Multi-scale Diffusion-based Salient Object Detection with Background and Objectness Seeds

Sai Yang[1], **Fan Liu[2]**, **Juan Chen[1]**, **Dibo Xiao[1]**, **Hairong Zhu[1]**
[1] School of Electrical Engineering, Nantong University
[e-mail: yangsai166@126.com]
Nantong, Jiangsu- China
[2] College of Computer and Information, Hohai University
Nanjing, Jiangsu- China
[e-mail: fanliu.njust@foxmail.com]
*Corresponding author: Fan Liu

Abstract

The diffusion-based salient object detection methods have shown excellent detection results and more efficient computation in recent years. However, the current diffusion-based salient object detection methods still have disadvantage of detecting the object appearing at the image boundaries and different scales.To address the above mentioned issues, this paper proposes a multi-scale diffusion-based salient object detection algorithm with background and objectness seeds. In specific, the image is firstly over-segmented at several scales. Secondly, the background and objectness saliency of each superpixel is then calculated and fused in each scale. Thirdly, manifold ranking method is chosen to propagate the Bayessian fusion of background and objectness saliency to the whole image. Finally, the pixel-level saliency map is constructed by weighted summation of saliency values under different scales. We evaluate our salient object detection algorithm with other 24 state-of-the-art methods on four public benchmark datasets, i.e., ASD, SED1, SED2 and SOD. The results show that the proposed method performs favorably against 24 state-of-the-art salient object detection approaches in term of popular measures of PR curve and F-measure. And the visual comparison results also show that our method highlights the salient objects more effectively.

1. Introduction

When dealing with large amounts of vision data, humans usually pay more attention to the most valuable information for priority processing. Inspired by this visual attention mechanism used in [1], computer vision employs saliency detection to automatically find the prominent regions in a scene to reduce the computational loads in the subsequent image processing and analysis. Early efforts focus on fixation prediction [2,3] which compute a probabilistic map of a image to predict the actual human eye gaze patterns. Alternatively, salient object detection [4,5] has been effectively applied to numerous computer vision tasks such as image segmentation, object retrieval, object recognition, content-aware image resizing and so on. Salient object detection has received intensive research attention in recent years, which can be further classified as either the top-down model or the bottom-up model. The top-down learns task-driven models through training process with specific prior knowledge. While the bottom-up is known as data driven, mainly detects saliency by visual contrast without any knowledge. In this paper, we focus on the bottom-up salient object detection.

All bottom-up salient object detection methods rely on some prior about salient objects or backgrounds. The contrast can be viewed as the most widely used prior, which usually assumes salient pixels or superpixels have strong feature distinctiveness in certain image areas. However, the rarity parts of the background are wrongly detected as targets only using contrast prior inevitably. Therefore, some salient object detection methods introduce center prior to solve above problems. Since salient objects do not always appear at the image center, the center-biased models may miss the foreground regions.

Recently, diffusion-based methods have attracted growing interest from the community. Almost these approaches explore the image boundaries as background prior, which map the image into a graph with image superpixels as nodes and edge strength is proportional to superpixels similarity. Then the saliency information of background seeds is spatially propagated to the whole image on this graph. The diffusion-based salient object detection algorithm has shown favorable detection result. However, similar to center-biased model, when salient object appear at the image boundary, this kind of models will lead to error results. In addition, detecting salient object in one scale hinder them from full performance since salient objects are likely to appear at different scales. To address the above mentioned issues, we propose a multi-scale diffusion-based salient object detection algorithm with background and objectness seeds.

2. Related Work

Because saliency detection is a highly ill-posed problem, the bottom up models almost rely on assumptions about the properties of objects and backgrounds. The most widely used is the contrast prior which assume the salient object has unique characteristics with their surroundings. Itti et al. [6] define image saliency as the center-surround contrast based on difference of Gaussians. Ma and Zhang [7] estimate saliency by calculating center-surrounding color differences in a small neighborhood. Achanta et al. [8] determine the saliency of each pixel as the distance between the average vectors in its inner region with the outer region of its neighborhood. Goferman et al. [9] further implement local contrast by incorporating positional information. Different from the local ones measure saliency by

various appearance contrast in the local neighborhood, global models aim to capture the holistic rarity from an image. Zhai and Shah [10] define pixel-level saliency based on a pixel's contrast to all other pixels only with luminance information. While Cheng et al. [11] measure the global contrast values for image pixels in the Lab color space. Achanta et al. [12] directly compute global saliency using a pixel's color difference with the average image color. The above methods define the saliency in pixel-level, which is computationally expensive in dealing with spatial relationships. Therefore, Cheng et al. [11] segment the image into superpixels the same as [13], and compute the saliency for each region as the weighted sum of the region's contrast to all other regions. Unlike RC method in [11] that rely on superpixels for image abstraction, Cheng et al. [14] further propose a GMM based abstract representation to get the efficient evaluation of global cues. Margolin et al. [15] integrate the pattern distinctness of patch via PCA with the color distinctness to achieve the final saliency.

While contrast prior based methods have achieved promising results, they still severed the limitations of mistaken detecting the rarity part as the salient object. Therefore, some salient object detection methods combine contrast prior with other priors such as center prior. Jiang et al. [16] both calculate the color distinction of each patch to their spatial neighbors and the average spatial distance to the image center. Yan et al. [17] combine the global color contrast of each patch and center prior as the saliency of each layer, and use hierarchical inference to fuse cues in different scales. Tong et al. [18] employ Bayesian inference to integrate the local contrast, integrity and center prior as saliency in each scale. The final saliency map is the weighted summation of saliency maps in different scales. Zhang et al. [19] combine the average of reconstruction errors and the location prior to get the initial saliency map. But the above center-biased models are not effective when salient objects do not appear at the image center. Therefore, Xie et al. [20] apply Harris points detection to get a coarse location of the salient region. And compute a prior map based on the rough region for the Bayesian model to achieve the final saliency map. While Jiang et al. [21] randomly sample certain number of windows over the image, and measure the probability of each window containing a complete object to get objectness prior.

Recently, more and more bottom-up methods prefer to use the image boundary as the background seeds. Wei et al. [22] propose measuring saliency of an image patch by calculating the length of its shortest path to the background node. Lu et al. [23] exploit image boundaries as the likely background regions to construct template. Based on this template, they reconstruct the entire image by dense and sparse appearance models from which errors are used as indication of saliency. Lu et al. [24] proposed a salient object detection method by exploiting both weak and strong models, in which a weak saliency model is constructed by the contrast between each region and the regions along the image border. Jia et al. [25] propose a detection method via a unified generative and discriminative model, where the generative model define image saliency as the sparse coding residual based on a learned background dictionary. Tong et al. [26] construct the bottom-up saliency map by integrating contrast prior, center prior and objectness prior, in which the contrast is indicated by the difference between each unit and the regions along the image border. Zhu et al. [27] propose a robust background measure which consider the region heavily connected to the image boundary as background. Zhou et al. [28] also exploit boundary information to remove the foreground noises from the superpixels located at four borders. Wang et al. [29] take both background and foreground into consideration, and the background seeds are collected using boundary information.

In order to exploit background prior to construct saliency detection model more efficiently, there rise up some diffusion-based methods to propagate saliency information of seeds to the whole image. Sun et al. [30] calculate the preliminary saliency map using the Markov

absorption probability on a weighted graph via partial image borders as background prior. Jiang et al. [31] choose the virtual boundary nodes as the absorbing nodes in a Markov chain and the saliency is computed by the absorbed time from each transient node to boundary absorbing nodes. Zhai et al. [32] formulate the saliency propagation process as random walks with two agents on a graph simultaneously. Zhang et al. [33] utilize the four boundaries of the input image as background prior to extract foreground queries for the final saliency map. The manifold ranking algorithm in [33] outperforms most of the state-of-the-art saliency detection methods and is more computationally efficient. Therefore, we improve this algorithm by propagating the saliency to the whole image with background and objectness seeds.Otherwise, we also extend the manifold ranking based saliency detection algorithm to multi-scale.
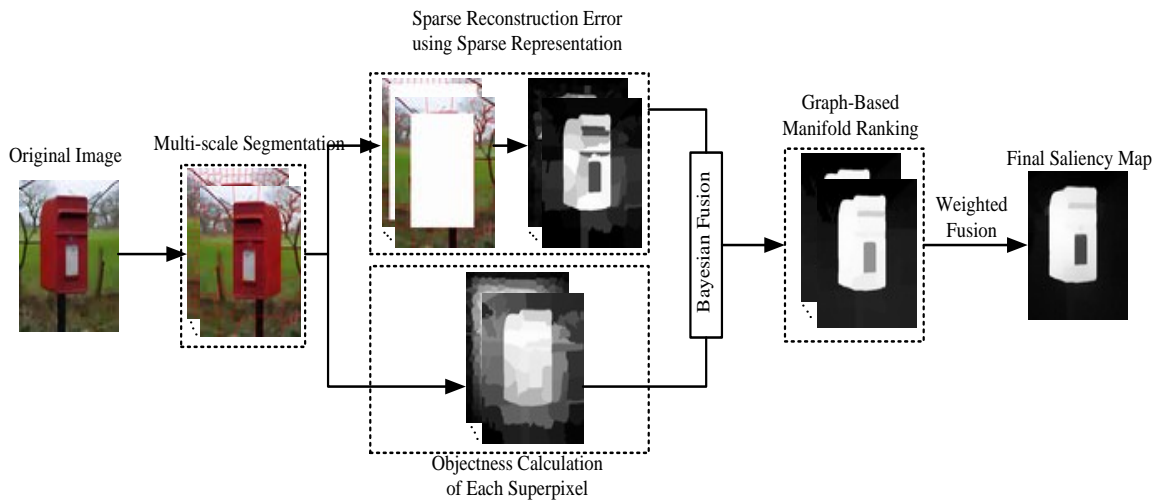

Fig. 1. Diagram of proposed salient object detection algorithm

3. Our Method

The diagram of our proposed salient object detection algorithm is shown in **Fig. 1**, which includes the following three main steps:(1)the background and objectness saliency of each superpixel is calculated and fused in each scale.(2)the saliency information of background and objectness seeds is propagated to the whole image using manifold ranking method in each scale. (3) the final saliency map is constructed by weighted summation of saliency values under different scales.The detailed description of each step is as follows.

3.1 Calculation of background and object seeds

For a given image, it is firstly over-segmented into superpixels at N scales using the simple linear iterative clustering(SLIC) algorithm. At any scale n, an image is segmented into superpixels $\{r_i^{(n)}\}, i = 1,2,...,R_n$, where R_n is the number of regions under scale n. Both Lab and RGB color spaces are used to extract the mean color features and coordinates of pixels to describe each superpixel. Each superpixel is then represented as $f_i^{(n)} \in R^{D\times1}$, where D is the feature dimension. We assume the superpixels locating at image boundary as the background prior. If the number of background superpixels is M, then the jth background superpixel is

denoted as $f_{Bj}^{(n)}$. We construct a background template set $\boldsymbol{B} = [f_{B1}^{(n)}, f_{B2}^{(n)}, ..., f_{BM}^{(n)}] \in R^{D \times M}$ as sparse representation, and encode the ith superpixel by

$$\alpha_i = \arg\min_{\alpha \in R^M} \left\| f_i^{(n)} - \boldsymbol{B}\alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1 \tag{1}$$

and the sparse reconstruction error is:

$$S_B(r_i^n) = \left\| f_i^{(n)} - \boldsymbol{B}\alpha_i \right\|^2 \tag{2}$$

Objectness state the probability of there being a complete object in a local window centered on each pixel. In order to compute the objectness of each pixel, we take any pixel in a given image as the center, and randomly sample W windows over the image. Each window w is assigned an objectness value $P(w)$ calculated by [34]. Thereafter, we overlap all the windows to obtain objectness $O(p)$ for each pixel p by

$$O(p) = \sum_{w \in W \cap p \in w} P(w_p) \tag{3}$$

where w denotes any window in W that contains pixel p. We set $W=10000$ in our experiment. For every superpixel $r_i^{(n)}$, we compute its region-level objectness by

$$S_o(r_i^n) = \frac{1}{p_i} \sum_{p \in r_i^n} O(p) \tag{4}$$

where p_i denotes the number of pixels in the ith superpixel. We use Bayesian inference to integrate the background and the objectness saliency measures by

$$S_{BI}(r_i^n) = \frac{S_o(r_i^n) S_B(r_i^n)}{S_o(r_i^n) S_B(r_i^n) + (1 - S_o(r_i^n))(1 - S_B(r_i^n))} \tag{5}$$

3.2 Saliency ranking with fused background and objectness seeds

In each single scale, we use the most popular way to construct a graph $G(V, E)$ in which each superpixel is the node and connected with its K-nearest neighbors just like the presentation in [35, 36]. The links between pairs of nodes are the edges E. In this work, the weight w_{ij} of the edge e_{ij} between adjacent nodes is defined as

$$w_{ij} = \begin{cases} e^{-\{\frac{\left\| c_i^{(n)} - c_j^{(n)} \right\|_2^2}{\sigma^2}\}}, & \text{if } l_j^{(n)} \in N(l_i^{(n)}) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $c_i^{(n)}$ and $c_j^{(n)}$ are the mean CIELAB color values of two superpixels, $l_i^{(n)}$ and $l_j^{(n)}$ are the mean coordinate values of two superpixels, $N(\cdot)$ is a neighborhood function, and σ^2 is a constant that controls the strength of the weight. The degree matrix $D = \text{diag}(d_1, ..., d_n)$ is then generated as

$$d_i = \sum_j w_{ij} \tag{7}$$

In order to manipulate manifold ranking to propagate the saliency to the whole image, we set the threshold as the mean saliency of $S_{BI}(r_i^n)$ to binary segment the saliency map. Accordingly, the binary query $S_q(r_i^n) \in \{0,1\}$ is obtained. Then the saliency of each node is computed as rank $S_r(r_i^n)$ in the graph with respect to a query S_q. The optimal ranking

minimizes the energy of the form

$$E(S_r) = \sum_{i,j} w_{ij} \left(\frac{S_r(r_i^n)}{\sqrt{d_{ii}}} - \frac{S_r(r_j^n)}{\sqrt{d_{jj}}} \right)^2 + \mu \sum_i (S_r(r_i^n) - S_q(r_i^n))^2 \right) \tag{8}$$

where μ is a controlling parameter. The optimal solution is given as:

$$S_r = (D - \alpha W)^{-1} S_q \tag{9}$$

where $\alpha = 1/(1+\mu)$.

3.3 Multi-Scale fusion

we propagate saliency values in different scales to pixels. The pixel-level saliency map is constructed by weighted summation of saliency values. The weights are measured using the Euclidean distance between the CIELab color of a pixel and the average of pixels within the superpixel. We define the overall saliency map S_p by,

$$S_p = \frac{\sum_{n=1}^{N} \sum_{i=1}^{R_n} S_r(r_i^{(n)}) (\|I_p - l_i^{(n)}\|_2 + \varepsilon)^{-1} \delta(p \in r_i^{(n)})}{\sum_{i=1}^{R_n} S_r(r_i^{(n)}) (\|I_p - l_i^{(n)}\|_2 + \varepsilon)^{-1} \delta(p \in r_i^{(n)})} \tag{10}$$

where i is the index of region, n is the index of superpixel scale, ε is a small constant, $l_i^{(n)}$ is the color center of region $r_i^{(n)}$, $\|\cdot\|$ is distance, $\delta(\cdot)$ is the indicator function.

The main procedure of the proposed method is summarized in Algorithm1.

Algorithm1 Multi-scale diffusion-based salient object detection with background and objectness seeds

---
Input: An image and required parameters

1.segment input image into superpixels at multi-scale

2.use the boundary superpixels as background template to compute saliency of each superpixel by Eq.(2) ;computer objectness of each superpixels by Eq.(4).

3.integrate the background and the objectness saliency measures by Eq.(5).

4.obtain the queries from fused saliency values, and compute saliency map by Eq.(9)

5.step 1-4 are repeated in other scales, and the saliency values of different scales is fused to get the final pixel-level saliency map using Eq.(10)

Output: the full resolution saliency map

4.Experiments

4.1 Database and evaluation metrics

Our experiments are conducted on four public datasets. The ASD dataset is a subset of MSRA database, and this dataset provides accurate human-labeled ground truth for each image. The second one is the SED dataset, which is further categorized as one single object sub-dataset SED1 and two objects sub-dataset SED2. Each sub-dataset contains 100 images and has accurate pixel-wise ground truth provided by three subjects. We also evaluate the proposed method on the more challenging SOD dataset, which contains 300 images taken from the well known Berkeley segmentation dataset. This dataset is more challenging than above three datasets as images usually contain multiple objects of different sizes and positions under more complex backgrounds. Seven subjects are asked to label the foreground salient object masks. For each object mask of the each subject, a consistency score is computed based on the labels

of the other six subjects. We select and combine the object masks whose consistency scores are higher than 0.7 as the final ground truth.

The first evaluation metric is PR curve. Each saliency map is segmented with a given threshold $T \in [0,255]$ to get a binary map, which is then compared with the ground truth mask to compute the precision and recall by

$$\text{Precision} = \frac{\sum\limits_{(x,y)} GT(x,y)BS(x,y)}{\sum\limits_{(x,y)} BS(x,y)} \tag{11}$$

$$\text{Recall} = \frac{\sum\limits_{(x,y)} GT(x,y)BS(x,y)}{\sum\limits_{(x,y)} GT(x,y)} \tag{12}$$

where *GT* and *BS* denote the ground truth and binary map respectively. Vary the threshold *T* from 0 to 255 to generate a sequence of precision-recall values for each image. Then the PR curve is plotted by taking average precision as horizontal coordinates and average recall as vertical coordinates. In the second evaluation, we use threshold t_α to binarize the saliency map to compute F-measure by

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

we set $\beta^2 = 0.3$ to emphasis precision more important than recall, the same as [11,17,21].The threshold t_α is set as proportional to the mean saliency of the image and computed by

$$t_\alpha = \frac{K}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x,y) \tag{15}$$

in which the empirical value of *K* is 2. In order to comprehensively report the F-measure characteristics, we uniformly sample a set of *K* in[0.1 6]with an interval 0.1 to compute the average F-measure for each K, and then plot the average F_β-K curve correspondingly.

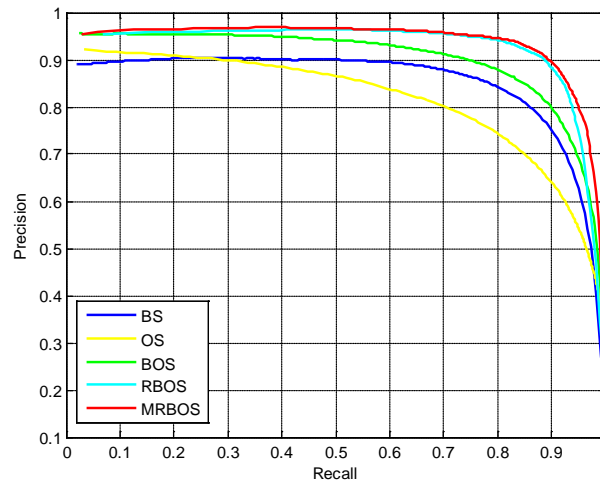## 4.2 Evaluation of each component in our method

Fig. 2. The PR curves of each component in our method

The seven scales in our proposed method is obtained by setting the superpixel numbers and compactness of SLIC as{50,100,150,200,250,300,350} and {10,10,20,20,25,30,30}. To clearly verify the effectiveness of each component in our method, we evaluate each individual component of the proposed algorithm on the ASD database in term of PR curve. Therefore, we quantitatively compare the background saliency map in one scale(BS), the object saliency map in one scale(OS), the Bayesian fusion of background saliency map and object saliency map(BOS), the ranking result with fused background and objectness seeds (RBOS), and the multi-scale fusion of the ranking saliency maps(MRBOS). The PR curve of each component is shown in **Fig. 2**.The result demonstrate that the BOS significantly outperforms BS and OS, which indicates background and objectness priors used in our method complement each other perfectly in detecting the saliency. We can also observe that RBOS achieves superior performance than BOS, which shows that the manifold ranking algorithm further refines the ahead result. We can also see that MRBOS consistently performs better than RBOS, which proves that the fusion of multi-scale detection contribute to the final result. All the above confirm that each component complements each other that leads to the best performance.

4.3Comparisons with state-of-art methods

We compare our method with state-of-art salient object detection algorithms, which is further classified into the following two categories. One class is the methods only use contrast priors, including IT [6], MZ [7], AC [8], CA [9], LC [10], HC [11], FT [12], RC [11], GC [14], PD [15], the other methods use other priors, including CBS[16], HS[17], MSS [18], CHB [20], UFO [21], GS [22], DSR [23], BL [24], GL [26], RBD [27], BFS [29], MAC [30], MAP [31], GBMR [33].We use the implementation provided by authors to get the saliency maps.
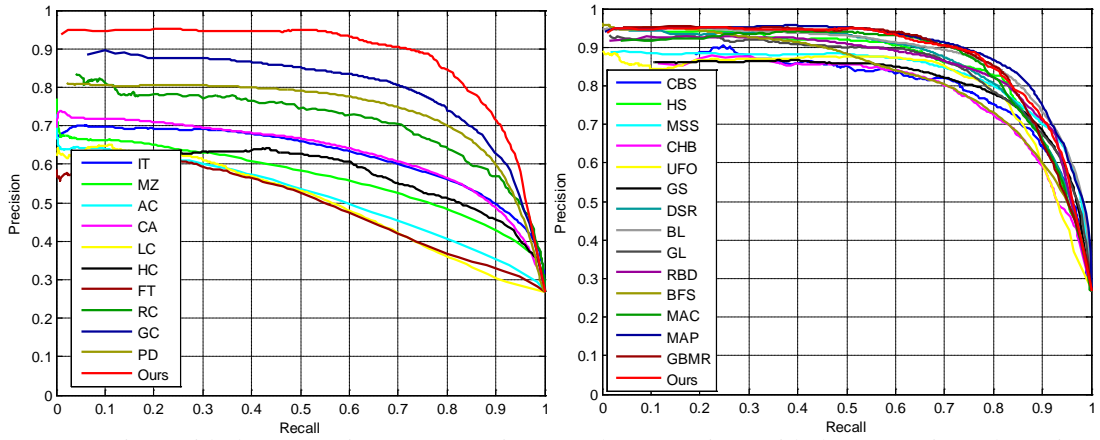
4.3.1 Quantitative Comparisons
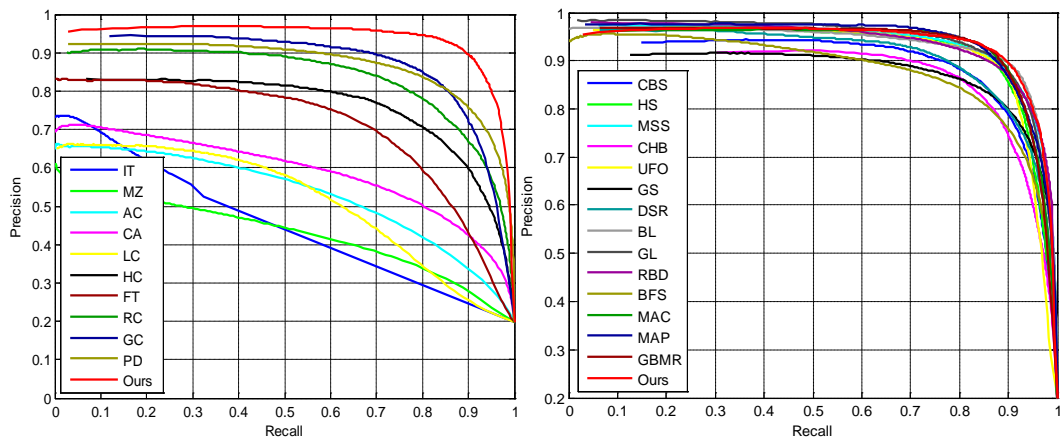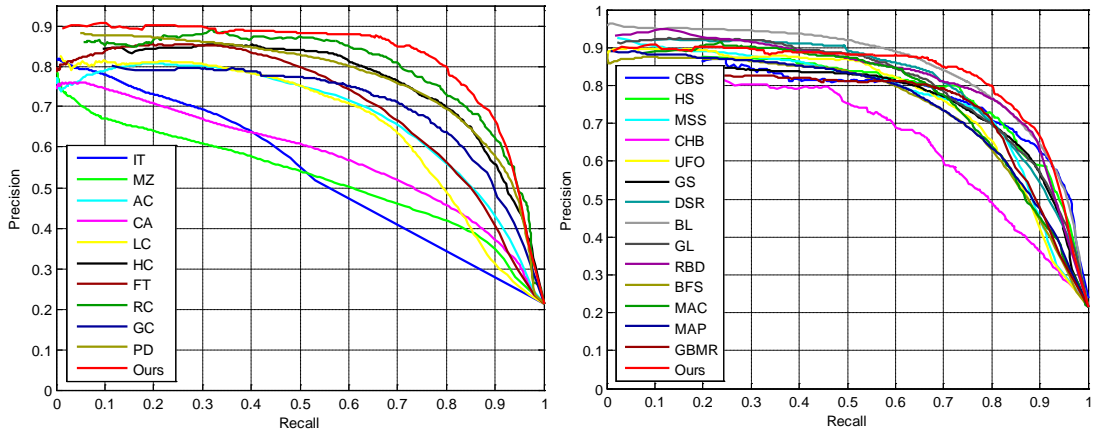

(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors
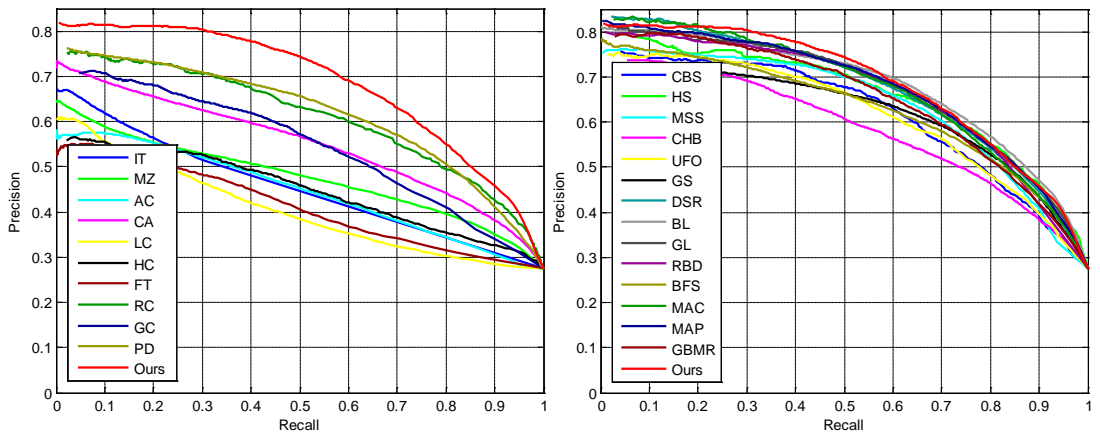
Fig. 3. Evaluation results on ASD dataset

(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors

Fig. 4. Evaluation results on SED1 dataset

(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors

**Fig. 5.** Evaluation results on SED2 dataset

(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors

Fig. 6. Evaluation results on SOD dataset

The PR curves of all salient object detection algorithms are presented in **Fig. 3**, **Fig. 4**, **Fig. 5** and **Fig. 6**. We can see that when the threshold T is equal to 0, the recall rate reaches the

maximum value, all methods have the same precision. The values are 0.1985, 0.2674, 0.2137, 0.2748 on ASD, SED1, SED2 and SOD respectively, which indicate each dataset has 19.85%, 26.74%, 21.37% , 27.48% of the pixels belong to the salient region. When the threshold T is equal to 255, the recall rate reaches the minimal value, the precision of our method has reached to 0.9549, 0.9215, 0.8917, 0.8467 on ADS, SED1, SED2, and SOD respectively. From **Fig. 3(a)**, **Fig. 4(a)**, **Fig. 5(a)**, and **Fig. 6(a)**, we see that our method significantly outperforms all the methods of the first class in almost the entire recall range on the four datasets. As observed from **Fig. 3(b)**, **Fig. 4(b)**, and **Fig. 6(b)**, our method achieves consistent and favorable performance against all the methods of the second class on ASD, SED1 and SOD. When the recall rate is above 0.85, our method can maintain the precision above 0.9 on ASD. When the recall rate is above 0.6, our method can maintain the precision above 0.9 on SED1. On SED2 dataset, the PR curve of our method is a little bit lower than the other methods but it still can maintain the recall rate at above 0.7 when the precision rate goes beyond 0.9.The performance improvement becomes obvious on the more complex SOD dataset especially when the recall rate locates at the range of [0.2,0.6]. The above analysis indicates our method can detect the salient region with a higher accuracy, and has a better detection performance.

(a)ASD dataset

(b)SED1 dataset

(c)SED2 dataset

(d)SOD dataset
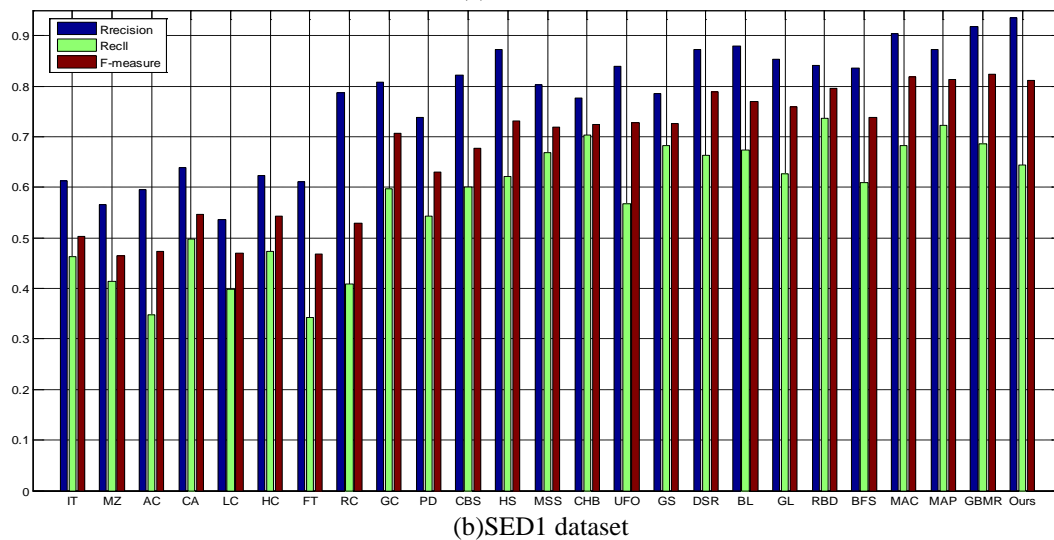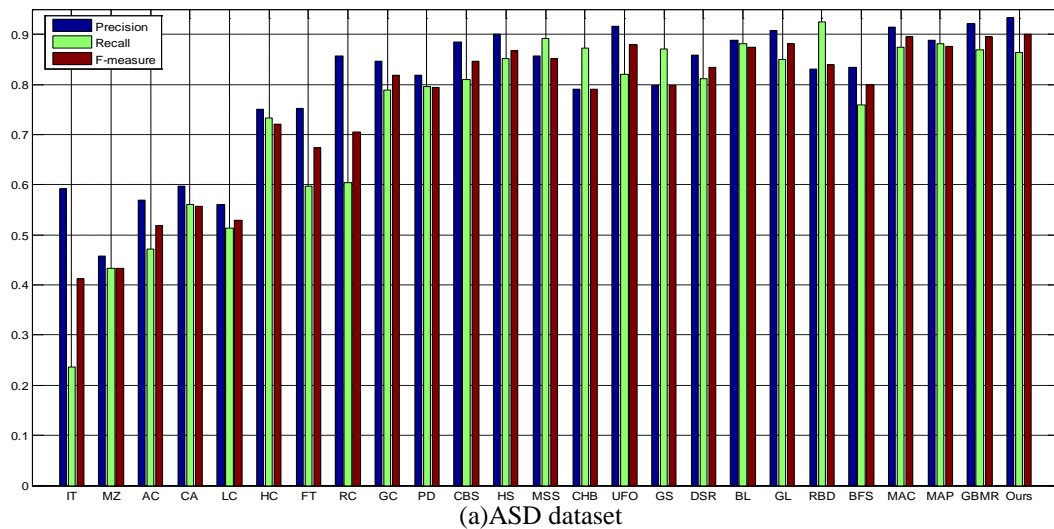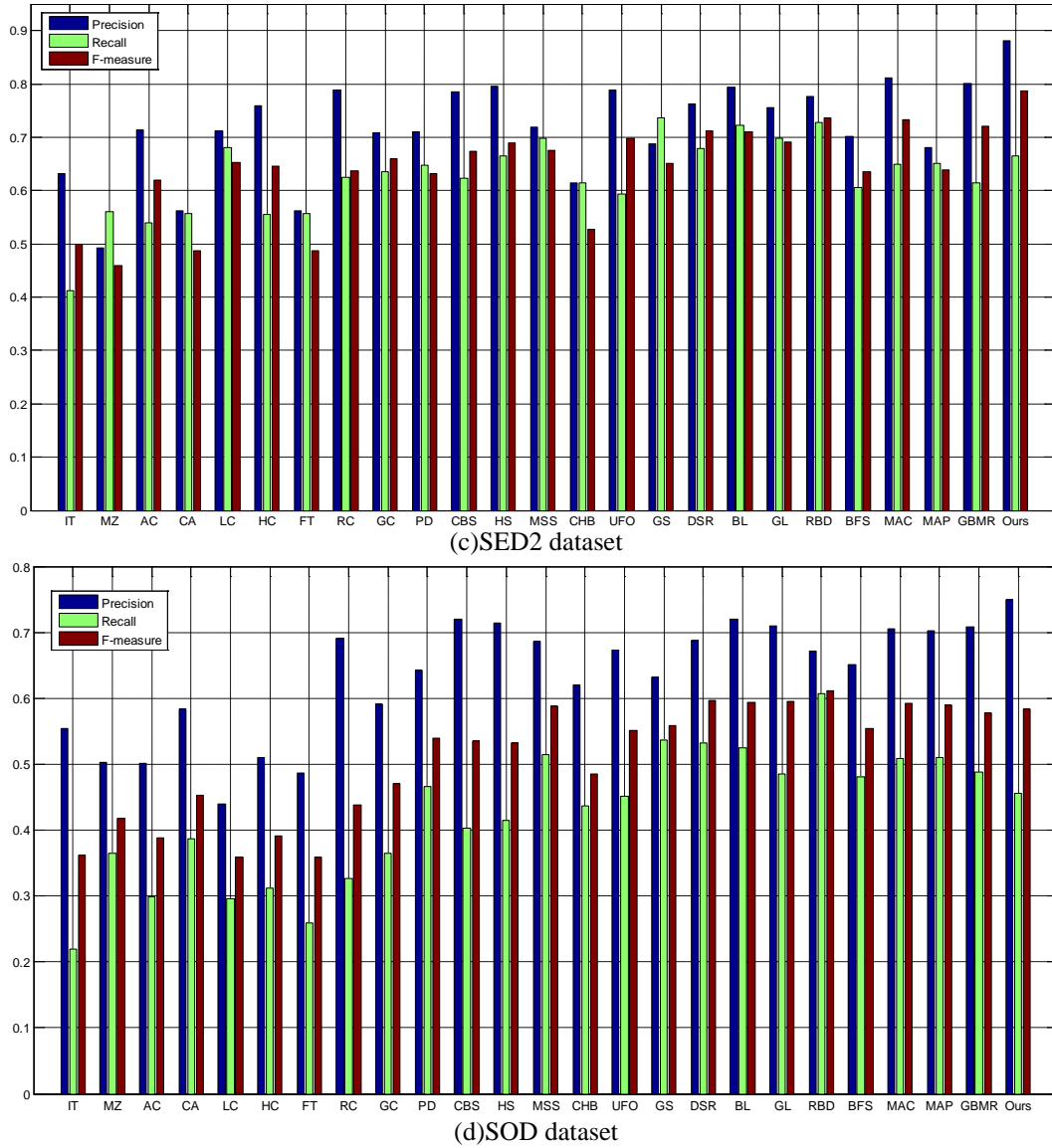
Fig. 7. Precision, recall and F-measure for adaptive thresholds

The bar graphs of average precision, recall and F-measure values using adaptive threshold are shown in **Fig. 7**. As observed from **Fig. 7(a)(b)(c)(d)**, we see that our method has the highest precision among all the methods on ASD, SED1, SED2, and SOD. Our method has the highest F-measure among all the methods on ASD and SED2, but our method only has the third and second F-measure among all the methods on SED1 and SOD. Otherwise, compared with most related work GBMR, our method increases the precision from 0.9207 to 0.9332, and increases F-measure from 0.8957 to 0.9004 on ASD. Meanwhile, our method achieves an increase of 1.83% in term of precision on SED1. Our method increases the precision from 0.8011 to 0.8813, and increases F-measure from 0.7209 to 0.7862 on SED2. Our method increases the precision from 0.7077 to 0.7497, and increases F-measure from 0.5774 to 0.5841 on SOD. The above analysis indicates that our method indeed improves GBMR by using background and objectness seeds and multi-scale detection. The F_{β}- K curves of all salient

object detection algorithms are shown in **Fig. 8**, **Fig. 9**, **Fig. 10**, and **Fig. 11**. The results show that our method has the top F_β values at most selections of the *K* value.
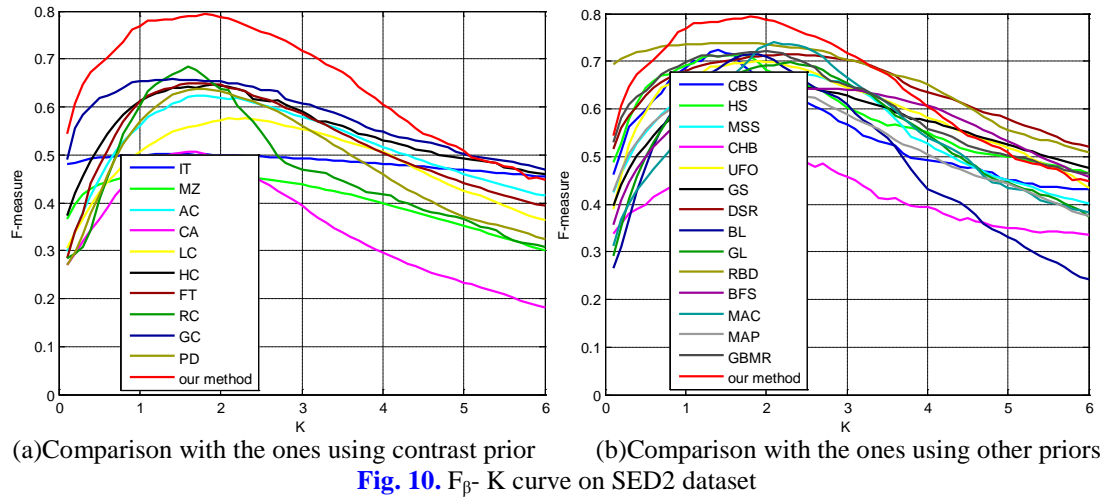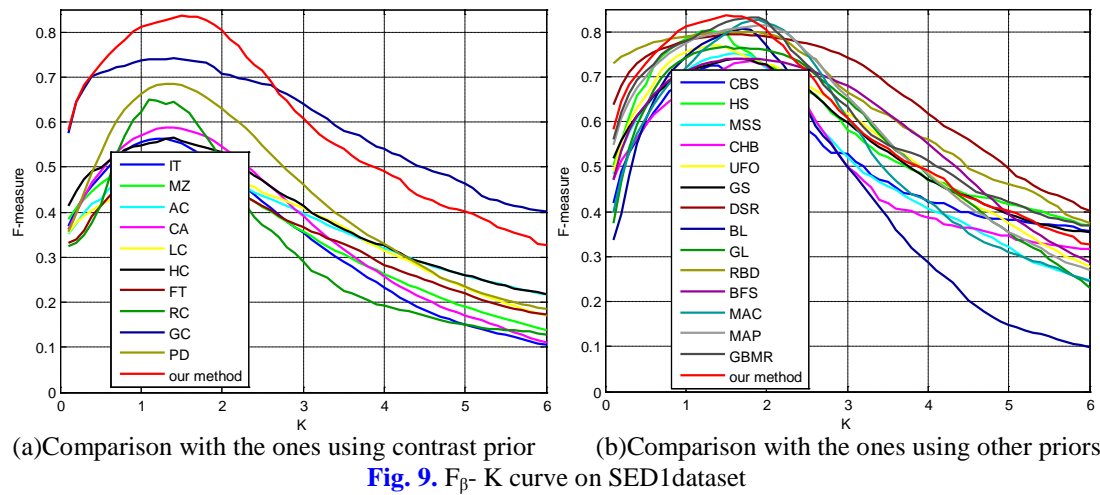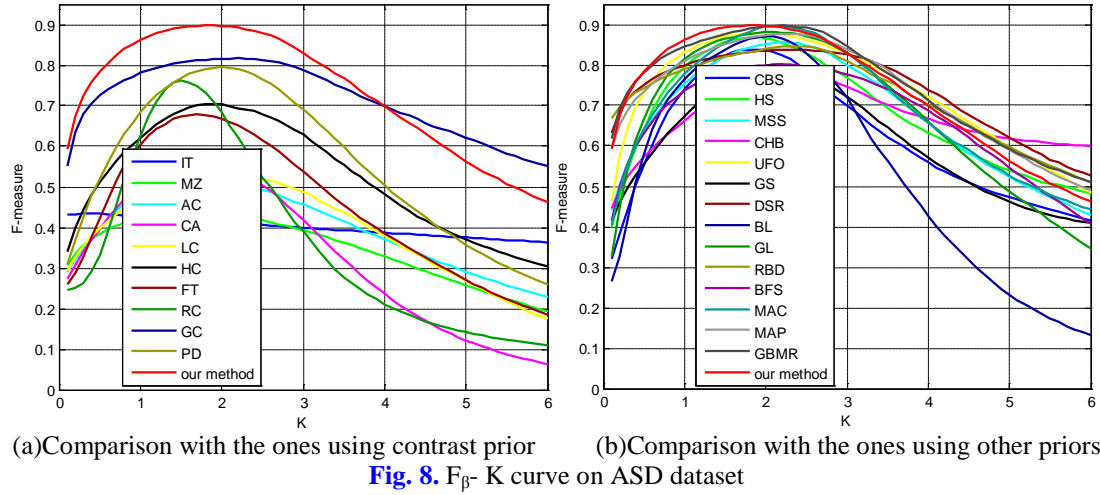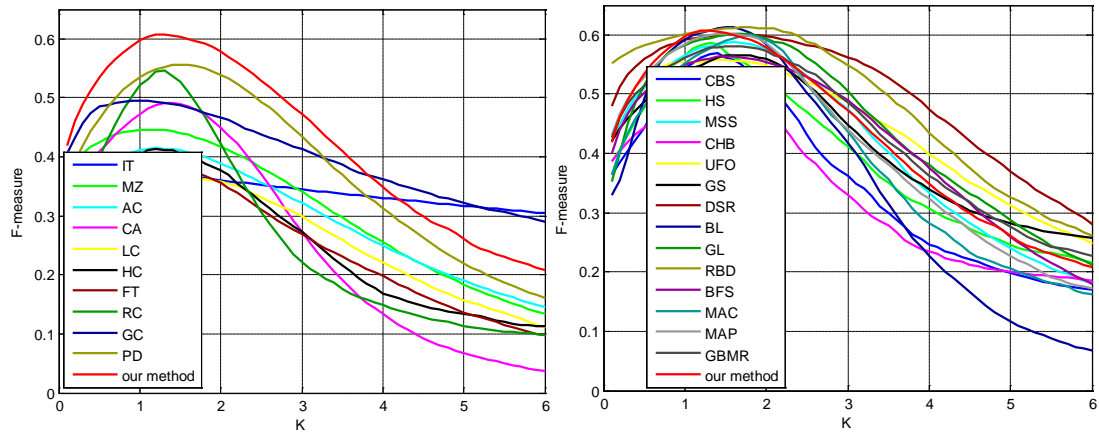
(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors

Fig. 8. F_β- K curve on ASD dataset

(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors

Fig. 9. F_β- K curve on SED1dataset

(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors
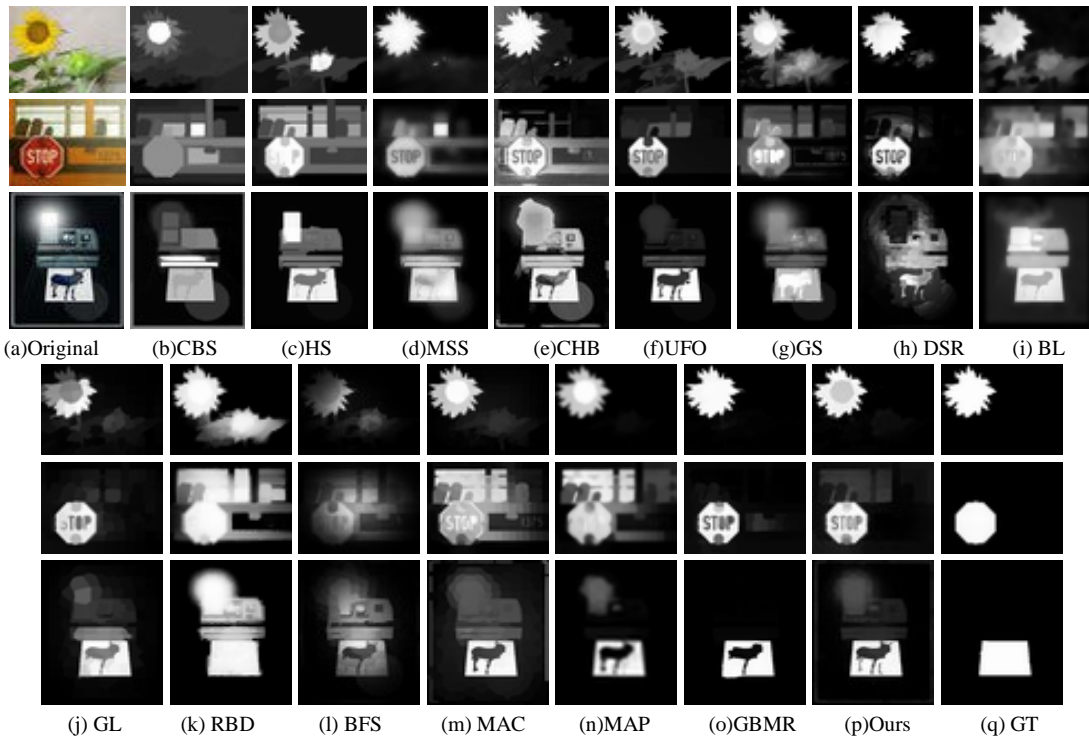
Fig. 10. F_β- K curve on SED2 dataset

(a)Comparison with the ones using contrast prior   (b)Comparison with the ones using other priors

**Fig. 11.** F$_\beta$- K curve on SOD dataset

4.3.2 Visual Comparisons

Some examples for visual comparison of all salient object detection algorithms are shown in **Fig. 12**, **Fig. 13**, **Fig. 14**, **Fig. 15**, **Fig. 16**, **Fig. 17**, **Fig. 18** and **Fig. 19**. As shown in **Fig. 12(b)**, **Fig. 14(b)**, **Fig. 16(b)**, and **Fig. 18(b)**, IT do not generate full resolution saliency maps because of using down sampling to realize multi-scale detection. As shown in **Fig. 12(c)(d)(e)**, **Fig. 14(c)(d)(e)**, **Fig. 16(c)(d)(e)**, and **Fig. 18(c)(d)(e)**, MZ, AC and CA tend to highlight the boundaries by virtue of calculating the local contrast in a small neighborhood. LC, HC, RC, GC estimate color contrast over the entire image, but they mistaken detecting the brightest color as the salient region as shown in the third row of **Fig. 12(f)(g)(i)(j)**. From **Fig. 12(k)**, **Fig. 14(k)**, **Fig. 16(k)**, and **Fig. 18(k)**, we can see that PD performs well in highlighting the boundaries but still cannot label all the pixels inside the salient objects. As shown in the first and second row of **Fig. 13(b)(c)(d)**, CBS, HS and MSS mistaken detecting the center region as the salient object. The accuracy of CHB depends on the location of convex hull. The result of **Fig. 19(e)** shows that when the background of image become complicated, CHB will fail as there appears more corner points on background regions. As shown in Fig.19(f)(j), UFO and GL do not work well in the more complex SOD dataset. The accuracy of GS, DSR, BL reduce when the object appearing at the boundary as shown in **Fig. 13(g)(h)(i)**. Compared to the saliency map generated by MAC, MAP, GBMR in **Fig. 13(m)(n)(o)**, Our method performs well in highlighting both the small and big salient objects. As shown in **Fig. 13(p)**, **Fig. 15(p)**, **Fig. 17(p)** and **Fig. 19(p)**, our method also work well in detecting multiple objects and the objects with complex background. Overall, the saliency maps of our method are more similar to their corresponding ground truth.
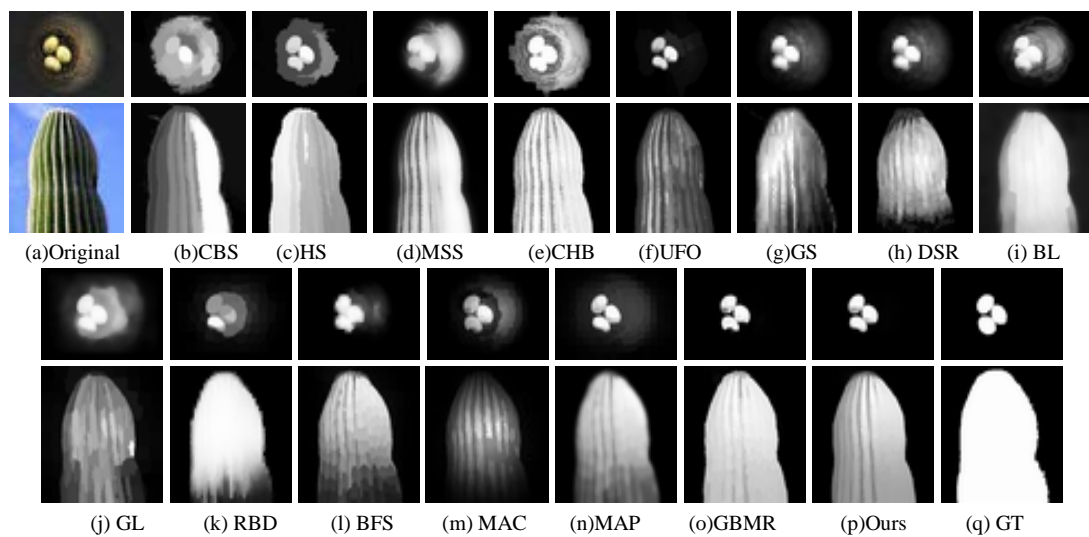

(a)Original (b)IT   (c)MZ   (d)AC   (e)CA   (f)LC   (g)HC   (h)FT   (i)RC   (j)GC   (k)PD   (l)Ours   (m)GT

Fig. 12. The visual comparisons with the detection algorithms only using contrast prior in ASD database

(a)Original   (b)CBS   (c)HS   (d)MSS   (e)CHB   (f)UFO   (g)GS   (h) DSR   (i) BL

(j) GL   (k) RBD   (l) BFS   (m) MAC   (n)MAP   (o)GBMR   (p)Ours   (q) GT

Fig. 13. The visual comparisons with the detection algorithms using other priors in ASD database

(a)Original   (b)IT   (c)MZ   (d)AC   (e)CA   (f)LC   (g)HC   (h)FT   (i)RC   (j)GC   (k)PD   (l)Ours   (m)GT
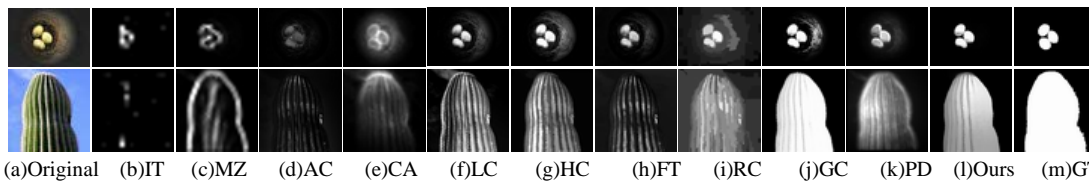
Fig. 14. The visual comparisons with the detection algorithms only using contrast prior in SED1 database

(a)Original   (b)CBS   (c)HS   (d)MSS   (e)CHB   (f)UFO   (g)GS   (h) DSR   (i) BL

(j) GL   (k) RBD   (l) BFS   (m) MAC   (n)MAP   (o)GBMR   (p)Ours   (q) GT

Fig. 15. The visual comparisons with the detection algorithms using other priors in SED1 database

(a)Original (b)IT   (c)MZ   (d)AC   (e)CA   (f)LC   (g)HC   (h)FT   (i)RC   (j)GC   (k)PD   (l)Ours   (m)GT

Fig. 16. The visual comparisons with the detection algorithms only using contrast prior in SED2 database
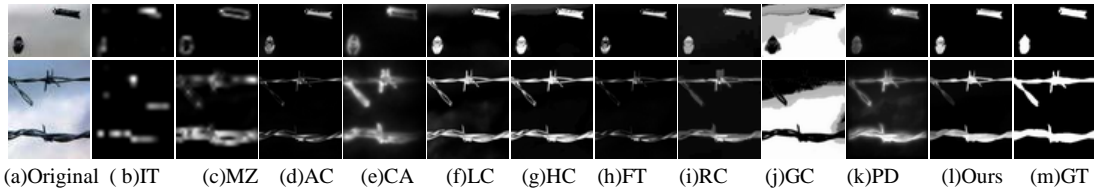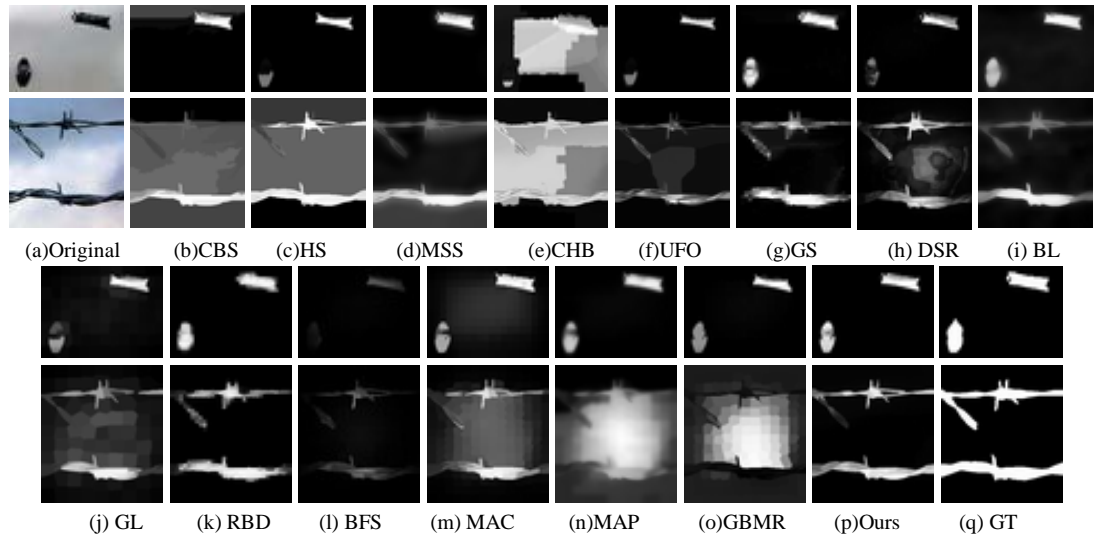

(a)Original   (b)CBS   (c)HS   (d)MSS   (e)CHB   (f)UFO   (g)GS   (h) DSR   (i) BL

(j) GL   (k) RBD   (l) BFS   (m) MAC   (n)MAP   (o)GBMR   (p)Ours   (q) GT

Fig. 17. The visual comparisons with the detection algorithms using other priors in SED1 database

(a)Original (b)IT   (c)MZ   (d)AC   (e)CA   (f)LC   (g)HC   (h)FT   (i)RC   (j)GC   (k)PD   (l)Ours   (m)GT
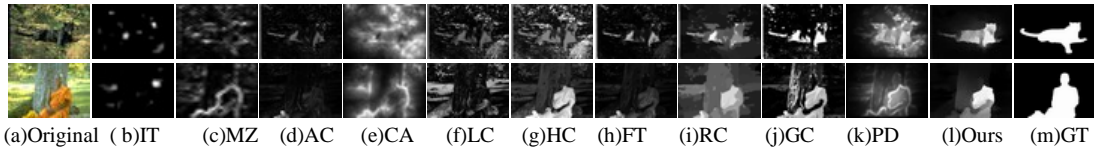
Fig. 18. The visual comparisons with the detection algorithms only using contrast prior in SOD database

(a)Original   (b)CBS   (c)HS   (d)MSS   (e)CHB   (f)UFO   (g)GS   (h) DSR   (i) BL

(j) GL   (k) RBD   (l) BFS   (m) MAC   (n)MAP   (o)GBMR   (p)Ours   (q) GT
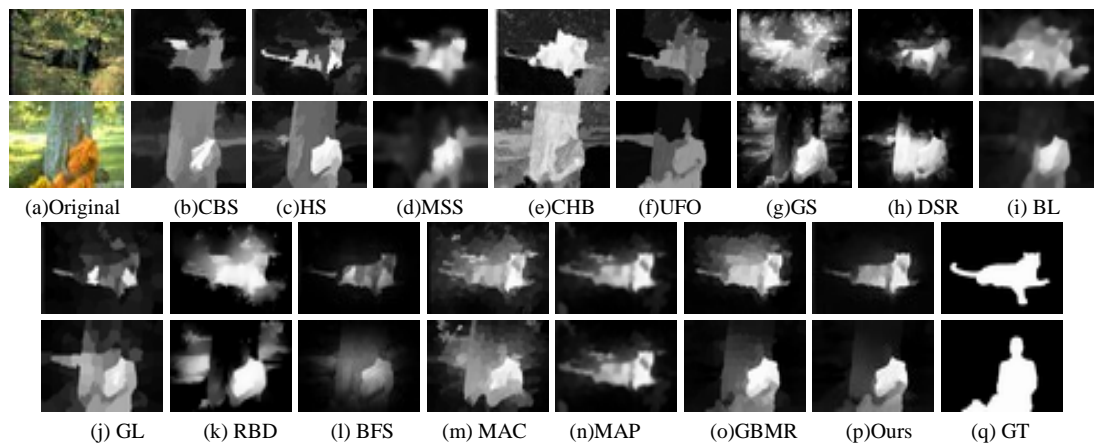
Fig. 19. The visual comparisons with the detection algorithms using other priors in SOD database

5. Conclusion

In order to improve the accuracy of diffusion-based salient object detection methods, this paper propose a multi-scale diffusion-based salient object detection algorithm with background and objectness seeds. The image is firstly over-segmented at several scales. The background and objectness saliency of each superpixel is then calculated in each scale. The Bayesian fusion of the background and objectness saliency is subsequently propagated to the whole image using manifold ranking algorithm. The final saliency is weighted summarization of the saliency under different scales. The results on ASD benchmark dataset confirm the efficiency of each component of our method. Extensive experiments against state-of-art methods were carried out on ASD, SED and SOD benchmark datasets. The quantitative comparison results show that the proposed methods performs favorably against 24 state-of-the-art salient object approaches. The visual comparison results show that our method can highlight the salient object more uniformly and brightly.

References

[1] Z. Guo, L. L. Gao, J. K. Song, X. Xu, J. Shao, H. T. Shen, "Attention-based LSTM with semantic consistency for videos captioning," in *Proc. of the 24th* ACM Conference on Multimedia, pp.357-361, Oct.15-19, 2016. Article(CrossRef Link)

[2] M. S. Gide, and L.J. Karam, "A locally weighted fixation density-based metric for assessing the quality of visual saliency predictions," *IEEE Transactions on Image Processing*, vol.25, no.8, pp.3852-3861, Aug., 2016. Article (CrossRef Link) .

[3] J. W. Wang, A. Borji, C. C. Jay Kuo, and L. Itti, "Learning a combined model of visual saliency for fixation prediction," *IEEE Transactions on Image Processing*, vol.25, no.4, pp.1566 - 1579, Apr., 2016. Article (CrossRef Link) .

[4] J. J. Lai, B. R. Wang, Y. M. Fang, W. S. Lin, P. L. Callet, N.Ling, and C. P. Hou, "A universal framework for salient object detection," *IEEE Transactions on Multimedia*, vol.18, no.9, pp.1783 - 1795, Sep., 2016. Article (CrossRef Link) .

[5] H. W. Peng, B. Li, H. B. Ling, W. M. Hu, W. H. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol.39, no.4, pp.818-832, Apr., 2017. Article (CrossRef Link) .

[6] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, Nov., 1998. Article (CrossRef Link) .

[7] Y. F. Ma, and H. J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. of the 11th ACM Conference on Multimedia*, pp.374-381, Nov.2-8, 2003. Article (CrossRef Link).

[8] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proc. of the 6th International Conference on Computer Vision Systems*, pp.66-75, May.12-15, 2008. Article (CrossRef Link).

[9] S. Goferman, L. Zelnik-Mamor, and A. Tal, "Context-aware saliency detection," in *Proc. of* the *23rd International Conference on Computer Vision and Pattern Recognition*, pp.2376-2383, Jun.13-18, 2010. Article (CrossRef Link).

[10] Y. Zhai, and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. of the 14th ACM Conference on Multimedia*, pp.815-824, Oct.23-27, 2006. Article (CrossRef Link).

[11] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, "Global contrast based salient region detection," in *Proc. of the 24th International Conference on Computer Vision and Pattern Recognition*, pp.409-416, Jun.20-25, 2011. Article (CrossRef Link).

[12] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. of the 22nd Conference on Computer Vision and Pattern Recognition*, pp.1597-1604, Jun.20-25, 2009. Article (CrossRef Link) .

[13] J. K. Song, L. L. Gao, M. M. Puscas, F. P. Nie, F. M. Shen, N. C. Sebe, "Joint graph learning and video segmentation via multiple cues and topology calibration," in *Proc. of the 24th* ACM Conference on Multimedia, pp.831-840, Oct.15-19, 2016. Article(CrossRef Link)

[14] M. M. Cheng, J. Warrell, and W. Y. Lin, "Efficient salient region detection with soft image abstraction," in *Proc. of the 14th International Conference on Computer Vision*, pp.1529-1536, Dec.1-8, 2013. Article (CrossRef Link) .

[15] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?," in *Proc. of the 26th International Conference on Computer Vision and Pattern Recognition*, pp.1139-1146, Jun.23-28,2013. Article (CrossRef Link) .

[16] H. Z. Jiang, J. D. Wang, Z. J. Yuan, and T. Liu. "Automatic salient object segmentation based on context and shape prior," in *Proc. of th*e *14th British Machine Vision Conference*, pp.111-122, Aug.29-Sep.2, 2011. Article (CrossRef Link) .

[17] Q. Yan, L. Xu, J. P. Shi, and J. Y. Jia. "Hierarchical Saliency Detection," in *Proc. of the 26th International Conference on Computer Vision and Pattern Recognition*, pp.1155 - 1162, Jun.23-28, 2013. Article (CrossRef Link) .

[18] N. Tong, H. C. Lu, L. H. Zhang, and X. Ruan, "Saliency detection with multi-scale superpixels," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp.1035-1039, Sep, 2014. Article (CrossRef Link).

[19] L. H. Zhang, S. F. Zhao, W. Liu, and H. C. Lu, "Saliency detection via sparse reconstruction and joint label inference in multiple features," *Neurocomputing*, vol. 55, no. C, pp.1-11, May., 2015. Article (CrossRef Link).

[20] Y. L. Xie, H. C. Lu, and M. H. Yang, "Bayesian saliency via low and mid level cues," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp.1689-1698, May, 2013. Article (CrossRef Link).

[21] P. Jiang, H. B. Ling, J. Yu, and J. L. Peng, "Salient region detection by UFO: uniqueness, focusness and objectness," in *Proc. of the 14th International Conference on Computer Vision*, pp.1976 - 1983, Dec.1-8, 2013. Article (CrossRef Link).

[22] Y.Wei, F.Wen,W. Zhu, and J. Sun, " Geodesic saliency using background priors," in *Proc. of the 12th European Conference on Computer Vision*, pp.29-42, Oct.7-13, 2012. Article (CrossRef Link).

[23] H. C. Lu, X. H. Li, L. H. Zhang, X. Ruan, and M. H. Yang, "Dense and sparse reconstruction error based saliency descriptor," *IEEE Transaction on Image Processing*, vol. 25, no. 4, pp.1592-1603, Apr., 2016. Article (CrossRef Link) .

[24] H. C. Lu, N. Tong, X. N. Zhang, J. Q. Qi, X. Ruan, and M. H. Yang, "Co-bootstrapping saliency,"*IEEE Transaction on Image Processing*, vol. 26, no. 1, pp.414-425, Jan., 2017. Article (CrossRef Link).

[25] C. Jia, J. Q. Qi, X. H. Li, and H. C. Lu, "Saliency detection via a unified generative and discriminative model," *Neurocomputing,* vol. 173, no. P2, pp.406-417, Jan., 2016. Article (CrossRef Link).

[26] N. Tong, H. C. Lu, L. H. Zhang, and X. Ruan, "Salient object detection via global and local cues," *Pattern Recognition*, vol. 48, no. 10, pp.3258-3267, Oct., 2015. Article (CrossRef Link).

[27] W. J. Zhu, S. Liang, Y. C. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. of the 27th International Conference on Computer Vision and Pattern Recognitio*n, pp.2814 - 2821, Jun.24-27, 2014. Article (CrossRef Link).

[28] J. B. Zhou, J.Y. Zhai, Y. F. Ren, A. L. Lu, "Background prior-based salient object detection via adaptive figure-ground classification," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 3, pp. 1264-1286, Mar., 2018. Article(CrossRef Link)

[29] J. P. Wang, H. C. Lu, X. H. Li, N. Tong, and W. Liu, "Saliency detection via background and foreground seed selection," *Neurocomputing*, vol. 152, no. C, pp. 359-368, Mar., 2015. Article (CrossRef Link) .

[30] J. G. Sun, H. C. Lu, and X. P. Liu, "Saliency region detection based on Markov absorption probabilities," *IEEE Transactions on Image Processing,* vol. 24, no. 5, pp. 1639-1649, May, 2015. Article (CrossRef Link) .

[31] B. W. Jiang, L. H. Zhang, H. C. Lu, C. Yang, and M. H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. of the 14th International Conference on Computer Vision*, pp.1665 - 1672, Dec.1-8, 2013. Article (CrossRef Link) .

[32] J. Y. Zhai, J. B. Zhou, Y. F. Ren, Z. J. Wang, "Salient object detection via multiple random walks," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 4, pp. 1712-1731, Apr., 2016. Article(CrossRef Link)

[33] L. H. Zhang, C. Yang, H. C. Lu, X. Ruan, and M. H. Yang, " Ranking saliency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1-15, Sep., 2016. Article (CrossRef Link) .

[34] B. Alexe, T. Deselaers, and F. Vittorio, "Measuring the objectness of image windows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2189-2202, Nov., 2012. Article (CrossRef Link) .

[35] J. K. Song, L. L. Gao, F. P. Nie, H. T. Shen, Y. Yan, N. Sebe, "Optimized graph learning using partial tags and multiple features for image and video annotation," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp.4999-5011, Nov., 2016. Article(CrossRef Link)

[36] L. L. Gao, J. K. Song, F. P. Nie, F. H. Zhou, N. C. Sebe, H. T. Shen, "Graph-without-cut:an ideal graph learning for image segmentation," in *Proc. of the 30th AAAI Conference on Artificial Intelligence*, pp.1188 - 1194, Feb.12-17, 2016.

Sai Yang is currently a lecturer of Nantong University. She received his M.S. degree from School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, China, in 2010, and a Ph.D. degree from School of Computer Science and Engineering, Nanjing University of Science and Technology,China, in 2015. Her research interests include computer vision, image processing, pattern recognition and machine learning.
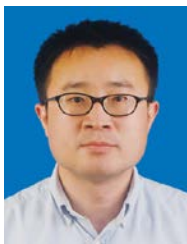
Fan Liu is currently a lecturer of HoHai University. He received his B.S. degree in network engineering from Nanjing University of Science and Technology (NUST) in June 2009. He received his Ph.D. degree from Nanjing University of Science and Technology in January 2015. From February 2014 to May 2014, he worked as an intern in Microsoft Research Asia. His research interests include computer vision, image processing, pattern recognition and deep learning. Dr. Liu also serves as reviewer of Information Sciences, Neurocomputing, KSII Transaction on Internet and Information Systems and Pattern Analysis and Application.

JuanChen is currently a senior experimentalist of Nantong University. She received her M.S.degree from School of Electrical Engineering, Guangxi University, China, in 2010, and is pursuiting her Ph.D.degree in School of Mechatronics Engineering and Automation, Shanghai University, China. Her research interests include fault diagnosis and fault tolerant control, analysis and control of power electronic system.

Dibo Xiao received his master and doctor degrees in Engineering from Nanjing University of Aeronautics and Astronautics. He is currently a lecturer at school of electrical engineering, Nantong University. His research interests are in the dynamics and control of flight vehicles, and robotics.

Hairong Zhu is currently an associate professor of Nantong University. He received his M.S. degree from School of Automation, Southeast University, China, in 2004, and a Ph.D. degree from School of Automation, Southeast University, China, in 2015. His research interests include control theory and application, servo control and robotics.