Salient object detection based on distribution-edge guidance and iterative Bayesian optimization

Chenxing Xia[1] · Xiuju Gao[2] · Kuan-Ching Li[3] · Qianjin Zhao[1] · Shunxiang Zhang[1]

Abstract

Salient object detection has witnessed rapid progress, despite most existing methods still struggling in complex scenes, unfortunately. In this paper, we propose an efficient framework for salient object detection based on distribution-edge guidance and iterative Bayesian optimization. By considering color, spatial, and edge information, a discriminative metric is first constructed to measure the similarity between different regions. Next, boundary prior embedded with background scatter distribution is utilized to yield the boundary contrast map, and then a contour completeness map is derived through a wholly closed shape of the object. Finally, the above both maps are jointly integrated into an iterative Bayesian optimization framework to obtain the final saliency map. Results from an extensive number of experimentations demonstrate that the promising performance of the proposed algorithm against the state-of-the-art saliency detection methods in terms of different evaluation metrics on several benchmark datasets.

Keywords Background scatter · Contour completeness · Iterative Bayesian · Salient object detection

1 Introduction

Saliency detection is concentrated on detecting the most attractive objects in an image. Recently, this area has witnessed rapid progress. As a preprocessing procedure, automatic saliency detection has been widely used in a variety of computer vision tasks such as image segmentation [1], object recognition [2], compression [3], image retrieval [4], de-blurring [5], and others.

Several saliency models have been proposed in the past years. Due to the lack of a uniform definition of salient objects, most salient object detection methods are based on effective assumptions. Contrast prior is one of the most popular principles adopted by various kinds of models from either a local or global view [12, 13]. Essentially,

local contrast-based methods [14] prefer to detect the high-frequency information such as edges, failing to pop out the salient holistic object, as shown in Fig. 1b. On the contrary, global contrast-based methods can locate the salient object while the performance of these methods is limited in such scenarios when the foreground regions are complex and with diverse appearance, as shown in Fig. 1c [13].

To address the limitation of the contrast cue, boundary prior is also applied to detect salient regions, where the image boundary areas are looked upon as background [15–17]. For example, Yang et al. [15] generated the saliency of image regions via manifold ranking, in which the regions on four boundary sides are treated as queries. However, such methods are fragile and limited when the object appears on the image boundary. To overcome this limitation, several improvement mechanisms have been proposed. Lu et al. [18] proposed an effective method based on dense and sparse reconstruction using the background templates. Xia et al. [19] exploited corner information together with the convex hull to extract background seeds. Li et al. [9] removed the top 30% border pixels with a considerable color difference among a border set. Wang et al. [20] dropped the untrustworthy superpixels with sharp edge out of the border set to obtain the reliable background regions. Zhou et al. [21] used diffusion-based techniques on the proposed sparse graph. All of them perform well

✉ Xiuju Gao
  xccgxj1314@163.com

1 College of Computer Science and Engineering,
   Anhui University of Science and Technology, Huainan, China

2 College of Electrical and Information Engineering,
   Anhui University of Science and Technology, Huainan, China

3 Department of Computer Science and Information
   Engineering, Providence University, Taichung, Taiwan

Fig. 1 Examples to illustrate the differences among various saliency models. (**a**) Input; (**b**) FT [6]; (**c**) PCA [7]; (**d**) GS [8]; (**e**) LPS [9]; (**f**) RR [10]; (**g**) MAP [11]; (**h**) Ours; (**i**) Ground-truth
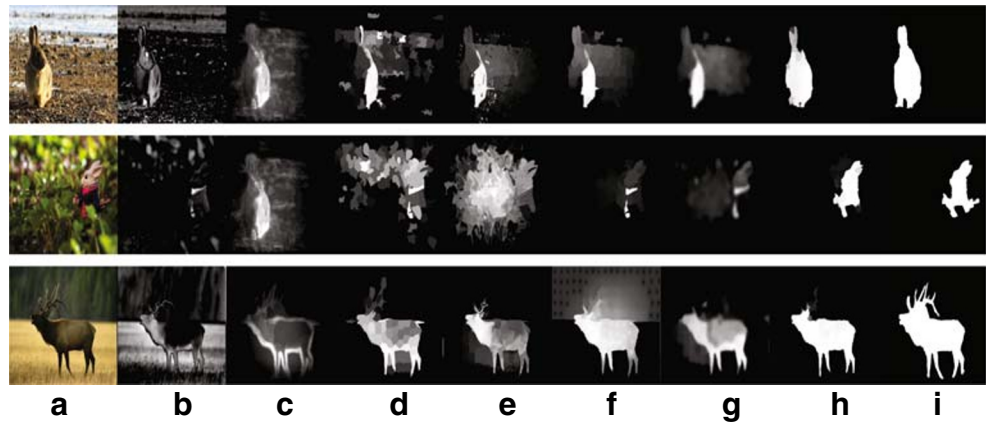
in easy cases, but they still struggle in complex scenes for two reasons: 1) the distinctiveness of their calculation is confined to the border set instead of the entire image; 2) their constructed similarity metric is weak, as shown in Fig. 1d-e.

Furthermore, it is insufficient for saliency detection based on simple boundary prior, especially when the image has an intricate background structure or the salient object has a similar appearance with non-salient. Accordingly, a variety of methods have been proposed. Some researchers committed to conducting a robust discriminative matrix based on high-level information to enhance the difference between object and background. Lee et al. [22] concatenated encoded low-level distance map and high-level features to compute saliency. Hou et al. [23] introduced short connections to the Holistically-Nested Edge Detector (HED), which can effectively combine rich multi-scale features to identify the whole salient object and accurately capture its boundary. Liu et al. [24] proposed Deeply-supervised Nonlinear Aggregation (DNA) to make better use of the complementary information on various side-outputs by using a nonlinear side-output prediction aggregation. Even though these models show promising performance on benchmark datasets, they need to collect large hand-labeled images, yet expensive to set up the

learning framework. On the contrary, based on the boundary prior, different propagation models have been put forward to improve the visual quality of saliency maps. Li et al. [10] estimated the saliency via regularized random walks ranking, Jiang et al. [11] formulated saliency detection based on Markov absorption probabilities on an image graph model, and Yu et al. [25] presented a cross-diffusion process for salient object detection. However, they still suffer from a saliency overestimation of the background when the background is clustered, and the similarity metric is inefficient. Besides, they may not be able to highlight salient objects fully and may even produce erroneous results, because the graph-based approaches ignore the overall consistency between different regions of the salient object, as depicted in Fig. 1f and g.

In this paper, we propose a novel saliency detection method, which is a hybrid of background scatter and foreground contour completeness. The pipeline of the proposed method is depicted in Fig. 2. First, we derive superpixels from an edge guided segmentation. Different from previous methods for the extraction of a robust background, we consider it from the global view. Namely, the background scatter of the whole image is introduced to remove the foreground noise from the border set (BS)
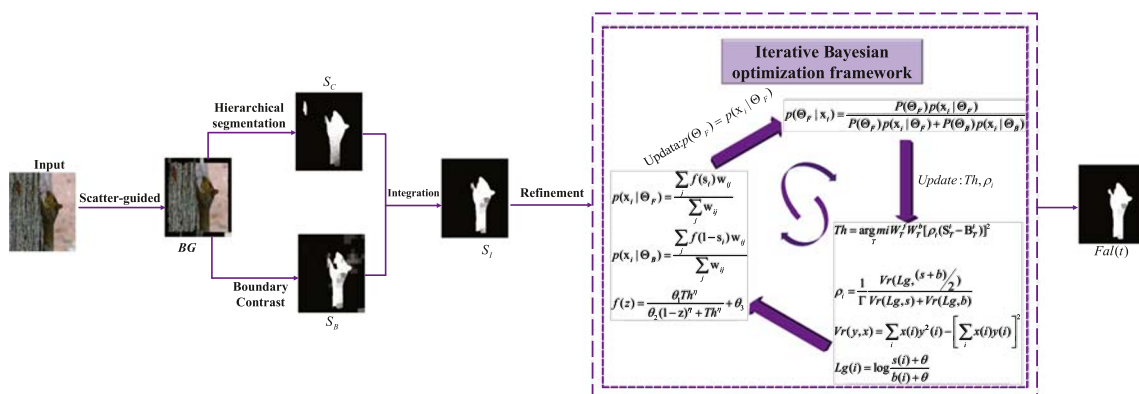

Fig. 2 Main pipeline of the proposed saliency detection algorithm on an example image

Salient object detection based on distribution-edge guidance and iterative Bayesian optimization...

2979

by setting different thresholds for different boundaries, so then the solid boundary background (BG) can be achieved. Based on the boundary prior, a boundary contrast saliency map is then generated. Besides, with the help of the BG together with the boundary contrast saliency map, a contour completeness map can be produced by considering the expectation of times of the regions being activated in a hierarchical segmentation space. Finally, we refine the integrated result to obtain a more smooth and accurate saliency map based on an iterative Bayesian optimization framework. Comprehensively, the main contributions of this proposed research lie in the following aspects:

(1) We proposed a novel model for saliency detection via background scatter and foreground contour completeness. A selective mechanism for robust background nodes is presented based on background scatter. To maintain the completeness of the salient object, the contour completeness saliency map is derived via the completely closed shape of the object.

(2) An iterative Bayesian optimization framework is proposed to achieve improved performance by optimizing the integration result of background-based saliency map and contour completeness saliency map into a more favorable result.

(3) Extensive experiments and comprehensive analysis are conducted on four public datasets that demonstrate the effectiveness of the proposed method against state-of-the-art saliency detection methods.

The work in this paper is a substantial extension of our preliminary study titled "A hybrid of background scatter and foreground contour completeness for salient object detection" [26] (referred to as BSFC). Three-fold extensions are made upon BSFC in this paper, namely algorithm, framework, and experiment. Specifically, this paper proposes to additionally develop a discriminative metric to measure the similarity between different nodes, which can improve the effectiveness of the proposed algorithm (as demonstrated in Section 2.1). Secondly, we improve the original framework by adding a refinement step, which can highlight the foreground while suppressing the background (as demonstrated in Section 2.5). Lastly, we conducted more plentiful experiments by comparing with more state-of-the-art methods and implementing the more comprehensive and in-depth analysis of the algorithm.

2 The proposed approach

The remainder of the proposed algorithm is organized as follows. First, a mechanism for extracting robust boundary background regions based on scatter prior is proposed. Next, two saliency cue maps (boundary contrast map and contour completeness map) are generated via two different computational schemes, and lastly, refinement to obtain the final saliency map is proposed.

2.1 Background scatter

The structured random forest edge detector [27] to derive the probability of boundary (PB) is adopted, so then superpixels are obtained by thresholding an ultrametric contour map (UCM) based on PB. Let us define the initial segmentation $P_0 = \{R_i\}$ and the number of regions is K_0. Based on the superpixels segmentation result, an undirected weighted graph $G = < V, E >$ is constructed, where V is the set of superpixels, and E is weighted by an affinity matrix.

Boundary prior is a handy cue that has been widely used in saliency detection that assumes the image boundary as background regions. However, it may lead to adverse effects if we directly use all border nodes as background when the object appears on the border. Accordingly, we perform boundary analysis to exclude salient superpixels on the four boundaries of the image. In an image, the salient regions usually represent a similar color and compact spatiality as well, while non-salient regions generally show a different appearance and loose scatter around. Hence, we introduce the scatter prior measured by spatial variance that takes the complete image as the research object instead of the border set (BS). In detail, it is defined as below:

$$Div_c(i) = normalize(\sum_{j=0}^{N} w_{ij} \cdot |s_j - \rho_i|), \qquad (1)$$

$$\rho_i = normalize(\sum_{j=0}^{N} w_{ij} \cdot s_j). \qquad (2)$$

Herein, s_i and ρ_i indicate the position and the weighted mean position of superpixel i, respectively; $N = K_0$ represents the number of superpixels; $normalize(x)$ is a function that normalizes x to [0,1]; w_{ij} is the edge weight used to measure the similarity between different nodes. At this point, a joint metric taking the color difference, spatial distance, and edge information into consideration is adopted to compute the edge weight w_{ij}. The following aspects are considered:

(1) According to the cognitive property of color similarity, image regions with similar colors usually belong to the same class.

(2) According to spatial proximity property, spatially adjacent image regions are likely to share the same label.

(3) In some cases, using the edge map can be better to highlight the outline between foreground and background than (1) and (2), as shown in Fig. 3.

Fig. 3 Effects of edge-weight.
(a) Input; **(b)** Only using color
affinity matrix; **(c)** Only using
spatial distance affinity matrix;
(d) Using both color and spatial
distance affinity matrix; **(d)**
Edge detection by [27]; **(e)** The
proposed affinity matrix

a   b   c   d   e   f
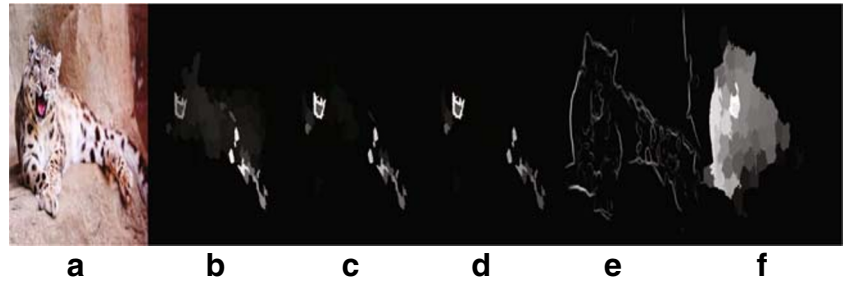
Based on the above considerations, the proposed edge weight is formulated as:

$$w_{ij} = exp(-\frac{d_c(i,j) + d_s(i,j) + d_e(i,j)}{2\sigma_w^2}), \quad (3)$$

where σ_w controls the strength of weight between a pair of nodes, $\sigma_w = 0.1$ is empirically set, $d_c(i,j)$, $d_s(i,j)$, and $d_e(i,j)$ represents the color difference, spatial distance, and contour magnitude, respectively. The color difference $d_c(i,j)$ is defined:

$$d_c(ij) = |R_i - R_j|, \quad (4)$$

where $|R_i - R_j|$ denotes the Euclidean distance between superpixel i and j in CIELab color space. For spatial distance $d_s(i,j)$, we combine geodesic distance and sine spatial distance to measure the distance between superpixel i and j:

$$s_{ij} = sqrt\left(\min\left(X_{ij}, I_x - X_{ij}\right)^2 + \min\left(Y_{ij}, I_y - Y_{ij}\right)^2\right),$$
$$s.t. \; X_{ij} = |sin(\pi \cdot |x_i - x_j|)|$$
$$Y_{ij} = |sin(\pi \cdot |y_i - y_j|)| \quad (5)$$

where x_i, y_i, I_x and I_y refer to the x- and y-coordinates of the region i and image, respectively. The calculation mechanism can realize the connection of the regions with the same structure well in an arbitrary shape and range. $d_e(i,j)$ is defined as the accumulated weight along the shortest path on the graph, i.e.

$$d_e(ij) = \min_{u_1 = i, \cdots, u_k = j} \sum_{j=1}^{k-1} e_{u_j u_{j+1}}, \quad (6)$$

where the weight of graph edge e_{ij} is assigned as the strength of image edge between i and j in terms of UCM.

As shown in Table 1, it is noted that there are different probabilities of the object appearing at different boundaries. Accordingly, we choose different threshold values for different sides instead of setting a single value, and a robust boundary background set BG can be obtained from the border set BS. As shown in Fig. 4, the proposed method is superior to other methods.

2.2 Boundary contrast map

According to the boundary prior, a saliency map can be obtained by measuring the contrast with the robust background BG. However, superpixels in the background may only be similar to a portion of the border set rather than all of the background nodes. Therefore, it is unwise to measure the saliency value of a superpixel by counting all the differences between corresponding superpixel and the background seeds. To solve it, several methods have been proposed in previous works [10, 28]. In this paper, we put forward a distinctive way of measuring the boundary contrast map. First, the $K - means$ algorithm is applied to divide the set BG into K clusters (empirically set $K = 3$). For each cluster k, we compute the mean color $\bar{c}_k^m, m \in \{L, a, b, Lab\}$ in CIELab color space and the color covariance matrix Var_k. Also, for each superpixel, we select one of the K clusters that have the minimum difference from the superpixel. Then the saliency of superpixel R_i can be obtained based on the Mahalanobis distance:

$$S_B(R_i) = \sum_{m \in \{L, a, b, Lab\}} v^m(R_i), \quad (7)$$

and

$$v^m(R_i) = \sum_{(w,h) \in R_i}^{|R_i|} P_{R_i} \sqrt{(c_g^m(w,h) - \bar{c}_g^m)(Var_g^m)^{-1}(c_g^m(w,h) - \bar{c}_g^m)^T}, \quad (8)$$

where g represents the index of the cluster that has the smallest difference with R_i; (w, h) indicates a pixel in superpixel R_i; $|R_i|$ is the cardinality of R_i; P_{R_i} provides the proportion of superpixel R_i compared with the whole

Table 1 Probability distribution on different datasets

Dataset	Top(%)	Down(%)	Left(%)	Right(%)
Dataset	Top(%)	Down(%)	Left(%)	Right(%)
ASD	0.2	1.6	0.3	0.5
ECSSD	5.3	22.4	6.8	7
DUT	2.48	14.4	5.2	4.3

Fig. 4 Robust boundary background extraction. **(a)** Input; **(b)** Original border set BS; **(c)-(e)** Boundary background extracted by different methods: BFS [20], LPS [9] and Ours
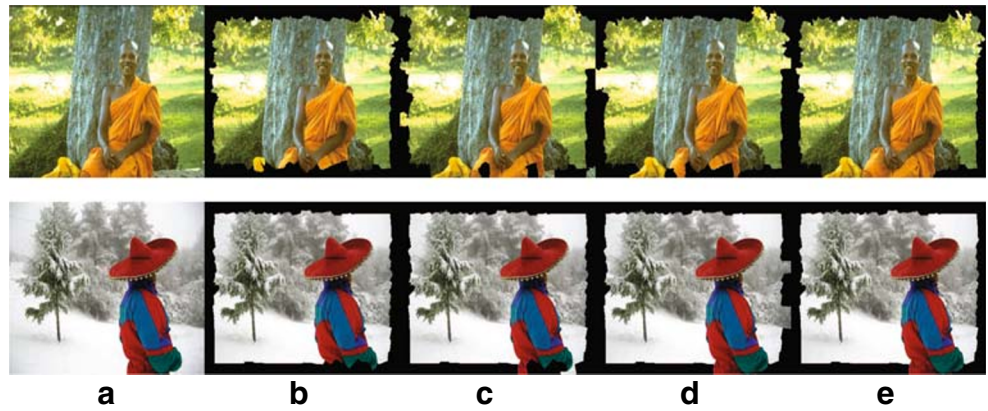
image. Then the background contrast saliency map S_B can be achieved by normalizing the obtained saliency values to [0,1]. Fig. 5 shows the results of S_B. It can be seen that S_B can highlight the salient regions.

2.3 Contour completeness map

Boundary contrast map S_B in collaboration with BG works well, whereas in some complex scenes. As is shown in the second example of Fig. 5b, relying solely on the boundary prior might lead to highlight the background regions erroneously. Generally, the salient object has a well-defined closed boundary [30]. As a consequence, the contour completeness cue is introduced to facilitate the salient object detection.

A hierarchical segmentation P_ξ can be obtained by setting different $\xi \in [0, \xi_N]$ for the derived hierarchical UCM. In this paper, we compute the expectation of times for which a region is superimposed in the whole hierarchical segmentation space to determine whether a region has a closed outline or not in segmentation $P_\xi = \{R_i\}$. That is, the more times the holistic homogeneous regions are activated,

the more salient they are. An indicator map Q at level ξ is defined:

$$Q_\xi(x) = \begin{cases} 1 & if x \in R, R \in P_\xi^{in} \\ s & if x \in R, R \in P_\xi^{out}, \end{cases} \quad (9)$$

where $P_\xi^{in} = \{R_i | R_i \cap BG = \emptyset\}$ is the set of inner regions, $P_\xi^{out} = P_\xi \setminus P_\xi^{in}$ is the set of outer regions. Different from [29], we adopt BG instead of BS in P_ξ^{in}. The rationale behind this is that it may yield a bad result if we adopt BS in P_ξ^{in} when the salient objects connect to the boundary (see the third row in Fig. 5d).

Because the accuracy of the saliency map is sensitive to the number of superpixels, the problem of an optimal threshold selection needs to be solved in thresholding segmentation. However, it is difficult for us to obtain an optimal threshold value by computing directly. To handle the problem, we first introduce saliency weights to assign a different probability to each pixel belonging to be the salient object. Here, we adopt the saliency value of S_B as the weights of each pixel, defined as s. By statistically averaging the expectation of the entire

Fig. 5 Boundary contrast map and contour completeness map. From left to right: **(a)** Input; **(b)** S_B; **(c)** Contour completeness map without processing [29]; **(d)** S_C; **(e)** Ground-truth
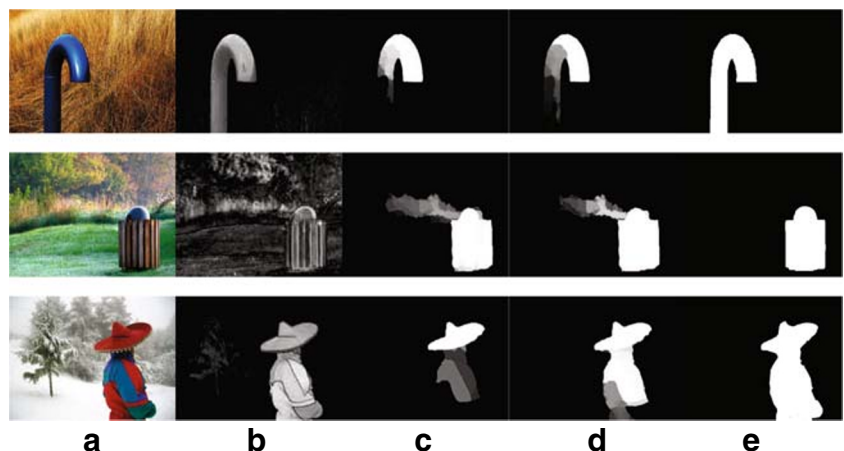
hierarchical segmentation space, we obtain the contour completeness saliency map:

$$S_C(x) = \int Q_\xi(x) \cdot p(\xi)d\xi, \qquad (10)$$

where ξ obeys a uniform distribution with the probability density function P_ξ. Besides, we compute the saliency map at different resolutions to further deal with the problem that salient objects are likely to appear at different scales, so then they are resized into the standard size and integrated to retrieve the strong map S_C, as shown in Fig. 5e. From observations, we note that the proposed method can work well in the cases where salient regions appear at the boundary.

2.4 Integration

The saliency maps measured by boundary contrast and contour completeness are complementary to each other: the former prefers to highlight the object while the latter aims to suppress noise. Thus, effective integration is a wise operation. Although previous works adopt linear combinations to fuse individual saliency maps using weight for simplicity, it is not feasible because such processing will either weaken the suppression of background noise or interfere with foreground detection.

In order to better achieve the purpose of seeking common ground while reserving differences, we have designed a fine-grained fusion method. The mathematical definition of the fusion mechanism can be expressed as follows:

$$\begin{aligned} S_I = {}&\Upsilon\{S_C^* \cdot S_B^* == 1\}max(S_C, S_B) \\ &+\Upsilon\{S_C^* + S_B^* == 0\}min(S_C, S_B) \\ &+\Upsilon\{S_C^* \neq S_B^*\}exp(S_C) \cdot exp(S_B). \end{aligned} \qquad (11)$$

Here $\Upsilon\{x\} = 1$ if x is true, and $\Upsilon\{x\} = 0$ otherwise. S^* is a saliency vector by binarizing a saliency map with a threshold calculated from S using the algorithm in [31]. Figure 6d demonstrates that the integration map is more accurate than the saliency map of Fig. 6b and c in the detection of foreground and background, which further proves the effectiveness of the proposed integration method.

2.5 Iterative Bayesian optimization framework

In order to optimize the integration results, an iterative Bayesian optimization framework is proposed. According to Bayesian inference, posterior probability $P(\Theta_k, k \in \{F, B\}|x_i)$ representing the probability that region x_i belongs to the salient Θ_F (background Θ_B) can be computed as:

$$P(\Theta_F|x_i) = \frac{P(\Theta_F)P(x_i|\Theta_F)}{P(\Theta_F)P(x_i|\Theta_F) + P(\Theta_B)P(x_i|\Theta_B)}, \qquad (12)$$

where $P(x_i|\Theta_k, k \in \{F, B\})$ is a conditional probability. $P(\Theta_F)$ and $P(\Theta_B) = 1- P(\Theta_F)$ are the prior probabilities of being salient and background, respectively. In this work, we set the fusion result S_I as the initial prior probability, i.e., $P(\Theta_F) = S_I$ and $P(\Theta_B)=1-S_I$. How to calculate the conditional probability will be described in detail as follows.

The key to the method based on the Bayesian framework is to obtain conditional probability. The current conventional method is to obtain a conditional probability by calculating the proportion of each quantized feature in a binary image. However, a simple thresholding operation may mislead further inferences, and it is also challenging to get a reliable threshold. Therefore, a soft approach is proposed based on iterative optimization to solve the problem. We define the probability of each region as the weighted average of all the saliency values. The conditional probabilities can be defined by [32]:

$$P(x_i|\Theta_F) = \frac{\sum_j f(SF_i)w_{ij}}{\sum_j w_{ij}}, \qquad (13)$$

$$P(x_i|\Theta_B) = \frac{\sum_j f(1 - SF_i)w_{ij}}{\sum_j w_{ij}}, \qquad (14)$$

Fig. 6 Integration and refinement. **(a)** Input; **(b)** S_B; **(c)** S_C; **(d)** S_I; **(e)** Final map without iteration $Fal(t0)$; **(f)** Final map $Fal(t)$; **(g)** Ground-truth

a   b   c   d   e   f   g
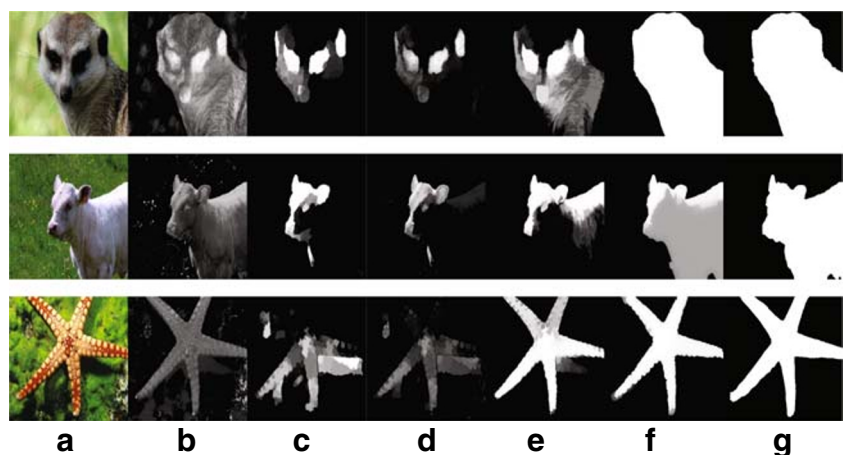
$$f(z) = \frac{\theta_1 Th^\tau}{\theta_2 (1-z)^\tau + Th^\tau} + \theta_3. \tag{15}$$

$f(z)$ is an enhancement function that makes the difference between foreground and background more apparent while keeping the saliency value of superpixel near the threshold (Th). The parameter θ_i is mainly to make sure that $f(z)$ can pass through three defined coordinate points (($0, 0$), ($1, 1$) and (Th, Th)), and $\tau = 4$. In this paper, we get the optimal threshold Th by size prediction and feature weight:

$$Th = \arg\min_T W_T^f W_T^b [\rho_i (S_T^i - B_T^i)]^2, \tag{16}$$

where $S_T^i (B_T^i)$ and $W_T^f (W_T^B)$ correspond to the mean value and the weight of the foreground (background) histogram in channel i of opponent color space [33], respectively. T denotes the percentage of the potential salient regions, and we set $T \in [0.1, 0.5]$. The rationality of this hypothesis lies in that the size proportion of salient objects is usually in a range from 0.05 to 0.6 [34]. ρ_i is feature weight, which is measured by computing its separating power based on variance ratio.

$$\rho_i = \frac{1}{\Gamma} \frac{Vr(Lg, \frac{s+b}{2})}{Vr(L, s) + Vr(L, b)}, \tag{17}$$

$$Vr(y, x) = \sum_i x(i) y^2(i) - [\sum_i x(i) y(i)]^2, \tag{18}$$

$$Lg(i) = log \frac{s(i) + \phi}{b(i) + \phi}, \tag{19}$$

where ϕ is a constant used to prevent the numerator and denominator from being 0. $\Gamma = \sum_i \rho_i$. $Lg(i)$ is the log-likelihood of foreground and background, where $s(i)$ and $b(i)$ denote the discrete probability distributions of foreground and background, respectively.

Based on the above formulas, we can get $P(\Theta_F | x_i)$. In order to optimize the results, we iteratively re-assign the prior function with $P(\Theta_F) = P(\Theta_F | x_i)$ and the feature weights ρ_i with new ρ_i. To eliminate the disconnectedness and small fragments and improve the accuracy of detection, we utilize the smooth result of $P(\Theta_F | x_i)$ to re-assign the prior function in the experiment. Herein, we adopt the guided filter [35] to smooth $P(\Theta_F | x_i)$, which can well preserve the integrity of the edge of the object. Finally, we denote $Fal(t)$, $t = \{t_0, t_1,$

$\cdots, t_n\}$ as the saliency map corresponding to various iterations (t), and $Fal(t_0)$ indicates the saliency map without iteration. Figure 6e-f show the result refined by the proposed iterative Bayesian optimization framework. Our model can always highlight the full salient objects while suppressing the background, which illustrates the effectiveness of our system.

3 Experimental results

In this paper, we extensively present comparison of the proposed algorithm against 15 state-of-the-art saliency detection methods that includes BFS [20], BL [36], BSCA [28], BSFC [26], FCB [37], CGV [38], GS [8], HCC [29], LPS [9], MR [15], MB [39], MILPS [34], RCRR [40], SF [41], SRD [42] on the ASD [6], DUT-OMRON [15], SOD [43], and ECSSD [44] datasets. For fair evaluation purposes, either the results provided by the original authors be directly used or execute their own implementations through the source codes publicly available. Admittedly, compared with the traditional methods, saliency methods using deep learning such as MDF [45] and ILS [46] can achieve better performance, although they require a large amount of training data. Considering that this paper is mainly focused on unsupervised learning, the proposed method is solely compared to traditional saliency methods.

3.1 Evaluation metrics

To compare the performance, several popular evaluation metrics are used in our experiments, which include the precision-recall curve (PR), average precision (AP), F-measure, overlapping ratio (OR) and mean absolute error (MAE).

We first use the Precision-Recall (PR) curves to evaluate the performance of the proposed method. For each saliency map, binary maps are generated by binarizing a saliency map with fixed thresholds varying from 0 to 255, and the PR curves are obtained by comparing the ground truth with binary maps.

For saliency detection, precision, and recall are often required to be high, though these two indicators are usually negatively correlated. To comprehensively assess the salient object detection model, the weighted harmonic mean of precision (P) and recall (R) called F-measure is calculated, given as:

$$F_\beta = \frac{(1+\beta) P \cdot R}{\beta^2 P + R}, \tag{20}$$

where β^2 is set to 0.3 to emphasize precision [13]. Since a high recall rate can easily be obtained by detecting the complete image as the salient object, it is not possible to guarantee such accuracy, unfortunately.

Noting that the above metrics mainly focus on the probability of salient pixels being correctly detected and ignore the effects of correct assignment of non-salient pixels, and therefore OR and MAE are adopted to address these issues. MAE represents the measure of the similarity between saliency map S and ground-truth G [41]: $MAE = mean(|S - G|)$. OR is defined as the overlapping ratio between the segmented object mask S' and ground truth G:

$OR = |S' \cap G|/|S \cup G|$, where S' can be obtained by binarization of S with an adaptive threshold, i.e., twice the mean values of S as in [47].

3.2 Quantitative results

We extensively present a comparison of the proposed algorithm against 15 recently proposed algorithms on four datasets, and the experimental results are summarized in Fig. 7 and Table 2. The results show that, in most cases, the algorithm ranks first or second in different evaluation indicators on the test datasets.

Specially, we can observe that our model achieves the best performance in terms of AP (0.9372, 0.8158, 0.7848), F-measure (0.9299, 0.7851, 0.7171), and OR (0.85159, 0.58105, 0.44333) on ASD, ECSSD, and SOD datasets.
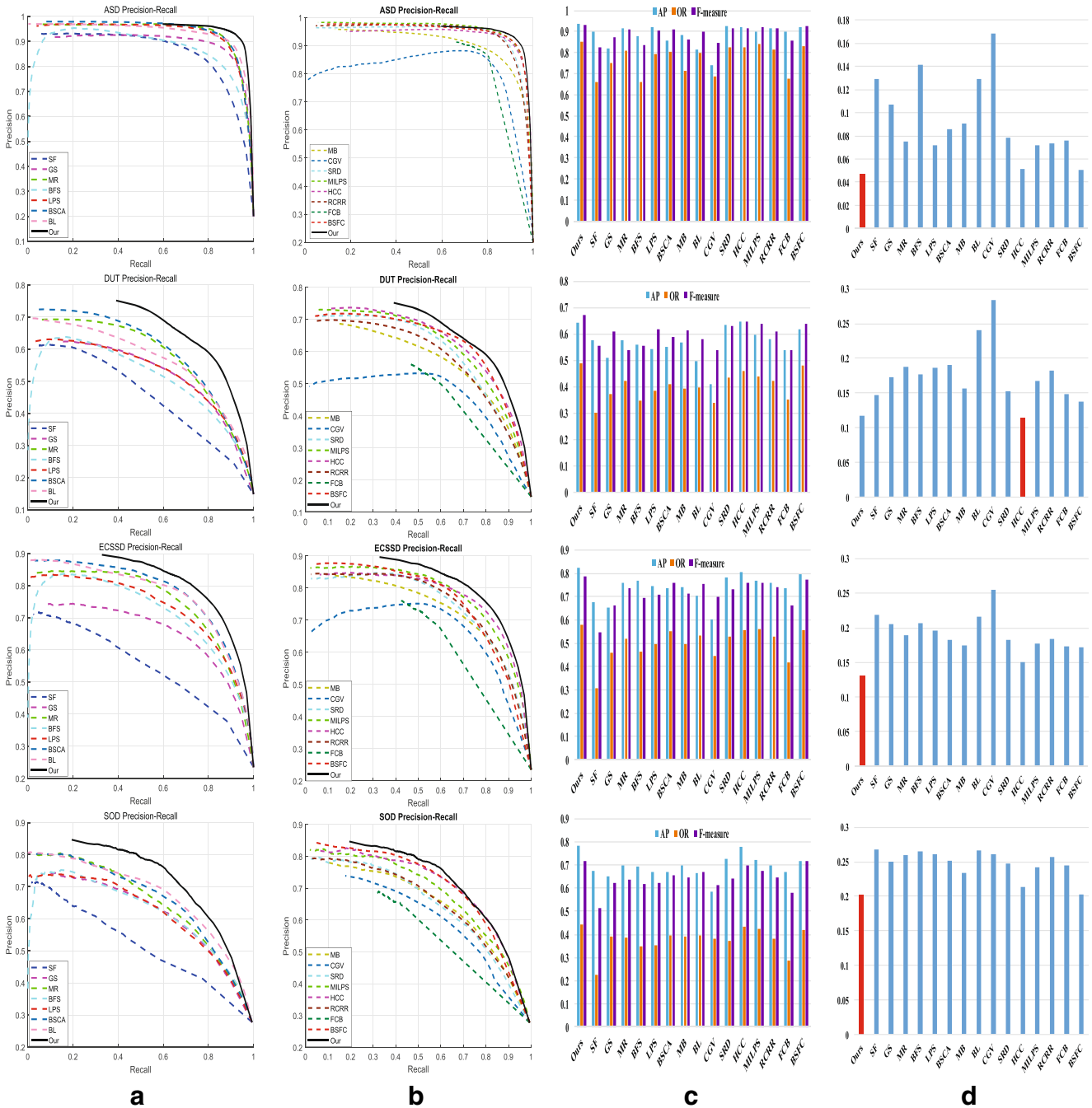

Fig. 7 Quantitative evaluations by precision-recall curves (left two columns), F-measure (third column), and mean absolute error (MAE) (right column) on four benchmark datasets: from top to bottom are ASD, DUT, ECSSD, and SOD

Table 2 Results on three datasets in terms of AP↑, F-Measure↑, OR↑, and MAE↓

	Method	Our	SF	GS	MR	BFS	LPS	BSCA	MB	BL	CGV	SRD	HCC	MILPS	RCRR	FCB	BSFC
(A)AP↑	ASD	0.9372	0.8972	0.8218	0.9146	0.8769	0.9232	0.8597	0.884	0.8146	0.7399	0.9245	0.9193	0.9003	0.9136	0.8981	0.9209
	DUT	0.64	0.5759	0.5079	0.5755	0.5597	0.5413	0.5497	0.5668	0.4953	0.4105	0.634	0.6473	0.5949	0.5796	0.5389	0.6182
	ECSSD	0.8158	0.6779	0.6531	0.7584	0.7665	0.7442	0.738	0.7406	0.7022	0.6012	0.7826	0.8041	0.7698	0.7591	0.7347	0.7958
	SOD	0.7848	0.6767	0.649	0.6962	0.6924	0.6689	0.6705	0.6977	0.6628	0.5848	0.7281	0.7797	0.7214	0.6974	0.6717	0.7149
(B)F-measure↑	ASD	0.9299	0.826	0.875	0.912	0.8373	0.905	0.909	0.865	0.902	0.845	0.9153	0.9157	0.9236	0.9136	0.8551	0.9287
	DUT	0.669	0.4953	0.5559	0.61	0.5391	0.5558	0.617	0.5888	0.5798	0.5388	0.6296	0.6469	0.6389	0.6105	0.5369	0.6386
	ECSSD	0.7851	0.548	0.661	0.736	0.6967	0.709	0.758	0.713	0.7544	0.6984	0.7312	0.7599	0.7581	0.74	0.663	0.7712
	SOD	0.7171	0.5155	0.6205	0.6363	0.618	0.6208	0.6541	0.6449	0.6679	0.6128	0.6398	0.6987	0.6744	0.6461	0.5812	0.7154
(C)OR↑	ASD	0.85159	0.66011	0.75312	0.81013	0.65886	0.79236	0.8022	0.71499	0.79825	0.68599	0.82757	0.82637	0.83868	0.81396	0.67523	0.82867
	DUT	0.48892	0.30194	0.3718	0.41937	0.34424	0.38191	0.4094	0.39236	0.3967	0.33747	0.43557	0.46065	0.43954	0.42112	0.35216	0.47742
	ECSSD	0.58105	0.30856	0.46057	0.52048	0.4641	0.49782	0.54913	0.49661	0.5319	0.4467	0.52703	0.5542	0.56081	0.52905	0.41613	0.55526
	SOD	0.44333	0.22702	0.39015	0.38425	0.34998	0.35056	0.39656	0.38881	0.3939	0.37913	0.36949	0.4316	0.42386	0.37876	0.28702	0.42027
(D)MAE↓	ASD	0.047386	0.12929	0.10725	0.07512	0.14168	0.071469	0.085774	0.090757	0.1291	0.16866	0.078317	0.051581	0.071659	0.073787	0.075832	0.05074
	DUT	0.11734	0.14681	0.17322	0.18745	0.17715	0.18697	0.19073	0.15664	0.2401	0.2844	0.15219	0.11475	0.1675	0.18231	0.1488	0.13741
	ECSSD	0.12973	0.21873	0.20581	0.18919	0.20717	0.19588	0.18243	0.17407	0.2169	0.25439	0.182	0.15024	0.17734	0.18404	0.17281	0.1721
	SOD	0.20278	0.26741	0.25066	0.25932	0.26464	0.26136	0.25128	0.23338	0.26703	0.26132	0.24782	0.21367	0.24263	0.25639	0.244	0.20262

[1] The ↑ indicates the larger value achieved, the better performance is, while the ↓ indicates the smaller, the better

[2] The best three results are highlighted with red, green and blue fonts, respectively

For the challenging DUT-OMRON dataset, our method performs the highest F-measure (0.669) and OR (0.48892), and the second-best MAE (0.11734) with a minor margin (0.00259) to the best results HCC (0.11475). Besides, except for the MAE on the SOD dataset, all indicators of the algorithm on other datasets are better than BSFC. Besides, our method can achieve the highest precision and recall simultaneously, which ensures that the results obtained by our method are closer to their ground truth, and the proposed method is more applicable to practical applications. Furthermore, our method achieves relatively superior performance compared to the other methods for PR curves on four datasets.

The above quantitative comparisons show that the proposed method consistently outperforms all the compared state-of-the-art approaches on various scenes.

3.3 Qualitative results

We provide some saliency maps of the proposed algorithm and other thirteen state-of-the-art algorithms in Fig. 8. From the results, we can see that the saliency maps generated by our algorithm are of the best similarity to the ground truth. For images with a single object, in the first two rows, our method can ultimately retain salient objects while background noise is well suppressed simultaneously. When

an image has complex background structures, the proposed algorithm can still obtain favorable results with a less noisy background. For example, in the 3rd and 4th rows, our saliency maps can uniformly pop out the foreground object of the images, but other algorithms fail to extract the salient object from the scattered background. Also, when salient objects and backgrounds share a similar appearance, our algorithm can detect salient regions accurately. At the same time, other methods either fail to identify the salient objects or incorrectly include background regions into detection results, as shown in the 5th and 6th two rows. Furthermore, as shown in the 7th and 8th two rows, our results perform well and can preserve the completeness while other methods are a weakness when salient objects are on the boundaries, which conclusively prove our high robustness of background selection. All the above results demonstrate the robustness and the effectiveness of our method in highlighting salient objects and restraining the background regions.

3.4 Analysis of our algorithm

(1) ***Effectiveness of the proposed boundary prior map***
To demonstrate the effectiveness of the proposed boundary prior map, we compute the PR curves and quantitative results of F-measure for boundary
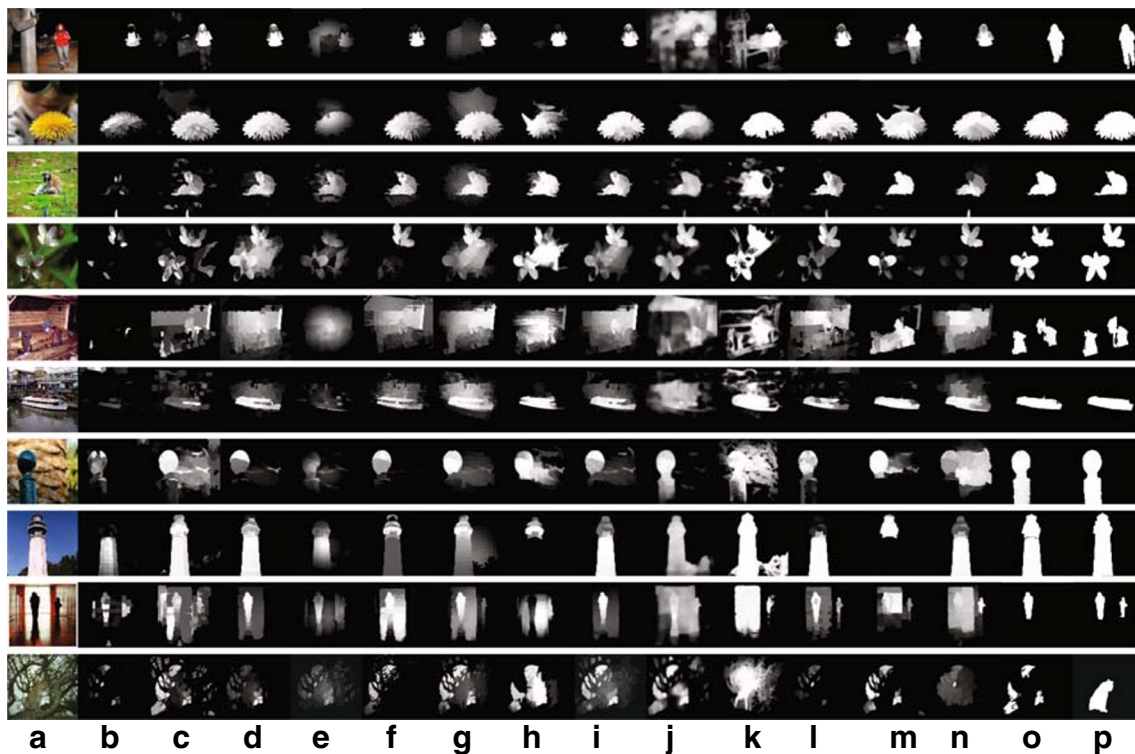

Fig. 8 Comparison of our saliency maps with other thirteen state-of-the-art algorithms. **(a)** Input; **(b)** SF; **(c)** GS; **(d)** MR; **(e)** BFS; **(f)** LPS; **(g)** BSCA; **(h)** MB; **(i)** RR; **(j)** BL; **(k)** CGV; **(l)** SRD; **(m)** HCC; **(n)** MILPS; **(o)** Ours; **(p)** Ground-truth

prior map on the ASD dataset. Different background seeds are received by three methods (BFS [20], MR [15], and LPS [9]) and construct their corresponding boundary-based saliency maps. The resulting curves and quantitative results are shown in Fig. 9a, where the dotted line shows the performance of each method using their own boundary set while the solid line provides the performance using our boundary nodes. Also, the red line shows the performance of our boundary contrast map. As shown by these curves, the proposed method outperforms those competitive methods. Yet, the performance of three saliency models has also been conspicuously improved, illustrating the effectiveness of the selection of boundary nodes and the computation of the background-based saliency map.

(2) *Validation of components in contour completeness map* Since both the robust boundary background set BG and saliency weights are introduced into the contour completeness map, we further examine the effectiveness of each component for the contour completeness map. Herein, we consider different components for comparative analysis, including:

(a) contour completeness map based on BS without saliency weights (SCBS),

(b) contour completeness map based on BG and without saliency weights (SCBG),

(c) contour completeness map based on BS with saliency weights (SCBSW),

(d) contour completeness map based on BG with saliency weights (SCBGW).

The resulting curves on SOD and the corresponding quantitative results are shown in Fig. 9b. We note that the performance of SCBG (SCBGW) is higher than the performance of SCBS (SCBSW) since the former takes the issue of the object connecting to boundary into consideration while the latter directly use all boundary nodes as background, effectively demonstrating the effectiveness of BS in improving the contour completeness map. Besides, we observe that the performance of SCBSW (SCBGW) is higher than the performance of SCBS (SCBG), due to the recent introduction of saliency weights that assign a different probability to each pixel belonging to be a salient object. Furthermore, Compared with SCBS, SCBG, and SCBSW, SCBGW achieves the best performance, which conclusively proves the necessity of considering the two indexes (BG and saliency weights) in the calculation of the contour completeness map.

(3) *Examination of refinement* Since the system optimization adopts an iterative mechanism, we first verify the stability of the system. To this end, we compute the performance of the proposed method on a different
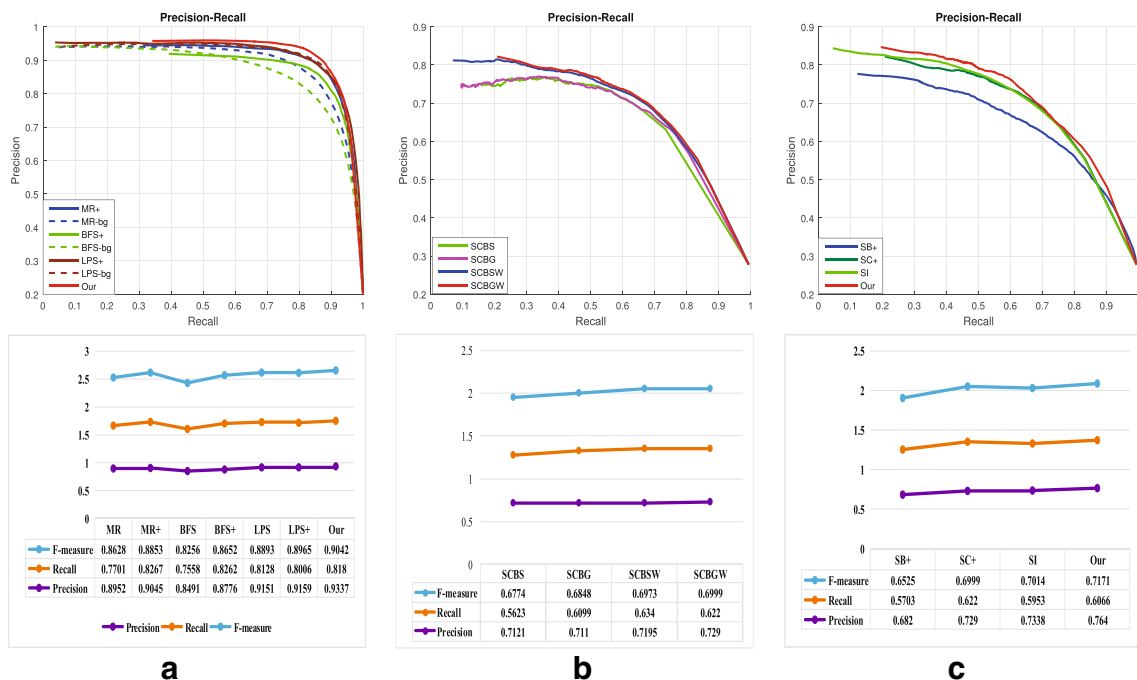

Fig. 9 Precision-recall curves and quantitative results of F-measure to measure the effectiveness of our algorithm. **(a)** Evaluation of our background-based saliency map on the ASD dataset; **(b)** Evaluation of our contour completeness saliency map on the SOD dataset; **(c)** Evaluation of each component in the proposed algorithm on the SOD dataset.
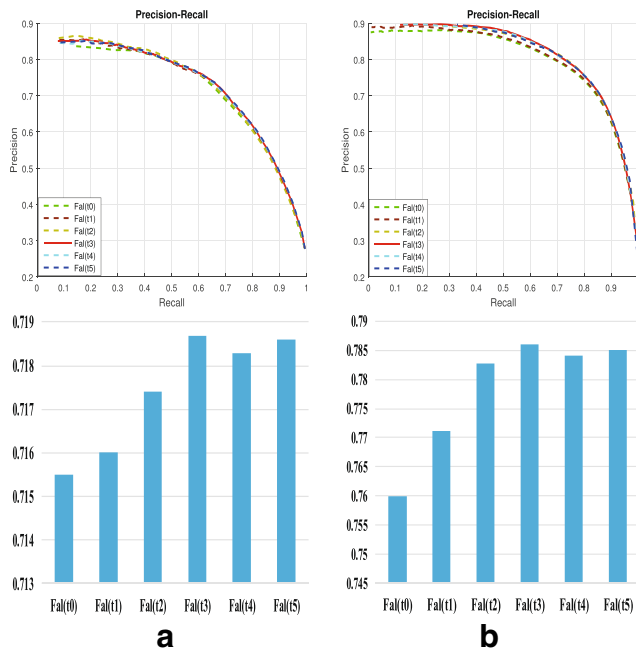
Fig. 10 Quantitative evaluation on two salient object datasets for our method with various iteration steps **(a)** SOD and **(b)** ECSSD

Table 3 Running average time test results (seconds per image) on ASD

Method	GS	LPS	MR	RR	BFS	Our
Times	0.425	3.376	0.715	3.56	7.513	1.637

RR, and BFS, even though being slower than MR and GS, the proposed method still outperforms them both considering the overall evaluation performance. Therefore, it can be attained that the proposed algorithm can achieve a reasonable balance between accuracy and efficiency.

3.6 Limitation and analysis

Although the proposed method can perform well in most of the cases, there remain some challenging scenarios that cannot accurately extract the complete salient objects. The last two rows of Fig. 8 show some failure cases. When an image contains multiple salient objects in complex scenes, the proposed method may not work well in highlighting all objects in an image due to the difference and diversity of multi-objects in terms of size and position. However, it is also tricky for state-of-art saliency detection methods. For an image with a complex background structure and small differences between object and background, it is inadequate for our method to achieve the satisfactory detection results by the proposed low-level and mid-level features. Hence, it is an excellent choice to design an even more discriminative similarity metric by incorporating the traditional low-level features and the high-level, in-depth learning features. Besides, the study on the reliability of edge and the selection of an optimal segmentation threshold can improve the performance of saliency detection as well.

4 Conclusions and future work

In this paper, we present an efficient salient object detection algorithm via background scatter and foreground contour completeness. To accurately extract robust background, background scatter is proposed and designed, and the corresponding boundary contrast map can be obtained via boundary prior. Besides, contour completeness embedded with boundary contrast map can be applied together with robust boundary nodes, to generate contour completeness saliency map, so finally, an optimization function is proposed to highlight salient objects uniformly. Experimental results not only demonstrate the superior performance of the proposed method but also the right balance between accuracy and computation cost.

As future work, we investigate efficient methods that incorporate high-level features with the support of deep

step of iteration with both PR curve and F-measures on two datasets (SOD and ECSSD), as shown in Fig. 10. We note that the system obtains a stable performance with only three steps of iterations (i.e., Fal(t3)), since the performance of the proposed system cannot further be improved by increasing the steps of iterations. Furthermore, to validate the effectiveness of the refinement, a comparison between the final result and other steps from both quantitative and qualitative aspects are processed and shown in Fig. 9c, where the quantitative comparison results of the saliency refinement mechanism on the SOD dataset present performance enhancement of the saliency map. Also shown in Fig. 6e and f, the salient object can be highlighted uniformly, and the background can be suppressed effectively due to the refinement, which qualitatively verifies the effectiveness of the refinement mechanism.

3.5 Running time

The execution time test is conducted on a 64-bit PC with Intel Core i5-4460 CPU @ 3.20GHz and 8G memory. All codes for the experimentation testings are provided by the corresponding authors and executed without any change in MATLAB R2015a with C++ mex implementation, and the average running time is computed on the ASD dataset.

We selected several competitive accuracy methods or those akin to the proposed algorithm, and the results are shown in Table 3. It is significantly faster than LPS,

learning to achieve higher performance. Taking into consideration that image depth information plays an essential role in saliency detection and challenging to collect image depth information via camera, ways how to use non-camera methods to obtain image depth information and improve saliency detection in also be investigated and validated.

References

1. Bai X, Wang W (2014) Saliency-svm: an automatic approach for image segmentation. Neurocomputing 136:243–255
2. Ren Z, Gao S, Chia L-T, Tsang IW-H (2014) Region-based saliency detection and its application in object recognition. IEEE Trans Circuits Syst Video Technol 24(5):769–779
3. Hadizadeh H, Bajic IV (2014) Saliency-aware video compression. IEEE Trans Image Process 23(1):19–33
4. Gao X, Shi X, Zhang G, Jin L, Liao M, Li KC, Li C (2018) Progressive Image Retrieval With Quality Guarantee Under MapReduce Framework. IEEE Access 6:44685–44697
5. Li Y, Lu H, Li KC, Kim H, Serikawa S (2018) Non-uniform de-scattering and de-blurring of underwater images. Mob Netw Appl 23(2):352–362
6. Achanta R., Hemami S., Estrada F., Susstrunk S. (2009) Frequency-tuned salient region detection. In: IEEE conference on computer vision and pattern recognition, pp. 1597–1604
7. Margolin R., Tal A., Zelnik-Manor L. (2013) What makes a patch distinct. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1139–1146
8. Wei Y., Wen F., Zhu W., Sun J. (2012) Geodesic saliency using background priors. In: Proceedings of the IEEE conference on european conference on computer vision, pp. 29–42
9. Li H, Lu H, Lin Z, Shen X, Price B (2015) Inner and inter label propagation: salient object detection in the wild. IEEE Trans Image Process 24(10):3176–3186
10. Li C., Yuan Y., Cai W., Xia Y., Dagan Feng D. (2015) Robust saliency detection via regularized random walks ranking. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2710–2717
11. Sun J, Lu H, Liu X (2015) Saliency region detection based on markov absorption probabilities. IEEE Trans Image Process 24(5):1639–1649
12. Cheng M-M, Warrell J, Lin W-Y, Zheng S, Vineet V, Crook N (2013) Efficient salient region detection with soft image abstraction. In: Proceedings of the IEEE international conference on computer vision, pp. 1529–1536
13. Cheng M-M, Mitra NJ, Huang X, Torr PH, Hu S-M (2015) Global contrast based salient region detection. IEEE Trans Patt Anal Mach Intell 37(3):569–582
14. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Patt Anal Mach Intell 20(11):1254–1259
15. Yang C, Zhang L, Lu H, Ruan X, Yang M (2013) Saliency detection via graph-based manifold ranking. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3166–3173
16. Jiang H, Wang J, Yuan Z, Wu Y, Zheng N, Li S (2013) Salient object detection: A discriminative regional feature integration approach. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2083–2090
17. Gong C, Tao D, Liu W, Maybank SJ, Fang M, Fu K, Yang J (2015) Saliency propagation from simple to difficult. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2531–2539
18. Lu H, Li X, Zhang L, Ruan X, Yang M.-H. (2016) Dense and sparse reconstruction error based saliency descriptor. IEEE Trans Image Process 25(4):1592–1603
19. Zhang H, Xia C, Gao X (2017) Robust saliency detection via corner information and an energy function. IET Comput Vis 11(6):379–388
20. Wang J, Lu H, Li X, Tong N, Liu W (2015) Saliency detection via background and foreground seed selection. Neurocomputing 152:359–368
21. Zhou L, Yang Z, Zhou Z, Hu D (2017) Salient region detection using diffusion process on a two-layer sparse graph. IEEE Trans Image Process 26(12):5882–5894
22. Lee G, Tai Y-W, Kim J (2016) Deep saliency with encoded low level distance map and high level features. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 660–668
23. Hou Q, Cheng M-M, Hu X, Borji A, Tu Z, Torr PH (2019) Deeply supervised salient object detection with short connections. IEEE Trans Patt Anal Mach Intell 41(4):815–828
24. Liu Y, Fan D-P, Nie G-Y, Zhang X, Petrosyan V, Cheng M-M Dna: Deeply-supervised nonlinear aggregation for salient object detection, arXiv:1903.12476
25. Yu J-G, Gao C, Tian J (2016) Collaborative multicue fusion using the cross-diffusion process for salient object detection. J Opt Society Am A Opt Image Sci Vis 33(3):404–415
26. Xia C, Zhao Q, Zhang S, Gao X (2019) A hybrid of background scatter and foreground contour completeness for salient object detection. In: International conference on applications and techniques in cyber security and intelligence, pp. 1266–1275
27. Dollár P, Zitnick CL (2013) Structured forests for fast edge detection. In: Proceedings of the IEEE international conference on computer vision, pp. 1841–1848
28. Qin Y, Lu H, Xu Y, Wang H (2015) Saliency detection via cellular automata. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 110–119
29. Liu Q, Hong X, Zou B, Chen J, Chen Z, Zhao G (2017) Hierarchical contour closure-based holistic salient object detection. IEEE Trans Image Process 26(9):4537–4552
30. Jiang H, Wang J, Yuan Z, Liu T, Zheng N, Li S (2011) Automatic salient object segmentation based on context and shape prior. In: Proceedings of the British Machine Vision Conference, pp. 9
31. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66
32. Huo S, Zhou Y, Xiang W, Kung S (2018) Semisupervised learning based on a novel iterative optimization model for saliency detection. IEEE Trans Neural Netw Learn Syst (99) pp. 1–17
33. Frintrop S, Werner T, Martin Garcia G (2015) Traditional saliency reloaded: A good old model in new shape. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 82–90
34. Huang F, Qi J, Lu H, Zhang L, Ruan X (2017) Salient object detection via multiple instance learning. IEEE Trans Image Process 26(4):1911–1922
35. He K, Sun J, Tang X (2013) Guided image filtering. IEEE Trans Patt Anal Mach Intell (6) pp. 1397–1409

36. Tong N, Lu H, Ruan X, Yang M-H (2015) Salient object detection via bootstrap learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1884–1892
37. Liu GH, Yang JY (2018) Exploiting color volume and color difference for salient region detection. IEEE Trans Image Process 28(1):6–16
38. Yang KF, Li H, Li CY, Li YJ (2016) A unified framework for salient structure detection by contour-guided visual search. IEEE Trans Image Process 25(8):3475–3488
39. Zhang J, Sclaro S, Lin Z, Shen X, Price B, Mech R (2015) Minimum barrier salient object detection at 80 fps. In: Proceedings of the IEEE international conference on computer vision, pp. 1404–1412
40. Yuan Y, Li C, Kim J, Cai W (2017) D. F Reversion correction and regularized random walk ranking for saliency detection. IEEE Trans Image Process 27(3):1311–1322
41. Perazzi F, Krähenbühl P, Pritch Y (2012) A. Hornung, Saliency filters: Contrast based filtering for salient region detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 733–740
42. Zhou L, Yang Z, Yuan Q, Zhou Z, Hu D (2015) Salient region detection via integrating diffusion-based compactness and local contrast. IEEE Trans Image Process 24(11):3308–3320
43. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the IEEE international conference on computer vision, pp. 416–423
44. Yan Q, Xu L, Shi J, Jia J (2013) Hierarchical saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1155–1162
45. Li G, Yu Y (2016) Visual saliency detection based on multiscale deep cnn features. IEEE Trans Image Process 25(11):5012–5024
46. Wang L, Lu H, Wang Y, Feng M, Wang D, Yin B, Ruan X (2017) Learning to detect salient objects with image-level supervision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 136–145
47. Li X, Li Y, Shen C, Dick AR, van den Henge A (2013) Contextual hypergraph modeling for salient object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3328–3335
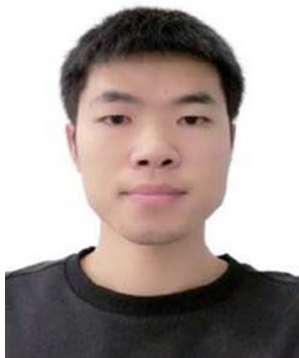
Xiuju Gao received the M.S. degree in Computer Science and Electronic Engineering from Hunan University, Changsha China, in 2016. She is currently an assistant at the College of Electrical and Information Engineering, Anhui University of Science and Technology. Her research interests include image/video processing and computer vision.

Kuan-Ching Li is a Distinguished Professor in the Dept of Computer Science and Information Engineering (CSIE) at Providence University, Taiwan, where he also serves as the Director of the High-Performance Computing and Networking Center. He published more than 250 scientific papers and articles and is author, co-author or editor of more than 25 books published by leading publishers. He is the Editor in Chief of the Connection Science and serves as an associate editor for several leading journals. Also, he has been actively involved, in a variety of capacities, in the organization of many national and international conferences. Professor Li is a Fellow of IET and a senior member of the IEEE. His research interests include parallel and distributed computing, Big Data, and emerging technologies.

Qianjin Zhao received the M.S. degree and the Ph.D. degree from Hefei University of Technology, China, in 2002 and 2006, respectively. He is currently a professor and dean at the School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, Anhui China. His current research interests include data mining, machine learning, etc.

Chenxing Xia received the Ph.D. degree from Hunan University in 2019. He is currently a lecturer at the College of Computer Science and Engineering, Anhui University of Science and Technology. His research interests include saliency detection, computer vision, image processing, and depth prediction.

Shunxiang Zhang received his Ph.D. degree from the School of Computing Engineering and Science, Shanghai University, Shanghai, in 2012. He is a professor at Anhui University of Science and Technology, China. His current research interests include Web Mining, Semantic Search, and Complex network.