

# Robust Visual Saliency Optimization based on Bidirectional Markov Chains

Fengling Jiang · Bin Kong · Jingpeng Li · Kia Dashtipour · Mandar Gogate

Received: date / Accepted: date

**Abstract Background** Saliency detection aims to automatically highlight most important area in an image. Traditional saliency detection methods based on absorbing Markov chain only take into account boundary nodes and often lead to incorrect saliency detection when the boundaries have salient objects. **Methods** In order to address this limitation and enhance saliency detection performance, this paper propose a novel task-independent saliency detection method based on the bidirectional absorbing Markov chain that jointly exploits not only the boundary information but also the foreground prior and background prior cues. More specifically, the input image is first segmented into number of superpixels, and the four boundary 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 the foreground possibility. Simultaneously, foreground prior (as the virtual absorbing nodes) is used to

calculate the absorption time and get the background possibility. In addition, the two aforementioned results are fused to form a combined saliency map which is further optimized by using a cost function. Finally, the superpixel-level saliency results are optimized by a regularized random walks ranking model at multi-scale. **Results** The comparative experimental results on four benchmark datasets reveal superior performance of our proposed method over state-of-the-art methods reported in the literature. **Conclusions** The experiments show that the proposed method is efficient and can be applicable to the bottom-up image saliency detection and other visual processing tasks.

**Keywords** Saliency detection · Bidirectional absorbing · Markov chain · Background and foreground possibility.

## 1 Introduction

Visual saliency detection simulates the human visual system aiming to automatically highlight the most important area in an image. The process of using an important part of an image instead of the full image reduces the computational costs for computer vision systems. Therefore, saliency detection has become a preprocessing stage in many applications including image segmentation [46], image retrieval [16], image classification [15], object detection [23,44,36], object recognition [31], object tracking [48] and video segmentation [38].

In the literature, extensive research has been carried out to develop saliency detection methods. The methods can be divided into three categories: bottom-up methods [12,49,9,28,6], top-down methods [39,14,8,27] and mixed methods [11,4,54,47]. The bottom-up methods are data-driven and mostly applied to real-time systems, utilizing color, edge, brightness, texture and other low-level features to obtain the saliency maps in an efficient and effective manner. In the

---

F. Jiang  
Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China  
University of Science and Technology of China, Hefei 230026, China  
School of Computer Science and Technology, Hefei Normal University, Hefei 230061, China  
E-mail: fljiang@mail.ustc.edu.cn

B. Kong(✉)  
Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China  
The Key Laboratory of Biomimetic Sensing and Advanced Robot Technology, Hefei 230031, China  
E-mail: bkong@iim.ac.cn

J. Li · K. Dashtipour  
Division of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, Scotland, UK

M. Gogate  
School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, Scotland, UK

bottom-up methods, there exists many methods such as random walk based methods [19,33,50], region contrast models [7,20] and so on. On the other hand, the top-down methods are goal-driven (i.e. task dependent), including some cnn-based deep learning methods[53,40,37,21], make use of high-level features. In general, top-down methods can get better performance but require large dataset for supervised training and are computationally expensive. The mixed methods integrate bottom-up and top-down methods to obtain final saliency results.

In our paper, we focus on the traditional bottom-up models. In the existing random walk models, one of an important work get saliency detection by using absorbing Markov chain(MC) [19], which can get a good results by considering boundary nodes as absorbing nodes to construct an absorbing Markov chain. Actually, background and foreground priors can play a complementary role in enhancing saliency detection. We consider to using both the boundary information and the foreground prior to duplicated as absorbing nodes, and propose a bidirectional absorbing Markov chains based saliency detection model to get final saliency maps. Firstly, we use boundary and foreground prior to get the background and foreground possibilities. Then, we use an optimized function to fuse two kinds of information. Finally, we get the final saliency map by pointing saliency values of superpixels to each pixels.

In summary, the contributions of our work include:

1. We present a novel bidirectional Markov chains model (BMC), which use background and foreground prior information to construct two absorbing Markov chain.
2. An optimization model is developed to combine both background and foreground possibilities, which are acquired through bidirectional absorbing Markov chains.

## 2 Related Works

In this section, we focus on the related state-of-the-art and other recently proposed models. The research of saliency detection is originated from biological disciplines. The Itti and Koch model [18] first computed saliency maps by using texture, orientation, intensity and color contrast features. Since then, various models have emerged [17,13,32,55,22,52,51]. Gopalakrishnan et al. [13] used the hitting time to calculate the most salient seed, then calculated the distance between the other nodes and the selected seed to get the final results. In addition, Sun et al. [32] exploited the relationship between the saliency detection and the Markov absorption probability for saliency detection. Jiang et al. [19] proposed a saliency detection model by random walk via absorbing Markov chain, where absorbing nodes are duplicated from the four boundaries. Based on geodesic distances, Zhu et al. [55] integrated boundary connectivity into a cost func-

tion to obtain the final optimized saliency map. Li et al. [22] used a regularized random walks ranking model to get the saliency maps. Additionally, Zhang et al. [52] presented a data-driven salient region detection model based on absorbing Markov chain via multi-feature. Similarly, Zhang et al. [51] proposed an approach to detection salient objects by exploring patch-level and object-level cues via absorbing Markov chain. Zhang et al. [50] proposed a learnt transition probability matrix taking into account the importance of the transition probability matrix based on the aforementioned work.

## 3 Traditional Saliency Detection via Absorbing Markov Chain

Formally, let  $X = \{x_1, x_2, \dots, x_n\}$  be a dataset containing  $n$  data points, saliency detection aims to solve the problem of determining the saliency values. One widely used approach to address this problem is to use Markov random walks. Jiang et al [19] first introduced an absorbing Markov chain for the visual saliency detection. The expected absorbing time of every transient node is computed to measures its similarity with all-absorbing nodes. The transient nodes that have similar appearance with absorbing nodes can be absorbed faster, i.e. have less expected absorbing time. The saliency detection process is outlined below:

**Step 1. Obtain the affinity matrix  $A$ .** The input image is segmented into superpixels through the method of simple linear iterative clustering (SLIC) algorithm [2] which is based on a spatially localized version of k-means clustering, decomposing the input image in visually homogeneous regions, namely superpixels (see Fig. 5(b)) and then, we construct a graph  $G = (V, E)$ , where  $V$  denotes the nodes and  $E$  denotes the edges between nodes. There are  $n$  original nodes and  $l$  duplicated nodes in  $V$ . The edge weight  $w_{ij}$  between node  $v_i$  and  $v_j$  can be calculated by feature vectors of two nodes. If node  $v_i$  is a transient node, the neighbor node or the neighbor's neighbor node  $v_j$  is connected to node  $v_i$ . Moreover, we acquire an affinity matrix  $A$ :  $a_{ij} = w_{ij}$  if  $v_i$  and  $v_j$  is connected;  $a_{ii} = 1, i = 1, 2, \dots, n$ ; otherwise  $a_{ij} = 0$ .

**Step 2. Compute the transition matrix  $P$ .** The degree matrix that records the sum of the weights connected to each node is written as  $D = \text{diag}(\sum_j a_{ij})$ . The transition matrix  $P$  is primitive [5]. Finally, the transition matrix  $P$  on the sparsely connected graph is given as  $P = D^{-1} \times A$ .

**Step 3. Renumber the nodes in transition matrix  $P$ .** There are two types of states in an absorbing Markov chain: one is called the transient state, and another is absorbing state. In an absorbing Markov chain, if a state is not an absorbing state, it is called a transient state. For an absorbing chain having  $n$  absorbing states and  $l$  transient states, its transfer matrix  $P$  can be written as:

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

where  $Q$  is an  $n \times n$  matrix giving transient probabilities between any transient states,  $R$  is a nonzero  $n \times l$  matrix giving these probabilities from transient state to any absorbing state,  $0$  is an  $l \times n$  zero matrix, and  $I_l$  is an  $l \times l$  identity matrix.

**Step 4. Compute the expected times.** For an absorbing chain  $P$ , all the transient states can enter the absorbing states in one or more steps, namely, the matrix  $N$  with invertible matrix, where  $n_{ij}$  denotes the average transfer times between transient state  $i$  to transient state  $j$ . Supposing  $e = [1, 1, \dots, 1]_{1 \times n}^T$  and  $I_n$  is an  $n \times n$  identity matrix, the absorbing time for each transient state can be expressed as:

$$Z = (I_l - Q)^{-1} \times e. \quad (2)$$

For each node  $v_i$ , the expected time is  $s_i = \sum_j Z_{ij}$ ,  $j = 1, 2, \dots, n$  [19].

## 4 The Proposed Approach

The above Markov chain model provides an effective saliency value for all data points. However, the main limitation of this method is that the output only depends on the superpixels of four boundaries as absorbing nodes. It may lead to incorrect prediction especially when the four boundaries have some salient objects. The limitation motivated us to develop a bidirectional absorbing Markov chain based method, that capture the absorption time of all nodes to the foreground and background more effectively. Therefore, we implement a saliency optimization on the results of two types of absorption time to obtain the saliency detection values of all superpixels. Finally, a regularized random walk ranking based on the pixel-wise graph is used to diffuse the saliency values from the superpixel level to pixel level. The pipeline is described in Fig. 1.

### 4.1 Construction of Three Graphs

**The Initial Graph:** Given an input image  $I$ , we use the SLIC [2] algorithm to segment the image into  $N$  superpixels. We extract visual features of average values of CIELAB color space and denote them as  $X = \{x_1, x_2, \dots, x_N\} \in R^{N \times 3}$ ,  $x_i = (L_i^*, a_i^*, b_i^*)$ , where  $L^*$  denotes the lightness,  $a^*$  and  $b^*$  indicating where the color falls along the green-red axis and blue-yellow axis. Next, we define an initial graph  $G = (V, E)$  on the dataset in Fig. 2, where  $V = \{V_1, V_2, \dots, V_N\}$  denotes the node set and  $E$  denotes the edges (weighted by a matrix  $W$ ) between two nodes.

The edge set  $E$  is determined as follows: (1) Each node is connected to its neighbors and also connected to the nodes that have the same neighbors; (2) All the boundary nodes

are connected. The weight of edge between node  $v_i$  and  $v_j$  is calculated by Eq. 3:

$$w_{ij} = e^{-\frac{\|x_i - x_j\|}{\sigma^2}}, i, j \in V \quad (3)$$

where  $\sigma$  is a constant, and  $x_i, x_j$  represents the feature vectors of graph nodes  $v_i$  and  $v_j$  respectively. The affinity matrix  $A$  is formulated as Eq. 4:

$$a_{ij} = \begin{cases} w_{ij}, & \text{if } j \in M(i) \quad 1 \leq i \leq j \\ 1, & \text{if } i = j \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where  $M(i)$  is a node set, in which the nodes are all connected to nodes  $i$ . The degree matrix is given as  $D = \text{diag}(\sum_j a_{ij})$ .

**The Second Graph:** We construct another graph  $G^b = (V^b, E^b)$  with  $N^b$  nodes including  $N$  primary nodes and  $b$  duplicated nodes in Fig. 3, where  $V^b = \{V_1, \dots, V_N, V_{N+1}, \dots, V_{N+b}\}$  denotes the node set and  $E^b$  denotes the edges (weighted by a matrix  $W^b$ ) between two nodes.

Then, we duplicate  $b$  boundary superpixels as background absorbing nodes, that display outside the blue box with pink dots (see Fig. 3). Edge  $E^b$  is determined as follows:

(1) The nodes (transient or absorbing) are associated with each other when superpixels in the image are adjacent or have the same neighbors. Additionally, the boundary nodes (i.e. superpixels on the boundary of the image) are fully connected to reduce the geodesic distance between similar superpixels;

(2) Any pair of absorbing nodes (which are duplicated from the boundary nodes) is not connected;

(3) The nodes duplicated from the four boundaries are also connected with the original duplicated nodes.

The weight of edge  $w_{ij}^b$  between nodes  $v_i$  and  $v_j$  are calculated by Eq. 3, the affinity matrix  $A^b$  is formulated as Eq. 4 and the diagonal (or degree) matrix is given as  $D^b = \text{diag}(\sum_j a_{ij}^b)$ .

**The Third Graph:** We construct one more graph  $G^f = (V^f, E^f)$  with  $N^f$  nodes including  $N$  primary nodes and  $f$  duplicated nodes in Fig. 4, where  $V^f = \{V_1, \dots, V_N, V_{N+1}, \dots, V_{N+f}\}$  denotes the node set and  $E^f$  denotes the edges (weighted by a matrix  $W^f$ ) between two nodes.

In order to obtain more effective results, we duplicate  $f$  foreground superpixels as absorbing nodes, which are shown in the blue points above the image (see Fig. 4). Edge  $E^f$  is determined as follows:

(1) Each transient or absorbing node is connected to the transient nodes which are the neighbors of it or have the same boundaries with its neighboring nodes;

(2) All transient nodes on the boundary are connected;

(3) Any pair of absorbing nodes (which are duplicated from the foreground) are unconnected.

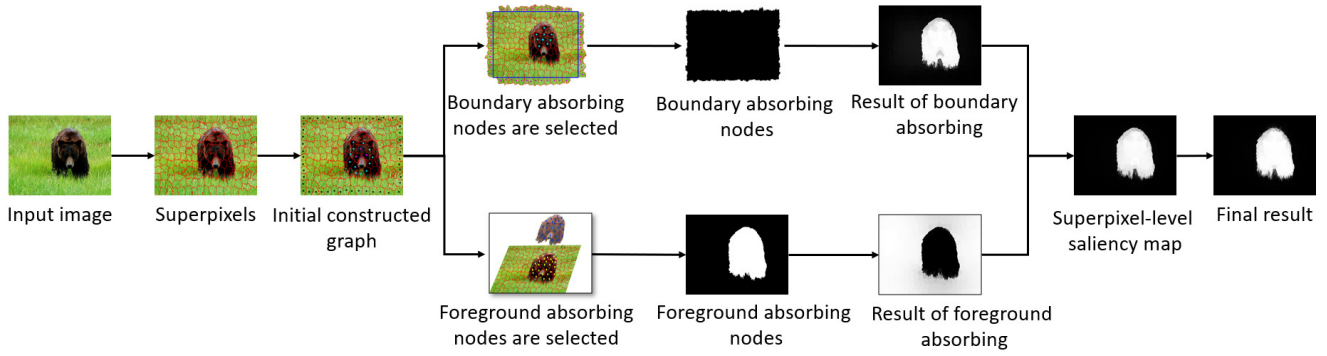


Fig. 1 The pipeline of our proposed method

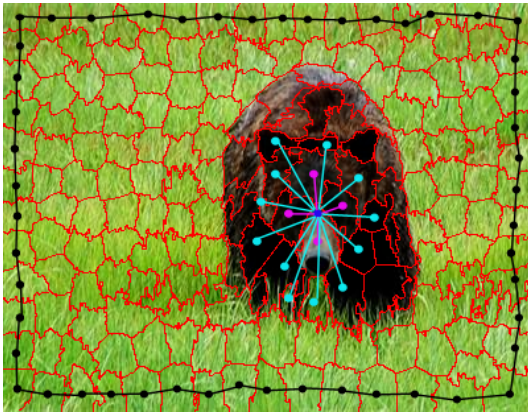


Fig. 2 Construction of the initial graph.

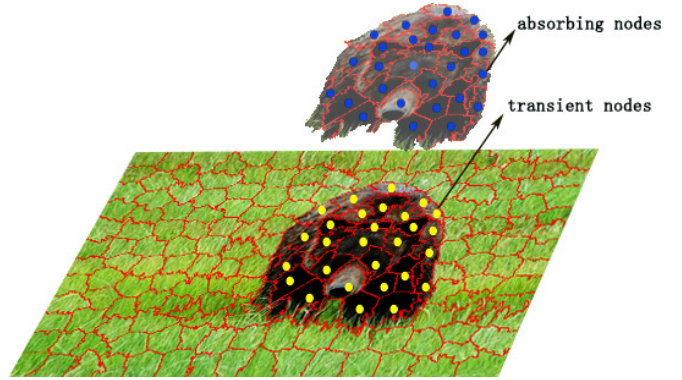


Fig. 4 Construction of the third graph. The superpixels above the image with blue dots are duplicated from the foreground superpixels as absorbing nodes. The superpixels in the image with yellow dots are transient nodes.

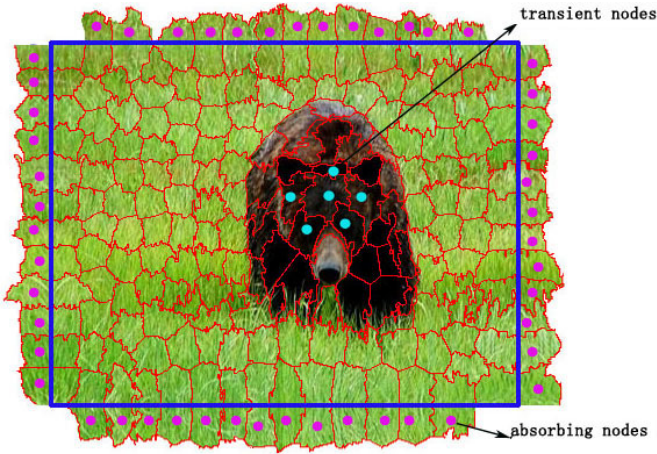


Fig. 3 Construction of the second graph. The superpixels with the pink dots are duplicated absorbing nodes, while the nodes in the blue box are transient nodes.

The weight of edge  $w_{ij}^f$  between nodes  $v_i$  and  $v_j$  are calculated by Eq. 3, the affinity matrix  $A^f$  is formulated as Eq. 4, and the degree matrix is given as  $D^f = \text{diag}(\sum_j a_{ij}^f)$ .

#### 4.2 Select Nodes by Foreground Prior

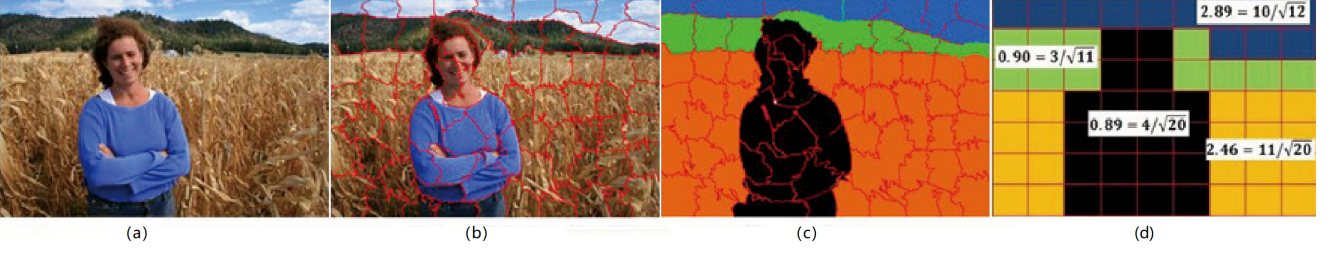
In the third graph, the duplicated nodes are selected by using the foreground information. The prior information is a significant cue in saliency detection and many other fields. There are many methods to obtain prior information. In our proposed method, we use boundary connectivity [55] to get the foreground prior. Boundary connectivity (BC) is the proportion of the boundary superpixels to the whole same cluster superpixels (see Fig. 5(c)), which is defined as follow:

$$BC_i = \frac{\sum_{j \in \mathcal{H}} a_{ij}}{\sqrt{\sum_{j=1}^N a_{ij}}} \quad (5)$$

where  $N$  is the number of superpixels,  $\mathcal{H}$  denotes the boundary area of image, and  $w_{ij}$  is the similarity between nodes  $i$  and  $j$ . We give an illustrative example of boundary connectivity in Fig. 5.

Let  $f_i$  be the foreground prior, it can be calculated by the following equation:

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



**Fig. 5** An illustrative example of boundary connectivity: (a) Input image; (b) The superpixels of input image; (c) The superpixels of similarity in each pitch; (d) The intuitive interpretation of boundary connectivity.

where  $\sigma_b = 1$ ,  $\sigma_s = 0.25$ ,  $d_a(i, j)$  and  $d_s(i, j)$  denote the CIELAB color feature distance and spatial distance between the  $i$ -th and  $j$ -th superpixels respectively.

If superpixel  $i$  has a high value of  $f_i$ , we can set it as a foreground prior node. Nodes with higher than average values (i.e.  $\{i | f_i > \text{avg}(f)\}$ ) are selected to form a set, which is duplicated as the set of absorbing nodes set (a subset of  $V^f$ ). The graph  $G^f$  is therefore constructed.

### 4.3 Foreground and Background Possibility

Following the aforementioned procedures, the initial input image is segmented into superpixels by the SLIC method to form an initial graph  $G$ . Moreover, we choose boundary nodes and foreground nodes, and duplicate them as absorbing nodes to obtain two graphs  $G^b$  and  $G^f$  respectively. Next, we use Eq. 3 and Eq. 4 to get the affinity matrix  $A$ . The degree matrix is written as  $D = \text{diag}(\sum_j a_{ij})$ . The transition matrix  $P$  is calculated as  $P = D^{-1} \times A$ , which can be reordered as  $n$  absorbing nodes and  $l$  transient nodes and broken down into four sub-matrices:  $Q$ ,  $R$ ,  $O$  and  $I_l$  (See Eq. 1). Supposing  $e = [1, 1, \dots, 1]_{1 \times n}^T$ , The absorbed time for each node  $v_i$  can be obtained by  $s_i = \sum_j ((I_l - Q)^{-1} \times e)_{ij}$ ,  $i, j = 1, 2, \dots, n$ , where  $e = [1, 1, \dots, 1]_{1 \times n}^T$  and  $I_l$  is an identity matrix.

In graph  $G^b$ , the absorbing nodes are selected from the boundary. Then for each node  $v_i$ , if the expected time is considerable that means it requires more time to transfer to the border, the node is more likely to be a foreground node; if the expected time is low, the node is more like to be a background one. We can set this value as the foreground possibility  $s^f$  of node  $v_i$ .

In graph  $G^f$ , the absorbing nodes are selected according to the foreground prior. Then for each node  $v_i$ , if the expected time is considerable, the node is more likely to be a background node; otherwise, the node is more like to be a foreground one. We can set this value as background possibility  $s^b$  of node  $v_i$ .

### 4.4 Superpixel-Level Saliency Optimization

We utilize background and foreground possibilities to obtain the superpixel-level saliency map. Towards the objective of assigning the foreground region value 1 and the background region value 0, the following optimization model presented in [55] is used:

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

where  $s_i$  is the superpixel level saliency value (i.e. the expected time of absorption), and  $a_{ij}$  is obtained from the initial graph. In order to achieve a minimization of the objective function (Eq.7), the first term is designed to encourage a superpixel  $i$  with large background probability  $s^b$  to obtain a small value  $s_i$  (close to 0), the second term is designed to encourage a superpixel  $i$  with large foreground probability  $s^f$  to obtain a large value  $s_i$  (close to 1), and the third term is designed to encourage the smoothness to acquire continuous saliency values.

In order to obtain the value  $s$  which is represented the optimal solution of Eq. 7, we rewrite Eq.7 as the matrix form:

$$s^T D^b s + (e - s)^T D^f (e - s) + 2s^T D s - 2s^T A s \quad (8)$$

where  $e = [1, 1, \dots, 1]^T$ ,  $D^b = \text{diag}(\sum_j a_{ij}^b)$ ,  $D^f = \text{diag}(\sum_j a_{ij}^f)$ ,  $D = \text{diag}(\sum_j a_{ij})$ .

Let  $F(s)$  denotes the function. To minimize  $F(s)$ , the derivative of  $F(s)$  with respect to  $s$  is:

$$\frac{\partial F}{\partial s} = 2D^b s + 2D^f s - 2D^f e + 4Ds - 4As \quad (9)$$

Supposing  $\frac{\partial F}{\partial s} = 0$ , we can obtain the superpixel-level saliency value  $s$  as follows:

$$s = (D^b + D^f + 2D - 2A)^{-1} e \quad (10)$$

### 4.5 Robust Pixel-Level Saliency Detection

From the above, we get the superpixel-level saliency vector  $s$ . If we assign the value of a superpixel to each pixel in the

superpixel, the most precise pixel-level results cannot be obtained. In order to obtain more efficient pixel-level saliency map, we employ a robust saliency ranking model [22] to get the saliency value. The optimal function of queries is computed by solving the following optimization problem:

$$\min_{s^p} \sum_{i,j} a_{ij} (s_i^p - s_j^p)^2 + \gamma \|s^p - \bar{s}\|^2 \quad (11)$$

where  $s^p$  is the pixel-level saliency value that needs to be optimized, and  $\bar{s}$  is pixelwise saliency value of  $s$ ,  $a_{ij}$  denotes the edge weight between pixel  $i$  and  $j$ . The first term encourages the smoothness to acquire continuous saliency values. The second term encourages the pixel-level saliency value  $s^p$  is close to the superpixel-level saliency value  $s$ ,  $\gamma$  is the parameter to adjust the second term, we let  $\gamma = \mu/2$ .  $\mu$  is the controlling parameter.  $Y$  is the pixel-wise vector obtaining from the  $s_{superpixel}$ .  $L$  is an  $n \times n$  Laplacian matrix,  $n$  is the pixel number of the input image. If  $Y > T_{high}$ , then  $k = 2$ , which means  $Y$  is foreground label. Otherwise, if  $Y < T_{low}$ , then  $k = 1$ , which means  $Y$  is background label.

Therefore, we use the SLIC method to segment the image into superpixels in this work. The number of superpixels influences the results. In order to keep within limits of this, we use multi-scale fusion  $S_{final} = \sum_h (s^p)^h$  to obtain the final saliency map, where  $h = 1, 2, \dots, n$ ,  $h$  means different scales. The whole algorithm of our proposed method is summarized in Algorithm 1.

---

**Algorithm 1** Saliency detection by bidirectional Markov chains.

---

**Require:** An image and required parameters as the input.

- 1: Use the SLIC method to segment the input image into superpixels, construct one initial graph  $G$  and two extended graphs  $G^b$  and  $G^f$ , and compute the affinity matrix  $A$  by using Eq. 3 and Eq. 4.
- 2: Compute three different transient matrices  $P$  by  $P = D^{-1} \times A$  on three graphs, where  $D = \text{diag}(\sum_j a_{ij})$ .
- 3: Rewrite  $P$  as Eq. 1 to get the sub-matrix  $Q$ .
- 4: Compute the expected time of absorption by  $s_i = \sum_j ((I - Q)^{-1} \times e)_{ij}$  for each node  $v_i$  in  $G^b$  and  $G^f$ , and get the foreground probability  $s^f$  and background probability  $s^b$  respectively.
- 5: Optimize the model by Eq. 7, and obtain superpixel-level saliency results by Eq. 10.
- 6: Optimize the model by Eq. 11, and compute the pixel-level saliency map  $s^p$  by Eq. 11.
- 7: Compute final saliency values with  $S_{final} = \sum_h (s^p)^h$ .

**Ensure:** Output a saliency map  $S_{final}$  with the same size of the input image.

---

## 5 Experiments

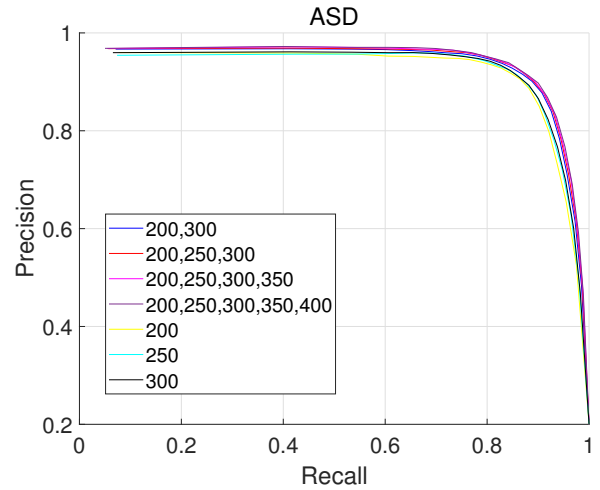
The proposed method is evaluated on four widely used benchmark datasets: ASD [1], CSSD [43], ECSSD [43] and SED [3].

We compare our model with the following seventeen related state-of-the-art saliency detection algorithms: CA [12], FT [1], SEG [30], BM [41], SWD [10], SF [29], GCHC [45], LMLC [42], HS [43], PCA [26], DSR [24], MC [19], MR [46], MS [34], RBD [55], RR [22], and MST [35].

### 5.1 Benchmark Datasets and Parameter Setting

The ASD dataset is a subset of the MSRA dataset [25], which contains 1000 images with the accurate human-labeled ground truth. The CSSD dataset, namely complex scene saliency detection, contains 200 complex images. The ECSSD dataset, an extension of CSSD dataset, contains 1000 images with the accurate human-labeled ground truth. The SED dataset has two parts of SED1 and SED2: images in SED1 contain one object, images in SED2 contain two objects, and in total there are 200 images.

From the experiments, experientially, the tuning parameters in the proposed algorithm are set as follows: the edge weight  $\sigma^2 = 0.1$  in Eq.3 to control the strength of weight between a pair of nodes, we let  $\mu = 0.99$ , then calculate  $\gamma$  in Eq.11 to obtain the final saliency value. We have conducted the experiment for the selection of superpixel numbers, the process is shown in Fig.6 that proves the used superpixel numbers  $N$  which are 200, 250, 300 and 350 in the superpixel element is the optimal choice.



**Fig. 6** The selection of superpixel numbers.

### 5.2 Evaluation of the Proposed Model

The Precision-Recall (PR) curves [1], F-measure curves [1] and the F-measure values are used as performance metrics. The precision is defined as the ratio of salient pixels correctly detected to all the pixels of extracted regions, while

the recall is defined as the ratio of salient pixels correctly detected to the ground-truth. They are formulated as:

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

where  $TP$ ,  $FP$ , and  $FN$  represent the true positive, false positive and false negative respectively. A PR curve is obtained by the threshold sliding from 0 to 255 to get the difference between the predicted saliency map and the manually labelled ground truth. F-measure is regarded as the overall performance measurement by calculating the weighted average between the precision and recall values, formulated as

$$F_{\beta} = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 Precision + Recall}, \quad (14)$$

where  $\beta^2 = 0.3$  is set to stress precision more than recall. According to different thresholds, we obtain the F-measure curve.

### 5.3 Quantitative Comparison

First, in ASD dataset, we give the comparison between our proposed method with the saliency detection via absorbing Markov chain based on background prior only and foreground prior only, respectively. From the Fig.7, the PR-curve is better than the others, which suggests the effective of our model.

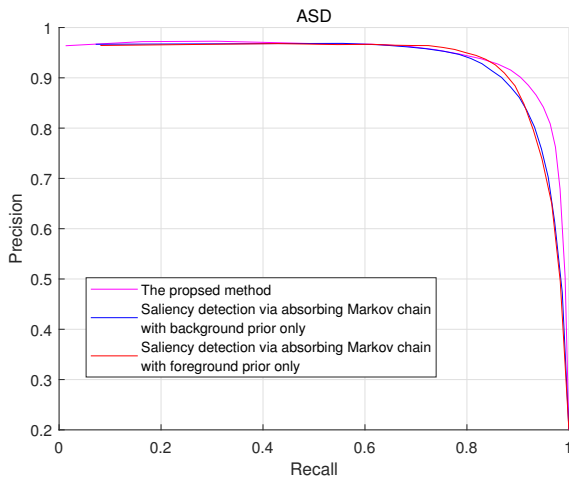


Fig. 7 The comparison with background prior and foreground prior.

We also compare the proposed method with superpixel-level saliency map and pixel-level saliency map on SED dataset, the process is shown in Fig.8 which proves the pixel-level saliency maps and superpixel-level saliency maps have

similar PR-curve, but the F-measure are 0.7877 and 0.7821 respectively, the pixel-level saliency maps are slightly better than the superpixel-level saliency maps.

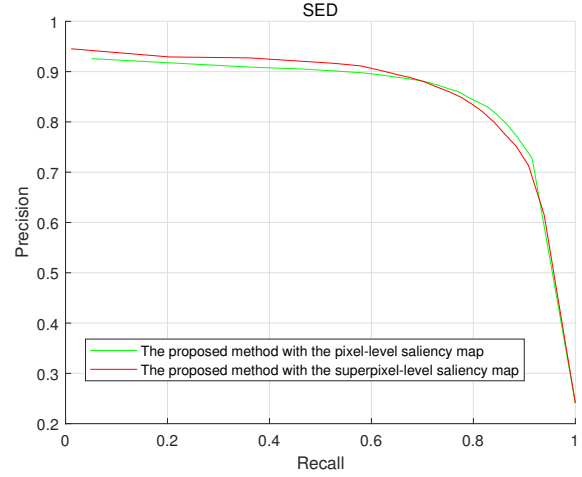


Fig. 8 The comparison of pixel-level saliency map and superpixel saliency map in our method.

Then, the PR-curves, F-measure curves and the F-measure values on four benchmark datasets are shown in Fig.9-12, where our proposed method outperform 17 state-of-the-art methods. Fig.9 and Fig.11 report the performance comparison on the ASD and CSSD dataset using the PR curve, F-measure curve and the F-measure values. Compared with the 17 state-of-the-art methods, our method achieves a competitive performance according to the 3 terms. Fig.10 shows the performance comparison with the 17 state-of-the-art methods on the ECSSD dataset, the proposed method achieves a high value of precision in the recall range from 0.5 to 0.9 and the highest F-measure value. Fig.12 shows the evaluation results comparison with the 17 state-of-the-art methods on the SED dataset, the proposed method achieves a high precision with recall ranging from 0.7 to 1 and the highest F-measure value.

Fig.13 presents output of the estimated saliency maps on sample images selected from the four datasets. It can be seen that, our proposed method can overall achieve the best saliency results as compared to other state-of-the-art methods.

### 5.4 Failure Cases Analysis

In this work, the idea of bidirectional absorbing Markov chains is proposed. The proposed method is effective for most images on the four benchmark datasets, and the final results are overall better than the superpixel-level saliency maps (Fig. 14(c)(d)). However, if the appearances of four boundaries and the foreground prior are similar to each other,

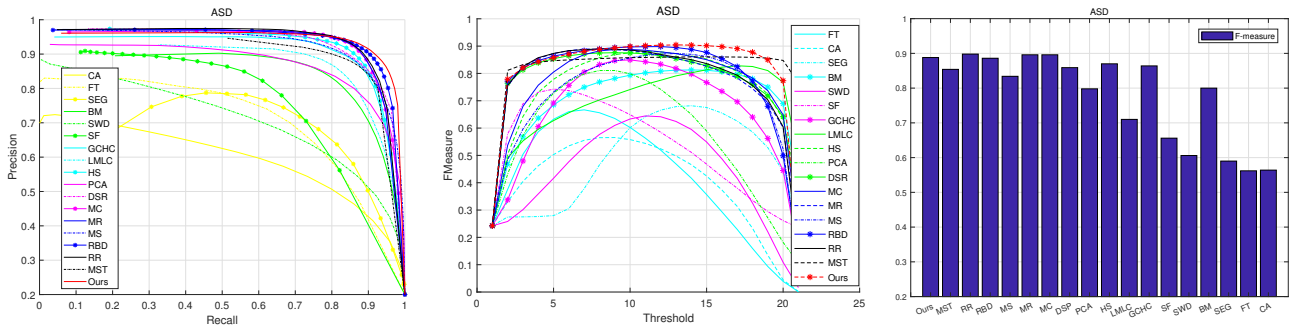


Fig. 9 The PR-curves, F-measure curves and F-measure values for different methods on ASD dataset.

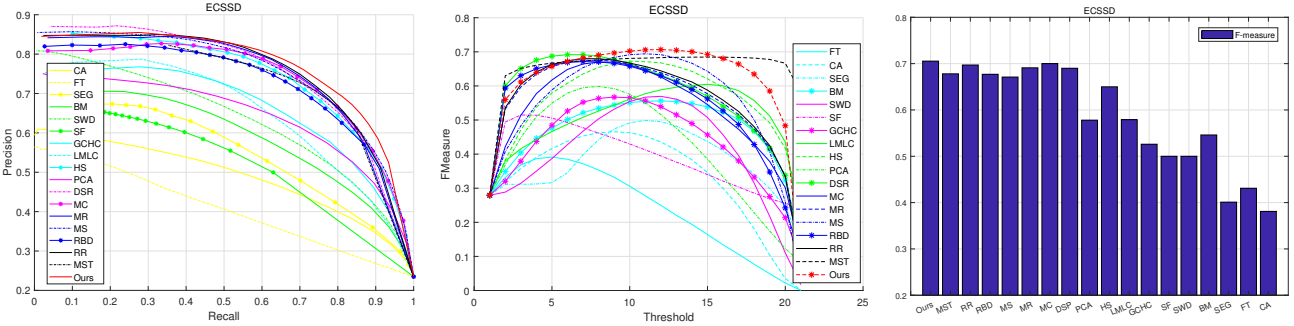


Fig. 10 The PR-curves, F-measure curves and F-measure values for different methods on ECSSD dataset.

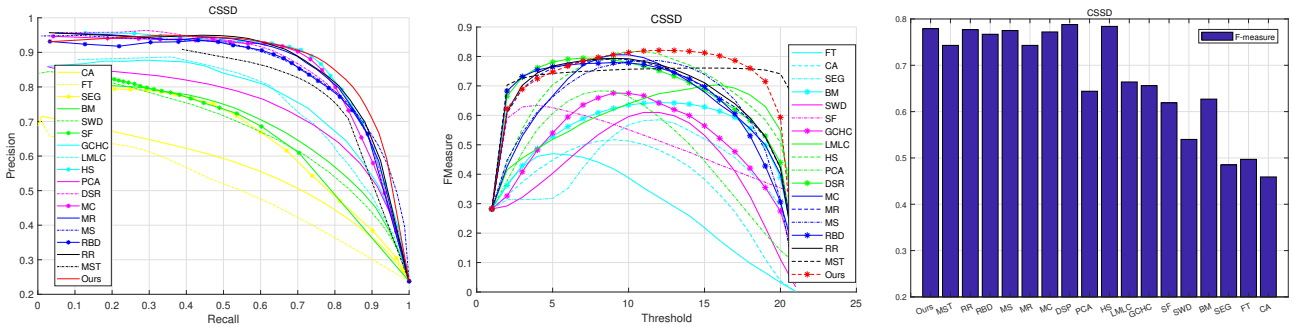


Fig. 11 The PR-curves, F-measure curves and F-measure values for different methods on CSSD dataset.

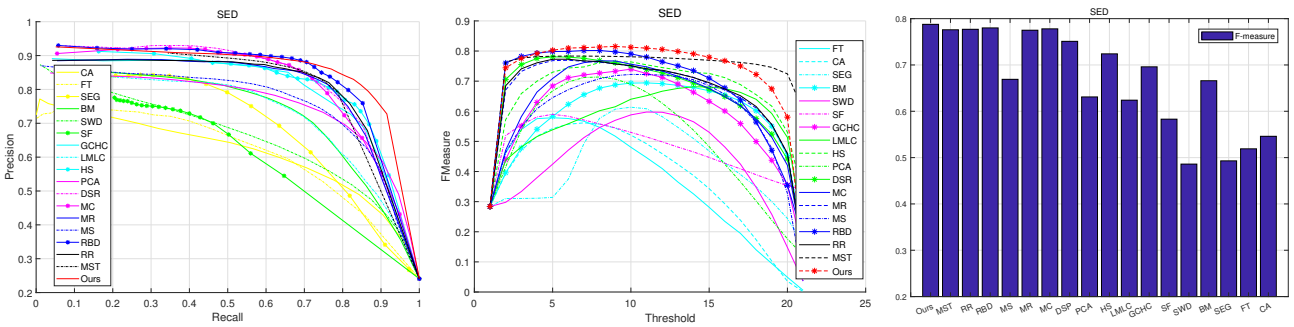


Fig. 12 The PR-curves, F-measure curves and F-measure values for different methods on SED dataset.

which forms an overlapping area that has the similar transient time between the two directions, the model fail to detect saliency maps with high precision. Furthermore, the small objects in the first image cannot be detected in the final

saliency map (as shown in Fig. 14), since the objects encompass a small number of pixels.



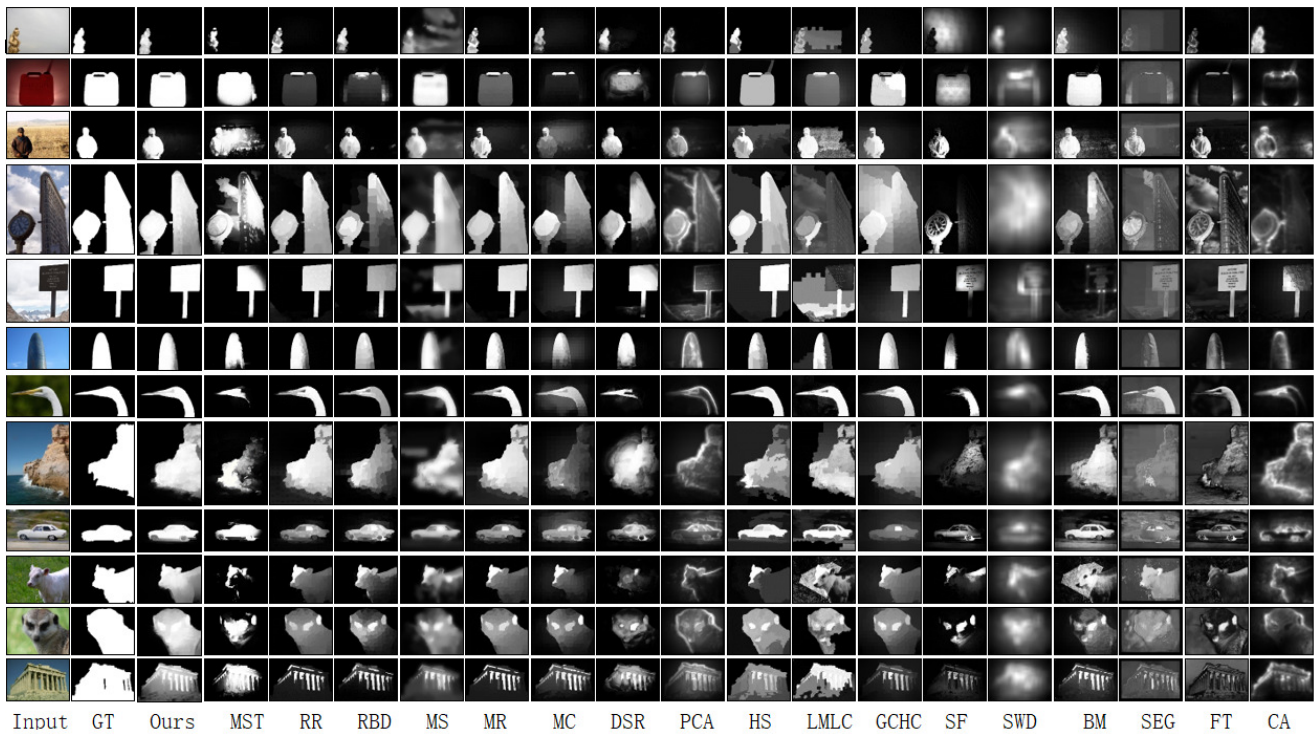


Fig. 13 Examples of output saliency maps results using different algorithms on the ASD, CSSD, ECSSD and SED datasets

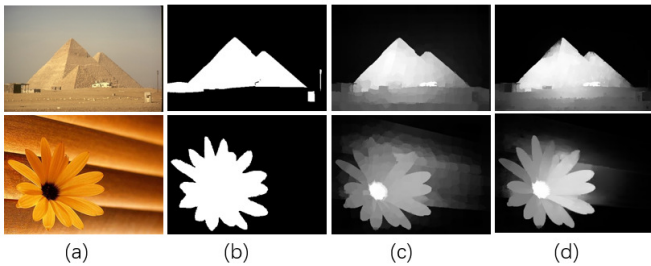


Fig. 14 Examples of our failure examples. (a) Input images. (b) Ground truth. (c) Superpixel-level saliency maps. (d) Final saliency maps.

## 6 Conclusion

In this paper, we propose a novel saliency detection method based on bidirectional absorbing Markov chains by taking into account both the boundary and foreground prior cues. An optimization model is developed to combine background and foreground possibilities, which are acquired through bidirectional absorbing Markov chains. The proposed approach outperforms seventeen recently proposed state-of-the-art approaches over four benchmark datasets as a whole. Although the model can get efficient results, but it just use the CIELab feature of image. In future, we intend to employ multimodal features to address this limitation and improve the overall performance. In addition, we intend to exploit our proposed saliency detection algorithm to other vision tasks including

video saliency detection, image segmentation and object detection.

**Acknowledgements** 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), the Fundamental Research Funds for the Central Universities of China(Grant No.ACAIM190302) and Universities Joint Key Laboratory of Photoelectric Detection Science and Technology in Anhui Province(Grant No.2019GDTCZD02).

## Conflict of interest

Compliance with Ethical Standards

Conflict of Interest

Fengling Jiang, Bin Kong, Jingpeng Li, Kia Dashtipour and Mandar Gogate declare that they have no conflict of interest.

Informed Consent

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5). Additional informed consent was obtained from all patients for which identifying information is included in this article.

Human and Animal Rights

This article does not contain any studies with human or animal subjects performed by the any of the authors

## References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on. 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 Analysis and Machine Intelligence (TPAMI)* **34**(11), 2274–2282 (2012)
3. Alpert, S., Galun, M., Brandt, A., Basri, R.: Image segmentation by probabilistic bottom-up aggregation and cue integration. *IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI)* **34**(2), 315–327 (2012)
4. Borji, A.: Boosting bottom-up and top-down visual features for saliency estimation. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 438–445. IEEE (2012)
5. Charles, M., Grinstead, J., Snell, L.: Introduction to probability. American Mathematical Society (1997)
6. Chen, C., Li, Y., Li, S., Qin, H., Hao, A.: A novel bottom-up saliency detection method for video with dynamic background. *IEEE Signal Processing Letters* **25**(2), 154–158 (2018)
7. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR. pp. 409–416 (2011)
8. Cholakkal, H., Johnson, J., Rajan, D.: Backtracking scspm image classifier for weakly supervised top-down saliency. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 5278–5287. IEEE (2016)
9. Du, J., Li, W., Xiao, B., Nawaz, Q.: Medical image fusion by combining parallel features on multi-scale local extrema scheme. *Knowledge-Based Systems* **113**, 4–12 (2016)
10. Duan, L., Wu, C., Miao, J., Qing, L., Fu, Y.: Visual saliency detection by spatially weighted dissimilarity. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 473–480. IEEE (2011)
11. Fang, Y., Lin, W., Lau, C.T., Lee, B.S.: A visual attention model combining top-down and bottom-up mechanisms for salient object detection. In: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. pp. 1293–1296. IEEE (2011)
12. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. *IEEE transactions on pattern analysis and machine intelligence* **34**(10), 1915–1926 (2012)
13. Gopalakrishnan, V., Hu, Y., Rajan, D.: Random walks on graphs for salient object detection in images. *IEEE Transactions on Image Processing* **19**(12), 3232–3242 (2010)
14. He, S., Lau, R.W., Yang, Q.: Exemplar-driven top-down saliency detection via deep association. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5723–5732 (2016)
15. Hu, W., Hu, R., Xie, N., Ling, H., Maybank, S.: Image classification using multiscale information fusion based on saliency driven nonlinear diffusion filtering. *IEEE transactions on image processing* **23**(4), 1513–1526 (2014)
16. Hussain, C.A., Rao, D.V., Masthani, S.A.: Robust pre-processing technique based on saliency detection for content based image retrieval systems. *Procedia Computer Science* **85**, 571–580 (2016)
17. Itti, L., Koch, C.: Computational modelling of visual attention. *Nature reviews neuroscience* **2**(3), 194 (2001)
18. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI)* **20**(11), 1254–1259 (1998)
19. Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M.H.: Saliency detection via absorbing markov chain. In: Proc. of IEEE Int. Conf. on Computer Vision (ICCV). pp. 1665–1672. IEEE (2013)
20. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: A discriminative regional feature integration approach. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 2083–2090. IEEE (2013)
21. Lee, G., Tai, Y.W., Kim, J.: Eld-net: An efficient deep learning architecture for accurate saliency detection. *IEEE transactions on pattern analysis and machine intelligence* **40**(7), 1599–1610 (2018)
22. Li, C., Yuan, Y., Cai, W., Xia, Y., Dagan Feng, D.: Robust saliency detection via regularized random walks ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2710–2717 (2015)
23. Li, G., Yu, Y.: Deep contrast learning for salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 478–487 (2016)
24. Li, X., Lu, H., Zhang, L., Ruan, X., Yang, M.H.: Saliency detection via dense and sparse reconstruction. In: Proc. of IEEE Int. Conf. on Computer Vision (ICCV). pp. 2976–2983. IEEE (2013)
25. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(2), 353–367 (2011)
26. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 1139–1146. IEEE (2013)
27. Murabito, F., Spampinato, C., Palazzo, S., Giordano, D., Pogorelov, K., Riegler, M.: Top-down saliency detection driven by visual classification. *Computer Vision and Image Understanding* (2018)
28. Peng, H., Li, B., Ling, H., Hu, W., Xiong, W., Maybank, S.J.: Salient object detection via structured matrix decomposition. *IEEE transactions on pattern analysis and machine intelligence* **39**(4), 818–832 (2017)
29. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 733–740. IEEE (2012)
30. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. In: Proc. of European Conf. on Computer Vision (ECCV). pp. 366–379. Springer (2010)
31. de San Roman, P.P., Benois-Pineau, J., Domenger, J.P., Paclet, F., Cataert, D., De Ruy, A.: Saliency driven object recognition in egocentric videos with deep cnn: toward application in assistance to neuroprostheses. *Computer Vision and Image Understanding* **164**, 82–91 (2017)
32. Sun, J., Xie, J., Liu, J., Sikora, T.: Image adaptation and dynamic browsing based on two-layer saliency combination. *IEEE Trans. on Broadcasting* **59**(4), 602–613 (2013)
33. Sun, J., Lu, H., Liu, X.: Saliency region detection based on markov absorption probabilities. *IEEE Transactions on Image Processing* **24**(5), 1639–1649 (2015)
34. Tong, N., Lu, H., Zhang, L., Ruan, X.: Saliency detection with multi-scale superpixels. *IEEE Signal Processing Letters* **21**(9), 1035–1039 (2014)
35. Tu, W.C., He, S., Yang, Q., Chien, S.Y.: Real-time salient object detection with a minimum spanning tree. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2334–2342. IEEE (2016)
36. Tu, Z., Zheng, A., Yang, E., Luo, B., Hussain, A.: A biologically inspired vision-based approach for detecting multiple moving objects in complex outdoor scenes. *Cognitive Computation* **7**(5), 539–551 (2015)
37. Wang, L., Wang, L., Lu, H., Zhang, P., Ruan, X.: Saliency detection with recurrent fully convolutional networks. In: European Conference on Computer Vision. pp. 825–841. Springer (2016)
38. Wang, W., Shen, J., Porikli, F.: Saliency-aware geodesic video object segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3395–3402 (2015)

39. Wang, X., Lv, Q., Wang, B., Zhang, L.: Airport detection in remote sensing images: a method based on saliency map. *Cognitive neurodynamics* **7**(2), 143–154 (2013)
40. 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)
41. Xie, Y., Lu, H.: Visual saliency detection based on bayesian model. In: Proc. of the 18th IEEE Int. Conf. on Image Processing (ICIP). pp. 645–648. IEEE (2011)
42. Xie, Y., Lu, H., Yang, M.H.: Bayesian saliency via low and mid level cues. *IEEE Trans. on Image Processing* **22**(5), 1689–1698 (2013)
43. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 1155–1162. IEEE (2013)
44. Yan, Y., Ren, J., Zhao, H., Sun, G., Wang, Z., Zheng, J., Marshall, S., Sorağan, J.: Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos. *Cognitive Computation* **10**(1), 94–104 (2018)
45. Yang, C., Zhang, L., Lu, H.: Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Processing Letters* **20**(7), 637–640 (2013)
46. Yang, W., Li, D., Wang, S., Lu, S., Yang, J.: Saliency-based color image segmentation in foreign fiber detection. *Mathematical and Computer Modelling* **58**(3-4), 852–858 (2013)
47. Yuan, Y., Li, D., Meng, M.Q.H.: Automatic polyp detection via a novel unified bottom-up and top-down saliency approach. *IEEE journal of biomedical and health informatics* **22**(4), 1250–1260 (2018)
48. Zhan, J., Zhao, H., Zheng, P., Wu, H., Wang, L.: Salient super-pixel visual tracking with graph model and iterative segmentation. *Cognitive Computation* pp. 1–12 (2019)
49. Zhang, J., Sclaroff, S.: Saliency detection: A boolean map approach. In: Proceedings of the IEEE international conference on computer vision. pp. 153–160 (2013)
50. Zhang, L., Ai, J., Jiang, B., Lu, H., Li, X.: Saliency detection via absorbing markov chain with learnt transition probability. *IEEE Transactions on Image Processing* **27**(2), 987–998 (2018)
51. Zhang, Q., Luo, D., Li, W., Shi, Y., Lin, J.: Two-stage absorbing markov chain for salient object detection. In: Image Processing (ICIP), 2017 IEEE International Conference on. pp. 895–899. IEEE (2017)
52. Zhang, W., Xiong, Q., Shi, W., Chen, S.: Region saliency detection via multi-feature on absorbing markov chain. *The Visual Computer* **32**(3), 275–287 (2016)
53. Zhao, R., Ouyang, W., Li, H., Wang, X.: Saliency detection by multi-context deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1265–1274 (2015)
54. Zhu, G., Wang, Q., Yuan, Y.: Tag-saliency: Combining bottom-up and top-down information for saliency detection. *Computer Vision and Image Understanding* **118**, 40–49 (2014)
55. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2814–2821. IEEE (2014)