Neurocomputing 383 (2020) 194–211

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Exploiting background divergence and foreground compactness for salient object detection

Chenxing Xia[a,b], Hanling Zhang[a,*], Xiuju Gao[c], Keqin Li[d]

[a] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
[b] College of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China
[c] School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China
[d] Department of Computer Science State University of New York New Paltz, New York 12561, USA

ARTICLE INFO

Article history:
Received 15 March 2019
Revised 8 September 2019
Accepted 23 September 2019
Available online 23 October 2019

Communicated by Dr. Shenghua Gao

Keywords:
Background divergence
Deep features
Foreground compactness
Manifold ranking
Salient object detection

ABSTRACT

In this paper, we propose an efficient and discriminative saliency method that takes advantage of background divergence and foreground compactness. Concretely, a graph is first constructed by introducing the concept of virtual node to effectively enhance the distinction between nodes along object boundaries and the similarity among object regions. A reasonable edge weight is defined by incorporating low-level features as well as deep features extracted from deep networks to measure the relationship between different regions. To remove incorrect outputs, two computational mechanisms are then developed to extract reliable background seeds and compact foreground regions, respectively. The saliency value of a node is calculated by fully considering the relationship between the corresponding node and the virtual background (foreground) node. As a result, two types of saliency maps are obtained and integrated into a uniform map. In order to achieve significant performance improvement consistently, we propose a robust saliency optimization mechanism, which subtly combine suppressed/active (SA) nodes and mid-level structure information based on manifold ranking. Extensive experimental results demonstrate that the proposed algorithm performs favorably against the state-of-art saliency detection methods in terms of different evaluation metrics on several benchmark datasets.

1. Introduction

Human visual system can effortlessly and accurately identify the most important regions in a scene. The selective attention mechanism can help with the high-level cognitive task. Therefore, saliency detection to model biological visual systems has received increasing interest from computer science, psychology and neurobiology in recent years [1]. As an essential pre-processing procedure, numerous saliency models have been applied to various computer vision fields, such as image segmentation [2], object recognition [3], forgery detection [4], image/video compression [5] and image retrieval [6], and so on.

Itti et al. [7] believed that human visual system paid more attention to high-contrast regions and computed saliency via local contrast. Based on this theory, contrast prior becomes one of the most used principles to be adopted by various kinds of saliency models [8,9] from either local or global view. For local methods [10], center-surround contrast is used to characterize saliency. Due to lack of global information, these methods tend to highlight the boundaries of the salient objects while failing to detect internal patches of the salient objects (see Fig. 1(c)). On the other hand, global contrast [11,12], which prefers to exploit the global information for saliency detection, can accurately determine the location of the salient objects. However, the effectiveness of these methods in uniformly detecting the salient objects is limited when background regions have similar appearance resulting from local information being ignored (see Fig. 1(d)).

Different from adopting contrast prior, many saliency methods [13–15] formulate their algorithms based on boundary prior, regarding that it is of high probability for image boundaries to be background. Although these methods perform well in some cases, they also have several drawbacks. Firstly, it is not appropriate to treat all regions on the boundary as background for the reason that sometimes the object may appear on the image boundary (see Fig. 1(e)). Secondly, most of these methods are effective in uncomplicated cases, but they still struggle in complex scenes due to the feeble low-level features.

For the shortcoming of low-level features, some methods [14,16] adopt task-driven models requiring supervised learning based on training samples with manual labels. The prime

* Corresponding author.
E-mail addresses: starry@hnu.edu.cn (C. Xia), jt_hlzhang@hnu.edu.cn (H. Zhang), gaoxiuju5@163.com (X. Gao), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.neucom.2019.09.096
0925-2312/

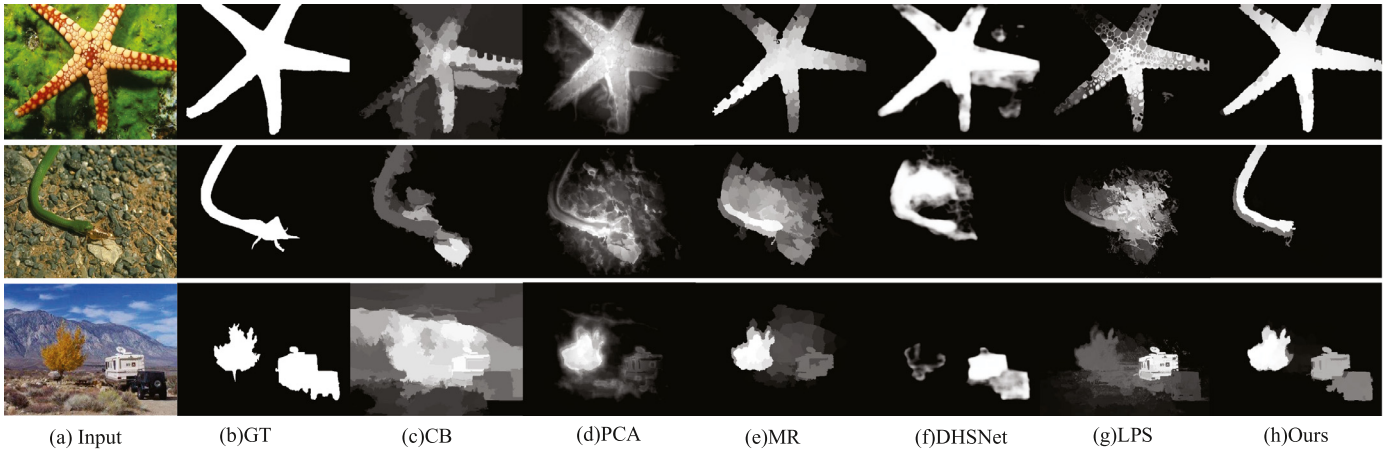


Fig. 1. The differences among various saliency models: including the local contrast based model CB [7], the global contrast based model PCA [8], the background prior based model MR [9], deep learning based model DHSNet [14] and the objectness based model LPS [13].

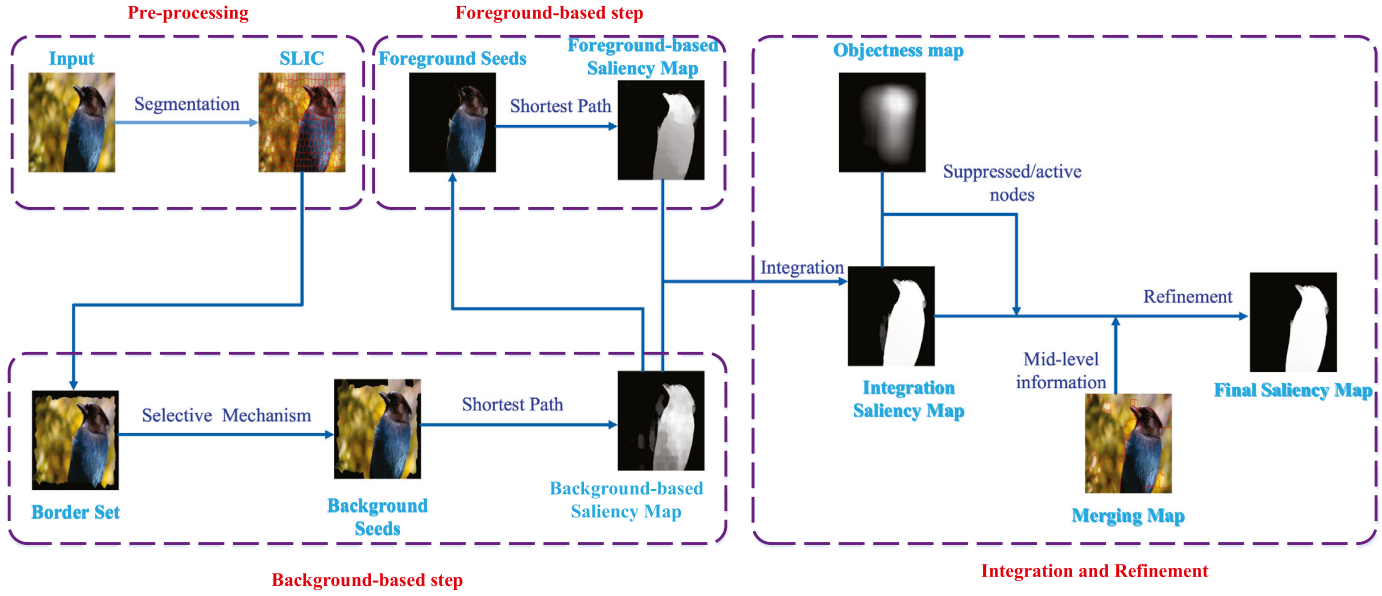


Fig. 2. Pipeline of our algorithm.

example is deep learning based methods, where deep features extracted from deep neural networks are exploited due to their powerful expressibility and strong capacity for discriminating the objects from the backgrounds. However, it is very expensive and time consuming to obtain an immense amount of manual label data. In addition, deep learning based methods often suffer from diluting image features due to convolutional and pooling layers [17] (see Fig. 1(f)). On the other hand, a recent trend is to incorporate high-level features [18,19] to facilitate detection. A type of high-level representation is the concept of objectness. The problem is that directly exploiting the objectness score to guide the saliency computation may lead to dissatisfactory results in complex scenes, especially provided that the correct salient objects cannot be predicted by the objectness value [18,20,21] (see Fig. 1(g)).

To this end, we put forward an unsupervised graph-based method for saliency detection. The pipeline of our method is presented in Fig. 2. First, based on the superpixel segmentation, we collect the image border superpixels to construct a border set. Then, a discriminative similarity metric is proposed by taking deep features and traditional handcrafted features into consideration. After that, we introduce an automatic selection scheme based on background divergence to get reliable background seeds instead of simply using border set as background seeds [13,22], and compute the saliency value of each node by finding the shortest path from the corresponding node to the virtual background node, thereby generating the background-based saliency map. To suppress the background noise, the foreground-based saliency map is constructed. Specifically, highly confident compact foreground seeds are generated by considering spatial compactness and visual rarity. Similar to the background-based saliency detection, the saliency value of each node is measured by finding the shortest path from the corresponding node to the virtual foreground node and the foreground-based saliency map then can be obtained. Furthermore, the two saliency maps are integrated by an unified function. Finally, based on manifold ranking, an improved saliency propagation mechanism, which introduces mid-level structure information and suppressed/active nodes mainly taking objectness map into consideration, is proposed to refine the integrated result.

The main contributions of this paper are as follows:

- We propose a saliency detection algorithm based on background divergence and foreground compactness on a novel graph structure, which introduces the background/foreground seeds and the concept of virtual background/foreground node. An automatic selection scheme based on background divergence is proposed to get reliable background seeds from image border. By taking spatial compactness and rarity of salient objects into consideration, a compact and coherent foreground region can be generated.
- A robust saliency propagation mechanism based on manifold ranking is proposed to refine saliency map. The experimental results demonstrate that it is of high versatility that can improve other methods based propagation when applying our propagation mechanism.
- We also devise a new object-touching-boundary (OTB[1]) saliency dataset comprising 1044 images that salient objects touch the image boundaries in each image. This dataset is specifically designed to evaluate the performance of salient object detection method on the scene that boundaries are adjacent to the salient objects.

The remainder of this paper is organized as follows. Section 2 describes the related work on salient object detection. Section 3 shows the detailed description of our proposed salient object detection. The experimental results and performance evaluation are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related work

Literature of salient object detection is huge and we just discuss the most related saliency detection methods in this section.

Recently, *boundary prior* is one of the most widely used principles to compute saliency. Yang et al. [13] computed the saliency of image regions according to the relevance to boundary nodes via manifold ranking. Wei et al. [23] estimated saliency value of each region by measuring the shortest-path distance to the boundary. Nevertheless, it may produce undesirable results if the object appears at image boundaries. In order to improve the effect, Wang et al. [24] dropped the superpixels with strong edge out of the border set to obtain the reliable background seeds. Li et al. [25] directly filtered out one boundary having the most distinctive color distribution and regarded the rest of three as background query seeds. Li et al. [20] removed the top 30% border pixels with high color difference among border set. However, their computational distinctiveness is confined to the border set instead of the entire image. Faced with the issue, we propose an effective way to remove the untrustworthy superpixels in the border set based on divergence background and define the rest of the border regions as background seeds.

A graph can be applied to represent relationships between image elements with affinity measure. Harel et al. [26] adopted multiple features to extract saliency information on their graph based visual saliency (GBVS) model. A hierarchical graph model was developed for saliency detection by combining context information [27]. Wang et al. [28] measured the saliency by site entropy rate on a full-connected graph. Gopalakrishnan et al. [29] exploited the Markov random walks on the fully connected graph and the sparsely connected graph to find the most salient seeds. Wang et al. [30] exploited a novel graph structure and background priors for saliency detection. Besides, graph-based methods are usually combined with diffusion processes. Thus far, more and more saliency methods have also been proposed by diffusing processes to propagate saliency information throughout a graph with different features and affinity measures. Zhang et al. [31] ranked the similarity of image elements with foreground or background cue via graph-based manifold ranking (MR). Li et al. [25] proposed a robust background measure to characterize the spatial layout of an image region with respect to the boundary regions and estimated the saliency via regularized random walks ranking (RRWR). Sun et al. [32] formulated saliency detection based on markov absorption probabilities (MAP) on an image graph model. Based on [32], an improved model for saliency detection is proposed via absorbing markov chain with learnt transition probability [33]. In order to make the obtained image regions have close spatial consistency, Kong et al. [34] proposed an extended random walk algorithm. Zeng et al. [35] formulated a Saliency Game among superpixels and proposed an iterative random walk for saliency detection on a graph model. All of them perform well, but they still struggle in complex scenes for the reason that these propagation methods are all under the assumption that it has a high color contrast between background and foreground. In order to reduce the propagation error and simultaneously improve the detection accuracy, Chen et al. [36] introduced the concept of sink points into manifold ranking for saliency detection. Inspired by these works, we propose an extended manifold ranking (EMR) algorithm, which introduces the mid-level information and suppressed/active nodes.

In addition, numerous generic object detection methods, which aim at generating the location of objects in an image, have been applied to salient region detection. In [37], a saliency measure was implemented by combining the objectness values of many overlapping windows. To improve saliency estimations, Chang et al. [18] presented an iterative optimization of energy functions, which combines saliency, objectness and interaction term. However, the background regions could be falsely detected as foreground if the objectness cannot predict the correct object. Recently, a co-transduction algorithm, namely label propagation saliency, is devised for saliency detection via incorporating low-level features and the objectness measure [20]. All of them demonstrate that objectness is helpful to saliency detection. Instead of directly using objectness to compute saliency, we integrate foreground regions based on saliency map with objectness to extract reliable suppressed nodes and active nodes resulting in boosting the performance of salient region detection.

With the rapid development of deep learning, more and more researchers have applied deep learning to the field of saliency detection. Zhao et al. [38] proposed a multi-context deep learning framework for salient object detection, which global context and local context are both taken into account. Li and Yu [39] predicted the saliency score for each superpixel by using multiscale CNN features. Lee et al. [40] designed a unified fully connected neural network to estimate saliency value by combining both high-level feature extracted from CNNs and low-level feature. Liu et al. [17] developed a architecture in a global to local and coarse to fine manner, in which a rough map was generated by a CNN and then another network was designed to refine the map hierarchically and progressively via integrating local context information. Li et al. [41] develop a fully convolutional multiscale refinement network, which combines bottom-up and top-down information to generate high-quality saliency map and salient object contours. Wang et al. [42] proposed a new saliency model using recurrent fully convolutional networks, which enables automatic learning to refine previous saliency map by correcting its previous errors. Hou et al. [43] proposed a new network by adding a series of short connections to the HED [44] architecture, which can effectively combine features of different levels and enable the activation of each side output layer to fully highlight the salient object and accurately locate its boundary. Considering that not all contextual information

[1] The dataset are available at our website: https://github.com/starry1614/OTB_saliency_dataset.

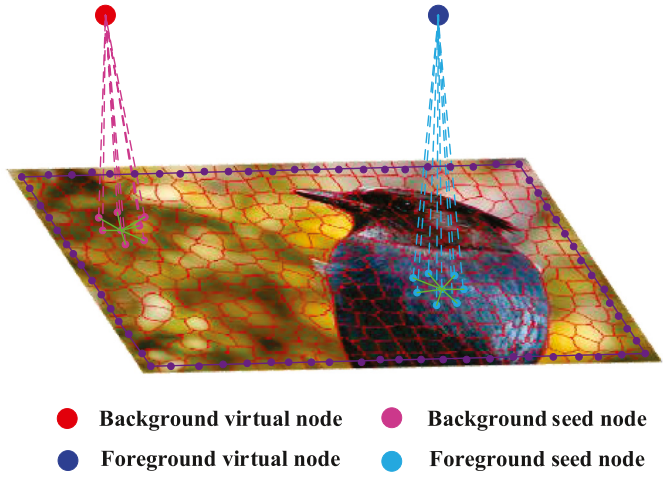


Fig. 3. Schematic illustration of proposed graph structure.

facilitates saliency detection, Liu et al. [45] designed a novel network to selectively attend to global or local contexts and construct information context features for each pixel to facilitate the final saliency detection. From the perspective of selectivity-invariance dilemma, Su et al. [46] proposed a novel boundary-aware network with successive dilation for salient object detection. Li et al. [47] developed a saliency inference module based on a multi-dilated depth-wise convolution architecture, which can quickly and directly infer saliency objects from multi-scale features. Sen et al. [48] designed a Richer and Deeper Supervision (RDS) network, which can combine the features of the output on each side without requiring too much extra computational space. In contrast to the previous approach of aggregating side output features in a linear fashion, Liu et al. [49] proposed a DNA module, which uses a nonlinear side-output prediction aggregation to make better use of the complementary information of various side-outputs. All of them had better ability of detecting saliency object due to deep learning which can extract high-level features. However, these methods based on deep learning perform weakness in obtaining the clear object boundary of an image. In addition, it is difficult for deep learning to learn some prior knowledge, such as background prior, compactness, which has proven to be effective in saliency detection. In this paper, we combine the deep features and low-level features to facilitate saliency detection.

3. The proposed approach

Firstly, we segment the image into N superpixels by the SLIC method [50] and use them as the minimum processing units for the reason that the superpixels cannot only capture the structural information of an image but also speed up the processing. Then, we construct a graph $G = (V, E)$, where each node $v_i \in V$ corresponds to a superpixel and each edge $e_{ij} \in E$ connects two superpixels.

Our graph is constructed by making two modifications on most existing methods [20,51]. First, they connect each node to its direct spatial neighbors which share a boundary, as well as its mediate spatial neighbors which share a common boundary with direct neighbors. However, each node is only connected to its direct spatial neighboring nodes in this paper (as shown in Fig. 3). Furthermore, inspired by [52], we define two virtual nodes in this graph, namely, virtual background node VT_{bp} and virtual foreground node VT_{fp}, which are used for predicting saliency value by computing the shortest path between the node with virtual node. Specifically, the edge weight between virtual background (foreground) node and well-defined background (foreground) seeds are allocated with zero. In order to reduce the geodesic distance of similar superpixels, we connect the well-defined seed nodes to the virtual node instead of associating boundary superpixels with each other. The rationale behind is that not all boundary superpixels are background regions and it may introduce errors for their mandatory connection.

3.1. Construction of edge weights of graphs

In this work, we adopt FCN-32s feature to encode object appearance, and choose the outputs of the conv5 as feature maps. The reasonable reason behind it is that the features from the last layers of CNNs are full of semantic information of objects and have strong powerful to discriminate the objects. Because of the existence of subsampling and pooling in the network, the resolution of the feature maps in each layer is not same. Herein, we use bilinear interpolation to make sure that the feature maps have the same resolution with input image. Hence, we can construct an affinity metric in deep-level feature space:

$$a_{ij}^{d} = exp\left(-\frac{d_{ij}^{d}}{2\sigma_w^2}\right) = exp\left(-\frac{|f_i^d - f_j^d|}{2\sigma_w^2}\right), \tag{1}$$

here f_i^d is the deep feature vector of superpixel i, σ_w controls the strength of weight between a pair of nodes.

Although the high-level semantic deep features have strong superiority in discriminating object, it cannot describe the low-level detailed information, which is also important in detecting. Therefore, we also employ some low-level features as a complement to deep features. In this paper, our constructed low-level feature space includes color difference, spatial distance and edge information. We denote the affinity between the ith superpixel and jth superpxiel in low-level feature space as a_{ij}^{l}, which is defined as follows:

$$a_{ij}^{l} = exp\left(-\frac{d_{ij}^{l}}{2\sigma_w^2}\right) = exp\left(-\frac{d_c(ij) + d_s(ij) + d_{edge}(ij)}{2\sigma_w^2}\right), \tag{2}$$

where $d_c(ij)$ indicates the color difference, $d_s(ij)$ represents the spatial distance between the nodes i and j, $d_{edge}(ij)$ is the intervening contour magnitude [53]. The color difference is defined as

$$d_c(ij) = |c_i - c_j|. \tag{3}$$

Here c_i and c_j are the mean value of superpxiel in CIELAB color space, which has been proved to be effective [13]. Most existing methods to compute the spatial distance generally adopt the Euclidean distance, which is prone to cause a large difference especially between opposite image borders owing to their largest distance. However, the image borders are of high probability to be background. To improve the detection performance, we measure the spatial distance by combining sine spatial distance [54] and geodesic distance:

$$\begin{aligned} &d_s(ij) = sqrt\left(\min(X_{ij}, I_x - X_{ij})^2 + \min(Y_{ij}, I_y - Y_{ij})^2\right), \\ &s.t.\ \ X_{ij} = |sin(\pi \cdot |x_i - x_j|)| \\ &\qquad Y_{ij} = |sin(\pi \cdot |y_i - y_j|)| \end{aligned} \tag{4}$$

where x_i and y_i represent the normalized coordinates of node i. The formulation makes sure that the spatial distance between boundary superpixels especially the nodes at the opposite borders of the image would be small.

As for complex scene, it is difficult to distinguish the foreground from the background just based on color feature or both color difference and spatial contrast (as shown in Fig. 4(b)–(c)). Fortunately, we observe that it can produce a good result due to texture distinction, which can pop out the entire contour of the tree and

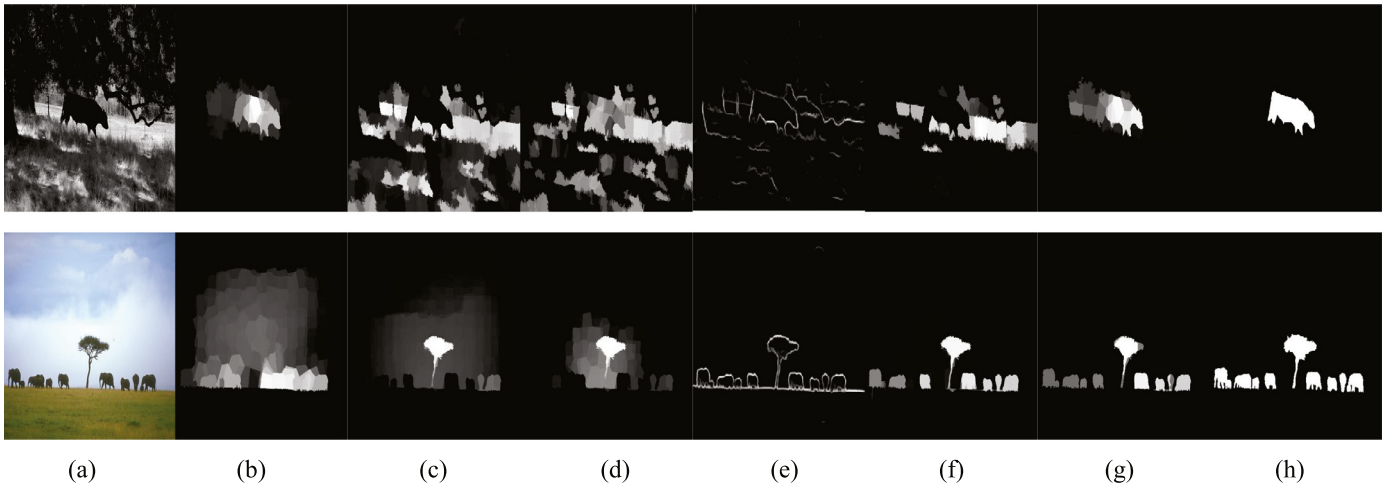


Fig. 4. Effects of edge weight by computing background saliency map. (a) Input; (b) Just using high-level feature; (c) Just considering color difference; (d) Considering color and spatial information; (e) Edge detection by [55]; (f) Just using low-level information; (g) The background saliency map by using our edge weight; (h) Ground-truth.

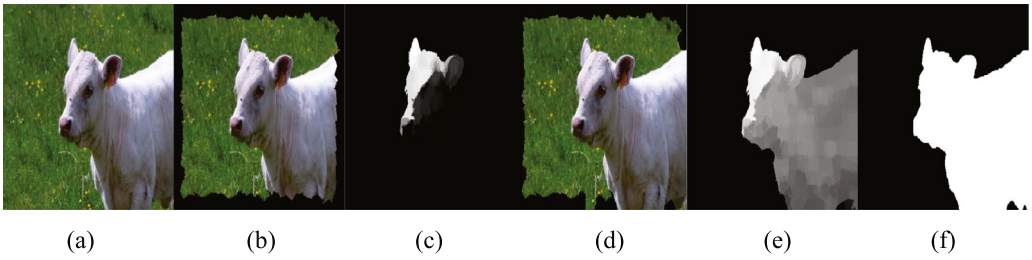


Fig. 5. Salient objects contact with image border: (a) Input; (b) Border set; (c) Use all boundary nodes based on Eq. 10; (d) Our background regions; (e) Our background-based saliency map; (f) Ground-truth.

bulls. Hence, we introduce the intervening image edge cues. The intervening contour magnitude $d_{edge}(ij)$ is defined as

$$d_{edge}(ij) = \max_{p \in line(loc_i, loc_j)} E(p), \tag{5}$$

where loc_i represents the centroid location of superpixel i, $line(loc_i, loc_j)$ is the line connecting node i and node j, p indicates the pixel on the line. $E(p)$ stands for the edge value on the edge map E. However, there also exists a problem that $d_{edge}(ij)$ can be extremely large for the opposite image borders. In order to mitigate this problem, we set $d_{edge}(ij) = \lambda d_c(ij) + (1-\lambda) d_s(ij)$ for $d_{edge}(ij)$ among all borders nodes. Here, we set $\lambda = 0.5$.

In this paper, we combine the low-level features and deep features with a simple linear combination to measure the similarity between superpixels i and j:

$$a_{ij} = exp\left(-\frac{\tau d_{ij}^d + (1-\tau) d_{ij}^l}{2\sigma_w^2}\right). \tag{6}$$

where τ is balance coefficient, and we set as 0.4.

3.2. Salient object probability with background information

3.2.1. Background seeds acquisition

The nodes along boundary are usually employed as the background seeds because object is likely to appear at or near the center of an image. Nevertheless, in some cases, the salient object may also appear at the border regions, misleading the saliency value to be 0 if the object is mistaken for background regions (as shown in Fig. 5(c)). Consequently, we propose an mechanism based on divergence information to select robust background regions from the border regions.

The background regions present such situation that they usually have diverse appearances and widely distributed. Based on the above findings, we extract background seeds via divergence information. [56] proposed the concept of color and texture scatter degree, where color and texture similarity are computed according to the form of Gaussian functions. Different from [56] using color and texture scatter to directly obtain saliency map, this paper gives a unique definition of divergence based on edge weight and center prior, and we use the divergence information to extract background seeds. According to edge weight, we define the divergence of superpixel i as

$$Div_c(i) = normalize\left(\sum_{j=1}^{N} a_{ij} \cdot |s_j - \rho_i|\right), \tag{7}$$

and

$$\rho_i = normalize\left(\sum_{j=1}^{N} a_{ij} \cdot s_j\right). \tag{8}$$

In Eqs. (7) and (8), s_j is the position of superpixel j; ρ_i indicates the mean position of superpixel i, which has been subjected to weighting; and N is the number of superpixels; $normalize(x)$ is a function that normalizes x. Considering that image center region is more likely to be salient object, we introduce the center prior guided divergence information, which is defined in the following formulation

$$Div_m(i) = normalize\left(\sum_{j=1}^{N} a_{ij} \cdot |s_j - M|\right), \tag{9}$$

Table 1
Probability distribution on different datasets.

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
PASCAL-S	7.17	23.8	16	14.8
HKU-IS	3.3	9.3	6.7	6.5

where M is the normalized spatial coordinate of the image center. Then, the divergence value of superpixel i is finally measured using

$$Div(i) = normalize(Div_c(i) + Div_m(i)). \tag{10}$$

In this formulation, the larger the $Div(i)$ is, the more the corresponding superpixel i is likely to be the background.

Based on the divergence information, we can remove the unreliable background regions in border set by a threshold value. As shown in Table 1, we observer that there is different probabilities of the object appearing at different boundaries. For example, the probability of the object connecting with the down side is larger than those of three sides. If we adopt the same threshold value for all border sides, the results may be not very accurate. Accordingly, we choose different threshold values for different sides instead of setting a single value. In specific, the superpixels are removed from the border sets if their divergence values are lower than the mean divergence over the entire map for down side. For top side, we set the threshold value is one third of the down side' s, which is also twice for the left and right sides' s for simplicity. Finally, we can get the robust background seeds. As shown in Fig. 5(d), our method can suppress effectively such a situation that the object on the border is mistaken as background.

3.2.2. Background-based saliency map

According to the graph G with the above defined edge weight and the well-defined background seeds, the confidence of each superpixel i as foreground can be computed by cumulating the edge weights of the shortest path from i to the virtual background node VT_{bp}:

$$con_i^{bp} = \min_{u_1=v_i,\ldots,u_k=VT_{bp}} \sum_{j=1}^{k-1} m_{u_j u_{j+1}}, \tag{11}$$

where $m_{ij} = 1 - a_{ij}$. The magnitude of con_i^{bp} represents the difference between superpixel i and the well-defined background region, in other words, the larger the con_i^{bp} is, the lower the probability of superpixel i belonging to background is. For a node representing a background superpixel, the confidence of superpixel i is normally lower resulting from the shorter of the path to the virtual background node and the smaller of the edge weights over the shortest path. Similarly, it is usually larger for a node representing a foreground node. As shown in Fig. 5(e), the method of the shortest path to the virtual node with the help of the select background nodes works well.

3.3. Salient object probability with foreground information

3.3.1. Foreground seeds acquisition

While the background-based saliency map can pop out the object well whereas in some cases due to some true background superpixels not being involved in the background seeds and the faint difference between background and foreground (as shown in Fig. 5(e)). Therefore, we exploit foreground seeds to construct foreground-based map to inhibit background noises.

To obtain foreground regions, many researchers select the superpixels, having larger saliency values than an adaptive threshold, as foreground seeds [13,24]. As we know, visual rarity captures the fact that human eyes are often attracted to the rare features in an image but not to the common features. Different from previous methods, we consider the spatial compactness of salient objects and visual rarity and used the parametric maximum flow method to obtain reliable foreground regions. The foreground regions can be obtained by solving the following optimization problem:

$$f^* = \arg\min_f \left(\sum_{i=1}^{N} (-\ln S_i^{bp} + \eta R_i) f_i + \sum_{1<i<j<N} b_{ij} f_i f_j \right), \tag{12}$$

where $f_i \in \{0, 1\}$ indicates whether the corresponding superpixel i belongs to the foreground region, S_i^{bp} is saliency value of superpixel i in background-based saliency map, R_i indicates visual rarity of superpixel i, $b_{ij} = exp(-(S_i^{bp} - S_j^{bp})/2\sigma_b^2)$ and $\sigma_b^2 = 0.1$. According to the visual rarity, image background have strong tendency to capture larger area than salient object in an image while superpixels with rare color features are more likely to be salient foreground. Based on above considerations, visual rarity of superpixel i can be defined as:

$$\begin{aligned} R(i) &= normal\left(\sum_{j=1}^{N} a_{ij}^1 + \sum_{j=1}^{N} a_{ij}^2 \right), \\ s.t.\quad a_{ij}^1 &= \begin{cases} a_{ij} & j \in N_i \\ 0 & otherwise \end{cases} \\ a_{ij}^2 &= \begin{cases} a_{ij} & d_c(i,j) < \phi \\ 0 & otherwise \end{cases} \end{aligned} \tag{13}$$

where N_i denotes the spatial neighbors of node i, $d_c(i, j)$ denotes the color distance between the nodes i and j, and ϕ is a constant. As shown in Fig. 6(e)–(f), our method can yield more compact foreground regions than those by adaptive threshold or using Eq. (12) without introducing rare term. In addition, our results can more accurately highlight the salient objects.

3.3.2. Foreground-based saliency map

Similarly, we can also compute the confidence of each superpixel i as background by accumulating the edge weights of the shortest path from i to the virtual foreground node VT_{fp}:

$$con_i^{fp} = \min_{u_1=v_i,\cdots,u_k=VT_{fp}} \sum_{j=1}^{k-1} m_{u_j u_{j+1}}. \tag{14}$$

If the con_i^{fp} is smaller, it represents a higher similarity between i and the reliable foreground regions, which also means that it has a higher confidence to be foreground. As shown in Fig. 6(g), the undesired highlighted background regions in the background-based saliency map are greatly suppressed in the foreground-based one owing to the contrast to the foreground seeds.

3.4. Saliency integration and refinement

3.4.1. Integration

The background-based saliency map can highlight the object, however, the background noise may be not strongly suppressed due to the incompleteness of the background type in the background node set (as shown in Fig. 7(b)). On the other hand, the foreground-based saliency map contains most of the foreground and can suppress the background noise well when the background super pixel and the foreground node set are compared, however, small feature differences are amplified in the foreground-based saliency map during comparison because the foreground object features are not exactly the same (as shown in Fig. 7(c)). This suggests that a complementary combination of them could generate

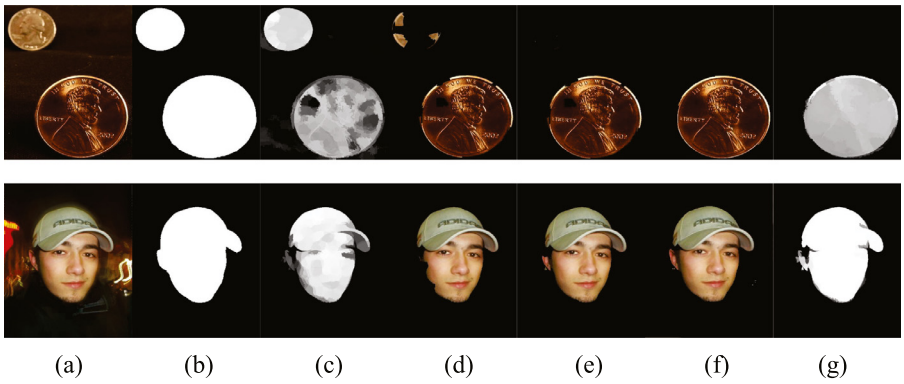


Fig. 6. The effect of the proposed foreground-based saliency map: (a) Input; (b) Ground-truth; (c) Background-based saliency map; (d) Foreground regions generated by adaptive threshold; (e) Foreground regions generated without introducing rare; (f) Foreground regions generated by the proposed method; (g) Foreground-based saliency map.

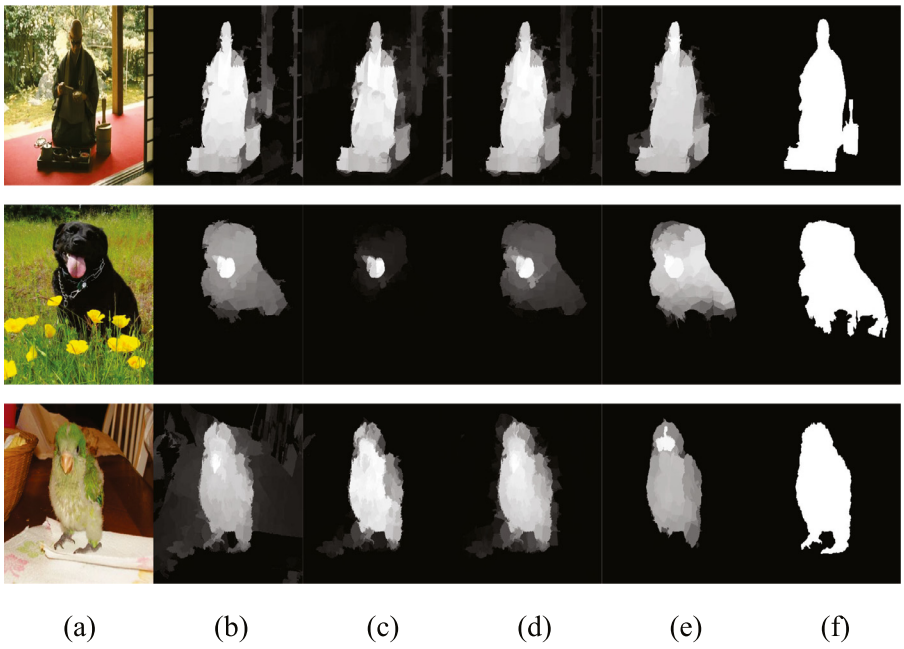


Fig. 7. Improvement effects of foreground regions: (a) Input; (b) Background-based saliency map; (c) Foreground-based saliency map; (d) Integration saliency map; (e) Refinement saliency map; (f) Ground-truth.

better results. Based on [24], we use the following formulation to incorporate them:

$$S_i^{IM} = con_i^{bp} \times (1 - exp(-\kappa \times con_i^{fp})), \tag{15}$$

where κ weights the two saliency maps. As shown in Fig. 7(d), the integrated saliency map can not only highlight the salient object uniformly but also significantly restrain the noises.

3.4.2. Refinement based on extended manifold ranking (EMR)

In order to further optimize saliency maps, we propose an Extended Manifold Ranking (EMR) on graphic model. To improve the performance, two extensions are made upon [13]. First, we introduce the suppressed nodes and active nodes into the data manifold. The suppressed nodes will never disseminate information to the neighboring nodes during diffusion, and active nodes otherwise. $\Omega_f = diag\{\xi_1, \ldots, \xi_N\}$, where $\xi_i = 0$ if the corresponding superpixel i is a suppressed node and an active node is represented by $\xi_i = 1$. Different from [36], which defined so-called sink points mainly based on center weight, we adopt objectness map [57] and integrated map to determine suppressed nodes and active nodes. Herein, the higher threshold of the two-level Ostu' s adaptive threshold method [55] is conducted on objectness map (OM) and the integrated map (IM), to respectively obtain fg_{OM} and fg_{IM}. After that, we assume that the union set $fg_T = fg_{OM} \bigcup fg_{IM}$ is highly confident object regions, in other words, the regions in the complement of fg_T are more likely to be background. Therefore, we set $\xi_i = 1$ if the superpixel $i \in fg_T$ and 0 otherwise.

In addition, the appearance of manifold ranking is based on the hypothesis that the background has a high contrast with objects, which determines that it is easy to imprecisely pop out salient objects when the scene is complex and low-contrast between foreground and background. Accordingly, we merge a quantity of superpixels together into much bigger regions using the mid-level clustering algorithm [58]. As a result, the nodes with the same clusters would have the similar value. Therefore, the mid-level similarity matrix $P = [p_{ij}]_{N \times N}$ can be defined as:

$$p_{ij} = w_{ij} + q_{ij}, \tag{16}$$

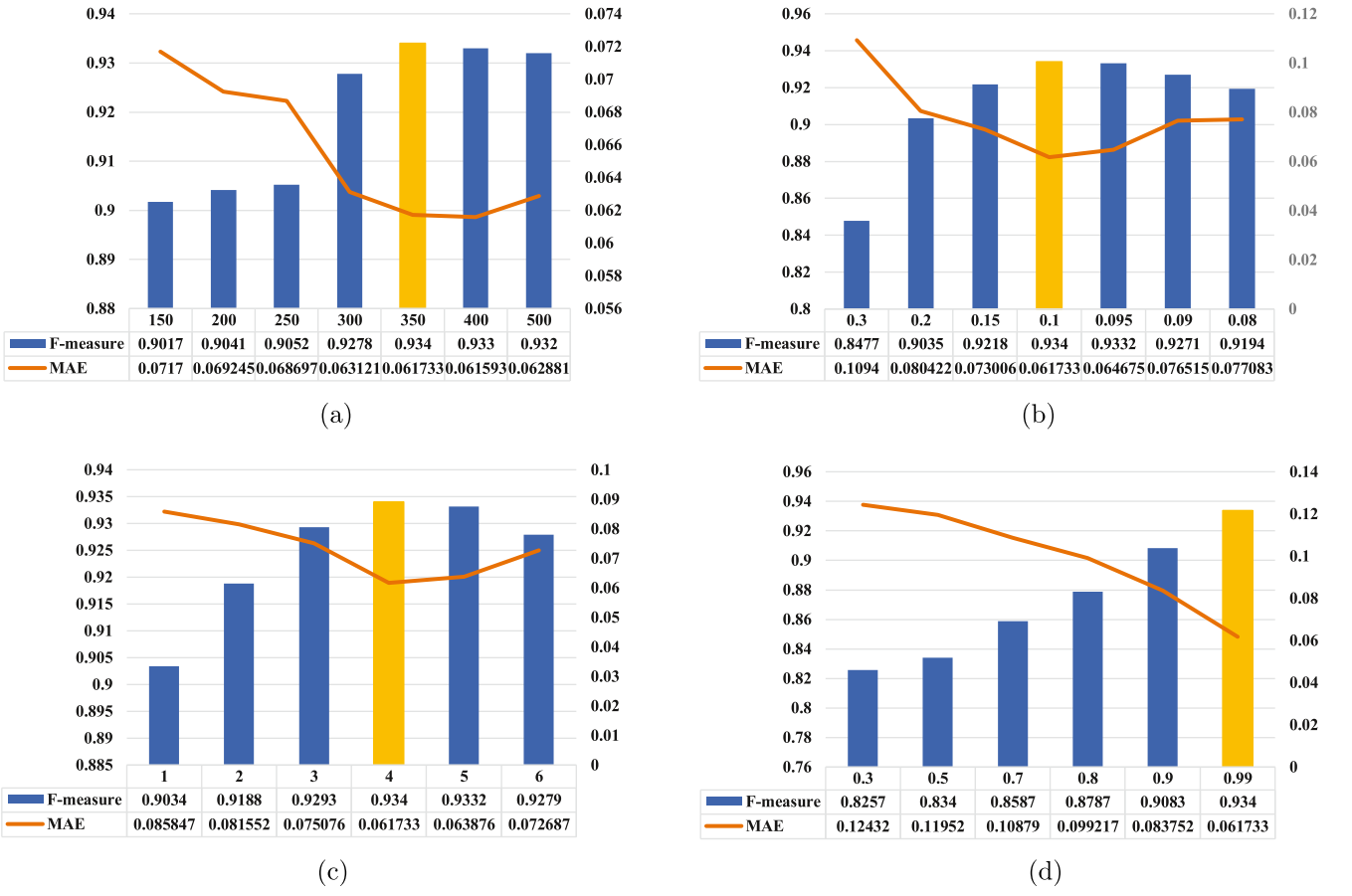


Fig. 8. Parameters Setting. (a) the number of superpixels N, (b) σ_w^2 (in Eq. (6)), (c) κ (in Eq. (15)), (d) ∂ (in Eq. (21)).

and

$$q_{ij} = \begin{cases} 1 & i \text{ and } j \text{ are in the same cluster,} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

$$w_{ij} = \begin{cases} a_{ij} & j \in N_i. \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

Where N_i denotes the spatial neighbors of node i. Based on above definitions, the optimal ranking of queries are computed by solving the following optimization problem:

$$f^* = \arg\min_f \frac{1}{2}\left(\sum_{i,j=1}^{n} p_{ij}\| \frac{\xi_i f_i}{\sqrt{d_i}} - \frac{\xi_j f_j}{\sqrt{d_j}}\|^2 + \mu \sum_{i=1}^{n} \|\xi_i f_i - S_i^{IM}\|^2\right), \tag{19}$$

where the parameter μ controls the balance of the smoothness constraint (the first term) and the fitting constraint (the second term). The minimum solution is computed by setting the derivative of the above function to be zero. The resulted ranking function can be written as:

$$f^* = (I - \partial T\Omega_f)^{-1} s, \tag{20}$$

where I is an identity matrix, $\partial = 1/(1+\mu)$. and T is the normalized Laplacian matrix, $T = D^{-1/2}PD^{-1/2}$. Here, $D = diag\{d_1, \ldots, d_N\}$ is a normalized degree matrix and $d_i = \sum_j p_{ij}$. We can get another ranking function by using the unnormalized Laplacian matrix in Eq. (20) inspired by [13] which can achieve better performance:

$$f^* = (D - \partial P\Omega_f)^{-1} s. \tag{21}$$

The saliency of each node is defined as its ranking score computed by Eq. (21) which is rewritten as

$$f^* = Bs. \tag{22}$$

The learnt optimal affinity matrix B is equal to $(D - \partial P\Omega_f)^{-1}$. The effectiveness of our refinement is illustrated in Fig. 7(e). Our method is able to detect the foreground uniformly and extract the well-defined object boundary.

4. Experimental results

To evaluate our saliency detection algorithm, we conduct a series of experiments on six benchmark datasets with pixel-wise manually labeled ground truth, including ASD [59], DUT-OMRON [13], ECSSD [60], SOD [61], PASCAL-S [62] and HKU-IS [39]. ASD includes 1000 images selecting from the MSRA salient object database [63], which covers a large variety of scenarios. This dataset usually has only one salient object in an image and there are strong contrast between backgrounds and foregrounds. DUT-OMRON consists of 5168 images natural images carefully labeled by five users. So far, none of existing saliency methods can achieve a high accuracy on this dataset. ECSSD contains 1000 semantically meaningful and structurally complex images acquired from the Internet with pixel-level saliency labeling. SOD consists of 300 images selected from the Berkeley Segmenttation dataset with labeled ground truth. The dataset is rather difficult with different biases such as number of salient objects, image clutter and center-biases and so on. PASCAL-S contains 850 natural images labeled by 12 users. The dataset was built on the validation set of the PASCAL

VOC 2010 segmentation challeng. HKU-IS contains 4447 challenging images with low contrast and multiple salient objects.

Although many datasets have been designed for saliency detection in different scenarios, there are few data sets on object-boundary contact. The detection on the scene where salient object touches the boundary has always been a big problem in the salient object detection community. Therefore, it is necessary to design such a dataset to test the performance of salient object detection method in handling such scenarios. In this paper, we devise a new OTB saliency dataset, which contains 1044 images. These images are collected from MSRA10K [9], ECSSD, PASCAL-S, Judd-A [64] and Imgsal [65]. All of the datasets come with human-labeled pixel-wise ground truth. To the best of our knowledge, the designed OTB is the largest dataset to evaluate the performance of salient object detection method on the scene that boundaries are adjacent to the salient objects.

4.1. Evaluation metrics and parameters

4.1.1. Evaluation metrics

To compare the performance, we adopt several evaluation criterions including the precision-recall (PR) curve, the F-measure, area under the ROC (receiver operating characteristic) curve (AUC), and mean absolute error (MAE).

A given saliency map can be converted into several binary maps by segmenting it with a threshold varying from 0 to 255 [59]. And then, a pair of precision and recall values can be generated by comparing the binary map and the ground truth. A PR curve is then obtained by varying the threshold from 0 to 1. Fore comprehensively assessing the salient object detection model, we also compute the maximal F-measure, which is a harmonic mean of precision and recall. Here the F-measure is defined as:

$$F_\beta = \frac{(1+\beta^2)\cdot precision \cdot recall}{\beta^2 \cdot precision + recall}, \tag{23}$$

where we set $\beta^2 = 0.3$ to emphasize precision [9]. To better evaluate the results, we also introduce the Ared Under Curve (AUC) which sums the area under of the receiver operating characteristic (ROC) curve. The ROC curve is achieved by true positive rates and false positive rates obtained when we compute the PR curve.

Note that the PR curve is limited to detecting the true positive pixels and neglecting the correct negative elements detected. To address such issue, we employ the mean absolute error (MAE) as another evaluation criterion, which is defined as the average pixel-wise absolute difference between the ground truth (G) and saliency map (S) [66].

$$MAE = \frac{1}{W \times H}\sum_{x=1}^{W}\sum_{y=1}^{H}|S(x,y) - G(x,y)|, \tag{24}$$

where W and H are the width and height of the saliency map S, respectively.

4.1.2. Parameters setup

There are four key parameters in our proposed method: the number of superpixels N, σ_w^2 (in Eq. (6)), κ (in Eq. (15)) and ∂ (in Eq. (21)). We test different values for these parameters on the 1000-image ASD database using MAE and F-measure. As shown in Fig. 8, We set $N = 350$, $\sigma_w^2 = 0.1$, $\kappa = 4$ and $\partial = 0.99$.

4.2. Comparison with state-of-the-art

4.2.1. Comparison with conventional methods

We extensively present comparison of the proposed algorithm against thirteen state-of-the-art conventional saliency detection methods including GS [23], MR [13], BFS [24], BSCA [67], LPS [20], MB [68], NCS [69], RR [25], SP [70], SRD [71], SBD [72], SG [73], RCRR [74]. The experiments are conducted on the six datasets ASD, ECSSD, DUT-OMRON, PASCAL-S, SOD, and HKU-IS. The results are shown in Fig. 9 and Table 2.

As reported in Fig. 9 and Table 2, we can note that our method significantly outperforms other state-of-the-art algorithms on all of the dataset in terms of all evaluation metrics. Take the challenging ECSSD dataset for example, our PR curve is better than others with dominant advantages, and its F-measure, AUC and MAE exceed the second best method with large gaps. Specifically, it improves by 6.91% and 2.04% over the second best method in terms of F-measure and AUC, respectively, and its MAE is 12.2% lower than the second best. Besides, some methods may have high precision value, but they suffer from low recall, such as MR on DUT-OMRON and SBD on HKU-IS. Comparatively, our method achieves more balanced relationship between precision and recall, thus, we can obtain the best F-measure and AUC on all the datasets. The promising results of the proposed method validate the strong robustness and capacity of our method contributed by the robust computational mechanism to extract reliable background seeds and compact foreground regions, the reasonable edge weight combining high-level semantic information and low-level information, as well as the appropriate refinement.

4.2.2. Comparsion with learning based methods

We also compare our method against thirteen state-of-the-art learning saliency detection methods, namely BL [75], HDCT [19], MILPS [76], DRFI [14], MDF [39], LEGS [77], MCDL [38], DLC [78], DHSNet [17], SCSD-HS [79], UCF [80], PAGR [81] and PiCANet [45], and the latter nine algorithms are deep learning based methods. Their experimental results are summarized in Fig. 10 and Table 3.

Compared with non-deep learning based methods, such as BL, HDCL, MILPS and DRFI, our method can still maintain significant advantages against the comparison learning based methods. Take the PASCAL-S as example, its F-measure is 3.37% higher than the best BL (0.7278 to 0.7041), its AUC is 0.99% higher than the best DRFI (0.79253 to 0.78474) and its MAE is 15.1% lower than the best DRFI (0.17829 to 0.21003). For deep learning based methods, unsurprisingly the overall performances of these methods are significantly higher than those of the conventional methods in Table 2, due to the high-level features involved in their learning processes. As shown in Table 3, the proposed method is competitive with deep learning based approaches. For example, the F-meausres and AUCs of our model are better than MCDL on all datasets except for ECSSD. In addition, we note that deep learning based methods seem sensitive to the training tactics used. For DCL, it gains 16.1% in F-measure and 3.4% in AUC than our method on SOD dataset, possibly due to their deep contrast network using the MSRA-B dataset for training. Nevertheless, DCL loses 0.52% in AUC compared with our proposed method on DUT-OMRON dataset. The comparative results can reveal the shortcoming of supervised approach while unsupervised method can make up for it.

4.2.3. Visual comparison

Fig. 11 shows some visual comparisons on different methods with several types of images. For single-object images shown in the first two rows, our method can uniformly detect the salient object with few thinly scattered regions. In addition, we can observe that the proposed method can also effectively handle the challenging cases where background is complex or shares similar pattern with foreground. For example, as shown in third and fourth rows, compared with other methods, our approach can perform better to distinguish between background and foreground regions with similar appearance. Due to the reasonable edge weight combining with robust background and foreground seeds, our algorithm can effectively assign different value to the dissimilar nodes resulting

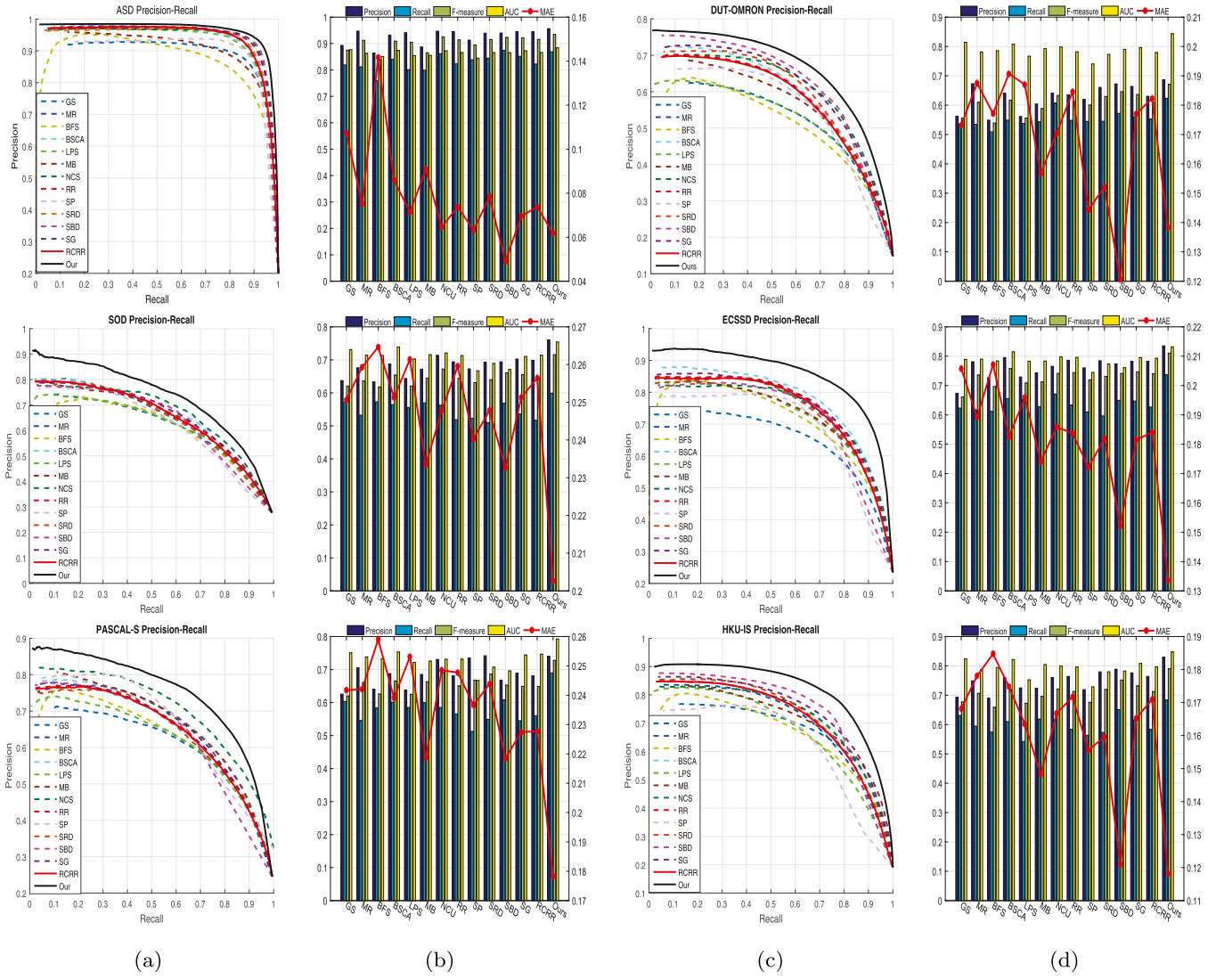


Fig. 9. Quantitative evaluations against the state-of-the-art conventional methods on six datasets.

in separating them successfully. For images with multiple objects shown in last two rows, some methods could not highlight the integrity of the objects or include undesired background noises into the final results. Comparatively, our method has a great ability to pop out all the salient objects successfully. The extended manifold ranking algorithm can further enhance the performance, thus detecting more accurate salient objects even in complex scene.

4.3. Validation of the proposed approach

In order to illustrate the effectiveness of the selective mechanism for the salient objects which are touching the border of images, we compute the performance for the final saliency maps on the proposed OTB dataset. We compare with MR [13], LPS [20], MAP [32]. For fair comparison, all their edge weights are set the same with our method. As shown the dotted lines in Fig. 12(a), our method outperforms those competitive methods. In addition, we further introduce our selective mechanism to these methods by replacing their background seeds with our background seeds to evaluate its generality. As can be seen from the solid lines in Fig. 12(a), their performance can be further improved, which illustrates the effectiveness of our selective mechanism. Furthermore, we also evaluate some deep learning baselines (MCDL, MDF) on the OTB dataset. It can be seen from Fig. 12(a) that we can get a similar effect to the deep learning algorithms, which further demonstrates the effectiveness of the proposed method. Fig. 12(b) provides the visualization results.

To further demonstrate the effectiveness of the graph, we compare the performance evaluation of the proposed graph with the existing graphs (2-ring, 3-ring, 4-ring) for saliency detection using the proposed background seeds. Quantitative evaluation on ECSSD dataset is shown in Fig. 13(a). In this paper, we set the edge weight as zero between virtual nodes and seed nodes. Hence, it may not be appropriate to choose too long graph connection. As can be observed, our method outperforms other methods, which proves the effectiveness of the proposed graph model.

We also examine different edge weight computation methods using different features in Fig. 13(b). We compute the performance evaluation of the final saliency maps on ECSSD dataset. According to Fig. 13(b), we can know that the integrated features outperform single feature. This is because the integrated features take both high-level semantic and low-level detailed information into consideration.

To evaluate the effectiveness of our method to obtain foreground seeds, we also compute the performance evaluation for final saliency maps on ECSSD dataset. As reported in Fig. 13(c), the

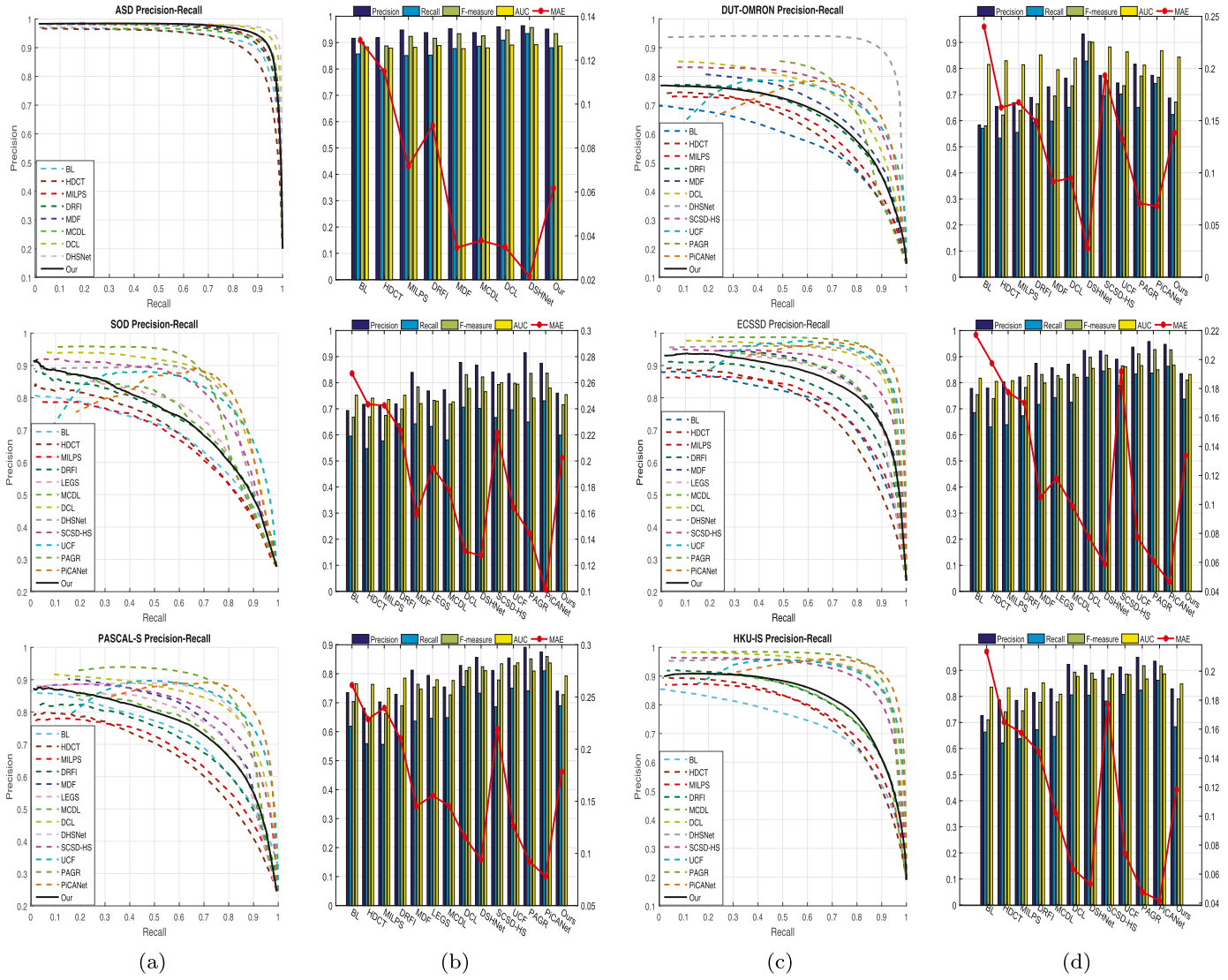


Fig. 10. Quantitative evaluations against the state-of-the-art learning based methods on six datasets.

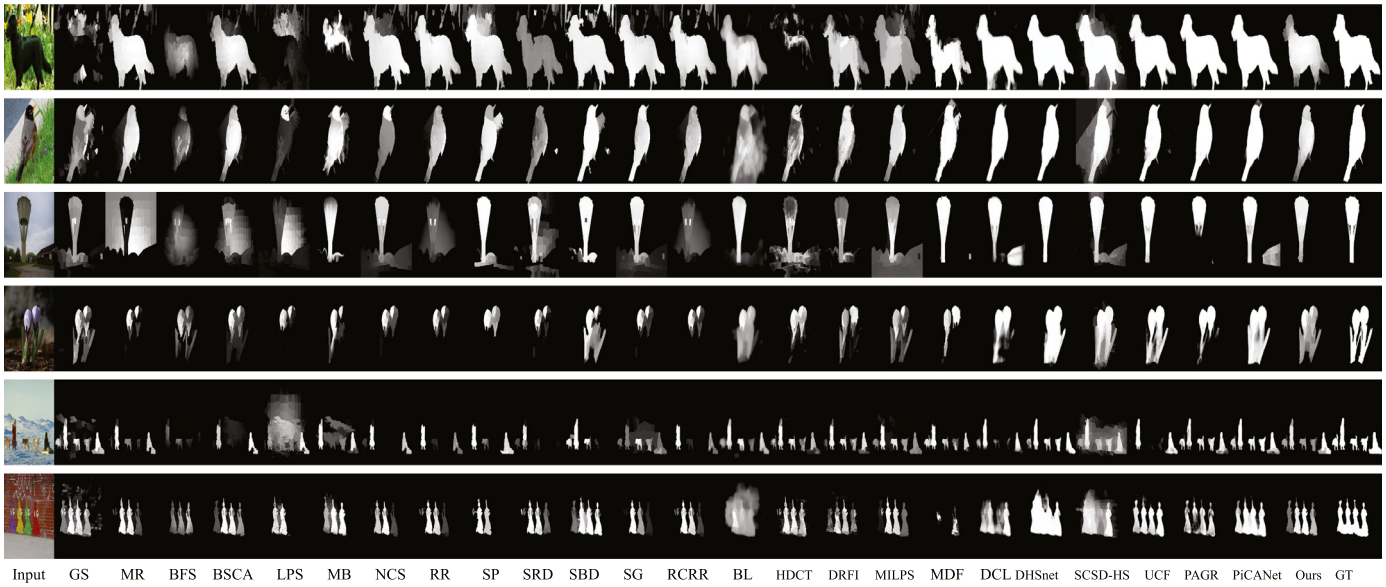


Fig. 11. Visual comparison on six datasets.

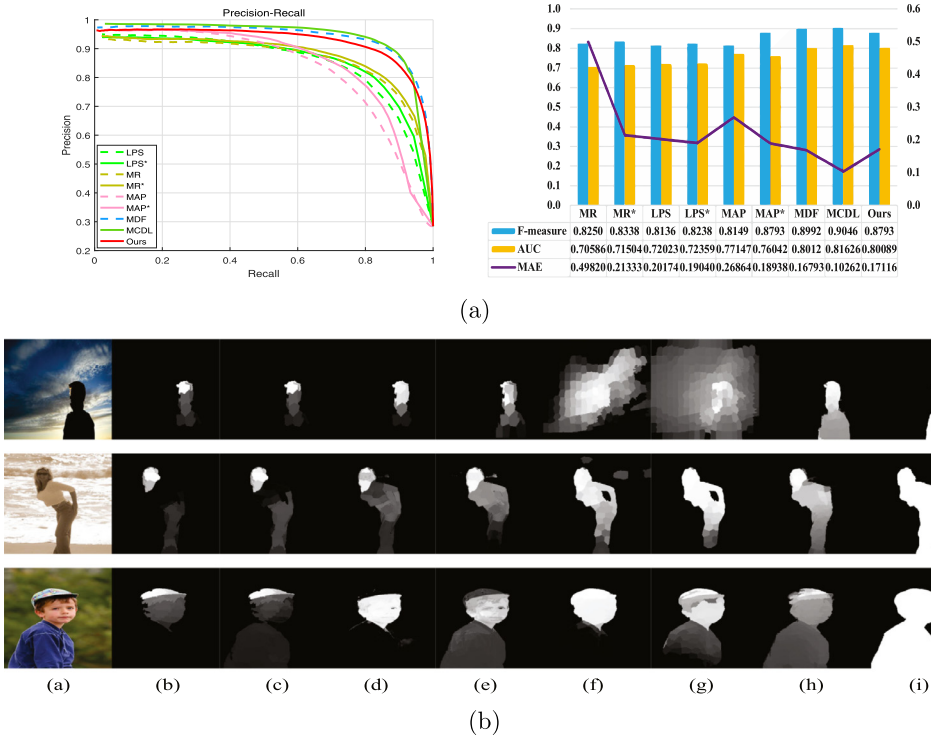


Fig. 12. Effects of our background selective mechanism. (a) Quantitative comparison results. (b) The visualization result, from left to right: Input, MR, MR*, LPS, LPS*, MAP, MAP*, Ours and Ground-truth.

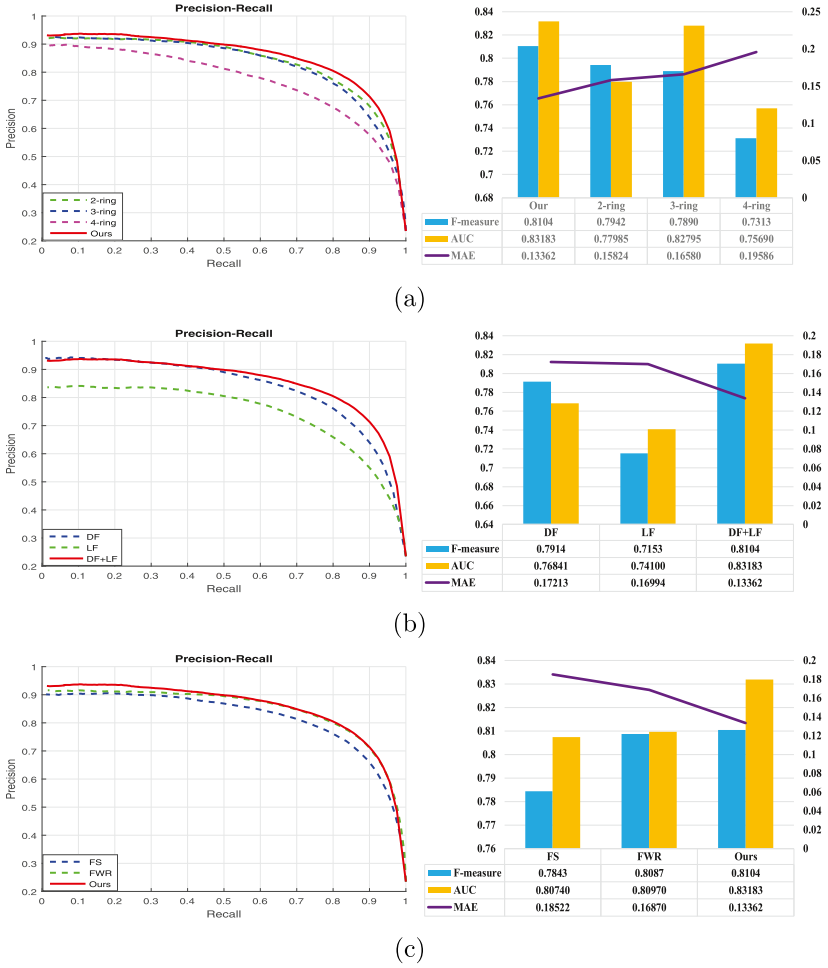


Fig. 13. (a) Effects of the proposed graph structure; (b) Effects of the edge weight; (c) Effects of the proposed foreground regions calculation mechanism.

Table 2
Quantitative results of proposed method against the state-of-the-art conventional methods in terms of F-measure, AUC and MAE.

	Metric	Ours	GS	MR	BFS	BSCA	LPS	MB	NCS	RR	SP	SRD	SBD	SG	RCRR
(A)ASD	F-measure ↑[a]	**0.934**	0.875	0.912	0.8373	0.909	0.905	0.865	*0.9255*	0.914	0.8944	0.9153	***0.9237***	0.9218	0.9157
	AUC ↑	**0.88699**	*0.87683*	0.8632	0.85076	*0.87436*	0.85409	0.85543	0.87262	0.8662	0.84456	0.8657	0.86565	0.8725	0.8662
	MAE ↓[a]	*0.061733*	0.10725	0.07512	0.14168	0.085774	0.071469	0.090757	0.06455	0.0738	***0.063158***	0.078317	**0.049042**	0.06947	0.073787
	Metric	Ours	GS	MR	BFS	BSCA	LPS	MB	NCS	RR	SP	SRD	SBD	SG	RCRR
(B)DUT	F-measure ↑[a]	**0.6711**	0.5559	0.61	0.5391	0.617	0.5558	0.5888	0.6331	0.6127	0.6007	0.6296	*0.646*	***0.6365***	0.6105
	AUC ↑	**0.84374**	*0.81397*	0.78133	0.78579	***0.80817***	0.76708	0.79304	0.79834	0.78179	0.74065	0.77313	0.79008	0.7963	0.77975
	MAE ↓[a]	*0.13822*	0.17322	0.18745	0.17715	0.19073	0.18697	0.15664	0.1704	0.18452	***0.14423***	0.15219	**0.12063**	0.177	0.18231
	Metric	Ours	GS	MR	BFS	BSCA	LPS	MB	NCS	RR	SP	SRD	SBD	SG	RCRR
(C)SOD	F-measure ↑[a]	**0.7155**	0.6205	0.6363	0.618	0.6541	0.6208	0.6449	*0.672*	0.6438	0.631	0.6398	***0.6604***	0.6553	0.6461
	AUC ↑	**0.75421**	*0.7313*	0.71406	0.71338	*0.7387*	0.70284	0.71543	0.72079	0.71296	0.6669	0.68933	0.67099	0.71049	0.71404
	MAE ↓[a]	**0.20263**	0.25066	0.25932	0.26464	0.25128	0.26136	***0.23338***	0.24785	0.25952	0.24031	0.24782	*0.23263*	0.25125	0.25639
	Metric	Ours	GS	MR	BFS	BSCA	LPS	MB	NCS	RR	SP	SRD	SBD	SG	RCRR
(D)ECSSD	F-measure ↑[a]	**0.8104**	0.661	0.736	0.6967	*0.758*	0.709	0.713	0.7413	0.744	0.7193	0.7312	0.7411	***0.7465***	0.74
	AUC ↑	**0.83183**	0.78894	0.78994	0.78359	*0.81523*	0.78247	0.78327	***0.7975***	0.7959	0.74447	0.77377	0.76085	0.7953	0.79317
	MAE ↓[a]	**0.13362**	0.20581	0.18919	0.20717	0.18243	0.19588	0.17407	0.18585	0.1837	***0.17206***	0.182	*0.15209*	0.18148	0.18404
	Metric	Ours	GS	MR	BFS	BSCA	LPS	MB	NCS	RR	SP	SRD	SBD	SG	RCRR
(E)PASCAL-S	F-measure ↑[a]	**0.7278**	0.6202	0.6608	0.6261	0.6655	0.6248	0.6634	*0.6908*	0.6506	0.6677	***0.6862***	0.6528	0.6489	0.6483
	AUC ↑	**0.79253**	***0.75055***	0.73829	0.73226	*0.75321*	0.72158	0.72565	0.73124	0.73158	0.66698	0.70734	0.69618	0.74427	0.7464
	MAE ↓[a]	**0.17829**	0.24168	0.24209	0.25901	0.2391	0.2531	***0.21878***	0.24849	0.24774	0.23663	0.24382	*0.21844*	0.22743	0.22768
	Metric	Ours	GS	MR	BFS	BSCA	LPS	MB	NCS	RR	SP	SRD	SBD	SG	RCRR
(F)HKU-IS	F-measure ↑[a]	**0.7903**	0.6772	0.7065	0.659	0.7192	0.6726	0.6962	0.7214	0.7131	0.6759	0.7196	*0.7516*	***0.7321***	0.7127
	AUC ↑	**0.8479**	*0.82389*	0.78699	0.79387	***0.82173***	0.75233	0.80432	0.7998	0.7963	0.7286	0.7798	0.78255	0.80841	0.7966
	MAE ↓[a]	**0.11818**	0.16814	0.17809	0.18474	0.17476	0.1635	***0.14819***	0.16667	0.17174	0.15553	0.15961	*0.12095*	0.16514	0.171

[a] The up-arrow ↑ indicates the larger value achieved, the better performance is, while the down-arrow ↓ indicates the smaller, the better. [b]The best three results are highlighted with bold, italic and bold-italic fonts, respectively.

Table 3
Quantitative results of proposed method against the state-of-the-art learning based methods in terms of F-measure, AUC and MAE.

	Metric	Ours	BL	HDCT	MILPS	DRFI	MDF*	LEGS*	MCDL*	DCL*	DSHNet*	SCSD-HS*	UCF*	PAGR*	PiCANet*
(A)ASD	F-measure ↑[a]	***0.934***	0.902	0.8874	0.9236	0.9166	0.9339	–	0.9256	*0.948*	**0.9572**	–	–	-	-
	AUC ↑	0.88699	0.8837	0.87922	0.88209	***0.88878***	0.87697	–	0.87996	*0.89086*	**0.89206**	–	–	-	-
	MAE ↓[a]	0.061733	0.1291	0.11487	0.071659	0.090207	*0.034564*	–	0.037854	***0.034822***	**0.021496**	–	–	-	-
	Metric	Ours	BL	HDCT	MILPS	DRFI	MDF*	LEGS*	MCDL*	DCL*	DSHNet*	SCSD-HS*	UCF*	PAGR*	PiCANet*
(B)DUT	F-measure ↑[a]	0.6711	0.5798	0.6214	0.6389	0.6641	0.6944	–	–	0.7336	**0.9029**	0.7539	0.7342	*0.7709*	***0.7663***
	AUC ↑	0.84374	0.8155	0.83864	0.81438	0.8521	0.79498	–	–	0.83939	**0.90066**	*0.88235*	0.86378	0.81294	***0.86812***
	MAE ↓[a]	0.13822	0.2401	0.16692	0.1675	0.14959	0.09157	–	–	0.094847	**0.027392**	0.19343	0.13222	***0.070943***	*0.067896*
	Metric	Ours	BL	HDCT	MILPS	DRFI	MDF*	LEGS*	MCDL*	DCL*	DSHNet*	SCSD-HS*	UCF*	PAGR*	PiCANet*
(C)SOD	F-measure ↑[a]	0.7155	0.6679	0.6692	0.6744	0.6992	0.784	0.7321	0.7178	***0.831***	0.8221	0.7932	0.7982	0.8358	**0.8364**
	AUC ↑	0.75421	0.75208	0.76335	0.73538	0.75192	0.72012	0.72929	0.72671	0.77796	0.7656	**0.80012**	0.79442	0.74096	***0.77962***
	MAE ↓[a]	0.20263	0.26703	0.22876	0.24263	0.22348	0.15875	0.1949	0.17794	***0.13084***	*0.12729*	0.22159	0.16428	0.14473	**0.10121**
	Metric	Ours	BL	HDCT	MILPS	DRFI	MDF*	LEGS*	MCDL*	DCL*	DSHNet*	SCSD-HS*	UCF*	PAGR*	PiCANet*
(D)ECSSD	F-measure ↑[a]	0.8104	0.7544	0.7392	0.7581	0.7819	0.8316	0.8274	0.832	0.8977	0.9055	0.8648	***0.9105***	**0.9268**	*0.9262*
	AUC ↑	0.83183	0.8173	0.81752	0.80768	0.82702	0.79911	0.81495	0.82063	0.85488	0.85364	***0.86158***	*0.86552*	0.84955	**0.86718**
	MAE ↓[a]	0.13362	0.2169	0.19632	0.17734	0.17019	0.10491	0.118	0.098632	0.077529	*0.058819*	0.19188	0.077651	***0.060868***	**0.046417**
	Metric	Ours	BL	HDCT	MILPS	DRFI	MDF*	LEGS*	MCDL*	DCL*	DSHNet*	SCSD-HS*	UCF*	PAGR*	PiCANet*
(E)PASCAL-S	F-measure ↑[a]	0.7278	0.7041	0.6483	0.6625	0.6898	0.7636	0.7543	0.727	0.8105	0.8242	0.7785	***0.8278***	*0.8513*	**0.8592**
	AUC ↑	0.79253	0.76545	0.76335	0.74986	0.78474	0.74726	0.7793	0.77667	0.82206	0.8106	***0.83439***	**0.83762**	0.8093	*0.83733*
	MAE ↓[a]	0.17829	0.26115	0.22876	0.23944	0.21003	0.14526	0.15553	0.14555	0.11521	***0.0938***	0.2197	0.12631	*0.092241*	**0.077885**
	Metric	Ours	BL	HDCT	MILPS	DRFI	MDF*	LEGS*	MCDL*	DCL*	DSHNet*	SCSD-HS*	UCF*	PAGR*	PiCANet*
(F)HKU-IS	F-measure ↑[a]	0.7903	0.7104	0.7409	0.745	0.7772	–	–	0.7786	***0.8929***	0.8895	0.8707	0.8856	**0.9176**	0.9171
	AUC ↑	0.8479	0.83546	0.83292	0.82925	0.85155	–	–	0.80835	0.87657	0.86588	**0.88636**	*0.8837*	0.86704	***0.88607***
	MAE ↓[a]	0.11818	0.21357	0.1647	0.15754	0.1445	–	–	0.10193	0.063428	***0.052966***	0.17727	0.07405	*0.047511*	**0.041715**

[a] The up-arrow ↑ indicates the larger value achieved, the better performance is, while the down-arrow ↓ indicates the smaller, the better. [b] The best three results are highlighted with bold, italic and bold-italic fonts, respectively. [c] The *: deep learning based method.

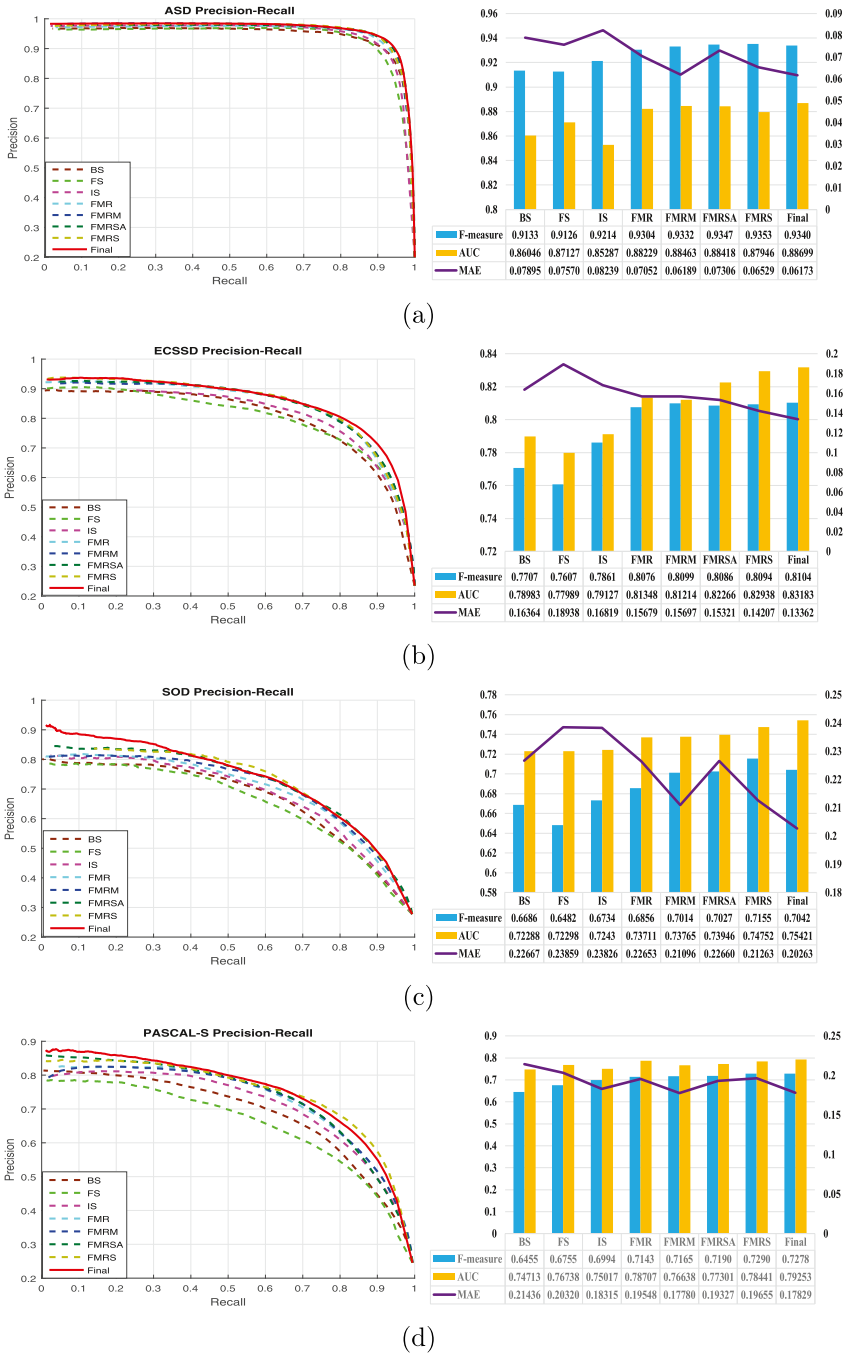


Fig. 14. Performance of the proposed method with different design options on different datasets.

blue curve provides the performance of the method with a single threshold (FS). Similarly, the green curve gives the performance of the method without the rarity term (FWR), and the red curve presents the performance of the final map with rarity term (Ours). The final result can produce better performance than FS and FWR, which demonstrates that our method with rarity term can facilitate the saliency detection.

Since we refine the saliency map based on the EMR algorithm, we evaluate the contributions of separate components in EMR algorithm on four datasets (ASD, ECSSD, SOD and PASCAL-S datasets). Fig. 14 shows the performance of each step in the proposed method, i.e., background-based saliency map (BS), foreground-based saliency map (FS), integrated saliency map (IS), refinement with manifold ranking (FMR), refinement with manifold ranking and mid-level information (FMRM), refinement with manifold ranking and suppressed/active nodes (FMRSA). Combining all these steps makes the system more robust and achieve better performance. In addition, we also compare with the refinement with manifold ranking and sink point (FMRS). Accordingly to experimental data, the overall performance of our method is better than FMRS, which proves the robustness of our method. Furthermore, in order to demonstrate that our refinement method is of high versatility that can improve other methods based propagation when applying our propagation mechanism, we apply our refinement mechanism to optimize state-of-the-art results on ASD and ECSSD datasets, including GS [23], SF [66], MC [82]. For

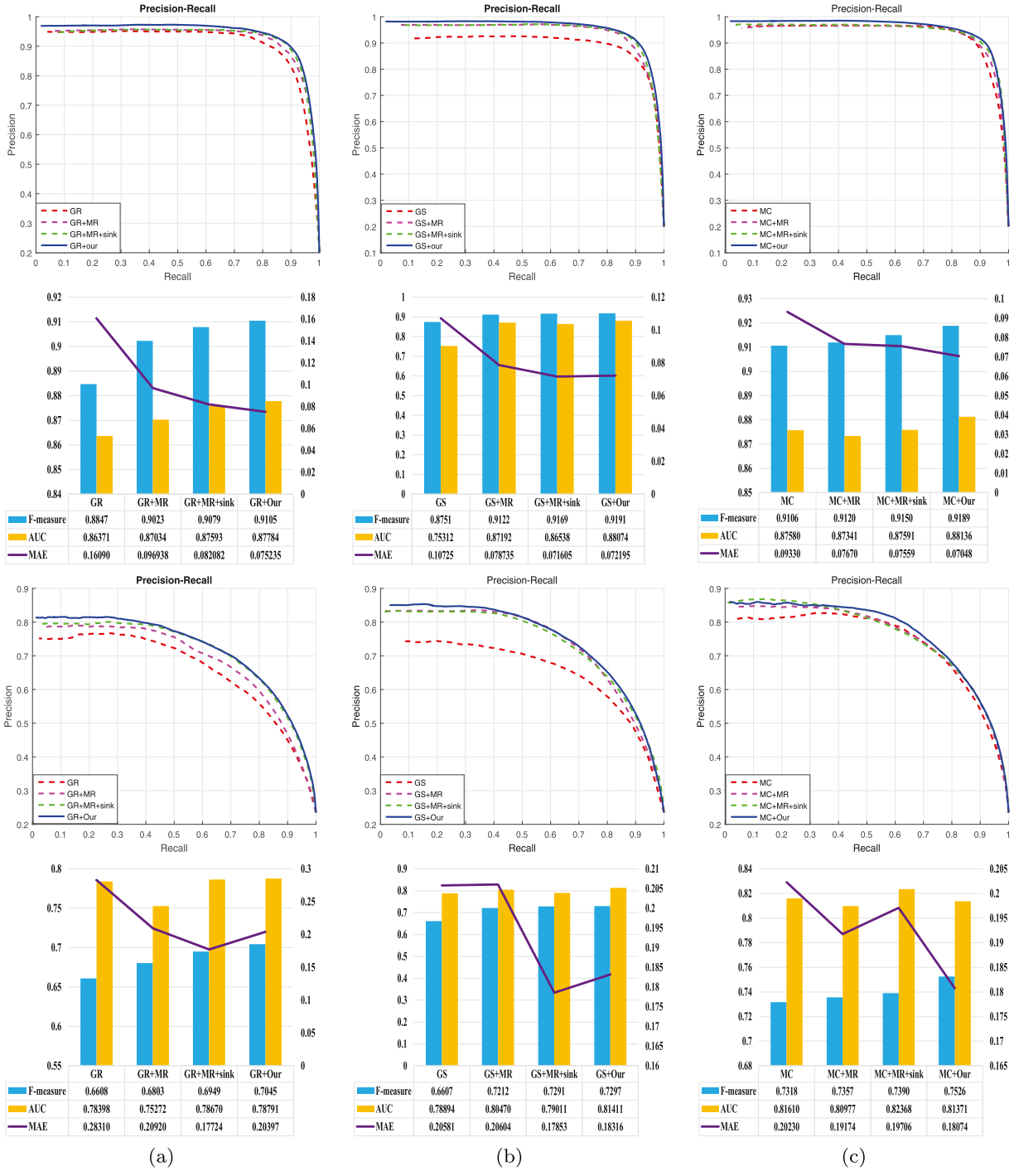


Fig. 15. Performance of different methods and their optimized version by different approaches on ASD and ECSSD dataset.

different methods, we use their final saliency map as the input coarse saliency map. We then apply the optimization methods to refine them. For fair comparison, we adopt the proposed foreground seeds selection method to obtain foreground seeds for manifold ranking (MR) [13]; for MR+sink, we mainly adopt [36] to refine the maps. Performance variation can be clearly observed in Fig. 15. We can see that all of them are significantly improved to a similar level after our method. In addition, we can also find that our performance is better than those optimized by MR [13] or the method introducing sink points into manifold ranking (MR+sink) [36]. The results illustrate the powerful versatility of our refinment mechanism.

4.4. Running time

The Running time test is conducted on a 64-bit PC with Intel Core i5-4460 CPU @ 3.20GHz and 8GB RAM. All the tested codes are provided by the authors and run unchanged in MATLAB R2015a with some C++ mex implementations. Average running time is computed on the ASD dataset. We choose several competitive

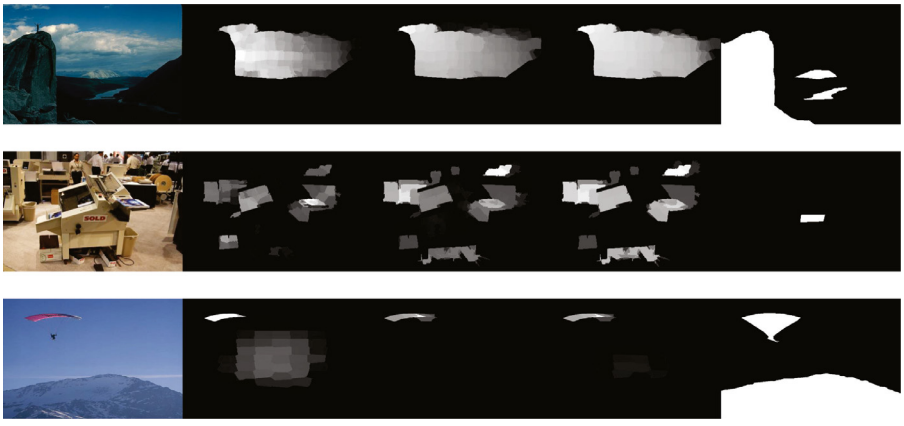


Fig. 16. Some failure examples. (a) Input; (b) background-based saliency maps; (c) foreground-based saliency map; (d) Final saliency maps; (e) Ground-truth.

Table 4
Running average time test results (seconds per image) on ASD.

Method	LPS	MR	RR	BFS	MDF	MCDL	Our
Times	3.376	0.715	3.56	7.513	112.472	6.415	$t1$: 3.1524 $t2$: 1.307

accuracy methods or those akin to ours, and the results are shown in Table 4. Note that Table 4 dispalys the average time of our method for extracting deep features $t1$ and the holistic method takes only 1.307s to process one image without considering extracting deep features. Although being slower than traditional methods, our method still outperforms them both considering the overall evaluation performances. In addition, we can clearly see that the proposed algorithm is faster than other deep learning methods.

4.5. Limitation and analysis

Although our method can perform well in most cases, it still suffers from accurately guaranteeing the completeness of the salient objects in some challenging scenarios in Fig. 16. This is because our model design emphasizes the high-level discriminative image knowledge of objects against the background. Since we adopt unsupervised scheme, we will investigate weakly supervised learning to achieve a better performance in our future work.

5. Conclusions

This paper proposes a novel graph structure for salient object detection in which both background divergence and foreground compactness are utilized. First, we construct a novel graph and define a reasonable edge weight, which considers low-level feature and high-level semantic information. Additionally, two maps are calculated by reliable background seeds and compact foreground regions with two automatic selection schemes. Finally, a robust optimization mechanism is proposed to further refine the final map. Experimental results demonstrate that the proposed method can achieve favorable performance.

Declaration of Competing Interest

None.

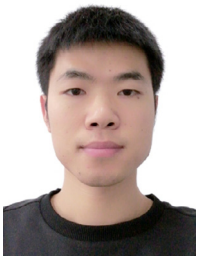
Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61672222, 61572183, 61472131, 61806006), Hunan Natural Science Foundation (Grant No. 2018JJ2054), Science and Technology Key Projects of Hunan province (Grant No. 2015TP1004), Natural Science Research Project of Colleges and Universities in Anhui Province (Grant No. KJ2018A0083).

References

[1] H. Katrin, M.K. Leila, C. Marisa, D.J. Heeger, When size matters: attention affects performance by contrast or response gain, Nat. Neurosci. 13 (12) (2010) 1554–1559.
[2] Z. Liu, R. Shi, L. Shen, Y. Xue, K.N. Ngan, Z. Zhang, Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut, IEEE Trans. Multim. 14 (4) (2012) 1275–1289.
[3] Z. Ren, S. Gao, L.T. Chia, W.H. Tsang, Region-based saliency detection and its application in object recognition, IEEE Trans. Circuits Syst. Video Technol. 24 (5) (2014) 769–779.
[4] Z. Liang, G. Yang, X. Ding, L. Li, An efficient forgery detection algorithm for object removal by exemplar-based image inpainting, J. Vis. Commun. Image Represent. 30 (2015) 75–85.
[5] C. Shi, J. Zhang, Y. Zhang, Content-based onboard compression for remote sensing images, Neurocomputing 191 (2016) 330–340.
[6] R. Li, J. Cai, H. Zhang, T. Wang, Aggregating complementary boundary contrast with smoothing for salient region detection, Vis. Comput. 33 (9) (2017) 1155–1167.
[7] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (2002) 1254–1259.
[8] M.M. Cheng, J. Warrell, W.Y. Lin, S. Zheng, V. Vineet, N. Crook, Efficient salient region detection with soft image abstraction, in: International Conference on Computer Vision, 2013, pp. 1529–1536.
[9] M.M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, S.M. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 569–582.
[10] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, S. Li, Automatic salient object segmentation based on context and shape prior., in: British Machine Vision Conference, 2011, p. 9.
[11] M. Ran, A. Tal, L. Zelnikmanor, What makes a patch distinct? in: Computer Vision and Pattern Recognition, 2013, pp. 1139–1146.
[12] Y. Fang, W. Lin, B.S. Lee, C.T. Lau, Z. Chen, C.W. Lin, Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum, IEEE Trans. Multim. 14 (1) (2012) 187–198.
[13] C. Yang, L. Zhang, H. Lu, R. Xiang, M.H. Yang, Saliency detection via graph-based manifold ranking, in: Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.
[14] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, Salient object detection: a discriminative regional feature integration approach, in: Computer Vision and Pattern Recognition, 2013, pp. 2083–2090.
[15] Z. Wang, D. Xiang, S. Hou, F. Wu, Background-driven salient object detection, IEEE Trans. Multim. 19 (4) (2017) 750–762.

[16] S. Lu, V. Mahadevan, N. Vasconcelos, Learning optimal seeds for diffusion-based salient object detection, in: Computer Vision and Pattern Recognition, 2014, pp. 2790–2797.
[17] N. Liu, J. Han, Dhsnet: deep hierarchical saliency network for salient object detection, in: Computer Vision and Pattern Recognition, 2016, pp. 678–686.
[18] K.Y. Chang, T.L. Liu, H.T. Chen, S.H. Lai, Fusing generic objectness and visual saliency for salient object detection, in: International Conference on Computer Vision, 2011, pp. 914–921.
[19] J. Kim, D. Han, Y.W. Tai, J. Kim, Salient region detection via high-dimensional color transform, in: Computer Vision and Pattern Recognition, 2014, pp. 883–890.
[20] H. Li, H. Lu, Z. Lin, X. Shen, B. Price, Inner and inter label propagation: salient object detection in the wild, IEEE Trans. Image Process. 24 (10) (2015) 3176–3186.
[21] Y. Jia, M. Han, Category-independent object-level saliency detection, in: International Conference on Computer Vision, 2013, pp. 1761–1768.
[22] M. Xu, H. Zhang, Saliency detection with color contrast based on boundary information and neighbors, Vis. Comput. 31 (3) (2015) 355–364.
[23] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: European Conference on Computer Vision, 2012, pp. 29–42.
[24] J. Wang, H. Lu, X. Li, N. Tong, W. Liu, Saliency detection via background and foreground seed selection, Neurocomputing 152 (C) (2015) 359–368.
[25] C. Li, Y. Yuan, W. Cai, Y. Xia, D.D. Feng, et al., Robust saliency detection via regularized random walks ranking., in: Computer Vision and Pattern Recognition, 2015, pp. 2710–2717.
[26] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: International Conference on Neural Information Processing Systems, 2006, pp. 545–552.
[27] Y. Lu, W. Zhang, H. Lu, X. Xue, Salient object detection using concavity context, in: International Conference on Computer Vision, 2011, pp. 233–240.
[28] W. Wang, Y. Wang, Q. Huang, W. Gao, Measuring visual saliency by site entropy rate, in: Computer Vision and Pattern Recognition, 2010, pp. 2368–2375.
[29] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, IEEE Trans. Image Process. 19 (12) (2010) 3232–3242.
[30] Q. Wang, W. Zheng, R. Piramuthu, Grab: visual saliency via novel graph model and background priors, in: Computer Vision and Pattern Recognition, 2016, pp. 535–543.
[31] L. Zhang, C. Yang, H. Lu, R. Xiang, M.H. Yang, Ranking saliency, IEEE Trans. Pattern Anal. Mach. Intell. 39 (9) (2017) 1892–1904.
[32] J. Sun, H. Lu, X. Liu, Saliency region detection based on markov absorption probabilities., IEEE Trans. Image Process. 24 (5) (2015) 1639–1649.
[33] L. Zhang, J. Ai, B. Jiang, H. Lu, X. Li, Saliency detection via absorbing markov chain with learnt transition probability, IEEE Trans. Image Process. 27 (2) (2018) 987–998.
[34] Y. Kong, L. Wang, X. Liu, H. Lu, X. Ruan, Pattern mining saliency, in: European Conference on Computer Vision, 2016, pp. 583–598.
[35] Y. Zeng, M. Feng, H. Lu, G. Yang, A. Borji, An unsupervised game-theoretic approach to saliency detection, IEEE Trans. Image Process. 27 (9) (2018) 4545–4554.
[36] S. Chen, L. Zheng, X. Hu, P. Zhou, Discriminative saliency propagation with sink points, Pattern Recognit. 60 (2016) 2–12.
[37] P. Jiang, H. Ling, J. Yu, J. Peng, Salient region detection by ufo: uniqueness, focusness and objectness, in: International Conference on Computer Vision, 2013, pp. 1976–1983.
[38] R. Zhao, W. Ouyang, H. Li, X. Wang, Saliency detection by multi-context deep learning, in: Computer Vision and Pattern Recognition, 2015, pp. 1265–1274.
[39] G. Li, Y. Yu, Visual saliency based on multiscale deep features, in: Computer Vision and Pattern Recognition, 2015, pp. 5455–5463.
[40] G. Lee, Y.-W. Tai, J. Kim, Deep saliency with encoded low level distance map and high level features, in: Computer Vision and Pattern Recognition, 2016, pp. 660–668.
[41] G. Li, Y. Xie, L. Lin, Y. Yu, Instance-level salient object segmentation, in: Computer Vision and Pattern Recognition, 2017, pp. 2386–2395.
[42] L. Wang, L. Wang, H. Lu, P. Zhang, X. Ruan, Saliency detection with recurrent fully convolutional networks, in: European Conference on Computer Vision, 2016, pp. 825–841.
[43] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, P.H. Torr, Deeply supervised salient object detection with short connections, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (4) (2019) 815–828.
[44] S. Xie, Z. Tu, Holistically-nested edge detection, in: Computer Vision and Pattern Recognition, 2015, pp. 1395–1403.
[45] N. Liu, J. Han, M.-H. Yang, Picanet: Learning pixel-wise contextual attention for saliency detection, in: Computer Vision and Pattern Recognition, 2018, pp. 3089–3098.
[46] J. Su, J. Li, C. Xia, Y. Tian, Selectivity or invariance: boundary-aware salient object detection, arXiv:1812.10066 (2018).
[47] Z. Li, C. Lang, Y. Chen, J. Liew, J. Feng, Deep reasoning with multi-scale context for salient object detection, arXiv:1901.08362 (2019).
[48] S. Jia, N.D. Bruce, Richer and deeper supervision network for salient object detection, arXiv:1901.02425 (2019).
[49] Y. Liu, D.-P. Fan, G.-Y. Nie, X. Zhang, V. Petrosyan, M.-M. Cheng, Dna: deeply-supervised nonlinear aggregation for salient object detection, arXiv:1903.12476 (2019).
[50] A. Radhakrishna, S. Appu, S. Kevin, L. Aurelien, F. Pascal, S. Sabine, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[51] H. Zhang, C. Xia, X. Gao, Robust saliency detection via corner information and an energy function, IET Comput. Vis. 11 (6) (2017) 379–388.
[52] L. Ye, Z. Liu, L. Li, L. Shen, C. Bai, Y. Wang, Salient object segmentation via effective integration of saliency and objectness, IEEE Trans. Multim. 19 (8) (2017) 1742–1756.
[53] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 898–916.
[54] J. Zhang, K.A. Ehinger, H. Wei, K. Zhang, J. Yang, A novel graph-based optimization framework for salient object detection, Pattern Recognit. 64 (2017) 39–50.
[55] N. Ohtsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybernet. 9 (1) (2007) 62–66.
[56] P. Hu, W. Wang, C. Zhang, K. Lu, Detecting salient objects via color and texture compactness hypotheses, IEEE Trans. Image Process. 25 (10) (2016) 4653–4664.
[57] C. Xia, H. Zhang, X. Gao, Combining multi-layer integration algorithm with background prior and label propagation for saliency detection, J. Vis. Commun. Image Represent. 48 (2017) 110–121.
[58] H. Sajid, S.-C.S. Cheung, Universal multimode background subtraction, IEEE Trans. Image Process. 26 (7) (2017) 3249–3260.
[59] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: Computer Vision and Pattern Recognition, 2009, pp. 1597–1604.
[60] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: Computer Vision and Pattern Recognition, 2013, pp. 1155–1162.
[61] V. Movahedi, J.H. Elder, Design and perceptual validation of performance measures for salient object segmentation, in: Computer Vision and Pattern Recognition Workshops, 2010, pp. 49–56.
[62] Y. Li, X. Hou, C. Koch, J.M. Rehg, A.L. Yuille, The secrets of salient object segmentation, in: Computer Vision and Pattern Recognition, 2014, pp. 280–287.
[63] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.Y. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 353–367.
[64] A. Borji, What is a salient object? A dataset and a baseline model for salient object detection, IEEE Trans. Image Process. 24 (2) (2014) 742–756.
[65] J. Li, M.D. Levine, X. An, X. Xu, H. He, Visual saliency based on scale-space analysis in the frequency domain, IEEE Trans. Pattern Anal. Mach. Intell. 35 (4) (2012) 996–1010.
[66] P. Krahenbuhl, Saliency filters: contrast based filtering for salient region detection, in: Computer Vision and Pattern Recognition, 2012, pp. 733–740.
[67] Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, in: Computer Vision and Pattern Recognition, 2015, pp. 110–119.
[68] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, R. Mech, Minimum barrier salient object detection at 80 fps, in: International Conference on Computer Vision, 2015, pp. 1404–1412.
[69] K. Fu, C. Gong, Y.H. Gu, J. Yang, Normalized cut-based saliency detection by adaptive multi-level region merging, IEEE Trans. Image Process. 24 (12) (2015) 5671–5683.
[70] C. Gong, D. Tao, W. Liu, S.J. Maybank, M. Fang, K. Fu, J. Yang, Saliency propagation from simple to difficult, in: Computer Vision and Pattern Recognition, 2015, pp. 2531–2539.
[71] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, D. Hu, Salient region detection via integrating diffusion-based compactness and local contrast, IEEE Trans. Image Process. 24 (11) (2015) 3308–3320.
[72] T. Zhao, L. Li, X. Ding, Y. Huang, D. Zeng, Saliency detection with spaces of background-based distribution, IEEE Signal Process. Lett. 23 (5) (2016) 683–687.
[73] L. Zhou, Z. Yang, Z. Zhou, D. Hu, Salient region detection using diffusion process on a 2-layer sparse graph, IEEE Trans. Image Process. 26 (12) (2017) 5882–5894.
[74] Y. Yuan, C. Li, J. Kim, W. Cai, D.D. Feng, Reversion correction and regularized random walk ranking for saliency detection, IEEE Trans. Image Process. 27 (3) (2018) 1311–1322.
[75] N. Tong, H. Lu, R. Xiang, M.H. Yang, Salient object detection via bootstrap learning, in: Computer Vision and Pattern Recognition, 2015, pp. 1884–1892.
[76] F. Huang, J. Qi, H. Lu, R. Xiang, R. Xiang, Salient object detection via multiple instance learning, IEEE Trans. Image Process. 26 (4) (2017) 1911–1922.
[77] L. Wang, H. Lu, R. Xiang, M.H. Yang, Deep networks for saliency detection via local estimation and global search, in: Computer Vision and Pattern Recognition, 2015, pp. 3183–3192.
[78] G. Li, Y. Yu, Deep contrast learning for salient object detection, in: Computer Vision and Pattern Recognition, 2016, pp. 478–487.
[79] J. Kim, V. Pavlovic, A shape preserving approach for salient object detection using convolutional neural networks, in: International Conference on Pattern Recognition, 2016, pp. 609–614.
[80] P. Zhang, D. Wang, H. Lu, H. Wang, B. Yin, Learning uncertain convolutional features for accurate saliency detection, in: International Conference on Computer Vision, 2017, pp. 212–221.
[81] X. Zhang, T. Wang, J. Qi, H. Lu, G. Wang, Progressive attention guided recurrent network for salient object detection, in: Computer Vision and Pattern Recognition, 2018, pp. 714–722.
[82] B. Jiang, L. Zhang, H. Lu, C. Yang, M.H. Yang, Saliency detection via absorbing markov chain, in: International Conference on Computer Vision, 2013, pp. 1665–1672.



Chenxing Xia received the B.E. degree from the Wuhan Textile University, Wuhan, China, in 2014. He is currently pursuing the Ph.D. degree at Hunan University, Changsha, China. His current research interests include saliency detection, image segmentation and object recognition.

Hanling Zhang received the M.s. degree in Applied Mathematics from University of Central South, Changsha, China, in 2000, and the Ph.D. degree in Signal and Information Process from University of Northwestern Polytechnic, Xi'an, China, in 2003. He is an Associated Professor with Hunan University, Changsha, China. He made an academic visit to western university, Canada, form 2007 to 2008. He is the PI of several projects such as the Natural Science Foundation of China, Changzhou talent plan of Jiangsu and National forestry public service projects. His research interests include image and video signal processing, computer vision, big data computing. He has published over 60 journal articles and conference papers.

Keqin Li is a SUNY Distinguished Professor of computer science in the State University of New York. He is also a Distinguished Professor of Chinese National Recruitment Program of Global Experts (1000 Plan) at Hunan University, China. He was an Intellectual Ventures endowed visiting chair professor at the National Laboratory for Information Science and Technology, Tsinghua University, Beijing, China, during 2011–2014. His current research interests include parallel computing and high-performance computing, distributed computing, energy-efficient computing and communication, heterogeneous computing systems, cloud computing, big data computing, CPU-GPU hybrid and cooperative computing, multi-core computing, storage and file systems, wireless communication networks, sensor networks, peer-to-peer file sharing systems, mobile computing, service computing, Internet of things and cyber-physical systems. He has published over 520 journal articles, book chapters, and refereed conference papers, and has received several best paper awards. He is currently or has served on the editorial boards of *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Computers*, *IEEE Transactions on Cloud Computing*, *IEEE Transactions on Services Computing*, and *IEEE Transactions on Sustainable Computing*. He is an IEEE Fellow.

Xiuju Gao received M.s. degree in Computer Science and Electronic Engineering from Hunan University, Changsha, China, in 2016. She is now a lecturer in the school of electrical and information engineering, Anhui university of science and technology, Huainan, China. Her current research interests include image/video processing and computer vision.