Pattern Recognition 60 (2016) 2–12

Contents lists available at ScienceDirect

Pattern Recognition

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

Discriminative saliency propagation with sink points

Shuhan Chen [a,b,*], Ling Zheng [a], Xuelong Hu [a], Ping Zhou [b]

[a] College of Information Engineering, Yangzhou University, Yangzhou, China
[b] Wanfang Electronic Technology Co., Ltd., Yangzhou, China

ARTICLE INFO

Article history:
Received 20 September 2015
Received in revised form
17 April 2016
Accepted 7 May 2016
Available online 20 May 2016

Keywords:
Saliency Detection
Saliency Propagation
Similarity Metric
Sink Points

ABSTRACT

Salient object detection is still very challenging especially in images with complex or cluttered background. In this paper, we present an efficient and discriminative framework to address it. In specially, a discriminative similarity metric is first proposed by measuring the chi-square distance in a new constructed feature space. Then, we apply it to calculate a background based coarse saliency map by introducing distribution prior to remove foreground noises in the image boundaries. Based on manifold ranking, a robust saliency propagation mechanism is further developed to highlight salient object and simultaneously suppress background region by setting appropriate sink points. Finally, several simple refinement techniques are utilized to generate pixel-wise and smooth saliency maps. Extensive experimental results show the superior performance of the proposed method in terms of different evaluation metrics. In addition, the proposed framework can be also applied to the existing saliency propagation methods for significant performance boosting. We also believe that it is a good choice for subsequent applications based on the achieved high performance and acceptable computational overhead.

1. Introduction

Quickly prioritizing external visual stimuli and localizing most interest in a scene is an astonishing capability of humans [1]. It is called visual attention that has become a very active topic in both neuroscience and computer vision. Its research mainly contains three aspects [2]: eye fixation prediction [3], saliency detection [4], and objectness estimation [5]. In this paper, we focus on saliency detection which aims to make certain objects or regions of an image stand out from their neighbors and catch immediate attention. It is still challenging to develop efficient saliency detection algorithms which can make great helpfulness to a wide range of computer vision tasks, such as object recognition [6], image retargeting [7], visual tracking [8], image compression [9] and so on.

The development of saliency detection can be roughly divided into two stages. The first stage can be summarized as low-level cue based saliency detection, which tries to exploit effective cues, such as contrast, and spatial distance. One of the most adopt principles, backgroundness prior, is to take the contrast versus a narrow border of the image (i.e. pseudo-background) as a region's saliency [10–14,47]. However, it will be imprecise and directly lead to undesirable results when the object is adjacent to one or more image boundaries. To improve it, Wang et al. [30] remove the superpixels with large edge feature calculated by the average probability of boundary value, which are regarded as the foreground noises. Li et al. [20] treat the top 30% pixels with high color difference empirically among boundaries as foreground noise. While Li et al. [24] drop the most distinctive boundary and keep the remaining three as background seeds. Nevertheless, it is still unreliable and not robust enough if only calculating distinctiveness among image boundaries. Contrast prior is another widely used principle by measuring the local or global contrast difference between the object region with respect to other regions in an image to compute saliency [4,11,15–17]. Spatial distribution prior [18], focusness prior [1,19] and objectness prior [1,20] have been also exploited to facilitate the detection of salient objects. All of them perform well in easy cases, but still struggle in complex scenes as shown in Fig. 1 (a), which is mainly caused by the powerless similarity metric, such as the Euclidean distance in CIELab color space.

The second stage we call propagation based saliency detection, which applies some optimization techniques to improve the visual quality of such coarse saliency maps. In detail, an input image is first represented by a graph with segmented superpixels as nodes which are connected by weighted edges. Then, saliency values are conducted by different propagation models diffused along these edges from labeled superpixels to unlabeled neighbors, such as random walks [14,22], manifold ranking [12]. Recently, we have witnessed a blossom of saliency propagation based methods which achieved state-of-the-art performance. Sun et al. [23] rank

* Corresponding author at: College of Information Engineering, Yangzhou University, Yangzhou, China.
E-mail addresses: c.shuhan@gmail.com (S. Chen), zlapgx@163.com (L. Zheng), xlhu@yzu.edu.cn (X. Hu), zp@wfdz.com.cn (P. Zhou).

http://dx.doi.org/10.1016/j.patcog.2016.05.016
0031-3203/

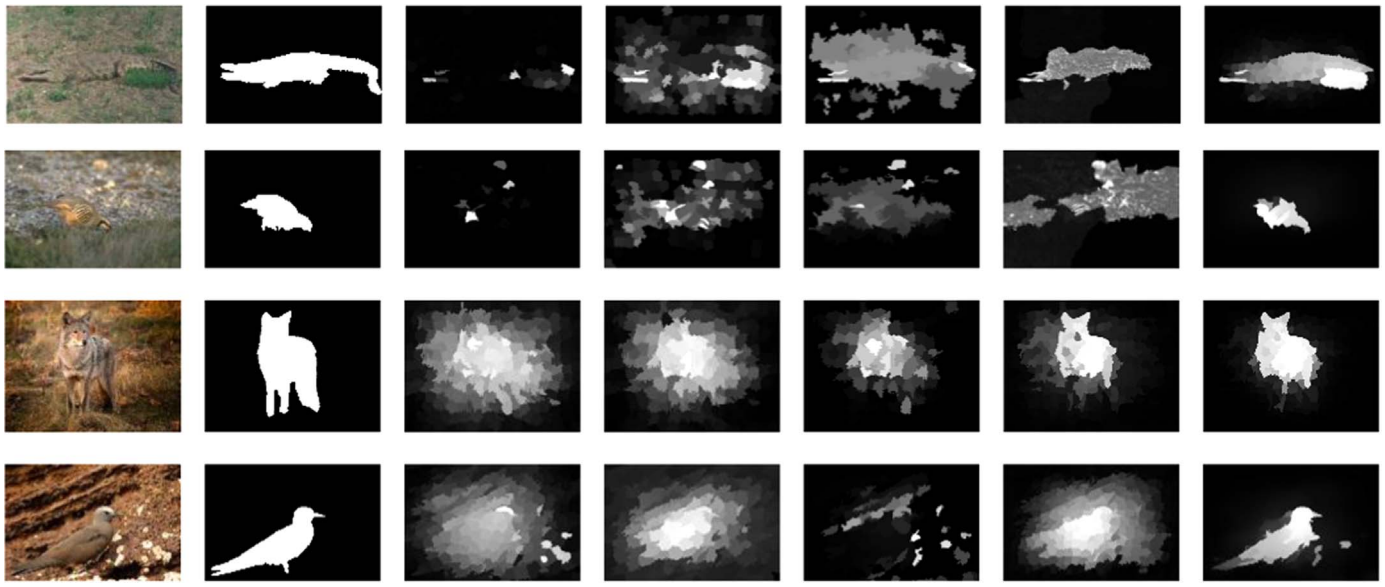


Fig. 1. Saliency detection in challenging images. The first two rows are low-level cues based methods, from left to right: input image, ground truth, SF [15], GS [10], RC [39], UFO [1], and our method; the last two rows are propagation based methods, from left to right: input image, ground truth, MC [14], MR [12], RBD [41], BSCA [31], and our method.

the similarities of all image elements with the foreground cues generated by the Markov absorption probability on a 2-ring graph model. Li et al. [20] propose an inner and an inter label propagation method combining with a co-transduction framework for saliency detection. In [31], a cellular automata based propagation mechanism is proposed to exploit the intrinsic relevance of similar superpixels by interacting with neighbors. To get pixel-wised saliency map, Li et al. [24] develop a regularized random walks ranking model to optimize the prior maps. Inspired by the theory of educational psychology, Gong et al. [25] present a novel propagation approach whose main idea is to postpone the propagation of difficult regions and meanwhile advance the less ambiguous simple regions. While Tong et al. [45] treat it as a classification problem and propose a bootstrap learning method to exploit multiple features for saliency detection. However, they may cause errors when the foreground seeds are not accurate enough. In other words, some background regions will also be diffused if they are falsely selected as foreground seeds, thus leading to inaccurate detection results. Although several seed selection mechanisms have been proposed in the literature [21,30], they are not accurate enough or time consuming. Furthermore, these propagation methods still can not highlight salient objects uniformly and suppress background regions well simultaneously in challenging images, especially when the background is very similar to the object, as can be seen in Fig. 1(b).

In a word, saliency propagation is an effective solution to get high quality saliency maps and also one of the main research directions in saliency detection recently. However, it is still not satisfactory enough especially in images with complex or cluttered background. In this paper, we present a discriminative saliency propagation (DSP) framework to overcome the drawbacks mentioned above. It can be also applied to other saliency propagation models for significant improvement. First, an integrated feature space is constructed by combing three color spaces (CIELab, HSV, and opponent color space [26]) and gradient magnitude channel for similarity measure. To the best of our knowledge, all the existing saliency propagation methods address it based on the Euclidean distance in CIELab color space. However, it is powerless to distinguish the salient regions and backgrounds with very similar color or appearance which happens frequently in real scenes. Second, we filter out the superpixels in the boundaries that most unlikely belong to the background by using distribution prior (e.g., when objects appear at the image boundary) and thus obtain more stable and reliable background seeds. Thus, a distribution guided backgroundness (DGB) saliency map can be obtained by measuring the difference with them for each superpixel. Third, to reduce propagation errors, we introduce sink points into manifold ranking, which are nodes whose rough saliency values are lower than a threshold during the manifold ranking process. This way, the ranking scores of other nodes close to the sink points (i.e., nodes sharing similar saliency values with the sink points) will be naturally penalized during the ranking process based on the intrinsic manifold. As a result, background regions will not be diffused if we set appropriate sink points, and objects can be well stood out accordingly. Iteration is not needed, thus, it is more simple and efficient. Finally, the sigmoid function and fast bilateral filter [43] are used to produce smooth and pixel-wise saliency maps, and a new two-level fusion approach is proposed for further improvement, which is a combination of saliency maps using two different superpixel segmentation methods to capture their complementary characteristic.

In a nutshell, the main contributions of this work include:

1) We propose an efficient and robust background based weak saliency model by integrating distribution prior and a new discriminative similarity metric, which are used to get desirable results even when the salient object is connected to one or more image boundaries and handle images with complex or cluttered background respectively.

2) Different with the seed selection methods, we develop a novel saliency propagation framework by introducing sink points into manifold ranking to reduce propagation errors especially in background regions. Less background regions will be falsely detected as salient objects when comparing with the existing propagation based approaches. It can be also applied to improve all the propagation based methods to a similar performance level.

3) Several refinement techniques are further proposed for superpixel based saliency detection, which can uniformly highlight the salient objects and generate smooth saliency maps.

By using the Precision-Recall (PR) curve, F-measure, Mean Absolute Error (MAE) and Area Under ROC Curve (AUC) as the evaluation criterions, we compare our approach with fifteen state-of-the-arts on five benchmark datasets. The experiment results show the impressive performance of the proposed method against all the other methods. Note that it is also comparable with supervised deep learning based method [27] both in quantitative and qualitative comparison. Furthermore, we achieve it with acceptable computational overhead (about 0.47 seconds per image) which makes it a good choice for subsequent applications. The remainder of the paper is organized as follows. Section 2 introduces the proposed saliency detection framework, including coarse map generation, saliency propagation, and refinements. Section 3 shows the comparison results with other algorithms and analysis of them. Finally, conclusions are drawn in Section 4.

2. Proposed algorithm

In this section, we detail our effective and robust saliency detection framework. At first, an input image is segmented into a fixed number of superpixels by efficient segmentation algorithm. Then, we construct our feature space by combining multiple color spaces and gradient magnitude for superpixel description, which can be applied both in coarse map generation and saliency propagation to distinguish very similar regions. After that, the background based weak saliency map is calculated by using distribution prior to remove foreground noises in the image borders. The obtained coarse map is further optimized by our propagation algorithm with sink points. Finally, we refine the generated saliency map by the sigmoid function and fast bilateral filter. A new two-level fusion is further used for performance improvement. The pipeline of the proposed algorithm is shown in Fig. 2.

2.1. Discriminative similarity metric

Similarity metric is essential for saliency detection, which is used to measure the difference between background and foreground region. Many discriminative unsupervised [11,34] and supervised [35] similarity metrics have been explored for saliency detection. To the best of our knowledge, all the existing methods simply use the Euclidean distance in CIELab color space as similarity measure in the affinity matrix construction of saliency propagation. However, it is not discriminative enough to separate background and foreground regions successfully, especial in complex images.

Different with [36], in which two color features are combined into saliency propagation by joint label inference, we directly concatenate multiple color histogram features to a discriminative feature vector. CIELab and HSV are the most used color spaces in previous works and their effectiveness have been demonstrated in [11]. In this paper, we further introduce opponent color space [26] whose three component channels are: black versus white obtained by $(R+G+B)/3$, red versus green obtained by $R–G$, and blue versus yellow obtained by $B-(R+G)/2$, where R, G, and B denote red, green, and blue channel in RGB color space respectively. Then, a concatenated histogram feature can be obtained in such three color spaces, which will be used in the following initial saliency map estimation and saliency propagation.

Nevertheless, it still may be struggle in complex images, especially when the background and foreground have very similar color. Fortunately, they usually have different texture structure which is measured by local binary pattern feature in most of the previous works [11,44,45]. Here, we simply use gradient magnitude for efficiency. By applying Gaussian filter with different window sizes (5×5, 9×9, and 15×15) and sigma values (1, 3, 5) to the gradient magnitude map, we can get a rough saliency map by adding them together. From Fig. 3, we can see that either background or foreground can be successfully highlighted when they have different texture structure. By concatenating it with the above color channels, we can get the final feature for superpixel description. Table 1 summarizes the features that we have used. In short, our superpixel feature vector consists of $256\times3+64=832$ dimensions. Finally, the similarity can be simply measured in the obtained feature space by the chi-square distance which is more discriminative than the previous Euclidean distance measure.

2.2. Distribution guided backgroundness

Backgroundness is a very effective cue and has been widely used in saliency detection, which assumes the image borders as prior background regions. However, it will lead to negative effects on saliency detection as mentioned above. Then, some works try to solve it by calculating distinctiveness only among image boundaries. Unfortunately, it will lose its effectiveness when the background is complex. In such case, some background regions may be distinctive and then be removed as foreground noise. Here, we solve it by using distribution prior of the whole image which is

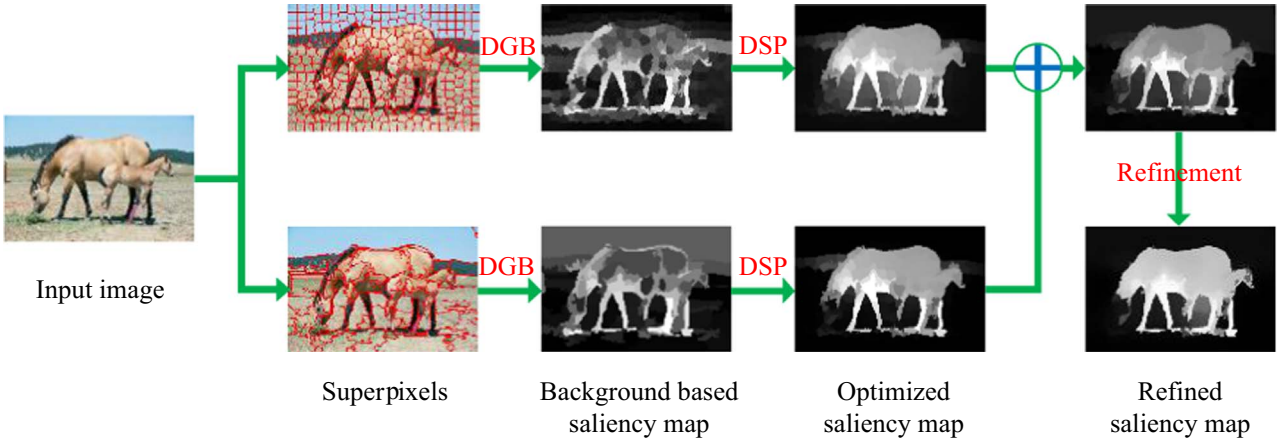


Fig. 2. Pipeline of the proposed saliency detection method. The input image is first segmented into superpixels by two different approaches. Then we compute coarse saliency maps by selecting appropriate background superpixels in the image boundaries for each of them. Such coarse saliency maps are further optimized by our discriminative saliency propagation. Finally, we combine them together as its final saliency map, and some refinement techniques are further applied to improve its visual quality.

Fig. 3. Input images (left) and their corresponding gradient maps (right).

Table 1
Features for each superpixel.

Feature descriptions	Dimensions
The CIELab color histogram	256
The HSV color histogram	256
The opponent color histogram	256
The gradient histogram in CIELab	64

measured by spatial variance. Superpixels with low variance indicate more salient than spatially widely distributed ones [15]. In detail, it is defined as below:

$$D_i = \frac{1}{Z_i}\sum_{j=1}^{N}\exp\left(-\frac{h_{ij}}{2\sigma_d^2}\right)\cdot\|p_j-\mu_i\|^2 \quad (1)$$

$$\mu_i = \frac{1}{Z_i}\sum_{j=1}^{N}\exp\left(-\frac{h_{ij}}{2\sigma_d^2}\right)\cdot p_j \quad (2)$$

where h_{ij} is the chi-square distance between two histograms of superpixel i and j in our constructed feature space, p_j is the position of superpixel j, Z_i is the normalization term, μ_i defines the weighted mean position of superpixel i, and N is the number of superpixels. The parameter σ_d controls the color sensitivity of the

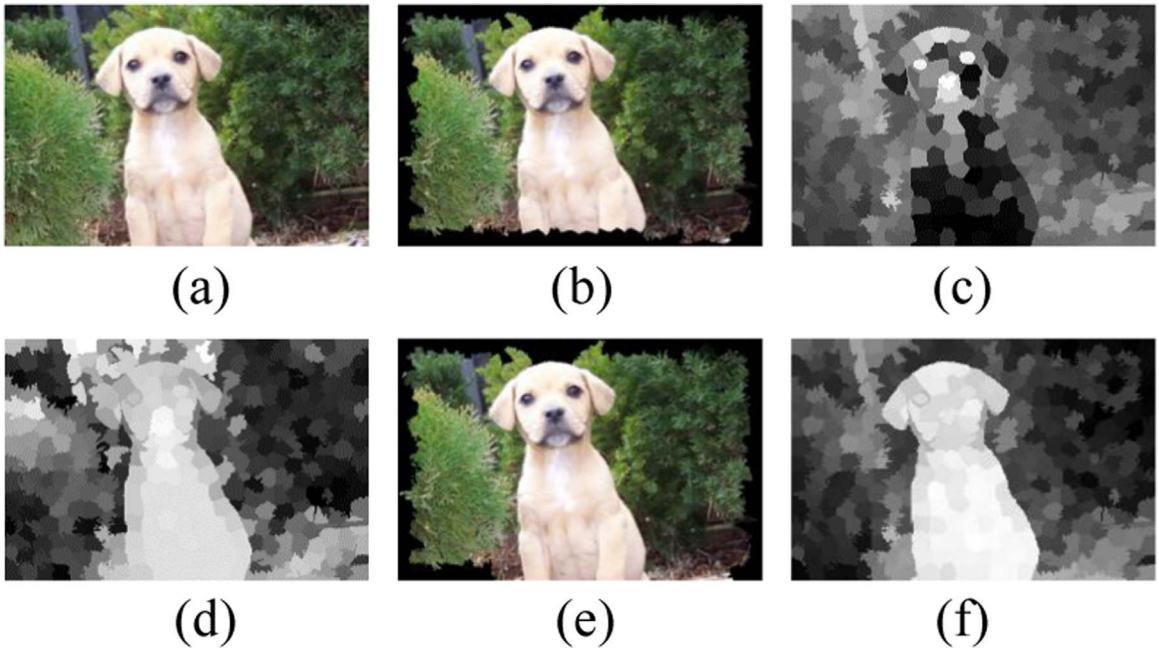


Fig. 4. The effect of our foreground noise removal in Section 2.2: (a) input image, (b) and (e) background seeds (the black regions along image borders), (c) and (f) background based saliency map using (b) and (e) respectively, (d) distribution based saliency map.

element distribution, which is set to 1 in this paper.

Based on the distribution prior, we can simply remove foreground noises in the image border regions by setting different thresholds for different border sides since their different probabilities of connecting with salient objects. In specific, superpixels are dropped from the background sets if their distribution based saliency values are larger than a threshold which is adaptively derived by Otsu [37] for bottom side and its twice for the other sides for simplicity. Then, the remaining superpixels in the image borders are selected as background set denoted as BG, which is stable and reliable as can be seen in Fig. 4.

Then, the background based saliency map can be measured by calculating the contrast versus the obtained background seeds. However, superpixels in the background seeds may be distinctive with each other if the input image is complex. In other words, a superpixel in the background will only be similar to some part of it rather than all of the superpixels in the background seeds, while a superpixel in foreground will be dissimilar to all of the background seeds. Thus, it is not a good choice to sum up all the differences between a superpixel and the background seeds to measure its saliency value. In previous works, saliency is measured alone in each image border [12] or three background parts obtained by K-means clustering [31]. In this work, we simply sum up the minimum L differences to measure the background based saliency, which is defined as:

$$B_i = \sum_{l=1}^{L} \tilde{h}_{il},\ i = 1, \dots, N \tag{3}$$

where $\tilde{h}_{il}$ denotes the ascending sorted distance between superpixel i and superpixel l belonging to the background set. Here, L is adaptively set as $|BG|/10$, where $|BG|$ is the cardinality of BG. After that, the coarse saliency map of our DGB can be obtained by normalizing B to [0, 1].

2.3. Saliency propagation with sink points

The coarse saliency maps generated by the proposed DGB need to be further optimized by saliency propagation whose task is to reliably and accurately transmit saliency values from the labeled superpixels to the remaining unlabeled ones. All the propagation approaches have a similar form as $f^* = As$, where A is a propagation matrix, f and s denote saliency values and foreground seeds respectively. Some works [28] focus on constructing discriminative propagation matrixes, while some others [21] try to learn optimal foreground seeds. Different with them, we improve it in a novel way by introducing sink points into manifold ranking to reduce propagation errors.

Given an input image segmented into N superpixels, the neighborhood of each superpixel is defined as prior works [12,31], which include superpixels neighboring it as well as sharing common boundary with its adjacent superpixels. In addition, those superpixels belonging to the BG are also considered as connecting with each other. The similarity of two superpixels is also measured by the chi-square distance in the above constructed feature space, which is defined as:

$$w_{ij} = \begin{cases} \exp\left(-\lambda \cdot h_{ij}\right) & j \in NB(i) \\ 0 & i = j \text{ orotherwise} \end{cases} \tag{4}$$

where $NB(i)$ is the set of neighbors of superpixel i, and λ is a parameter to control the strength of similarity. We adaptively set $\lambda = 8 \cdot \left(1 - \text{mean}(h)\right)$ in the experiments. It can be explained by the fact that high average value of h indicates large salient object, which we should set rather loose strength. Then, we can get a weight matrix $W = [w_{ij}]_{N \times N}$ and a normalized degree matrix $D = \text{diag}\{d_1, \dots, d_N\}$, where $d_i = \sum_j w_{ij}$. Based on this, a graph structure $G = (V, E)$ with nodes V and edges E can be established, where V corresponds to the coarse saliency values $s = [B_1, \dots, B_N]$ and E is weighted by W. Let f: $s \rightarrow \mathbf{R}^N$ be the ranking function assigning rank values $f = [f_1, \dots, f_N]^T$ to each input B_i. It is noted that we directly use the rough saliency value of each superpixel to replace the binary queries in the original manifold ranking [12] which will introduce errors in the binarization.

Based on the above definitions, we introduce the concept of sink points into the data manifold to derive the propagation matrix A. The sink points we defined here are the points which will never spread any ranking score to their neighbors during the ranking process. It can be intuitively imagined as "black holes" where ranking scores spreading to them will be absorbed and no ranking scores would escape from them [33]. Let $I_f = diag\{\delta_1, \dots, \delta_N\}$, where $\delta_i = 0$ if superpixel i is a sink point and 1 otherwise, the cost function associated with the ranking score vector f can be formulated as:

$$f^* = \arg\min_f \frac{1}{2}\left(\sum_{i,j=1}^{N} w_{ij} \left\| \frac{\delta_i}{\sqrt{D_{ii}}} f_i - \frac{\delta_j}{\sqrt{D_{jj}}} f_j \right\|^2 + \mu \sum_{i=1}^{N} \|\delta_i f_i - B_i\| \right) \tag{5}$$

The optimal solution of f then can be obtained by setting the derivative of the above function to be zero, which can be written as:

$$f^* = \left(I - \alpha D^{-1/2} W D^{-1/2} I_f\right)^{-1} s \tag{6}$$

where I is the identity matrix, and $\alpha = 1/(1 + \mu)$.

Let $P = D^{-1} W I_f$, then P is the similarity transformation of $D^{-1/2} W D^{-1/2} I_f$ as follows:

$$\begin{aligned} D^{-1/2} W D^{-1/2} I_f &= D^{1/2} D^{-1} W D^{-1/2} D^{1/2} I_f D^{-1/2} \\ &= D^{1/2} D^{-1} W I_f D^{-1/2} \\ &= D^{1/2} P D^{-1/2} \end{aligned}$$

Hence, they have the same eigenvalues. Then, we rewrite Eq. (6) as:

$$f^* = \left(I - \alpha D^{-1} W I_f\right)^{-1} s \tag{7}$$

Similar as [12], we get another ranking function by using the unnormalized Laplacian matrix in Eq. (7) which has been demonstrated to achieve better performance [12]:

$$f^* = \left(D - \alpha W I_f\right)^{-1} s \tag{8}$$

Thus, the propagation matrix A is equal to $\left(D - \alpha W I_f\right)^{-1}$. In addition, to reduce the influence of self-similarity of the superpixels in the process of propagation, we set the diagonal elements of A to 0 as [12] does. Finally, saliency can be measured by the normalized ranking score using the foreground seeds given by the coarse saliency map, which can be written as: $S(i) = \bar{f}^*(i),\ \ i = 1, 2, \dots, N$.

As the salient object usually lies near the center of an image, known as center prior which has been widely applied for salient object detection. In this work, center weight is used for setting sink points, which is defined as:

$$c_i = \frac{1}{Z_i} \exp\left(-\frac{\|\bar{p}_{i,x} - 0.5\|^2}{2\sigma_x^2} - \frac{\|\bar{p}_{i,y} - 0.5\|^2}{2\sigma_y^2} \right) \tag{9}$$

where $\bar{p}_{i,x}$ and $\bar{p}_{i,y}$ are the normalized coordinates of superpixel i, σ_x and σ_y are set to one third of the image width and height respectively. Then we have a center-weighted coarse saliency map $S_{CB} = [c_1 B_1, \dots, c_N B_N]$, and simply set these superpixels as sink points if their saliency values are less than or equal to a threshold T_a

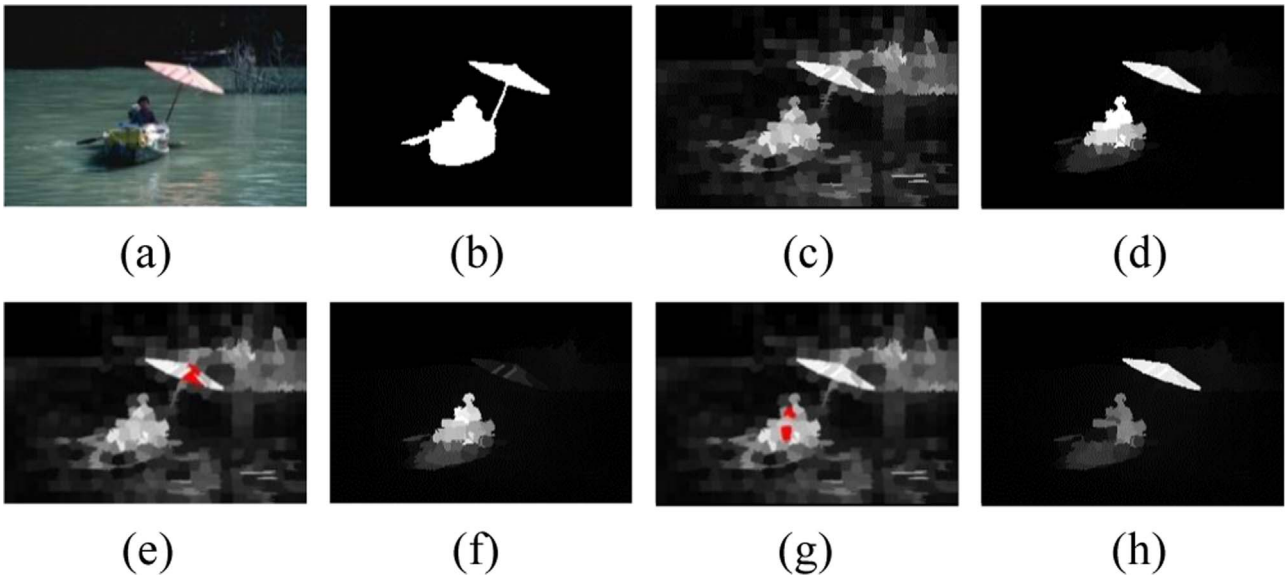


Fig. 5. Illustration of the proposed sink points: (a) input image, (b) ground truth, (c) coarse saliency map, (d) saliency propagation result of our method, (e) and (g) coarse saliency map with sink points in red, (f) and (h) saliency propagation result using sink points in (e) and (g) respectively.

which is simply set to 0 in this paper.

The effectiveness of the proposed sink points is illustrated in Fig. 5. From Fig. 5(e) and (f), we can see that the umbrella region will be suppressed after propagation if we set a sink point in it. It is the same in the body and boat regions as can be seen in Fig. 5 (g) and (h). Thus, we can suppress background regions in the process of propagation though setting appropriate sink points.

2.4. Saliency refinements

Multi-level segmentation can be further applied to improve performance such as [11,13,16,45] do. Different with them, we try to integrate different segmentation algorithms with complementary characteristic to increase the robustness to scale variation. SLIC [32] is a widely used superpixel segmentation method which can generate superpixels with similar sizes. While we can obtain superpixels with varying sizes by another efficient graph-based segmentation [29] approach. For example, a smooth area will be segmented into different superpixels by SLIC while treated as one region by graph-based segmentation as can be seen in Fig. 2. Thus, we can use them to produce two complementary saliency maps for each image, and the final saliency output is simply generated by averaging them. Obviously, our two-level fusion refinement can be also applied to the existing superpixel based methods to obtain further improvement.

After saliency propagation, some foreground regions may still have not been fully highlighted, or some backgrounds have not been deeply suppressed. To further resolve this problem, we first construct a sigmoid function to refine the saliency maps, which is defined as:

$$S_R = \frac{1}{2}\left(S + \frac{1}{1+\exp\left(-\Phi(S)\right)}\right) \tag{10}$$

$$\Phi_i(S) = \begin{cases} 10\cdot\left(\dfrac{S_i}{T_b} - 1\right), & S_i < T_b \\ 10\cdot\dfrac{S_i - T_b}{1 - T_b}, & S_i \geq T_b \end{cases} \tag{11}$$

where T_b is a threshold to segment S and is also computed by Otsu [37] in this paper. Then, we integrate it with S by linear summation to improve the visual quality. Finally, we smooth the integrated saliency map to produce pixel-wise result by the fast bilateral filter [43] which is an efficient edge-preserving smoothing operator and performs better than the guided filter used in [23].

3. Experimental results

We evaluate our method on five typical benchmark datasets: MSRA-10k [39], ECSSD [16], SOD [40], PASCAL-S [42], and DUT-OMRON [12]. MSRA-10k is a descendant of the MSRA dataset and contains 10,000 annotated images. ECSSD contains 1000 images, most of which are semantically meaningful and structurally complex. SOD consists of 300 images collected from the Berkeley segmentation dataset and most of which have multiple salient objects either with low contrast or overlapping with the image boundary. PASCAL-S is arguably one of the most challenging saliency dataset which is ascended from the validation set of PASCAL VOC 2010 segmentation challenge and contains 850 images with complex background. It is also believed to eliminate the design bias, e.g., center bias and color contrast bias. The last DUT-OMRON is another challenging dataset that contains 5168 images with complex background.

We compare the proposed method with fifteen newest state-of-the-art methods including GS [10], DRFI [11], MR [12], DSR [13], MC [14], SF [15], LPS [20], RRWR [24], TLLT [25], LEGS [27], BSCA [31], MCA [31], RC [39], RBD [41], and BL [45]. Note the extended version of RC and multi-level DRFI are used in our comparison, and MCA is integrated by HS [16], DSR [13], MR [12], RBD [41], and BSCA [31] as described in [31]. The results of the above methods are achieved by running the source codes or provided by authors, which can guarantee a fair comparison.

3.1. Parameters and evaluation metrics

3.1.1. Experimental setup

Superpixel number N is adaptively set as height $\times$ width/300 for all the superpixel based approaches using SLIC, $\sigma=0.8$, and $k=100$ for graph-based segmentation, $\mu=0.01$, all the parameters are unchanged in the following experiments.

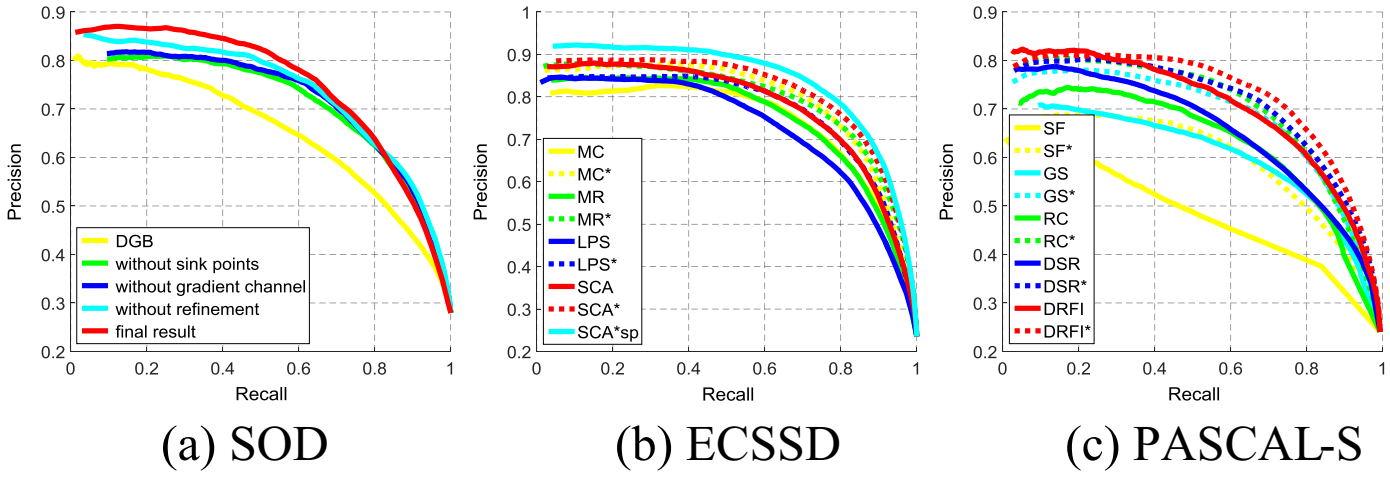


Fig. 6. PR curve results: (a) is the comparison of each individual component, (b) shows improvement of existing saliency propagation approaches by the proposed similarity matrix or sink points (*sp), (c) shows improvement of state-of-the-art methods by our saliency propagation algorithm. Note that * indicates the improved version.

3.1.2. Evaluation metrics

We evaluate the performance of all the methods using PR curve, F-measure, AUC score, and MAE. Similar as prior works, the PR curves are computed by comparing the ground truth with binary maps obtained by binarizing the saliency map using thresholds in the range of 0 and 255. The F-measure is an overall performance measurement which is defined as:

$$F_\beta = \frac{(1+\beta^2)\cdot Precision\cdot Recall}{\beta^2\cdot Precision + Recall} \tag{12}$$

where *Precision* and *Recall* are computed by an adaptive threshold which is twice the mean saliency value of the input image, and β^2 is set to 0.3 to emphasize the precision [38]. The AUC represents the area under the receiver operating characteristics curve and can effectively reflect the global properties of different methods. Higher score indicates better performance, and around 0.5 will be obtained by random guessing [46]. Thus, we show it by AUC – 0.5 in our experiment for the clarity of the figures. The MAE is the average difference between the saliency map and the ground truth in pixel level, which indicates the similarity of them and is complementary to PR curves.

3.2. Quantitative results

3.2.1. Individual component analysis

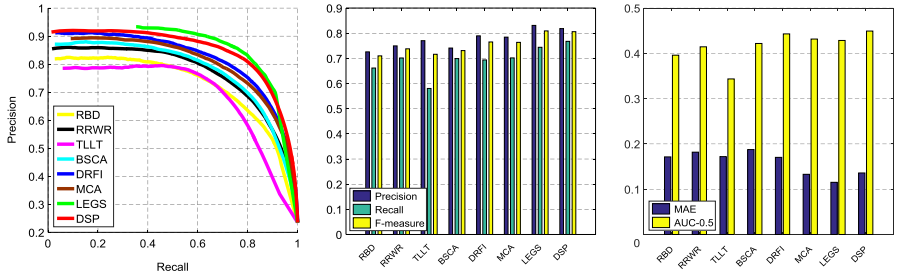
In order to evaluate the contributions of separate components in our approach, we only show evaluation results of PR curves in Fig. 6 (a) due to limited space. We can observe that performance of our method can be improved by introducing gradient magnitude channel and sink points on the challenging SOD dataset. In specific, 1.59% and 0.89% improvement can be obtained in terms of F-measure respectively. We analyze that SOD contains many complex natural images and most of them have different texture structures which can be well captured by the gradient statistic. With the help of two-level fusion enhancement, we further improve the performance while still keep the efficiency as shown in Table 2. For instance, it improves by 1.74% on SOD and 1.18% on MSRA-10k in terms of F-measure as can be seen in Fig. 7(e). After refinement, our method lowers the MAE by 2.35% on SOD which indicates the good similarity between the predicted saliency map and the ground truth. Thus, all the individual components make their contributions independently to the final performance improvement.

Table 2
Comparison of average running time (seconds per image) on ECSSD. {M=MATLAB, C=C/C++}.

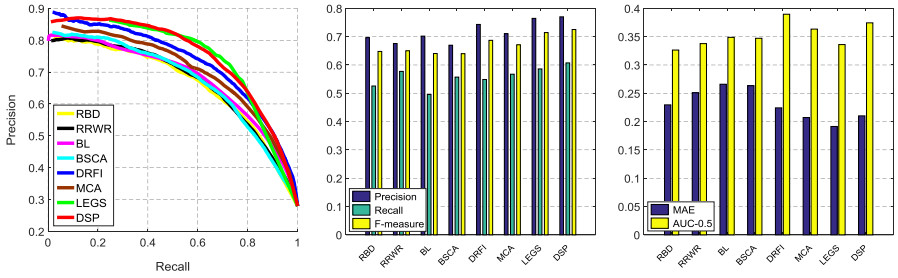
Method	Code	Time(s)
MC	M+C	0.12
MR	M	0.58
RBD	M+C	0.27
LPS	M+C	2.56
RRWR	M	1.23
BSCA	M+C	1.01
DRFIs	M+C	1.81
DRFI	M+C	7.14
DSPs	M+C	0.47
DSP	M+C	0.69

3.2.2. Comparison with state-of-the-arts

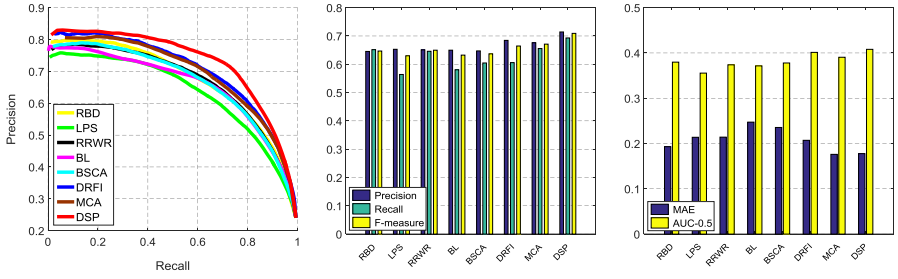
Quantitative comparison results of different saliency propagation methods are shown in Fig. 7, including RBD [41], LPS [20], TLLT [25], BSCA [31], BL [45], and RRWR [24], most of them are published very recently. Additionally, we also report the performance of DRFI [11], MCA [31], and LEGS [27] for comparison. Note that the results of some methods are not reported on some datasets because they are not provided by authors. As can be seen, the proposed DSP significantly outperforms the other saliency propagation methods throughout all the evaluation metrics and datasets, which indicates the superior overall performance of the proposed approach. Take the challenging PASCAL-S dataset for example, it improves by 2.97%, 4.14%, and 4.46% over the second best method according to the precision, recall, and F-measure values, respectively. Besides, some methods tend to detect accurate salient objects at the expense of low recall, such as TLLT on ECSSD and LPS on SOD, which cause the imbalance between precision and recall. Comparatively, our DSP produces more balanced results, thus, best F-measure is obtained on all the tested datasets. Furthermore, we observe that the proposed DSP has a competitive high precision value in the high range of recall, which indicates the strong capability to suppress the background regions contributed by the appropriate sink points. It is also noted that the proposed method is not only better than MCA which is an aggregation model, but also slightly better than the multi-level segmentation based DRFI which achieves top performance in a recent benchmark of saliency detection [46]. More surprisingly, it also gets similar performance with the deep learning based LEGS and even achieves higher AUC score as seen from Fig. 7 (a)–(b). In addition, it is interesting to find that almost all the methods get extremely larger recall values comparing with precision on the DUT-OMRON dataset, which might be caused by its noisy labeling as described in [46].



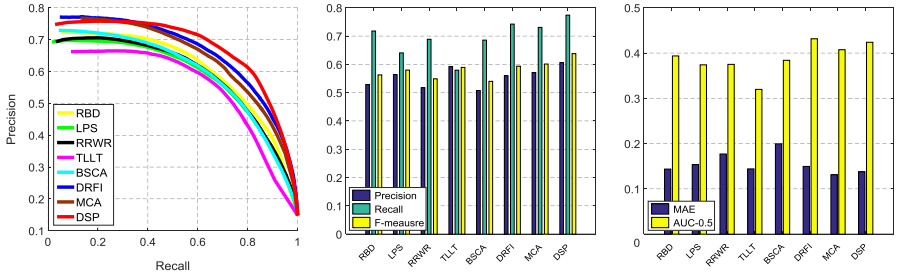
(a) ECSSD



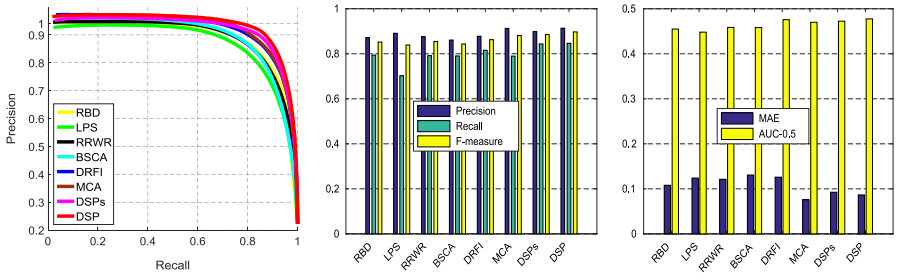
(b) SOD



(c) PASCAL-S



(d) DUT-OMRON



(e) MSRA-10k

Fig. 7. Quantitative comparison results of different saliency propagation methods on different datasets. From left to right: PR curves, F-measure, MAE and AUC values. DSPs indicates the single level version of the proposed method.

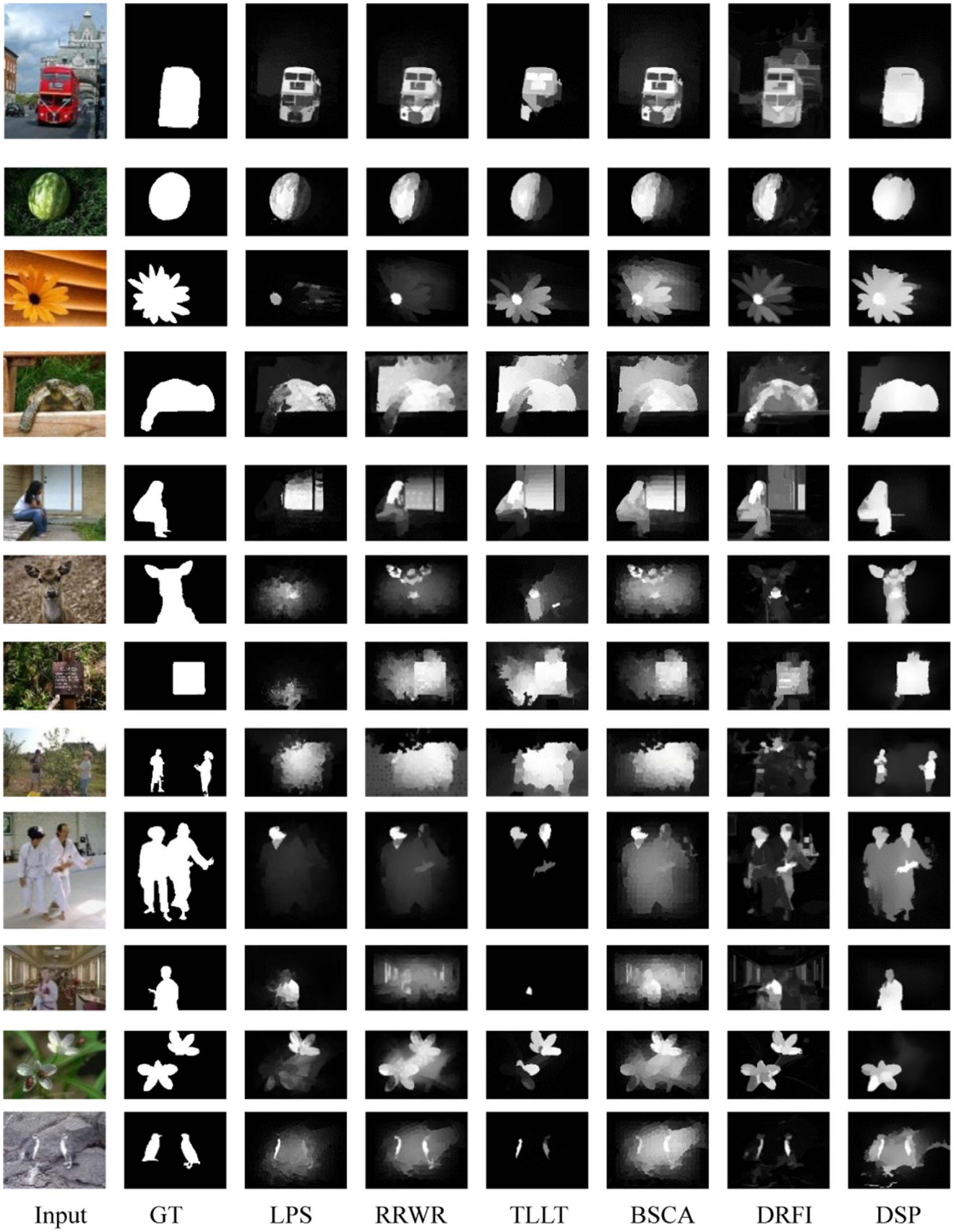


Fig. 8. Visual comparison of the saliency maps. The last two rows are failure cases where DSP powerlessly or abundantly detects the salient objects.

3.2.3. Improvement of the-state-of-arts

The proposed similarity matrix W combining multiple color spaces and gradient channel can be used to enhance the existing saliency propagation methods (including MR [12], MC [14], LPS [20], and SCA [31]) to a similar performance level. For fair comparison, we use our DGB as the input coarse saliency map for MR and SCA. PR curves are shown in Fig. 6(b), which strongly proves the effectiveness of the proposed similarity metric which can greatly improve their performance by a large margin. We further introduce our sink points to the recent proposed SCA by replacing its impact factor matrix with $D^{-1}WI_f$ to evaluate its generality. As can be seen from Fig. 6(b), the performance can be further enhanced which demonstrates the robustness of the proposed sink points. We also apply our saliency propagation to optimize state-of-the-art results which are used as the input coarse saliency maps, including RC [39], GS [10], SF [15], DSR [13], and DRFI [11]. From the results in Fig. 6(c), we can see that all of them are significantly improved to a similar level after our saliency propagation even though the input saliency maps are not satisfying.

3.3. Qualitative results

We further show some qualitative comparison results of different approaches in Fig. 8. As can be seen, the salient objects are fully highlighted by our method thus very similar to the ground truth, which indicates lower MAE value as mentioned above. Besides, we can observe that the proposed method can deal well with the challenging cases where the background is complex or very similar to the foreground. For example, in the 3rd, 6th, and 8th rows, all the other approaches are not discriminative enough to distinguish the background and foreground regions with very close color while our method can separate them successfully. We also find that our approach shows better capability in background suppression contributed by the proposed sink points. As shown in the images at 4th, 5th, and 7th rows in Fig. 8, background regions are well eliminated while salient objects are completely retained simultaneously by our method. In addition, our method produces smoother saliency map with more uniformly highlighted objects and well defined boundaries comparing with others due to the contribution of our refinements, e.g., the bus in the first row and the watermelon in the second row can be fully highlighted by our approach, and the people in the 9th row are more smooth than the others. It is also worth pointing out that the proposed method can perform well when the salient object is connected with the image boundary as seen in the 6th and 10th rows in Fig. 8, which demonstrates the effectiveness of the proposed DGB.

3.4. Running time

The running time test is carried out in a 64-bit PC equipped of an i7-4790 k 4.00 GHz CPU and 16 GB RAM. All the tested codes are provided by the authors and run unchanged in MATLAB R2015a with some C++ mex implementations (superpixel segmentation and distance measure), and only single thread is utilized. The average running time per image of different approaches on the ECSSD dataset are listed in Table 2. As can be seen, both of our single-level and two-level algorithms are much more efficient than the listed saliency propagation methods. Note that our single-level version is also faster than MR, which is contributed by our efficient DGB. Although it is slower than MC and RBD, better results can be obtained at the cost of acceptable computational time. Thus, it is a reasonable trade-off between accuracy and efficiency.

3.5. Limitation and analysis

The last two rows of Fig. 8 show some failure cases where the background and foreground is very similar or contains multiple disconnected salient objects. The first case is caused by the small difference between background and foreground which is inadequate to distinguish them by the proposed similarity metric. However, it is also challenging for the state-of-the-arts even deep learning based methods. The second case is due to the center prior used in the sink points setting which will suppress some salient object regions far away from the image center in saliency propagation. Objectness measure could be a possible solution. However, the current objectness measure methods are not accurate and efficient enough and even worse than the center prior in our experiments. It would be interesting to exploit more discriminative similarity metric or deep feature and efficient objectness measure to further enhance the current results, which are the two directions of our future research.

4. Conclusions

In this paper, we propose a simple unsupervised bottom-up saliency detection method in which both weak and strong saliency models are integrated. Based on the distribution prior, foreground noises in the image borders are significantly reduced and reliable background seeds are obtained to produce background based saliency map. By setting appreciate sink points in manifold ranking, propagation errors are reduced especially in the background regions. The proposed similarity metric is very discriminative for the coarse saliency map estimation and saliency propagation even in complex images. In addition, both of them can be applied to the existing saliency propagation models for great performance enhancement. Finally, pixel-wise and smooth saliency maps are generated simply by the sigmoid function and fast bilateral filter. Moreover, further improvement is obtained by our new two-level fusion combining the advantages of SLIC and graph-based segmentation. Experimental results not only show the superior performance of the proposed method but also a good balance between saliency detection accuracy and computational cost. Thus, we believe that it is a good choice for subsequent or real applications, such as object detection or semantic segmentation [48].

Acknowledgments

The authors would like to express their sincere thanks to the anonymous reviewers for their invaluable suggestions and comments to improve this paper. This work was supported by University Science Research Project of Jiangsu Province (15KJB510032), and “Lv Yang Jin Feng” Program of Yangzhou.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.patcog.2016.05.016.

References

[1] P. Jiang, H. Ling, J. Yu, J. Peng, Salient region detection by UFO: uniqueness, focusness and objectness, ICCV (2013) 1976–1983.
[2] A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a survey, arXiv (2014).

[3] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, CVPR (2007) 1–8.
[4] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu, Global contrast based salient region detection, CVPR (2011) 569–582.
[5] M.-M. Cheng, Z. Zhang, W.-Y. Lin, P. Torr, BING: binarized normed gradients for objectness estimation at 300 fps, CVPR (2014) 3286–3293.
[6] U. Rutishauser, D. Walther, C. Koch, P. Perona, Is bottom-up attention useful for object recognition, CVPR (2004) II-37–II-44.
[7] X. Li, Y. Li, C. Shen, A. Dick, A. Hengel, Contextual hypergraph modeling for salient object detection, ICCV (2013) 3328–3335.
[8] V. Mahadevan, N. Vasconcelos, Saliency-based discriminant tracking, CVPR (2009) 1007–1013.
[9] L. Itti, Automatic foveation for video compression using a neurobiological model of visual attention, IEEE Trans. Image Process. 13 (10) (2004) 1304–1318.
[10] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, ECCV (2012) 29–42.
[11] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, Salient object detection: a discriminative regional feature integration approach, CVPR (2013) 2083–2090.
[12] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, CVPR (2013) 3166–3173.
[13] X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, ICCV (2013) 2976–2983.
[14] B. Jiang, L. Zhang, H. Lu, C. Yang, M.-H. Yang, Saliency detection via absorbing Markov chain, ICCV (2013) 1665–1672.
[15] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, CVPR (2012) 733–740.
[16] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, CVPR (2013) 1155–1162.
[17] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, N. Crook, Efficient salient region detection with soft image abstraction, ICCV (2013) 1529–1536.
[18] T. Liu, J. Sun, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, CVPR (2007) 1–8.
[19] N. Li, J. Ye, Y. Ji, H. Ling, J. Yu, Saliency detection on light fields, CVPR (2014) 2806–2813.
[20] H. Li, H. Lu, Z. Lin, X. Shen, B. Price, Inner and inter label propagation: salient object detection in the wild, IEEE Trans. Image Process. 24 (10) (2015) 3176–3186.
[21] S. Lu, V. Mahadevan, N. Vasconcelos, Learning optimal seeds for diffusion-based salient object detection, CVPR (2014) 2790–2797.
[22] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, IEEE Trans. Image Process. 19 (12) (2010) 3232–3242.
[23] J. Sun, H. Lu, X. Liu, Saliency region detection based on Markov absorption probabilities, IEEE Trans. Image Process. 24 (5) (2015) 1639–1649.
[24] C. Li, Y. Yuan, W. Cai, Y. Xia, D.D. Feng, Robust saliency detection via regularized random walks ranking, CVPR (2015).
[25] C. Gong, D. Tao, W. Liu, S.J. Maybank, M. Fang, K. Fu, J. Yang, Saliency propagation from simple to difficult, CVPR (2015).
[26] S. Frintrop, T. Werner, G.M. Garcia, Traditional saliency reloaded: a good old model in new shape, CVPR (2015).
[27] L. Wang, H. Lu, X. Ruan, M.-H. Yang, Deep networks for saliency detection via local estimation and global search, CVPR (2015).
[28] M. Donoser, H. Bischof, Diffusion processes for retrieval revisited, CVPR (2013) 1320–1327.
[29] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vis. 59 (2) (2004) 167–181.
[30] J. Wang, H. Lu, X. Li, N. Tong, W. Liu, Saliency detection via background and foreground seed selection, Neurocomputing 152 (25) (2015) 359–368.
[31] Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, CVPR (2015).
[32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, Slic superpixels, Technical report, 2010.
[33] X.-Q. Cheng, P. Du, J. Guo, X. Zhu, Y. Chen, Ranking on data manifold with sink points, IEEE Trans. Knowl. Data Eng. 25 (1) (2013) 177–191.
[34] L. Zhu, D.A. Klein, S. Frintrop, Z. Cao, A.B. Cremers, A multisize superpixel approach for salient object detection based on multivariate normal distribution estimation, IEEE Trans. Image Process. 23 (12) (2014) 5094–5107.
[35] S. Li, H. Lu, Z. Lin, X. Shen, B. Price, Adaptive metric learning for saliency detection, IEEE Trans. Image Process. 24 (11) (2015) 3321–3331.
[36] L. Zhang, S. Zhao, W. Liu, H. Lu, Saliency detection via sparse reconstruction and joint label inference in multiple features, Neurocomputing 155 (1) (2015) 1–11.
[37] N. Otsu, A threshold selection method from gray level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.
[38] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, CVPR (2009) 1597–1604.
[39] M.-M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, S.-M. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell 37 (3) (2015) 569–582.
[40] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, ICCV (2001) 416–423.
[41] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, CVPR (2014) 2814–2821.
[42] Y. Li, X. Hou, C. Koch, J. Rehg, A. Yuille, The secrets of salient object segmentation, CVPR (2014) 280–287.
[43] S. Paris, F. Durand, A fast approximation of the bilateral filter using a signal processing approach, Int. J. Comput. Vis. 81 (1) (2009) 24–52.
[44] N. Tong, H. Lu, Y. Zhang, X. Ruan, Salient object detection via global and local cues, Pattern Recognit. 48 (10) (2015) 3258–3267.
[45] N. Tong, H. Lu, M. Yang, Salient object detection via bootstrap learning, CVPR (2015).
[46] A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5706–5722.
[47] X. Zhang, C. Xu, M. Li, R.K.F. Teng, Study of visual saliency detection via non-local anisotropic diffusion equation, Pattern Recognit. 48 (4) (2015) 1315–1327.
[48] Y. Wei, X. Liang, Y. Chen, X. Shen, M.-M. Cheng, Y. Zhao, S. Yan, STC: a simple to complex framework for weakly-supervised semantic segmentation, arXiv (2015).

Shuhan Chen is currently working as a lecturer at College of Information Engineering in Yangzhou University. Before that, he received his BS and Ph.D. degrees from Chongqing University, China, in 2009 and 2013, respectively. His research interests include object detection and image classification.

Ling Zheng now is a Master candidate in Yangzhou University, China. She received BE degree from Yangzhou University, China, in 2014. Her current research interests are object detection and visual attention.

Xuelong Hu is a Professor and discipline leader of Information and Communication Engineering in Yangzhou University. He is also an IEEE member, executive member of Chinese Society of Image and Graphics (CSIG). His current research areas include computer vision, image processing, and Internet of Things.

Ping Zhou is a senior engineer in Wanfang Electronic Technology Co., Ltd., Yangzhou, China. He is interested in parallel and distributed computing for big data.