Liu, J., Sikora, T.: Image adaptation and dynamic browsing based on two-layer saliency combination. *IEEE Trans. Broadcast.* **59**(4), 602–613 (2013)
24. Tong, N., Lu, H., Zhang, L., Ruan, X.: Saliency detection with multi-scale superpixels. *IEEE Signal Process. Lett.* **21**(9), 1035–1039 (2014)
25. Tu, W.C., He, S., Yang, Q., Chien, S.Y.: Real-time salient object detection with a minimum spanning tree. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2334–2342. IEEE (2016)
26. Wang, W., Shen, J., Yang, R., Porikli, F.: Saliency-aware video object segmentation. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **40**(1), 20–33 (2018)
27. Wang, Z., Ren, J., Zhang, D., Sun, M., Jiang, J.: A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. *Neurocomputing* **287**, 68–83 (2018)
28. Xie, Y., Lu, H.: Visual saliency detection based on Bayesian model. In: Proceedings of the 18th IEEE International Conference on Image Processing (ICIP), pp. 645–648. IEEE (2011)
29. Xie, Y., Lu, H., Yang, M.H.: Bayesian saliency via low and mid level cues. *IEEE Trans. Image Process.* **22**(5), 1689–1698 (2013)
30. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1155–1162. IEEE (2013)
31. Yan, Y., et al.: Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. *Pattern Recogn.* **79**, 65–78 (2018)

32. Yang, C., Zhang, L., Lu, H.: Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Process. Lett.* **20**(7), 637–640 (2013)
33. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3166–3173. IEEE (2013)
34. Yang, J., Yang, M.H.: Top-down visual saliency via joint CRF and dictionary learning. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2296–2303. IEEE (2012)
35. Yang, J., Yang, M.H.: Top-down visual saliency via joint CRF and dictionary learning. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **39**(3), 576–588 (2017)
36. Yang, X., Qian, X., Xue, Y.: Scalable mobile image retrieval by exploring contextual saliency. *IEEE Trans. Image Process.* **24**(6), 1709–1721 (2015)
37. Zhao, R., Ouyang, W., Li, H., Wang, X.: Saliency detection by multi-context deep learning. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1265–1274. IEEE (2015)
38. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2814–2821. IEEE (2014)