Salient Region Detection using Diffusion Process on a 2-Layer Sparse Graph

Li Zhou, *Member, IEEE,* Zhaohui Yang, Zongtan Zhou, and Dewen Hu, *Senior Member, IEEE*

***Abstract*—Diffusion-based salient region detection has recently received intense research attention. In this paper, we present some effective improvements concerning two important aspects of diffusion-based methods: the construction of the diffusion matrix and the seed vector. First, we construct a 2-layer sparse graph, which is generated by connecting each node to its neighboring nodes and the most similar node that shares common boundaries with its neighboring nodes. Compared with the most frequently used 2-layer neighborhood graph, our graph not only effectively uses local spatial relationships, but also removes dissimilar redundant nodes. Second, we use the spatial variance of superpixel clusters to obtain the seed vector and, compared with the previously most-used boundary prior, our approach can better distinguish saliency seeds from the background seeds, especially when salient objects appear near the image boundaries. Finally, we calculate two preliminary saliency maps using the saliency and background seed vectors, and more accurate results are obtained using the manifold ranking diffusion method. Integrating these two diffusion-based saliency maps, we obtain the final saliency map. Extensive experiments in which we compare our method with 20 existing state-of-the-art methods on five benchmark datasets: ASD, DUT-OMRON, ECSSD, MSRA5K, and MSRA10K, show that the proposed method performs better in terms of various evaluation metrics.**

***Index Terms*—Salient region detection, 2-layer sparse graph, compactness, diffusion process, manifold ranking.**

I. Introduction

VISUAL attention is an important mechanism of the human visual system by which a person filters out redundant visual information and selects the most relevant information from a scene. With the popularity of digital cameras and smartphones, in addition to the increasing maturity of social networks and photo sharing on the Internet, the volume of images is explosively growing, and the work to quickly and efficiently find the most informative regions in images is becoming increasingly important. Hence, the computational visual attention has become an important research problem in recent years.

From the perspective of information processing mechanisms, existing visual attention methods can be broadly categorized as either bottom-up (stimulus driven) [1]-[10] or top-down (task driven) [11], [12]. Bottom-up methods usually exploit low-level cues, such as features, colors, and spatial distances, to construct saliency maps. In contrast, top-down methods are often based on high-level information, such as prior knowledge of the task, and usually require supervised learning. In addition to the dissociation between bottom-up and top-down methods, visual attention methods can be categorized as fixation prediction [1] or salient region detection [13]-[15]. Fixation prediction methods aim to predict points that people look at, whereas salient region detection methods aim to completely highlight entire object(s) of interest and sufficiently suppress background regions. In this work, we focus on bottom-up salient region detection.

The output of salient region detection methods can be used to solve numerous computer vision problems, such as image classification [16], [17], object detection and recognition [18], [19], image compression [20], and image segmentation [21], [22]. As a fundamental computer vision task, salient region detection has been extensively studied over the past few years and a number of algorithms have been proposed [23]-[25]; however, it remains a challenging issue, especially for images with complex patterns or rare background distractors.

Over the past several years, contrast-based methods have become the dominant approach for salient object detection. Despite that, these methods have several limitations that tend to highlight a salient object's edges instead of uniformly propagating the saliency to the interior [3] or incorrectly suppress some salient regions when the foreground objects and background are similar [7]. To mitigate these weaknesses, some studies have used deep Convolutional Neural Networks (CNNs) to extract more powerful features [26]-[28], and some other studies have introduced diffusion-based methods [29]-[36] to propagate saliency information throughout a graph. In this work, we focus on the diffusion-based method.

These diffusion-based methods have two important aspects: the construction of the diffusion matrix and seed vector. The diffusion matrix is particularly associated with the spatial relationships of nodes. Most recent diffusion-based methods [29]-[35] construct a 2-layer neighborhood graph by connecting each node to its neighboring nodes and the nodes sharing common boundaries with its neighboring nodes to exploit the spatial relationship. However, as shown in Fig. 1, if the 2-layer neighborhood nodes of salient objects are inhomogeneous or incoherent, the diffusion processes may lead to errors and incorrectly suppress the foreground region. Additionally, several recent diffusion-based methods [29], [30], [32], [35], [36] have calculated their saliency seeds based

The work of L. Zhou was supported by the National Natural Science Foundation of China under Grant 61502519. The work of D. Hu was supported in part by the National Basic Research Program of China under Grant 2013CB329401 and in part by the National Natural Science Foundation of China under Grant 61375034.

Li Zhou and Zhaohui Yang are with Naval Academy of Armament, Beijing, 100036, China. E-mail: lizhounaa@aliyun.com, zhyangnaa@sina.com

Zongtan Zhou and Dewen Hu are with the Department of Automatic Control, Collage of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan, 410073, P.R. China. E-mail: narcz@163.com, dwhu@nudt.edu.cn

Manuscript received April 23, 2016; revised

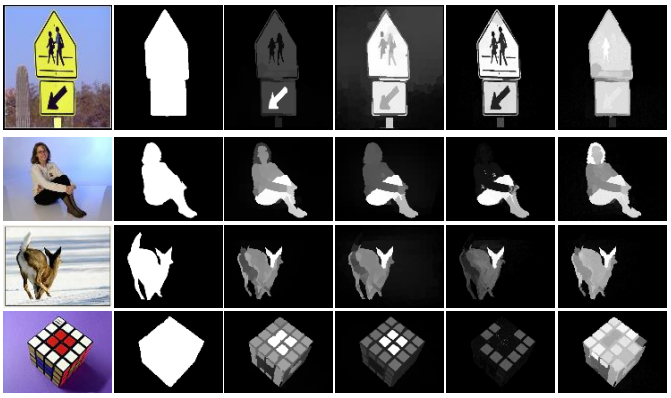


Fig. 1. Visual comparison of the 2-layer neighborhood graph and our 2-layer sparse graph. (a) Input image. (b) Ground truth salient regions. (c) Saliency maps produced by the method in [32]. (d) Saliency maps produced by the method in [29]. (e) Saliency maps produced by the method in [35]. (f) Our method.

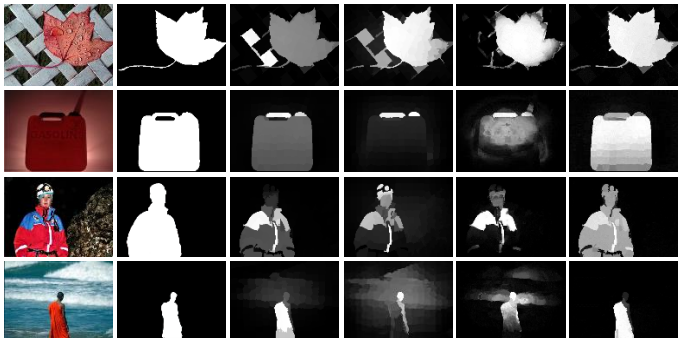


Fig. 2. Saliency maps when objects touch the image boundaries. (a) Input image. (b) Ground truth. (c) Saliency maps produced by the method in [32]. (d) Saliency maps produced by the method in [29]. (e) Saliency maps produced by the method in [10]. (f) Our method.

on the boundary prior using the nodes on each side of the image as labeled background queries. Although it is highly possible that the image border could be the background, if the salient objects touch the image boundary, as illustrated in Fig. 2, the chosen background seeds will be imprecise and could incorrectly suppress salient object regions that touch the image boundaries.

In this paper, we propose effective methods to address these problems. First, we consider that not all 2-layer neighborhood nodes are suitable for representing the spatial relationship, especially for neighborhood nodes that are inhomogeneous or incoherent to the labeled nodes. Therefore, in contrast the 2-layer neighborhood graph, we construct a sparser graph called a 2-layer sparse graph, that is generated by connecting each node to its neighboring nodes and the most similar node that shares a common boundary with its neighboring nodes. This small improvement is extremely critical because the proposed 2-layer sparse graph not only effectively use the local spatial relationship, it also removes dissimilar redundant nodes. The 2-layer sparse graph performs better than the 2-layer neighborhood graph, which can be seen in Fig. 1(f). Second, to highlight the salient objects that touch image boundaries more completely, we do not directly use the boundary prior to calculate saliency and background seeds. We first apply the K-means algorithm to cluster image superpixels into clusters via their three-dimensional Lab color features and use the compactness of superpixel clusters to obtain the saliency and background seeds. This approach is based on the observation that the clusters in the background have a larger spread than the salient clusters, that is, more compact clusters are, more salient. As shown in Fig. 2(f), our approach can better distinguish the salient regions from the background regions. Third, we propose a graph-based post-process to highlight the salient regions more completely and better suppress the background regions.

We evaluate the proposed approach using various evaluation metrics against 20 state-of-the-art methods on five image benchmarks. The experimental results show that our method outperforms these existing state-of-the-art methods.

The remainder of this paper is organized as follows. Section II contains a review of work related to graph-based salient region detection. We introduce our proposed salient region detection method in Section III. The experimental setup and results for the ASD [4], DUT-OMRON [32], ECSSD [37], MSRA5K [38], and MSRA10K [39] datasets are provided in Section IV. Finally, in Section V, we conclude the paper.

II. Related Work

Our work focuses on graph-based detection, and, in particular, diffusion-based bottom-up salient region detection. A comprehensive survey of visual attention can be found in [40], and a quantitative analysis of salient object detection methods was provided in [41].

Harel et al. [2] proposed a graph-based visual saliency method for nonlinearly combining local uniqueness maps from different feature channels to concentrate conspicuity. Liu et al. [38] trained a conditional random field to combine different features for salient object detection. Chang et al. [42] proposed the construction of a graphical model to explore the relatedness of objectness and saliency. Mai et al. [43] presented a data-driven approach to saliency aggregation that integrates saliency analysis results from multiple individual saliency analysis methods using a conditional random field.

Recently, there has been a growing interest in using diffusion processes to propagate saliency information throughout a graph. Ren et al. [44] applied a Gaussian mixture model (GMM) to cluster superpixels based on their color similarity, and calculated the saliency value for each cluster using a compactness metric together with a modified PageRank propagation. Gopalakrishnan et al. [45] formulated the object detection problem as a binary segmentation or labeling task on a graph. They identified the most salient seed and several background seeds using the behavior of random walks on a complete graph and k-regular graph. Yang et al. [32] cast saliency detection as a graph-based ranking problem. They ranked the similarity of the image superpixels with foreground cues or background cues via graph-based manifold ranking.

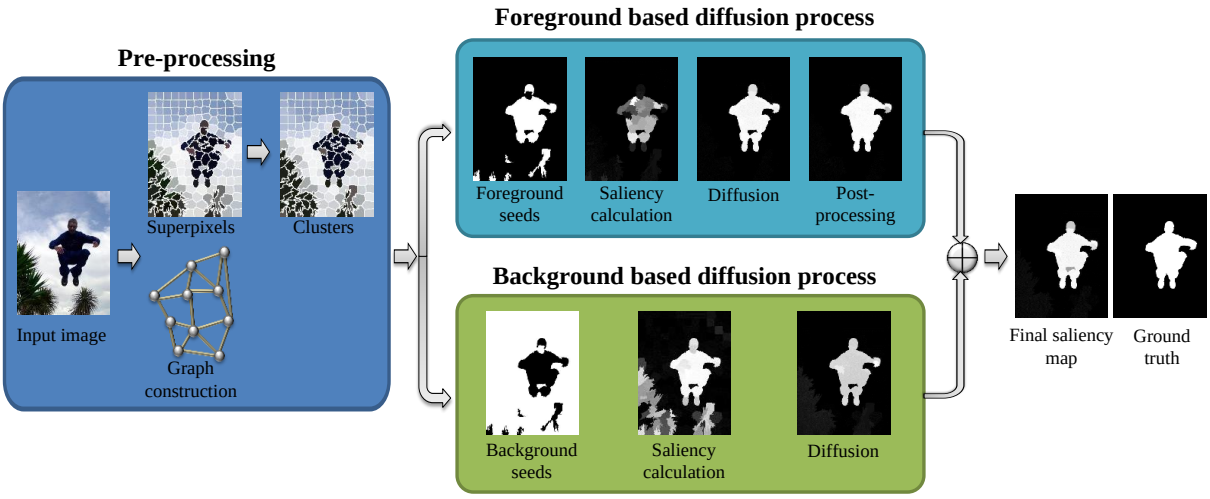


Fig. 3. Main steps of the proposed salient region detection algorithm on an example image.

Jiang et al. [29] formulated saliency detection via an absorbing Markov chain on an image graph model. They detected salient regions by determining the random-walk time of each element to the image boundary. In [33], Lu et al. proposed a method for learning the optimal seeds for object saliency using a diffusion process. Sun et al. [30] presented a salient object detection approach by exploiting the relationship between saliency detection and the Markov absorption probability. Jiang et al. [31] proposed a scheme to promote any diffusion-based salient object detection algorithm using original methods to resynthesize the diffusion matrix and construct the seed vector. Li et al. [46] proposed a regularized random-walk ranking to formulate pixel-wise saliency maps from a superpixel-based background and foreground saliency estimations. In [34], a novel propagation mechanism dependent on cellular automata was proposed to exploit the intrinsic relevance of similar regions through neighbor interactions. Li et al. [35] devised a co-transduction algorithm to fuse both boundary and objectness labels based on an inter-propagation scheme. In [36], a propagation algorithm employing teaching-to-learn and learning-to-teach strategies was proposed to improve propagation. Zhou et al. [47] proposed a salient region detection approach by integrating diffusion-based compactness and local contrast.

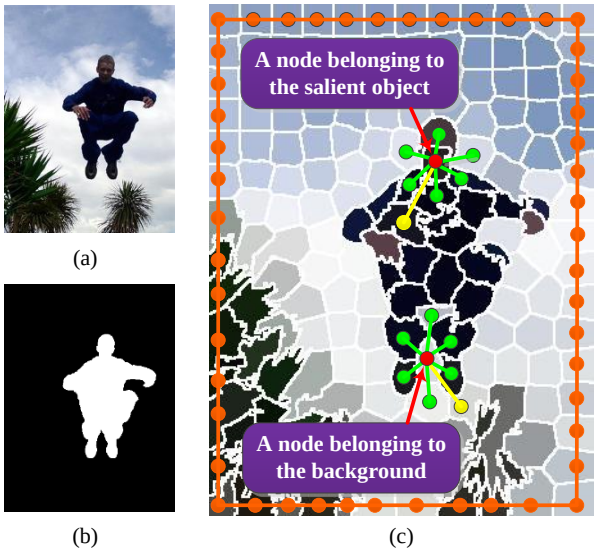


Fig. 4. The 2-layer sparse graph model. (a) Input image. (b) Ground truth. (c) Edge connection between nodes. A node (illustrated by a red dot) connects to both its adjacent nodes (green dot and connection) and the most similar node (yellow dot and connection) sharing common boundaries with its adjacent nodes. Additionally, each pair of boundary nodes are connected to each other (orange dot and connection).

III. Proposed Method

In this section, we present an efficient and effective saliency region detection method based on the diffusion process of manifold ranking on the 2-layer sparse graph, as shown in Fig. 3. We first abstract the image into superpixels and construct a graph. Next, we apply the K-means algorithm to group image superpixels into clusters and use the compactness of superpixel clusters to obtain the foreground and background seeds. We further compute two complementary saliency maps based on the foreground and background seeds, and the resulting saliency maps are propagated using a diffusion process on the constructed graph. Finally, we integrate the two computed saliency maps to generate a pixel-wise saliency map.

A. Graph Construction and Clustering

1) Abstraction: Following the observation of Perazzi et al. [7] that abstracting an input image into homogeneous superpixels can improve the performance of salient object detection, we used the SLIC model [48] to abstract the input image into uniform and compact regions.

2) Graph Construction: After abstracting the image, we construct a graph $G = (V, E)$ with N nodes $V = \{v_i \,|1 \leq i \leq N\}$, and edges $E = \{e_{ij} \,|1 \leq i, j \leq N\}$. Node v_i corresponds to the i^{th} image superpixel and edge e_{ij} link nodes v_i and v_j to each other. Most existing diffusion-based methods [29]-[35] construct a 2-layer neighborhood graph by connecting each node to neighboring nodes and nodes

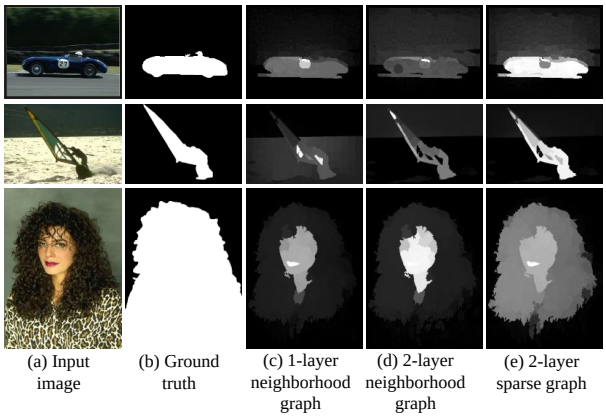


Fig. 5. Visual comparison of different graph construction methods. (a) Input image. (b) Ground truth. (c) Saliency maps produced by a 1-layer neighborhood graph. (d) Saliency maps produced by a 2-layer neighborhood graph. (e) Saliency maps produced by our 2-layer sparse graph.

that share common boundaries with the neighboring nodes. However, in this paper, a 2-layer sparse graph is proposed. As shown in Fig. 4, the proposed graph is generated by connecting each node to neighboring nodes and the most similar node sharing a common boundary with its neighboring nodes. Additionally, nodes on the four edges of the image are connected to each other to reduce the geodesic distance between similar superpixels, as proposed in [29], [30], [32], and [47]. Fig. 5 shows some motivating examples that highlight the advantage of our 2-layer sparse graph compared with the 1-layer and 2-layer neighborhood graphs. The proposed graph not only effectively uses the local spatial relationship but also removes dissimilar redundant nodes, so it performs better than 1-layer and 2-layer neighborhood graphs.

The weight w_{ij} of edge e_{ij} is defined as

$$w_{ij} = \begin{cases} e^{-\frac{\|\boldsymbol{l}_i - \boldsymbol{l}_j\|_2}{\sigma^2}} & \text{if } v_i \text{ and } v_j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $\boldsymbol{l}_i$ and $\boldsymbol{l}_j$ are the mean of the superpixels corresponding to nodes v_i and v_j, respectively, in the Lab color space, σ is a constant that controls the strength of the weight, $\boldsymbol{W} = [w_{ij}]_{N\times N}$ is the affinity matrix of graph G, and $\boldsymbol{D}$ is the degree matrix, defined as $\boldsymbol{D} = diag\{d_{11}, d_{22}, \cdots, d_{NN}\}$, where $d_{ii} = \sum_j w_{ij}$.

3) K-means Clustering: We use the K-means algorithm to cluster N image superpixels into K clusters via their three-dimensional Lab color features and let $\boldsymbol{C} = [\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_K]$ denote the K cluster centers.

B. Compactness-Based Seed Calculation

Several recent diffusion-based methods [29], [30], [32], [35], [36] have calculated their saliency and background seeds based on the boundary prior using the nodes on each side of the image as labeled background queries. However, these methods may fail when the salient objects touch the image boundary.

Salient objects generally correspond to real objects and consequently are grouped into connected regions. Therefore, salient objects typically have compact spatial distributions, whereas background regions have a wider distribution over the entire image. Motivated by this, we calculate the saliency and background seeds via the spatial variance of the K clusters.

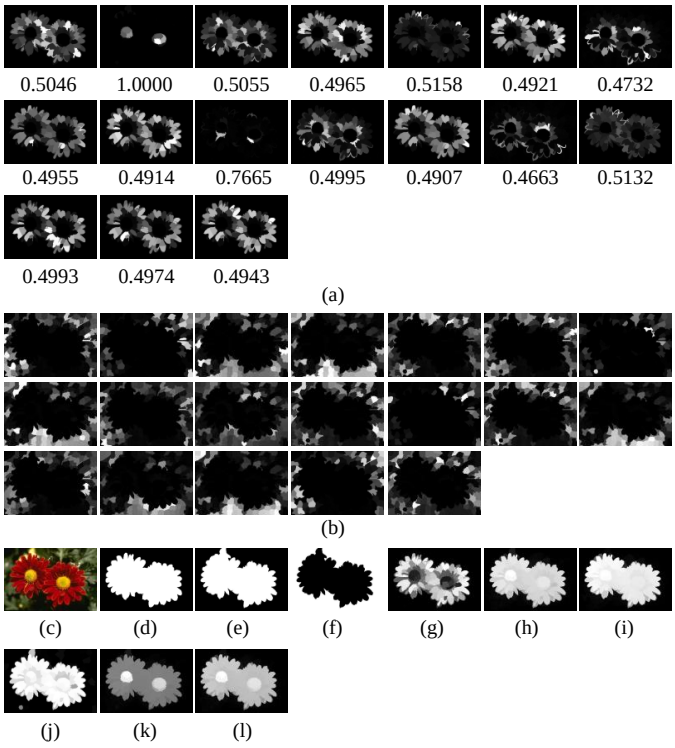


Fig. 6. Main phases of the proposed salient region detection model on an example image. (a) Visual similarity a_{ij} between nodes and saliency seeds and the probabilities that corresponding clusters are chosen as saliency seeds. (b) Visual similarity a_{ij} between nodes and background seeds. (c) Input image. (d) Ground truth. (e) Saliency seeds. (f) Background seeds. (g) $\boldsymbol{S}_{fg\_p}$ saliency map. (h) $\boldsymbol{S}_{fg\_mr}$ saliency map. (i) $\boldsymbol{S}_{fg}$ saliency map. (j) $\boldsymbol{S}_{bg\_p}$ saliency map. (k) $\boldsymbol{S}_{bg}$ saliency map. (l) Final saliency map $\boldsymbol{S}$.

We first define the similarity a_{ij} between a pair consisting of node v_i and cluster center $\boldsymbol{c}_j$ using

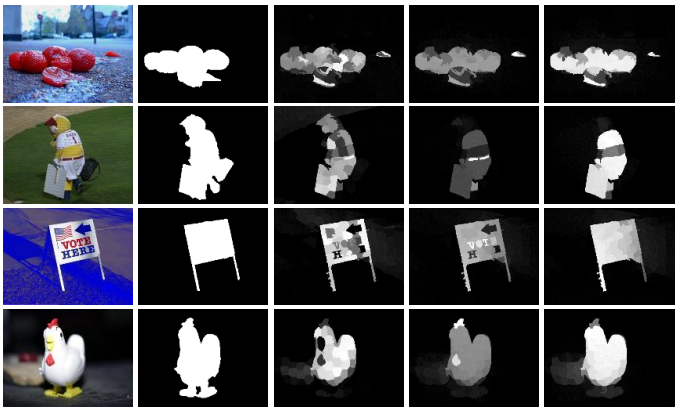
$$a_{ij} = e^{-\frac{\|\boldsymbol{l}_i - \boldsymbol{c}_j\|_2}{\sigma^2}} \quad (2)$$

To describe the similarity between nodes and clusters more precisely, we propagate the similarity using manifold ranking [32], [33] through the constructed graph; that is,

$$\boldsymbol{H} = (\boldsymbol{D} - \alpha\boldsymbol{W})^{-1}\boldsymbol{A} \quad (3)$$

where $\boldsymbol{A} = [a_{ij}]_{N\times K}$, $\boldsymbol{H} = [h_{ij}]_{N\times K}$ is the similarity matrix after the diffusion process, and α is a constant that specifies the relative contributions to the ranking scores from neighbors and the initial ranking scores.

Salient objects are generally surrounded by background regions. Thus, in the spatial domain, the clusters of background regions typically have a larger spread over the whole image when compared with salient clusters; that is, more compact clusters are more salient. We calculate the spatial variance of cluster $\boldsymbol{c}_j$ using



(a) (b) (c) (d) (e)

Fig. 7. Saliency map refined by post processing. (a) Input image. (b) Ground truth. (c) $\boldsymbol{S}_{fg\_p}$ saliency map. (d) $\boldsymbol{S}_{fg\_mr}$ saliency map. (e) $\boldsymbol{S}_{fg}$ saliency map produced by post processing.

$$\boldsymbol{sv}(j) = \frac{\sum_{i=1}^{N} h_{ij} \cdot n_i \cdot \|b_i - \mu_j\|}{\sum_{i=1}^{N} h_{ij} \cdot n_i} \quad j = 1, 2, \cdots, K \tag{4}$$

where n_i is the number of pixels that belong to superpixel v_i, $b_i = [b_i^x, b_i^y]$ is the centroid of superpixel v_i, and the spatial mean $\mu_j = \left[\mu_j^x, \mu_j^y\right]$ is defined as

$$\mu_j^x = \frac{\sum_{i=1}^{N} h_{ij} \cdot n_i \cdot b_i^x}{\sum_{i=1}^{N} h_{ij} \cdot n_i} \tag{5}$$

and

$$\mu_j^y = \frac{\sum_{i=1}^{N} h_{ij} \cdot n_i \cdot b_i^y}{\sum_{i=1}^{N} h_{ij} \cdot n_i} \tag{6}$$

The probabilities that clusters are saliency seeds can be defined as

$$\boldsymbol{p} = 1 - Norm(\boldsymbol{sv}) \tag{7}$$

where $Norm(\boldsymbol{x})$ is a function that normalizes vector $\boldsymbol{x}$ to the range between 0 and 1.

The K clusters are divided into saliency seeds and background seeds based on probability $\boldsymbol{p}$ using an adaptive threshold, which we set as the mean value of $\boldsymbol{p}$. The two cluster sets of saliency and background seeds are referred to as FG and BG, respectively. The visual similarity a_{ij} between nodes and saliency seeds ($\boldsymbol{c}_j$ in FG) is shown in Fig. 6(a), and the visual similarity a_{ij} between nodes and background seeds ($\boldsymbol{c}_j$ in BG) is shown in Fig. 6(b). The saliency seeds and background seeds are shown in Fig. 6(e) and Fig. 6(f).

C. Diffusion-Based Saliency Calculation

1) Saliency Seed-Based Diffusion Process: Motivated by the compactness cue (more compact clusters are more salient), we calculate the saliency map using the weighted similarity a_{ij} between nodes and saliency seeds, which is denoted as

$$\boldsymbol{S}_{fg\_p}(i) = \sum_{\boldsymbol{c}_j \in FG} \boldsymbol{p}(j) \cdot a_{ij} \tag{8}$$

Here, $\boldsymbol{S}_{fg\_p}$ are normalized to the range between 0 and 1. The visual similarity between a_{ij} and the corresponding probability $\boldsymbol{p}$ is shown in Fig. 6(a), and normalized saliency maps $\boldsymbol{S}_{fg\_p}$ are shown in Fig. 6(g) and Fig. 7(c). We then propagate the $\boldsymbol{S}_{fg\_p}$ saliency map using manifold ranking as follows:

$$\boldsymbol{S}_{fg\_mr} = (\boldsymbol{D} - \alpha \boldsymbol{W})^{-1} \boldsymbol{S}_{fg\_p} \tag{9}$$

where $\boldsymbol{S}_{fg\_mr}$ is an N-dimensional column vector that is normalized to the range between 0 and 1. The $\boldsymbol{S}_{fg\_mr}$ saliency maps are shown in Fig. 6(h) and Fig. 7(d). To produce a saliency map that more uniformly covers the salient objects, as shown in Fig. 7(e), we propose an efficient graph based post processing operation to enhance the quality of the saliency map, which is defined as

$$m(i,j) = w_{ij} \cdot d_{jj} \tag{10}$$

$$q(i,j) = m(i,j) \Big/ \sum_{k=1}^{N} m(i,k) \tag{11}$$

$$\boldsymbol{S}_{fg}(i) = \sum_{j=1}^{N} q(i,j) \cdot \boldsymbol{S}_{fg\_mr}(j) \tag{12}$$

where $m(i,j)$ is the weight of $\boldsymbol{S}_{fg\_mr}(j)$ when calculating the saliency value $\boldsymbol{S}_{fg}(i)$ of superpixel v_i during the post processing calculation and $q(i,j)$ is the normalized weight for $m(i,j)$. The weight is related to two factors: w_{ij} of edge e_{ij} and degree d_{jj} of super-pixel v_j. First, when saliency value $\boldsymbol{S}_{fg}(i)$ of superpixel v_i is calculated, the superpixel that is more similar to v_i has a larger weight, which is easy to understand; that is, $m(i,j)$ is proportional to w_{ij}. Second, the larger the degree d_{jj} of superpixel v_j, the more similar v_j is to the superpixel connected to it, and the more likely that v_j is a salient region according to the compactness prior of saliency regions; that is, $m(i,j)$ is proportional to d_{jj}.

After calculation, $\boldsymbol{S}_{fg}$ are normalized to the range between 0 and 1. The $\boldsymbol{S}_{fg}$ saliency maps are shown in Fig. 6(i) and Fig. 7(e). As can be observed from Fig. 7, the proposed post processing can highlight the salient regions more completely.

2) Background Seed-Based Diffusion Process: Background nodes are generally similar to some other background seeds, whereas salient nodes are often dissimilar to all the background seeds. Thus, we calculate the saliency map using the similarity a_{ij} between nodes and background seeds, which can be expressed as

$$\boldsymbol{S}_{bg\_p}(i) = \prod_{\boldsymbol{c}_j \in BG} (1 - a_{ij}) \tag{13}$$

Visual similarity a_{ij} can be seen in Fig. 6(b) and the normalized $\boldsymbol{S}_{bg\_p}$ saliency map is shown in Fig. 6(j). We then propagate the $\boldsymbol{S}_{bg\_p}$ saliency map using manifold ranking as follows:

$$\boldsymbol{S}_{bg} = (\boldsymbol{D} - \alpha \boldsymbol{W})^{-1} \boldsymbol{S}_{bg\_p} \tag{14}$$

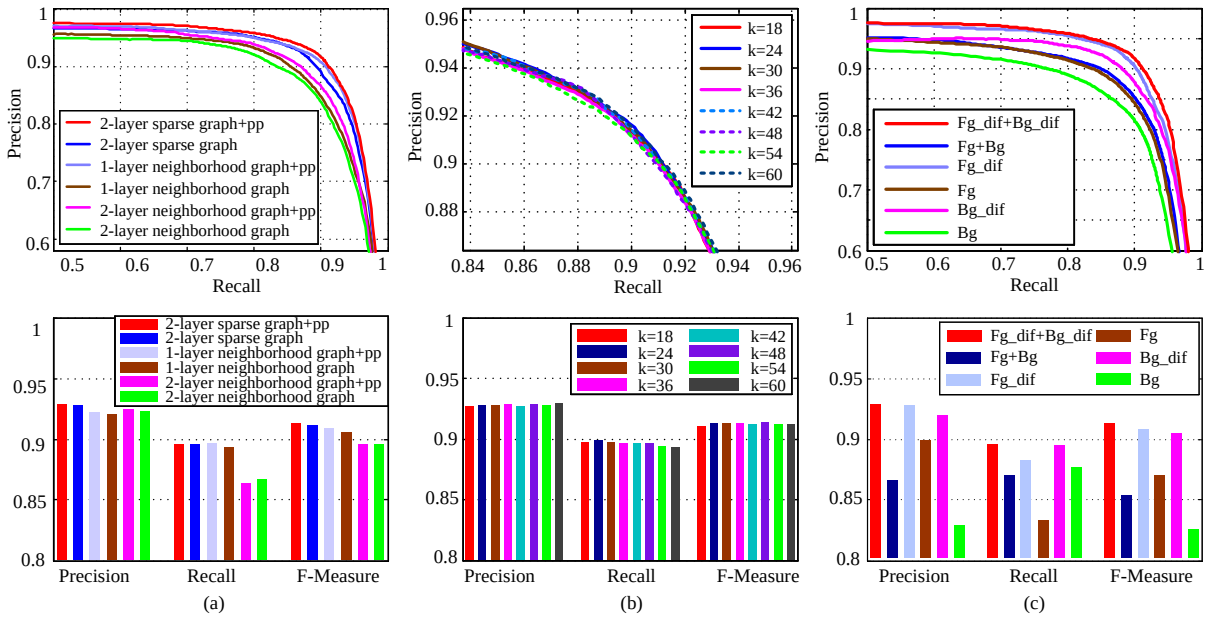


Fig. 8. Influence of various factors on saliency detection performance. Top: Precision-recall curves. Bottom: Precision, recall, and F-measure. (a) Influence of different graph construction methods and post processing (pp). (b) Influence of K. (c) Influence of different components. Fg: saliency seed-based saliency calculation without the diffusion process; Fg_dif: saliency seed-based saliency calculation with the diffusion process; Bg: background seed-based saliency calculation without the diffusion process; Bg_dif: background seed-based saliency calculation with the diffusion process.

where $\boldsymbol{S}_{bg}$ is an N-dimensional column vector normalized to the range between 0 and 1. The $\boldsymbol{S}_{bg}$ saliency map is shown in Fig. 6(k).

D. Saliency Map Integration

The saliency and background seed-based diffusion process efficiently produce two saliency maps, $\boldsymbol{S}_{fg}$ and $\boldsymbol{S}_{bg}$, respectively. These maps are complementary to each other. We directly integrate them to obtain the final saliency map,

$$\boldsymbol{S} = Norm(\boldsymbol{S}_{fg} + \boldsymbol{S}_{bg}) \tag{15}$$

The final saliency map $\boldsymbol{S}$ is shown in Fig. 6(l).

IV. Experimental Results

We conducted several experiments to evaluate the proposed saliency region detection method using five public datasets:

1) ASD. The ASD dataset [4] contains 1000 images selected from the MSRA database [38]. Each image is manually segmented into foreground and background regions. This dataset has been extensively used to test recently developed methods [7], [10], [29], [32].
2) DUT-OMRON. The DUT-OMRON dataset [32] consists of 5168 high-quality images manually selected from more than 140,000 images. Additionally, the pixel-wise ground truth annotations were produced by five subjects.
3) ECSSD. To represent natural image situations, Yan et al. [37] extended their CSSD dataset to the larger ECSSD, which contains 1000 images. It includes many semantically meaningful but structurally complex images with pixel-wise ground truth.
4) MSRA5K. The MSRA5K dataset [38] contains 5000 images from the MSRA dataset, with refined manually segmented ground-truth.
5) MSRA10K. The MSRA10K dataset [39] is the largest dataset and contains 10,000 images with pixel-level saliency labeling for 10,000 images from the MSRA dataset.

Using these five datasets, we compared our method with 20 state-of-the-art approaches, including local contrast-based approaches (IT [1], GB [2], and CSP [6]), global contrast-based approaches (FT [4], RC [5], MR [14], and CA [13]), compactness-based approach (SF [7]), background-based approaches (GS [8], RBD [24], MB [25], and MBA [25]), multiple visual cue integration approaches (UFO [15], NDE [23], and HS [37]), and diffusion-based approaches (MC [29], GBMR [32], BSCA [34], ILP [35], and IDCL [47]). For fair evaluation, we either directly used the results provided by the original authors or created our own implementations using source code publicly available online.

A. Experimental Setup and Evaluation Criteria

1) Experimental Setup: There are four parameters in the proposed method: N, the number of superpixel nodes used in the SLIC model, K, the number of clusters in the K-means algorithm, σ in Equations (1) and (2), which controls the fall-off rate of the exponential function, and α in Equations (3), (9), and (14), which balances the smooth and fitting constraints of the manifold ranking algorithm. For all five datasets, we experimentally set parameter $K = 36$, and empirically set $N = 200$, $\sigma^2 = 0.1$, and $\alpha = 0.99$, as provided in [32] and [47].

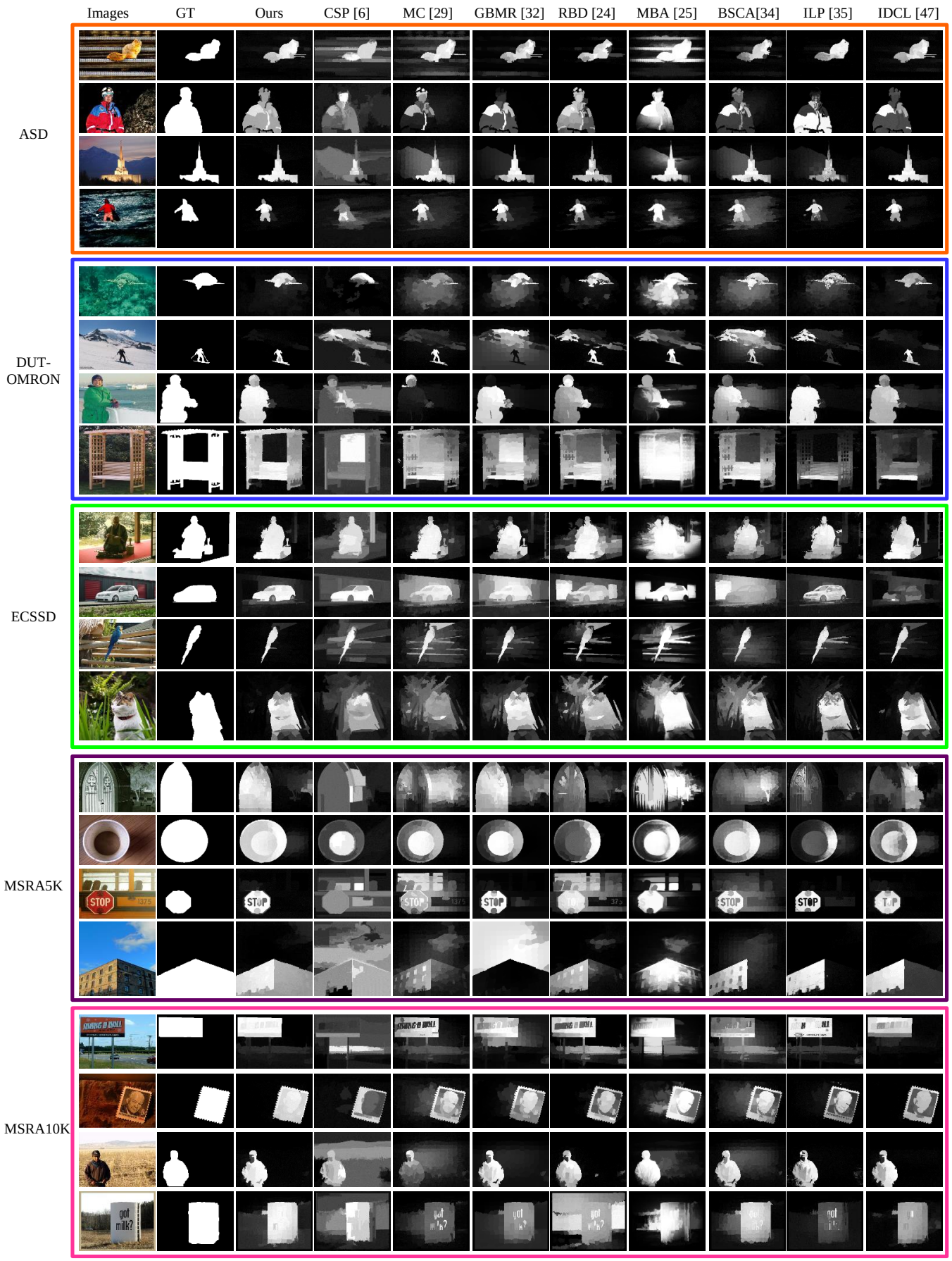


Fig. 9. Our visual image saliency maps compared with recently developed, state-of-the-art approaches on the ASD, DUT-OMRON, ECSSD, MSRA5K, and MSRA10K datasets.

TABLE I
PERFORMANCE STATISTICS OF THE ALGORITHMS ON THE FIVE CRITERIA AND FIVE DATASETS. FOR EACH DATASET AND CRITERION, THE TOP THREE RESULTS ARE HIGHLIGHTED IN **RED**, **BLUE**, AND **GREEN**, RESPECTIVELY.

Dataset	Criterion	**Our**	IDCL[47]	MBA[25]	MB[25]	BSCA[34]	ILP[35]	RBD[24]	GBMR[32]	MC[29]	CSP[6]
ASD	Precision	**0.9286**	**0.9253**	0.8732	0.8766	0.8865	**0.9254**	0.8893	0.9225	0.9176	0.8798
	Recall	**0.8961**	0.8839	0.8808	0.8010	**0.8860**	0.8372	**0.9029**	0.8641	0.8703	0.7954
	F-measure	**0.9132**	**0.9074**	0.8634	0.8441	0.8754	0.8875	0.8803	0.8954	**0.8956**	0.8361
	MAE	**0.0697**	**0.0481**	0.0763	0.0912	0.0863	0.0722	**0.0690**	0.0758	0.0938	0.1433
	MOR	**0.8432**	0.7582	0.7770	0.7166	0.8012	0.7915	**0.8146**	0.8084	**0.8121**	0.7200
DUT-OMRON	Precision	**0.5950**	**0.5720**	0.5181	0.5226	0.5162	**0.5780**	0.5280	0.5547	0.5492	0.5600
	Recall	0.6434	0.6352	**0.7507**	0.6723	**0.6728**	0.5906	**0.7141**	0.6243	0.6698	0.5281
	F-measure	**0.5631**	**0.5481**	0.5215	0.5094	0.5087	**0.5381**	0.5250	0.5287	0.5326	0.4985
	MAE	0.1763	0.1592	0.1679	**0.1566**	0.1908	**0.1446**	**0.1464**	0.1868	0.1863	0.2069
	MOR	**0.4552**	**0.4394**	0.4233	0.3946	0.4094	0.4176	**0.4292**	0.4201	0.4250	0.3743
ECSSD	Precision	**0.7498**	**0.7462**	0.6977	0.7007	0.7196	0.7270	0.7021	**0.7370**	0.7348	0.7261
	Recall	0.5930	0.5905	**0.6701**	0.5997	**0.6370**	0.5160	**0.6169**	0.5965	0.5990	0.5202
	F-measure	**0.6617**	**0.6593**	0.6558	0.6332	**0.6615**	0.6118	0.6411	0.6513	0.6556	0.5947
	MAE	0.2315	0.2318	**0.2255**	**0.2309**	0.2334	0.2373	**0.2254**	0.2353	0.2513	0.2665
	MOR	**0.5076**	0.5018	**0.5086**	0.4638	**0.5130**	0.4423	0.4944	0.4908	0.4955	0.4279
MSRA5K	Precision	**0.8649**	**0.8576**	0.8115	0.8088	0.8183	**0.8499**	0.8168	0.8461	0.8468	0.8187
	Recall	0.7853	0.7709	**0.8051**	0.7210	**0.7886**	0.7053	**0.7911**	0.7570	0.7647	0.6690
	F-measure	**0.8242**	**0.8178**	0.7923	0.7650	0.7936	0.7838	0.7912	0.8011	**0.8055**	0.7370
	MAE	**0.1181**	0.1195	**0.1116**	0.1241	0.1313	0.1211	**0.1136**	0.1288	0.1459	0.1847
	MOR	**0.7071**	**0.6932**	0.6715	0.6103	0.6767	0.6412	**0.6787**	0.6723	0.6768	0.5864
MSRA10K	Precision	**0.9075**	**0.9039**	0.8630	0.8579	0.8698	**0.8925**	0.8709	0.8914	0.8914	0.8519
	Recall	0.7843	0.7721	**0.7885**	0.7022	**0.7847**	0.7015	**0.7883**	0.7501	0.7524	0.6795
	F-measure	**0.8562**	**0.8528**	0.8296	0.7966	0.8335	0.8126	0.8331	0.8331	**0.8359**	0.7622
	MAE	**0.1143**	0.1185	**0.1074**	0.1255	0.1254	0.1241	**0.1113**	0.1258	0.1453	0.1778
	MOR	**0.7294**	**0.7183**	0.6954	0.6264	0.7066	0.6572	**0.7106**	0.6926	0.6938	0.6062

2) Evaluation Criteria: We evaluated the performance of the salient region detection methods using four popular evaluation criteria: average precision-recall curve; F-measure [4], [7], [15], [32], mean absolute error (MAE) [7], [24], [34], [35], and mean overlap rate (MOR) [15], [31].

For the average precision-recall curve, we produced a binary mask of the salient object for a given saliency map using a threshold of $T_f \in [0, 255]$. To compare the quality of the saliency maps, we varied this threshold from 0 to 255 and computed the precision and recall at each value of the threshold by comparing the binary mask and ground truth. We then plotted a precision-recall curve using the sequence of precision-recall pairs. Combining the results from all the images of each dataset, we obtained an average precision-recall curve.

We computed the F-measure using

$$F_\beta = \frac{\left(1+\beta^2\right) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \tag{16}$$

Following [4], we set $\beta^2 = 0.3$ and applied an adaptive threshold T_a to the saliency map before calculating F_β. The threshold is defined as twice the mean saliency of the image, that is,

$$T_a = \frac{2}{W \cdot H} \sum_{i=1}^{W} \sum_{j=1}^{H} S\left(i,j\right) \tag{17}$$

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

In the third evaluation criterion, we computed the MAE as

$$MAE = \frac{1}{W \cdot H} \sum_{i=1}^{W} \sum_{j=1}^{H} \left| S\left(i,j\right) - GT\left(i,j\right) \right| \tag{18}$$

where S is a continuous saliency map (normalized to the range between 0 and 1) and GT is the binary ground truth.

Finally, we defined the MOR as

$$MOR = \frac{F_g \cap G_t}{F_g \cup G_t} \tag{19}$$

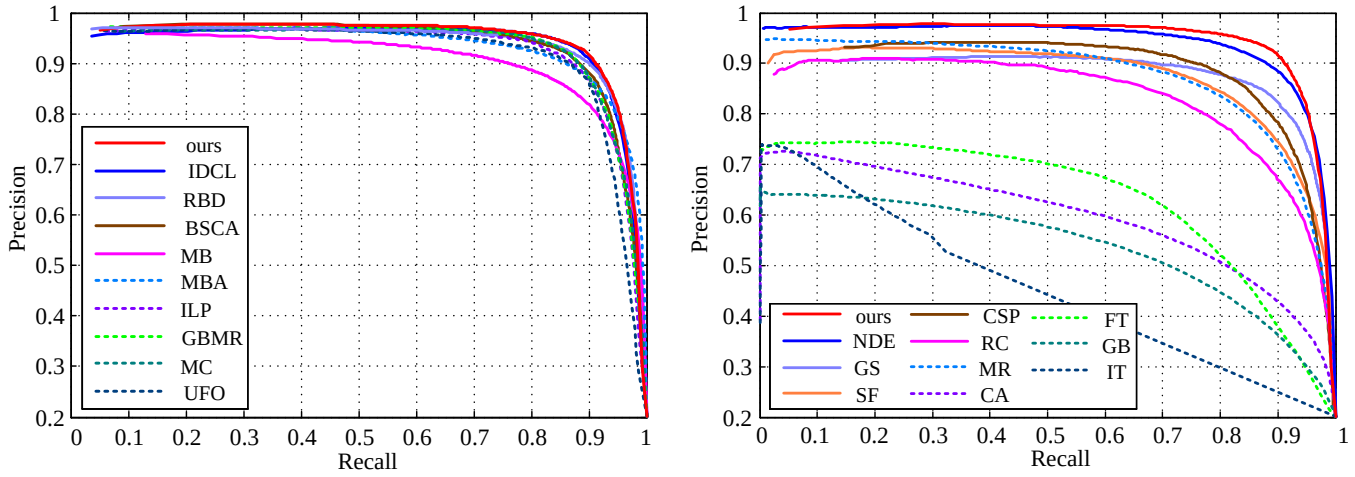


Fig. 10. Average precision-recall curves of the proposed method compared with 19 state-of-the-art methods for the ASD dataset.

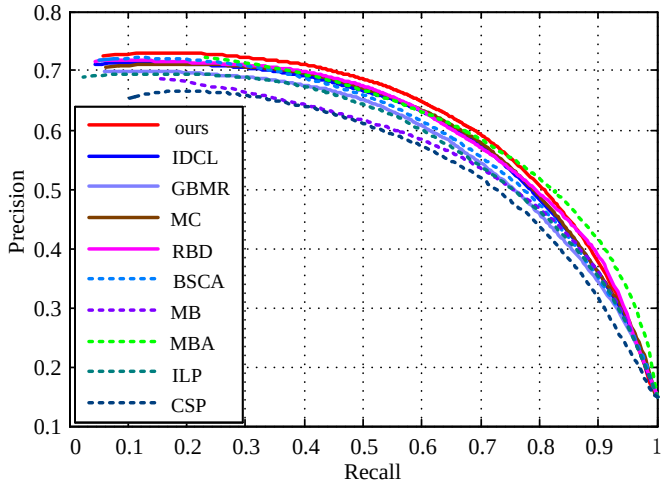


Fig. 11. Average precision-recall curves of the proposed method compared with nine state-of-the-art methods for the DUT-OMRON dataset.

where F_g and G_t are the areas of the detected salient region and marked ground truth, respectively.

B. Parameter Analysis

We conducted a series of experiments to investigate the influence of various factors on saliency detection. These factors include the node connection mode on the graph, postprocessing operation, values of K, and integration of the saliency seed-based diffusion process and background seed-based diffusion process. The experiments used the ASD dataset.

1) Graph Construction Methods and Postprocessing: Our method used diffusion processes to propagate saliency information throughout the proposed 2-layer sparse graph. To validate it, we first investigated how saliency detection performance was affected by the graph construction methods. We compared the performance using three node connection modes (1-layer neighborhood graph, 2-layer neighborhood graph, and the proposed 2-layer sparse graph) with and without postprocessing.

The precision-recall curves are plotted at the top of Fig. 8(a), the average precisions, recalls, and F-measures using an adaptive threshold are shown at the bottom of Fig. 8(a). The performance using the proposed 2-layer sparse graph is superior to that using the 1-layer neighborhood graph and 2-layer neighborhood graph. Additionally, the experimental results in Fig. 8(a) show that the proposed postprocessing operation improves performance, and Fig. 7 shows that the proposed postprocessing operation enhances the quality of the saliency map efficiently.

2) Parameters K: Parameter K is a parameter used in the K-means algorithm to control the number of clusters. However, the values of K may affect performance. Thus, we experimentally examined the influence of this parameter.

We used the K-means algorithm to group image superpixels into clusters, where K controls the number of clusters. If K is too small, many different objects will be mapped to the same cluster, which decreases the detection of salient objects. If K is too large, salient objects will be mapped to many different clusters, which may incorrectly suppress salient regions. Thus, we investigated whether saliency detection was affected by K. To determine an appropriate value for K, we varied it from 18 to 60 in intervals of six.

The precision-recall curves and the average precisions, recalls, and F-measures using an adaptive threshold are shown in Fig. 8(b) for different values of K, where the performance changes a small amount when the value of K was varied. This means that performance is not sensitive to K. We used 36 as the value for K in all subsequent experiments.

3) Component Analysis: We combined the saliency seed-based diffusion processes and background seed-based diffusion processes to obtain the final saliency map in this study. Thus, we experimentally evaluated the effectiveness of the two seed vector-based methods and the effectiveness of the diffusion processes.

The precision-recall curves are plotted at the top of Fig. 8(c) and the average precisions, recalls, and F-measures using an

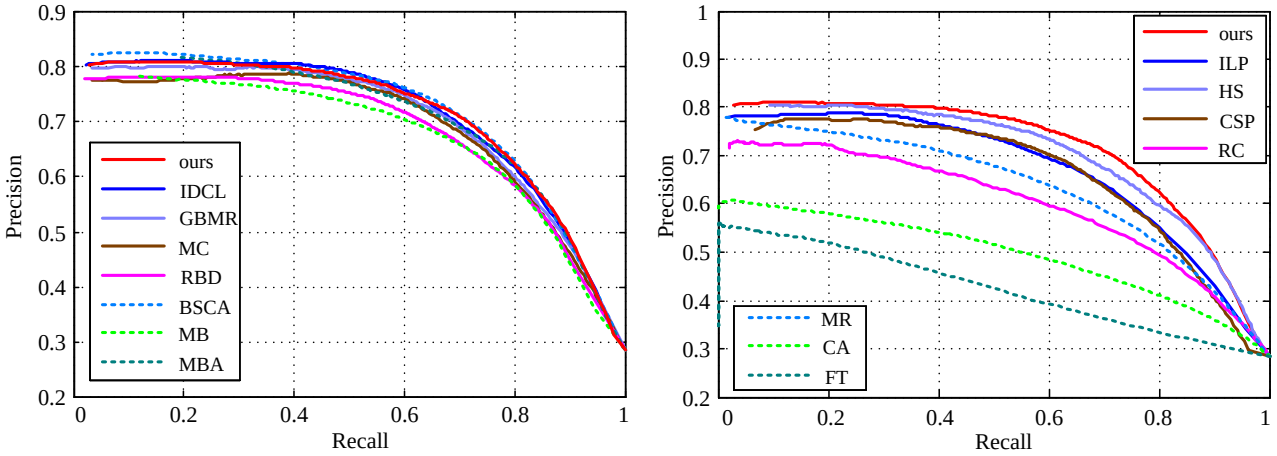


Fig. 12. Average precision-recall curves of the proposed method compared with 14 state-of-the-art methods for the ECSSD dataset.

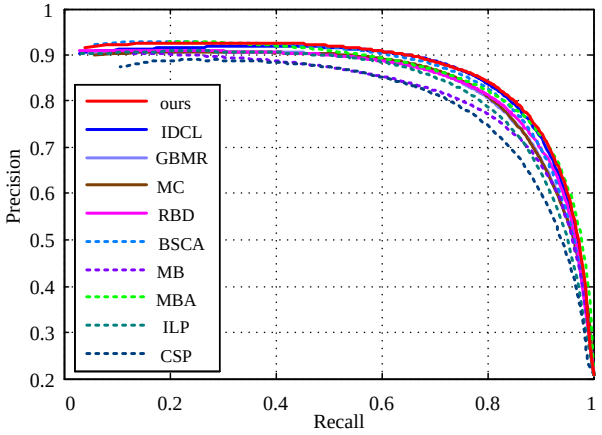


Fig. 13. Average precision-recall curves of the proposed method compared with nine state-of-the-art methods for the MSRA5K dataset.

adaptive threshold are shown at the bottom of Fig. 8(c). These results demonstrate that the saliency seeds and background seed-based saliency calculations complement each other, and the diffusion processes improve the quality of both the saliency seeds and background seed-based methods. Therefore, all the components of the saliency calculations integration and diffusion processes contribute to the final results, and help to precisely detect the salient region.

C. Visual Comparison

Several natural example images with complex backgrounds are shown in Fig. 9 for a visual comparison of our method with recently developed, state-of-the-art approaches on the ASD, DUT-OMRON, ECSSD, MSRA5K, and MSRA10K datasets.

The example images in Fig. 9 show that most salient region detection methods can effectively manage cases with relatively simple backgrounds and homogenous objects, such as the third image of the MSRA10K dataset. However, these methods fail when analyzing complicated images, including those with cluttered backgrounds, highly textured regions, and low contrast between the object and background. In contrast, our model can manage these complicated scenarios more effectively. In the following, we compare four aspects of the salient region detection methods: (1) the effectiveness of background suppression, (2) objects touching the image boundary, (3) the integrity of salient objects, and (4) the boundary preservation of salient objects.

First, as Fig. 9 shows, CSP [6] and MBA [25] cannot suppress the background regions adequately: they incorrectly highlight many background regions as salient in almost all the example images. Second, when salient objects touch the image boundary to some extent, the boundary prior based methods RBD [24], MBA [25], MC [29], and GBMR [32] may incorrectly suppress these salient object regions as the background, such as in the second image of the ASD dataset and the fourth image of the MSRA5K dataset. Third, when salient objects contain multiple heterogeneous regions in textured and cluttered backgrounds, BSCA [34], ILP [35], and IDCL [47] cannot highlight the salient objects completely and uniformly, such as in the fourth image of the DUT-OMRON dataset and second image of the MSRA10K dataset. Finally, as the first and fourth images of the DUT-OMRON dataset and first image of ECSSD dataset show, MBA [25] cannot preserve the boundary of salient objects well. However, Fig. 9 shows that our method can manage all these complicated scenarios effectively. Our saliency maps can uniformly highlight salient object regions with well-defined boundaries and sufficiently suppress background regions.

D. Quantitative Comparison

We compared the performance of our method with 20 previously published results. We applied the comparative state-of-the-art methods to the ASD, DUT-OMRON, ECSSD, MSRA5K, and MSRA10K datasets and compared the results using four evaluation criteria: the average precision-recall curve, F-measure, MAE, and MOR.

1) ASD Dataset: For the ASD dataset, we compared our method with 19 state-of-the-art approaches: the well-known IT [1], GB [2], FT [4], RC [5], CSP [6], SF [7], GS [8], CA [13], MR [14], UFO [15], MC [29], and GBMR [32]

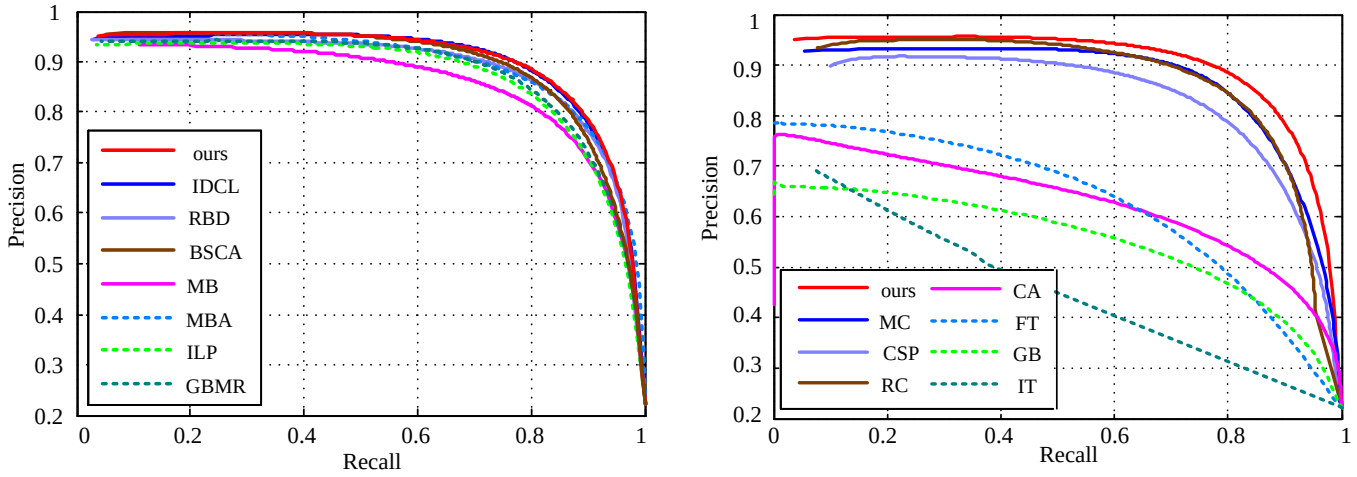


Fig. 14. Average precision-recall curves of the proposed method compared with 14 state-of-the-art methods for the MSRA10K dataset.

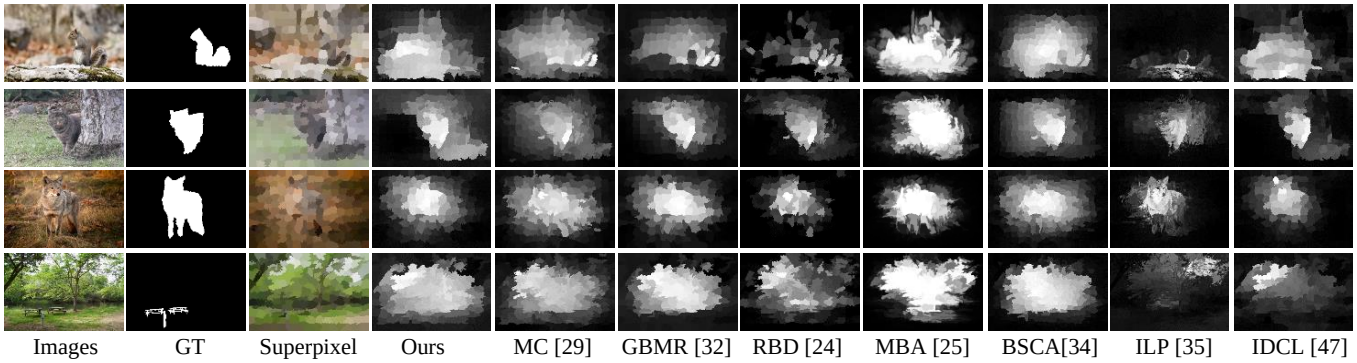


Fig. 15. Failure cases generated by our method and recently developed state-of-the-art approaches.

approaches, and seven recently developed saliency models: NDE [23], RBD [24], MB [25], MBA [25], BSCA [34], ILP [35], and IDCL [47]).

The average precision-recall curves in Fig. 10 show that the proposed method performed better than the other approaches for the ASD dataset. Note that the proposed method outperformed the MC [29], GBMR [32], BSCA [34], ILP [35], and IDCL [47] methods, which are also based on diffusion processes.

The average precision, recall, F-measure, MAE, and MOR using different approaches are shown in Tab. I. The proposed method achieved the highest precision, F-measure, and MOR values. These evaluation criteria consistently show that the proposed method outperformed all the other approaches. The MC [29], GBMR [32], and RBD [24] methods were the top-performing approaches for saliency detection in a recent benchmark study [41], and MBA [25], BSCA [34], ILP [35], and IDCL [47] are more recently developed, state-of-the-art approaches.

2) DUT-OMRON Dataset: Using the DUT-OMRON data set, we qualitatively and quantitatively compared the proposed method with nine recently developed state-of-the-art approaches: CSP [6], MC [29], GBMR [32], RBD [24], MB [25], MBA [25], BSCA [34], ILP [35], and IDCL [47].

Figure 11 compares our precision-recall curve with the these nine state-of-the-art approaches. As shown in the figure, the proposed method performed better than the other approaches for recall values from 0 to 0.75, but it performed poorly compared with the MBA [25] approach for recall values from 0.75 to 1. Table I shows that our method achieved the highest precision, F-measure, and MOR values.

3) ECSSD Dataset: Using the ECSSD dataset, we quantitatively compared the proposed method with 14 state-of-the-art approaches: FT [4], RC [5], CSP [6], CA [13], MR [14], MC [29], GBMR [32], HS [37], RBD [24], MB [25], MBA [25], BSCA [34], ILP [35], and IDCL [47].

Figure 12 compares our precision-recall curve with the 14 state-of-the-art approaches. As shown in the figure, the proposed method achieved comparable performance with the best results for recall values from 0.6 to 1, and the precision of the proposed approach was slightly lower than that of the BSCA [34] and IDCL [47] approaches for recall values from 0 to 0.6. However, as shown in Tab. I, the proposed method achieved the highest precision and F-measure.

4) MSRA5K and MSRA10K Datasets: Using the MSRA5K data set, we quantitatively compared the proposed method with nine recently developed state-of-the-art approaches (CSP [6], MC [29], GBMR [32], RBD [24], MB [25], MBA [25], BSCA

[34], ILP [35], and IDCL [47]), whereas for the MSRA10K data set, we compared our method with 14 state-of-the-art approaches (IT [1], GB [2], FT [4], RC [5], CSP [6], CA [13], MC [29], GBMR [32], RBD [24], MB [25], MBA [25], BSCA [34], ILP [35], and IDCL [47]).

The average precision-recall curves of Fig. 13 and Fig. 14 show that the proposed method achieved comparable performance with the best results reported thus far on these two data sets. Specifically, as shown in Tab. I, the precision, F-measure, and MOR values are the highest on both the MSRA5K and MSRA10K datasets.

5) Brief Summary: Table I shows the performance statistics of different algorithms on the five criteria and five datasets. For this total of 25 evaluating indicators, we achieved 14 first, one second, and four third rankings, and only six items did not enter the top three. Note that because of many false positive salient detections of MBA [25] (see Fig. 9), its approach has the highest recall value among the four datasets. However, as is generally accepted, compared with the recall value, a high F-measure value is more important in the saliency community because it is required to balance recall and precision well.

E. Failure Cases

Our method performed better than most of the state-of-the-art algorithms in terms of precision, F-measure, and MOR. However, as previously stated, the proposed method mainly depends on color information; therefore, it may fail for images that do not have much color variation, especially when foreground and background objects have similar colors. Figure 15 shows some failures generated by our method and recently developed state-of-the-art approaches. Figure 15 also shows that the estimated salient object is not salient according to its color contrast, and the abstraction processing further weakens the color contrast between the foreground and background, as shown in the superpixel maps. This limitation could be overcome by incorporating more features such as shape, texture, or even high-level knowledge. Note that, as shown in Fig. 15, the other state-of-the-art approaches also did not perform well in these cases.

V. Conclusion

In this paper, we proposed a bottom-up method for detecting salient regions in images by propagating the saliency and background seed vectors calculated via a compactness measure. After considering the limitations of the constructions of the diffusion matrix and seed vector in recent diffusion process methods, we proposed a new graph construction and seed vector generation method. Additionally, we calculated two preliminary saliency maps using the saliency and background seed vectors, and to produce a saliency map that more uniformly covered the salient objects, we propagated the saliency information using a manifold ranking diffusion process on the proposed 2-layer sparse graph. Our experimental results using five benchmark datasets demonstrated the effectiveness of the proposed method; it produced more accurate saliency maps with better precision-recall curves and higher precision, F-measure, and MOR values than 20 state-of-the-art approaches when applied to the ASD, DUT-OMRON, ECSSD, MSRA5K, and MSRA10K datasets.

References

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 20, no. 11, pp. 1254-1259, 1998.

[2] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, 2006, pp. 545-552.

[3] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2007, pp. 1-8.

[4] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2009, pp. 1597-1604.

[5] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2011, pp. 1063-6919.

[6] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li. "Automatic salient object segmentation based on context and shape prior," in *Proc. Brit. Mach. Vis. Conf*., 2011, pp. 1-12.

[7] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 733-740.

[8] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. Eur. Conf. Comput. Vis*., 2012, pp. 29-42.

[9] Y. Xie, H. Lu, and M.-H. Yang, "Bayesian saliency via low and mid level cues," *IEEE Trans. Image Process*., vol. 22, no. 5, pp.1689-1698, 2013.

[10] X.-H. Li, H.-C. Lu, L.-H. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis*., 2013, pp. 2976-2983.

[11] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "Sun: Topdown saliency using natural statistics," *Visual Cognit*., vol. 17, no. 8, pp. 979-1003, 2009.

[12] J. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 2296-2303.

[13] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2010, pp. 2376-2383.

[14] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 853-860.

[15] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by UFO: uniqueness, focusness and objectness," in *Proc. IEEE Int. Conf. Comput. Vis*., 2013, pp. 1976-1983.

[16] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 29, no. 2, pp. 300-312, Feb. 2007.

[17] G. Sharma, F. Jurie, and C. Schmid, "Discriminative spatial saliency for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 3506-3513.

[18] D. Walthera, U. Rutishausera, C. Kocha, and P. Peronaa, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," *Comput. Vis. Image Understand*, vol. 100, pp. 41-63, 2005.

[19] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, no. 11, pp. 2189-2202, 2012.

[20] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process*., vol. 19, no. 1, pp. 185-198, Jan. 2010.

[21] L. Wang, J. Xue, N. Zheng, and G. Hua, "Automatic salient object extraction with contextual cue," in *Proc. IEEE Int. Conf. Comput. Vis*., 2011, pp. 105-112.

[22] C. Jung and C. Kim, "A unified spectral-domain approach for saliency detection and its application to automatic object segmentation," *IEEE Trans. Image Process*., Vol. 21, pp. 1272-1283, 2012.

[23] L. Zhu, D.A. Klein, S. Frintrop, Z.-G. Cao, and A.B. Cremers, "A Multisize Superpixel Approach for Salient Object Detection Based on Multivariate Normal Distribution Estimation," *IEEE Trans. Image Process*., vol.23, no.12, pp. 5094-5107, Dec. 2014.

[24] X. Zhang, Z. Wang, C. Yan, H. Zou, Q. Peng, and X. Jiang, "Saliency Optimization from Robust Background Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2014, pp. 2814-2821.

[25] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Měch, "Minimum Barrier Salient Object Detection at 80 FPS," in *Proc. IEEE Int. Conf. Comput. Vis*., 2015.
[26] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 5455-5463.
[27] L. Wang, H. Lu, R. Xiang, and M.H. Yang, "Deep networks for saliency detection via local estimation and global search," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3183-3192.
[28] G. Li, and Y. Yu, "Visual Saliency Detection Based on Multiscale Deep CNN Features," IEEE Trans. Image Process., vol. 25, no. 11, pp. 5012-5024, 2016.
[29] B.-W. Jiang, L.-H. Zhang, H.-C. Lu, C. Yang, and M.-H Yang, "Saliency detection via absorb Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis*., 2013, pp. 1665-1672.
[30] J. Sun, H. Lu, and X. Liu, "Saliency Region Detection based on Markov Absorption Probabilities," *IEEE Trans. Image Process*., vol. 24, no. 5, pp. 1639-1649, May 2015.
[31] P. Jiang, N. Vasconcelos, and J. Peng, "Generic Promotion of Diffusion-Based Salient Object Detection," in *Proc. IEEE Int. Conf. Comput. Vis*., 2015, pp. 217-225.
[32] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 3166-3173.
[33] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2014, pp. 2790-2797.
[34] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via Cellular Automata," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2015, pp. 110-119.
[35] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and Inter Label Propagation: Salient Object Detection in the Wild," *IEEE Trans. Image Process*., vol. 24, no. 10, pp.3176-3186, Oct. 2015.
[36] C. Gong, D. Tao, W. Liu, S. J. Maybank, M. Fang, K. Fu and J. Yang, "Saliency propagation from simple to difficult," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2015, pp. 671-680.
[37] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 1155-1162.
[38] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 33, pp. 353-367, 2011.
[39] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach*. Intell., vol. 37, no. 3, pp. 569-582, Mar. 2015.
[40] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 35, no. 1, pp. 185-207, 2013.
[41] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient Object Detection: A Benchmark," *IEEE Trans. Image Process*., vol. 24, no. 12, pp.5706-5722, Dec. 2015.
[42] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis*., 2011, pp. 914-921.
[43] L. Mai, Y. Niu, and F. Liu, "Saliency aggregation: A data-driven approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 1131-1138.
[44] Z. Ren, Y. Hu, L.-T. Chia, and D. Rajan, "Improved saliency detection based on superpixel clustering and saliency propagation." in *Proceedings of the international conference on Multimedia*, 2010, pp. 1099-1102.
[45] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Trans. Image Process*., vol. 19, no. 12, pp. 3232-3242, Dec. 2010.
[46] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. Feng, "Robust saliency detection via regularized random walks ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2015.
[47] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, and D. Hu, "Salient region detection via integrating diffusion-based compactness and local contrast," *IEEE Trans. Image Process*., vol. 24, no. 11, pp.3308-3320, Nov. 2015.
[48] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, no. 11, pp. 2274-2282, 2012.

Li Zhou received his B.S. and M.S. degrees from Dalian Navy Academy, China, in 2004 and 2006, respectively. He got his Ph.D. from the National University of Defense Technology in 2012. From 2012, he was with the Naval Academy of Armament. His research interests include computer/biological vision, visual navigation, and machine learning.

Zongtan Zhou received his B.S., M.S., and Ph.D. degrees from the National University of Defense Technology, China, in 1990, 1994 and 1998, respectively. From February 2010 to February 2011, he was a Visiting Scholar with the Eberhard Karls Universitt Tübingen. He was promoted to Professor in 2007. His research interests include image/signal processing, computer/biological vision, neural networks, cognitive neuroscience and brain-computer interfaces.

Dewen Hu received his B.S. and M.S. degrees from Xi'an Jiaotong University, China, in 1983 and 1986, respectively. From 1986, he was with the National University of Defense Technology. From October 1995 to October 1996, he was a Visiting Scholar with the University of Sheffield, UK. He received his Ph.D. degree from the National University of Defense Technology in 1999. He was promoted to Professor in 1996. His research interests include image processing, system identification and control, neural networks, and cognitive science. He is an action editor of Neural Networks.