Saliency Detection via Bi-directional Propagation

Yingyue Xu, Xiaopeng Hong, Xin Liu and Guoying Zhao

Center for Machine Vision and Signal Analysis
P.O.Box 4500, 90014, University of Oulu, Finland

Abstract

Recent saliency models rely on propagation to compute the saliency map. Previous propagation methods are single directional, where foreground propagation and background propagation are separate (*e.g.*, only foreground propagation, or background propagation after foreground propagation). Different from the previous approaches, we propose a bi-directional propagation model (BIP) for saliency detection. The BIP model propagates from the labeled foreground superpixels and the labeled background superpixels to the unlabeled ones in the same iteration. A difficulty-based rule is adopted to manipulate the prorogation sequence, which considers both the distinctness of the superpixel to its neighboring ones and its connectivity to the labeled sets. The BIP model outperforms fourteen state-of-the-art saliency models on four challenging datasets, and largely enhances the propagation efficiency compared to single directional propagation models.

Keywords: saliency detection, bi-directional, propagation

1. Introduction

Recently, saliency detection [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25], aiming at identifying salient regions on a scene with biologically plausible cues, has aroused broad attention for its applications, such as image and video segmentation [26], video compression [27], image cropping [28], human behaviour analysis [29], *etc*. Generally, existing saliency models can be categorized into two types, including top-down models and bottom-up models.

Top-down saliency models, on one hand, depend on high-level features with various semantics (face detector [5], text representation [9], *etc*.), on the other, are task-driven which require supervised learning, for instance, support vector machine [5], AdaBoost [6], CRF [7], multiple kernel learning [8, 9], and deep convolutional neural networks [10, 11, 12], *etc*.

Different from top-down models, bottom-up saliency models compute saliency maps with low-level cues and are usually learning free. Thus, a variety of saliency models have been proposed with different strategies such as coarse-to-fine saliency estimation [13, 14, 30], local or global feature extraction [15, 16, 17, 18, 19, 31, 20, 21, 22, 32], making different assumptions, for example, boundary prior assumption [22, 33, 34], *etc*.

Propagation, as a bottom-up saliency modeling methodology, has been widely employed in recent years. The input image is firstly over-segmented into superpixels and is constructed as an undirected graph, which comprises of a set of vertices of the superpixels together with a set of edges representing the similarity between adjacent vertices. Then, the initial saliency values (labeled superpixels), that are obtained by selecting propagation seeds from a coarse saliency map [35, 36, 37, 38], are spatially diffused to the whole graph within several iterations. Traditional schemes involve all the superpixels of the image into each iteration, however, Gong *et al.* [13] argued that not all the superpixels are suitable to participate in the propagation in every iteration, especially when some of them are apparently different from the labeled ones. Thus, a TLLT saliency model [13] was proposed that measures the difficulty of each unlabeled superpixel based on the knowledge of the labeled set and only propagates those "simple" ones that are easy to judge as salient or unsalient in each iteration. Such a "propagation from simple to difficult" strategy optimizes the propagation quality by manipulating the propagation sequence.

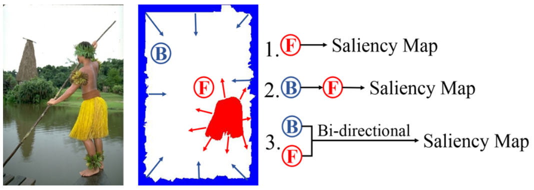


Figure 1: There are two directions in which the propagation methods spread the selected seeds to the whole image, including foreground propagation Ⓕ from the most salient seeds and background propagation Ⓑ from the most unsalient seeds. Most previous propagation methods are single directional with 1. only foreground propagation or 2. background propagation followed by foreground propagation. Different from previous methods, we propose a model that 3. propagates with both foreground seeds and background seeds in each iteration, which is bi-directional.

There are two types of propagation: foreground propagation and background propagation. Foreground propagation(Figure 1.1) selects the most salient values from a coarse

Email address: `{Yingyue.Xu, Xiaopeng.Hong, Xin.Liu, Guoying.Zhao}@oulu.fi` (Yingyue Xu, Xiaopeng Hong, Xin Liu and Guoying Zhao)

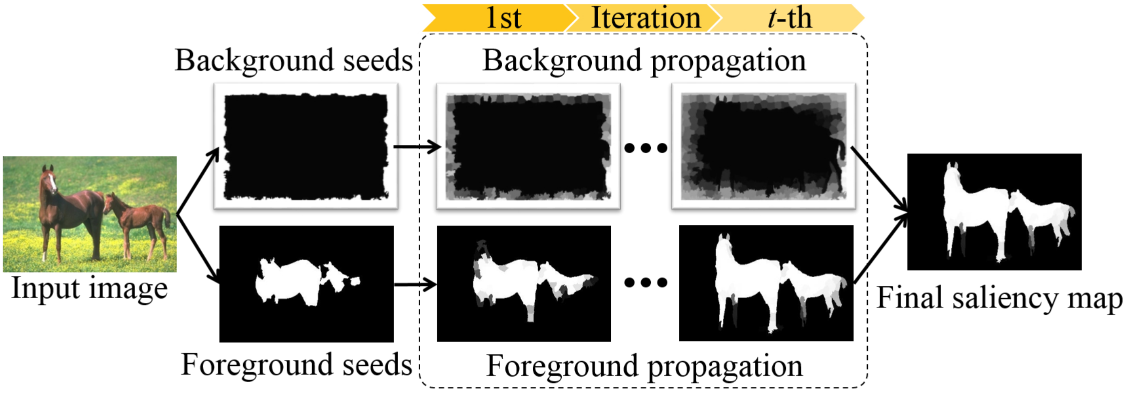


Figure 2: Framework of the proposed BIP model. Given an input image, both foreground seeds and background seeds are chosen as two initial labeled sets. In each iteration, the unlabeled superpixels are evaluated according to their difficulties to the labeled foreground set and the labeled background set respectively, only those with the lowest difficulties to each labeled set are selected and are accordingly spread to refine the foreground set or the background set. After all the unlabeled superpixels are labeled, the results from foreground propagation and background propagation are combined as the final saliency map.

saliency map as foreground seeds to propagate a saliency map. It is a direct approach to identify the salient regions on the input image, but the propagation performance is heavily influenced by the quality of seeds selection. In contrast, background propagation(Figure 1.2) selects background seeds based on boundary or background assumptions to propagate an unsaliency map. Generally, background seeds are much easier in selection than foreground ones and can better identify the unsalient regions of the image, but background propagation lacks the ability to judge the distinctness inside the salient regions.

Most previous propagation methods only select foreground seeds for propagation [35, 36, 37], but recent saliency models [13, 38] firstly compute a coarse saliency map by background propagation to obtain foreground seeds and calculate the final saliency map through foreground propagation. Obviously, the propagation is always single directional.

In this work, we propose a bi-directional propagation (BIP) model that efficiently performs foreground propagation and background propagation in one iteration(Figure 1.3). The BIP model manipulates the prorogation sequence with a difficulty-based rule. More specifically, we only choose the relatively simple superpixels instead of all for either foreground propagation or background propagation in each iteration, by measuring the difficulty of the unlabeled superpixels to the labeled foreground set and the labeled background set respectively. The framework of the proposed BIP model is illustrated in Figure 2.

The contributions of this paper are two folds:

1. We propose a bi-directional propagation model (BIP) for salient object detection. Different from previous single directional methods that perform foreground propagation and background propagation separately, the BIP model performs both foreground propagation and background propagation in every individual iteration with a difficulty-based rule.

2. We compare the proposed BIP model to fourteen state-of-the-art saliency models over four challenging datasets. Evaluation results show that the BIP model results in the best performance in both F-measure and MAE. Moreover, experimental results confirm that the BIP model largely reduces both the iteration numbers and the computational time compared to the previous single directional propagation methods.

2. Bi-directional Saliency Propagation

Superpixel algorithms group pixels in an image with similar appearance features into perceptually consistent units, and thus can efficiently reduce the computational complexity of subsequent image processing tasks. In this work, we over-segment the input image into N superpixels.

In this section, we will introduce the details of our proposed bi-directional propagation saliency model. Firstly, we depict the seeds selection for propagation including foreground seeds and background seeds. Secondly, we detail the bi-directional propagation with the difficulty-based rule.

2.1. Foreground Seeds and Background Seeds

In recent saliency detection tasks, it is widely accepted that the boundaries of an image are most likely to be the background regions. Wei *et al.* [39, 40] pointed out that the most background regions, other than salient ones, are easily connected to the image boundaries. Also, a number of saliency models [22, 41] generated a coarse saliency map with the compactness of image boundaries. Besides, several supervised saliency models [42, 14] also extracted the appearance features of boundaries for model training.

We firstly compute a coarse foreground map S_F based on the boundary prior. We assume that the more discrepant a superpixel is from the boundary ones, the more salient the superpixel is. Thus, we select the superpixels along the image boundaries as background seeds, and grouped them into K clusters by K-means algorithm. The number of superpixels belonging to the k-th cluster is denoted as N_k, $k = 1, \cdots, K$. If the n-th superpixel is still quite different from its most similar cluster, it is more likely to be salient. In this way, we compute the coarse foreground map S_F as follow:

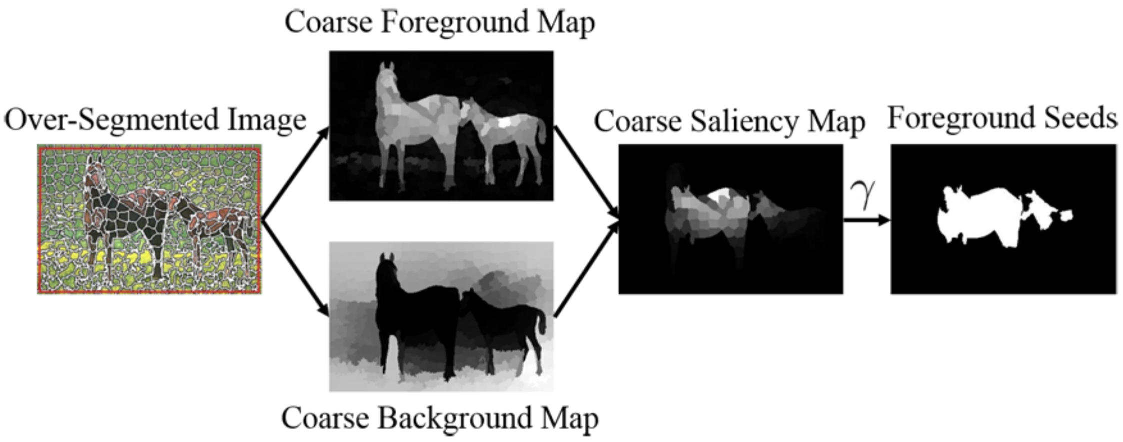


Figure 3: Foreground seeds selection. The input image is firstly over-segmented into superpixels. A coarse foreground map and a coarse background map are computed based on the boundary prior, and are integrated to obtain a coarse saliency map. The coarse saliency map is then thresholded with γ to obtain the foreground seeds.

$$S_F(n) = \min_{k\in\{1,...,K\}}(\frac{1}{N_k}\sum_{m=1}^{N_k}||\varphi_n, \varphi_m||), \tag{1}$$

where $||\varphi_n, \varphi_m||$ computes the Euclidean distance between the n-th superpixel and the m-th superpixel on CIELab features.

Still using the boundary prior, we compute a coarse background map S_B with a basic propagation method. The over-segmented image can be regarded as an undirected graph $G = (V, E)$, which comprises a set V of the superpixels together with a set E of edges representing the similarity between adjacent superpixels. The constructed graph G can be described as an adjacent matrix $W = [w_{nm}]_{N\times N}$. In this work, the similarity between two superpixels is computed as follow,

$$w_{nm} = exp(-||\mu_n, \mu_m||^2/(2\theta^2)), \tag{2}$$

where $||\mu_n, \mu_m||$ computes the Euclidean distance between superpixel μ_n and μ_m on CIELab-XY features, where $\mu_n = [\varphi_n^T; x_n; y_n]^T$, x_n and y_n are the coordinates of the n-th superpixel in X-Y space.

Again, we extract the superpixels along the image boundaries as background seeds and the propagation function is as follow:

$$S^{t+1} = I \cdot D^{-1} \cdot W \cdot S^t, \tag{3}$$

where I is the identity matrix and D is the diagonal degree matrix with $D_{nm} = \sum_m w_{nm}$, and the initial S^0 is computed based on the boundary prior as follow:

$$S^0(n) = \begin{cases} 1, & \text{the } n\text{-th superpixel is a boundary one} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

After T_1 times of iterations, the final propagated S^{T_1} is computed as the coarse background map S_B. Then we obtain a coarse saliency map S_{Coarse} as follow:

$$S_{\text{Coarse}} = \frac{S_F}{S_B + \alpha}, \tag{5}$$

where α is set as 0.001 to avoid the division-by-zero problem.

Finally, we threshold S_{Coarse} with γ to obtain the foreground seeds and the superpixels on the four boundaries of the image are background seeds.

2.2. Bi-directional Propagation

We propose a bi-directional propagation approach that spreads the labeled foreground superpixels and the labeled background superpixels to the unlabeled ones with a difficulty-based rule. At time t (the t-th iteration), the unlabeled set of superpixels are denoted as $\mathcal{U}^t$ and the labeled set as $\mathcal{L}^t$. $\mathcal{L}^t = \mathcal{L}^t_F \cup \mathcal{L}^t_B$, where $\mathcal{L}^t_F$ refers to the set of superpixels labeled by foreground propagation, while $\mathcal{L}^t_B$ is propagated by background seeds. Every superpixel on the image has two measures, one is its *saliency* value f^t_n for foreground propagation and the other is its *unsaliency* value b^t_n for background propagation. At time $t = 0$, $\mathcal{L}^0_F$ (or $\mathcal{L}^0_B$) is composed of foreground seeds (or background seeds) obtained in Section 2.1 with saliency (or unsaliency) values $f^0_n = 1$ (or $b^0_n = 1$).

At time t, according to how difficult to assign a superpixel to $\mathcal{L}^t_F$ and $\mathcal{L}^t_B$ respectively, we have three subsets $\mathcal{P}^t_F$, $\mathcal{P}^t_B$ and $\mathcal{P}^t_D$. $\mathcal{P}^t_F$ (or $\mathcal{P}^t_B$) contains the superpixels with the lowest difficulties to $\mathcal{L}^t_F$ (or $\mathcal{L}^t_B$) and will be propagated by $\mathcal{L}^t_F$ (or $\mathcal{L}^t_B$) at time t. For those superpixels belonging to $\mathcal{P}^t_D$, they are regarded as ambiguous ones which will not be involved in the t-th iteration of the propagation. The unlabeled superpixels in $\mathcal{U}^t$ are iteratively labeled under such a scheme until $\mathcal{U}^t$ is completely labeled.

We choose a set of $\mathcal{P}^t_F$ (or $\mathcal{P}^t_B$) from those superpixels C^t_F (or C^t_B) that are directly connected to the labeled set $\mathcal{L}^t_F$ (or $\mathcal{L}^t_B$) on the undirected graph G. If the n-th superpixel is under consideration, its difficulty to the labeled set $\mathcal{L}^t_F$ (or $\mathcal{L}^t_B$) is d^F_n (or d^B_n). $\mathcal{P}^t_F$ and $\mathcal{P}^t_B$ are two sets of superpixels with the lowest difficulties selected from C^t_F and C^t_B respectively.

To measure the difficulty that an unlabeled superpixel is related to a labeled set, we consider two aspects: *distinctness* to its neighborhood and *connectivity* to the labeled set. Distinctness computes the appearance difference between the unlabeled superpixel and its neighbors. If the superpixel is apparently similar to its surrounding ones, the superpixel is more likely to have similar saliency intensity as its neighbors and thus is regarded as a simple one. Connectivity measures the strength that the unlabeled superpixel is connected to the labeled set. If the superpixel is strongly connected to the labeled set, it is simple to propagate. Thus, d_n^F and d_n^B are computed as follows:

$$
\begin{aligned}
d_n^F &= \frac{1}{|\mathcal{N}(\varphi_n)|} \sum_{m \in \mathcal{N}(\varphi_n)} \|\varphi_n, \varphi_m\| + \frac{1}{|\mathcal{L}_F^t|} \sum_{m \in \mathcal{L}_F^t} \mathcal{G}(\mu_n, \mu_m), \\
d_n^B &= \frac{1}{|\mathcal{N}(\varphi_n)|} \sum_{m \in \mathcal{N}(\varphi_n)} \|\varphi_n, \varphi_m\| + \frac{1}{|\mathcal{L}_B^t|} \sum_{m \in \mathcal{L}_B^t} \mathcal{G}(\mu_n, \mu_m)
\end{aligned} \tag{6}
$$

where $\mathcal{N}(\varphi_n)$ contains all the neighboring superpixels of φ_n, $|\mathcal{N}(\varphi_n)|$ is the number of φ_n's neighbors, $\|\varphi_n, \varphi_m\|$ computes the Euclidean distance between superpixel φ_n and φ_m, $|\mathcal{L}_F^t|$ and $|\mathcal{L}_B^t|$ are the numbers of superpixels in $\mathcal{L}_F^t$ and $\mathcal{L}_B^t$ respectively. $\mathcal{G}(\mu_n, \mu_m)$ computes the geodesic distance between μ_n and μ_m as follows:

$$
\mathcal{G}(\mu_n, \mu_m) = \min_{v_1 = n, v_2, \ldots, v_r = m} \sum_{k=1}^{r-1} \max(\|v_k, v_{k+1}\| - a, 0) \tag{7}
$$

s.t. $v_k, v_{k+1} \in V$, v_k and v_{k+1} are connected in the undirected graph G, $\|v_k, v_{k+1}\|$ computes the Euclidean distance between v_k and v_{k+1}, and a is an adaptive threshold preventing the "small-weight-accumulation" problem [39, 13]. Thus, the $\mathcal{G}(\mu_n, \mu_m)$ measures the shortest path (geodesic) between μ_n and μ_m in G.

After computing the difficulty of all the candidate superpixels in C_F^t and C_B^t to their corresponding labeled sets $\mathcal{L}_F^t$ and $\mathcal{L}_B^t$, we pick up two sets of superpixels, $\mathcal{P}_F^t$ for foreground propagation and $\mathcal{P}_B^t$ for background propagation. The difficulty scores d_n^F of the superpixels in C_F^t are sorted in ascending order, and the first q_F superpixels are selected to $\mathcal{P}_F^t$. q_F at time t is computed by

$$
q_F^t = \lceil |C_F^t| \times \delta_F^t \rceil \tag{8}
$$

where δ_F^t is computed as follow:

$$
\delta_F^t = 1 - \frac{2}{q_F^t} \sum_{n=1}^{q_F^{(t-1)}} \min(f_n^{(t-1)}, 1 - f_n^{(t-1)}) \tag{9}
$$

δ_F^t is learned from the labeled set $\mathcal{L}_F^{t-1}$ at time $t-1$, which determines the percentage of superpixels we will select from C_F^t. δ_F^t is high if the saliency values of the superpixels in $\mathcal{L}_F^{t-1}$ are close to 1 (foreground) or 0 (background). When the saliency values are close to 0.5, it becomes ambiguous to judge whether the values are salient or not. Thus, δ_F^t is set small to avoid choosing ambiguous superpixels from C_F^t.

In a similar way, a set of superpixels $\mathcal{P}_B^t$ for background propagation can be selected by ranking the difficulty scores d_n^B of C_B^t in ascending order and choose the first q_B superpixels.

$$
\begin{aligned}
q_B^t &= \lceil |C_B^t| \times \delta_B^t \rceil, \\
\delta_B^t &= 1 - \frac{2}{q_B^t} \sum_{n=1}^{q_B^{(t-1)}} \min(b_n^{(t-1)}, 1 - b_n^{(t-1)})
\end{aligned} \tag{10}
$$

There is a special case when the n-th superpixel belongs to a set $\mathcal{P}^t$, where $p_n^t \in \mathcal{P}^t = \mathcal{P}_F^t \cap \mathcal{P}_B^t \neq \emptyset$. In such case, we classify the n-th superpixel by comparing its difficulty scores d_n^F and d_n^B. If $d_n^F \geq d_n^B$, $\mathcal{P}_F^t = \mathcal{P}_F^t \setminus p_n^t$; otherwise, $\mathcal{P}_B^t = \mathcal{P}_B^t \setminus p_n^t$.

As the superpixels for foreground propagation $\mathcal{P}_F^t$ and background propagation $\mathcal{P}_B^t$ are both determined, we need to spread the saliency values in $\mathcal{L}_F^t$ to $\mathcal{P}_F^t$ and the unsaliency values in $\mathcal{L}_B^t$ to $\mathcal{P}_B^t$ respectively by

$$
\begin{aligned}
f^{t+1} &= A_F^t \cdot D^{-1} \cdot W \cdot f^t, \\
b^{t+1} &= A_B^t \cdot D^{-1} \cdot W \cdot b^t.
\end{aligned} \tag{11}
$$

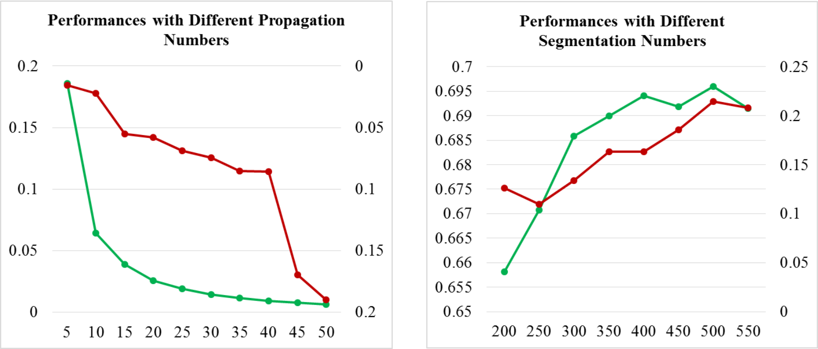
where A_F^t is a diagonal matrix with $A_{nn}^t = 1$ if the n-th superpixel belongs to $\mathcal{L}_F^t \cup \mathcal{P}_F^t$, otherwise $A_{nn}^t = 0$. Similarly, A_B^t is diagonal with $A_{nn}^t = 1$ if the n-th superpixel belongs to $\mathcal{L}_B^t \cup \mathcal{P}_B^t$, otherwise $A_{nn}^t = 0$. After the t-th iteration, the labeled set is $\mathcal{L}^{t+1} = \mathcal{L}^t \cup \mathcal{P}_F^t \cup \mathcal{P}_B^t$, and the unlabeled set is $\mathcal{U}^{t+1} = \mathcal{U}^t \setminus (\mathcal{P}_F^t \cup \mathcal{P}_B^t)$.

When all the unlabeled superpixels are labeled after T_2 times iterations, we have a set of superpixels involved in foreground propagation $\mathcal{F}$ and a set of superpixels involved in background propagation $\mathcal{B}$. If the n-th superpixel is involved in foreground propagation, its saliency value is now $f_n^{T_2}$; otherwise, its unsaliency value is $b_n^{T_2}$. Then, we transfer the unsaliency values $b_n^{T_2}$ into saliency values based on $\mathcal{F}$ by $\bar{b}_n^{T_2} = \min(\mathcal{F}) \times (1 - b_n^{T_2})$. Finally, we map the saliency values in $\mathcal{F}$ and the transferred saliency values in $\mathcal{B}$ to the original image to obtain the final saliency map.

3. Experiments

We investigate the bi-directional propagation model (BIP) by evaluating it over four challenging datasets: DUT-OMRON [38], ECSSD [43], PASCAL-S [44], and ASD [45]. The ASD dataset is one of the most widely used datasets with 1000 images from the MSRA-5000 Saliency Object Database[46], with distinct salient objects on the scenes. PASCAL-S is a dataset of 850 images from PASCAL VOC 2010 [47] with multiple salient objects on the scenes. The ECSSD dataset contains 1000 images with complex salient objects on the scenes, and the objects on the images are semantically meaningful. The DUT-OMRON dataset contains a large number of 5168 more difficult and challenging images.

We compare the proposed BIP model with fourteen state-of-the-art saliency models including BSCA [22], COV [20], DRFI [42], GBVS [16], GC [48], GP [49], HS [43], LR [50], MB [51], MR [38], PCAS [21], RB [39], TLLT [13], and UFO [42]. The implementations of the chosen approaches are directly from the corresponding authors.



(a) (b)

Figure 4: (a) The mean absolute difference between the current propagated result S^t and the previous propagated result S^{t-5} (green line) and the computational time (red line). The x-axis enumerates the iteration numbers t. The major y-axis shows the mean absolute difference, while the second y-axis is the computational time (seconds). The figures are average values evaluated on ECSSD dataset. θ=0.25. (b) The F-measure (green line) and the corresponding computational time (red line) with different settings of N, with T_1=20, θ=0.2, γ=0.02 and K=4. The x-axis enumerates the value of N for over-segmentation. The major y-axis represents the F-measure, while the second y-axis is the computational time (seconds). The figures are average values evaluated on ECSSD dataset.

3.1. Parameters

In this section, we discuss the tuning of the parameters of the proposed BIP model, including N, K, T_1, θ and γ.

The given image is over-segmented into N superpixels for the bi-directional propagation. Then, foreground seeds and background seeds are selected.

The foreground seeds are obtained from the coarse saliency map S_{Coarse}. Firstly, the BIP model computes a coarse foreground map S_F. The superpixels along the image boundaries are grouped into K clusters to perform a similarity measurement as shown in Eq. 1 to obtain S_F. Secondly, a coarse background map S_B is computed with a propagation method of T_1 iterations that involves a similarity function as Eq. 2, where the parameter θ is involved. Thirdly, the coarse saliency map S_{Coarse} is computed based on S_F and S_B as Eq. 5. Lastly, we threshold S_{Coarse} with γ to obtain the foreground seeds.

The background seeds are superpixels along the image boundaries. Finally, bi-directional propagation are performed the compute the final saliency map.

To explore the optimal settings of the parameters, we firstly tune T_1 as it only determines that with how many iterations the propagation results can be stable. Then, we regard θ and γ as pairwise parameters to explore an optimal combination based on the performances. Lastly, we tune the parameters K and N to approach a best performance for the BIP model.

3.1.1. T_1

In Section 2.1, we extract the superpixels along the image boundaries as background seeds and propagate T_1 times for the coarse background map S_B. We choose the parameter T_1 by considering two factors. Firstly, T_1 should be adequate to propagate a relatively stable coarse background map. Secondly, as the propagation time will increase in accordance with T_1, we need to control T_1 for computational efficiency.

We evaluate the propagation status by calculating the mean absolute difference between the current propagated result S^t and the previous propagated result S^{t-5}, as well as the recorded computational time for propagating S^t, as shown in Figure 4-(a). After 15 iterations, the mean absolute difference falls below 0.05 and the propagation results become relatively stable. The propagation time increases dramatically after 40 iterations. Further, Figure 5 shows the propagated results with different settings of T_1. The propagated results become relatively stable based on the intensity maps after 15 iterations.

In practice, we set T_1 as 20 to balance the propagation quality and the computational efficiency.

3.1.2. θ and γ

While computing the coarse background map S_B in Section 2.1, the parameter θ is involved in measuring the similarity between two superpixels. Then we obtain a coarse saliency map S_{Coarse} using the coarse background map S_B and the coarse foreground map S_F as in Eq. 5. Finally, we threshold S_{Coarse} with γ to obtain the foreground seeds.

Since different settings of θ produce the S_B of varying qualities, the resulted S_{Coarse} can be different. Accordingly, the setting of the threshold γ should be adjusted based on S_{Coarse} of different qualities. To obtain the optimal final result, we regard θ and γ as pairwise parameters and evaluate the final saliency map with different combinations of θ and γ as is shown in Table 1. From experimental results, it can be perceived that θ=0.2 and γ=0.02 is the optimal combination that produces the best result.

3.1.3. K and N

When computing the coarse foreground map S_F, we over-segment the image into N superpixels and group the superpixels along the image boundary into K clusters by K-means algorithm.

We firstly tune the value of K by setting it as 2, 3, 4, 5 and evaluate their corresponding performances as in Table 2. It can be perceived that when K is set as 2, 3 and 4, the evaluation results are with similarly good performances. In this work, we set K as 4 which receives the best performance.

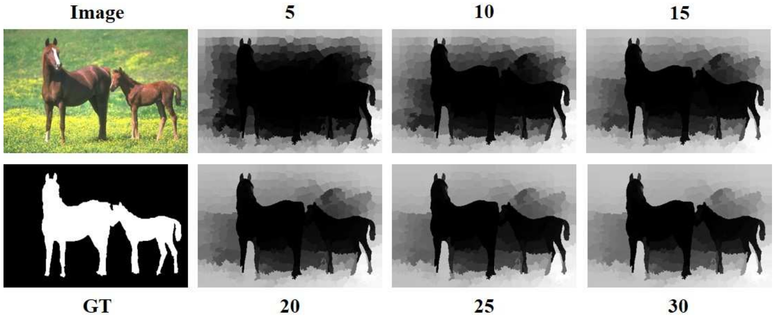


Figure 5: Examples of computed coarse background maps with different numbers of iterations with θ=0.25.

γ θ	0.005	0.01	0.02	0.05	0.1	0.15	0.2	0.25	0.3
0.05	0.401	0.4200	0.450	0.514	0.568	0.588	0.589	0.578	0.558
0.1	0.511	0.535	0.561	0.602	0.624	0.625	0.619	0.600	0.578
0.15	0.611	0.638	0.660	0.677	0.671	0.653	0.632	0.607	0.583
0.2	0.670	0.689	**0.692**	0.676	0.637	0.606	0.581	0.556	0.533
0.25	0.680	0.683	0.662	0.609	0.559	0.523	0.495	0.474	0.455
0.3	0.664	0.647	0.612	0.545	0.485	0.456	0.430	0.412	0.393
0.35	0.640	0.612	0.567	0.496	0.443	0.415	0.396	0.380	0.366
0.4	0.616	0.583	0.536	0.469	0.421	0.394	0.371	0.355	0.348

Table 1: The F-measure of the final saliency maps on ECSSD dataset with different combinations of θ and γ. T_1=25, N=400 and K=3. The five best results are underlined and the best result is in bold.

Lastly, we evaluate the performances of BIP model with different settings of N. Figure 4-(b) shows that the performances are relatively high when N is set in the range of [400, 550]. Thus, we recommend to set the over-segmented number N larger than 400. Taking the running time into consideration, we set N as 400 for rational computation.

3.2. Experimental Performance

We over-segment the images into $N = 400$ superpixels with the simple linear iterative clustering (SLIC) algorithm [52]. In practice, K is set as 4, T_1 is 20, θ is 0.2 and γ is 0.02. We employ two types of evaluation metrics to evaluate the performance of saliency maps: F-measure and mean absolute error (MAE). When a given saliency map is slidingly thresholded from 0 to 255, a precision-recall (PR) curve can be computed based on the ground truth. F-measure is computed to count for the saliency maps with both high precision and recall:

$$F = \frac{\left(1+\beta^2\right)\cdot precision \cdot recall}{\beta^2 \cdot precision + recall}, \tag{12}$$

where $\beta^2 = 0.3$ [45] to emphasize the precision.

MAE measures the overall pixel-wise difference between the saliency map *sal* and the ground truth *gt*:

$$MAE = \frac{1}{H}\sum_{h=1}^{H} |sal(h) - gt(h)|, \tag{13}$$

where H is the number of pixels on the map.

Figure 7 further plots four bar charts about the performance enhancement in F-measure and MAE by comparing the BIP model to every selected saliency model. Obviously, the BIP model outperforms every selected saliency model in both F-measure and MAE over all the four datasets. Figure 8 illustrates some examples of the fourteen state-of-the-art saliency models and the proposed BIP model. In the last two columns of Figure 8, we present two examples when parts or all of the salient objects are similar to the background in appearance. In such extreme cases, the BIP model may lose detection to the salient parts that are alike the background. However, it still keeps the shapes of the salient parts and eliminates the unsalient parts of the image more effectively than the other saliency models.

We compare the F-measure and MAE scores of the proposed BIP model to that of the fourteen state-of-the-art saliency models. Table 3 4 lists the average F-measure and MAE scores of the proposed BIP model as well as the fourteen saliency models over four datasets including ECSSD, ASD, DUT-OMRON and PASCAL-S datasets. It can be perceived that the proposed BIP model results in the best performance compared to all the selected saliency models over the four datasets in both F-measure and MAE score. More specifically, the proposed BIP model increases the highest F-measure of the fourteen saliency models by 7.5%, 4.2%, 4.2% and 2.4% on ECSSD, ASD, DUT-OMRON and PASCAL-S datasets respectively, and reduces the lowest MAE scores of the fourteen saliency models by 3.3%, 1.4%, 1.4% and 0.6% on ECSSD, ASD, DUT-OMRON and PASCAL-S datasets respectively.

In addition, we plot four bar charts of the average preci-

K	2	3	4	5
F-measure	0.637	0.692	0.693	0.691

Table 2: The F-measure of the final saliency maps with different settins of K. The evaluation results are on ECSSD dataset with T_1=20, θ=0.2, γ=0.02 and N=400.

Model	ECSSD	ASD	OMRON	PASCAL
GP	0.619	0.819	0.471	0.580
MB	0.616	0.785	0.501	0.579
BSCA	0.603	0.796	0.479	0.550
DRFI	0.589	0.773	0.507	0.510
MR	0.571	0.803	0.480	0.516
HS	0.567	0.770	0.473	0.533
RB	0.562	0.809	0.501	0.528
UFO	0.481	0.710	0.397	0.431
GC	0.484	0.729	0.403	0.460
PCAS	0.425	0.590	0.382	0.400
LR	0.419	0.566	0.349	0.385
GBVS	0.399	0.432	0.321	0.391
COV	0.355	0.346	0.256	0.315
TLLT	0.583	0.828	0.507	0.483
BIP	**0.694**	**0.870**	**0.549**	**0.604**

Table 3: F-measure scores of fourteen state-of-the-art saliency models and the proposed BIP model on ASD, ECSSD, DUT-OMRON and PASCAL-S datasets. The best results for each measurement over different datasets are in bold, while the second best ones are underlined.

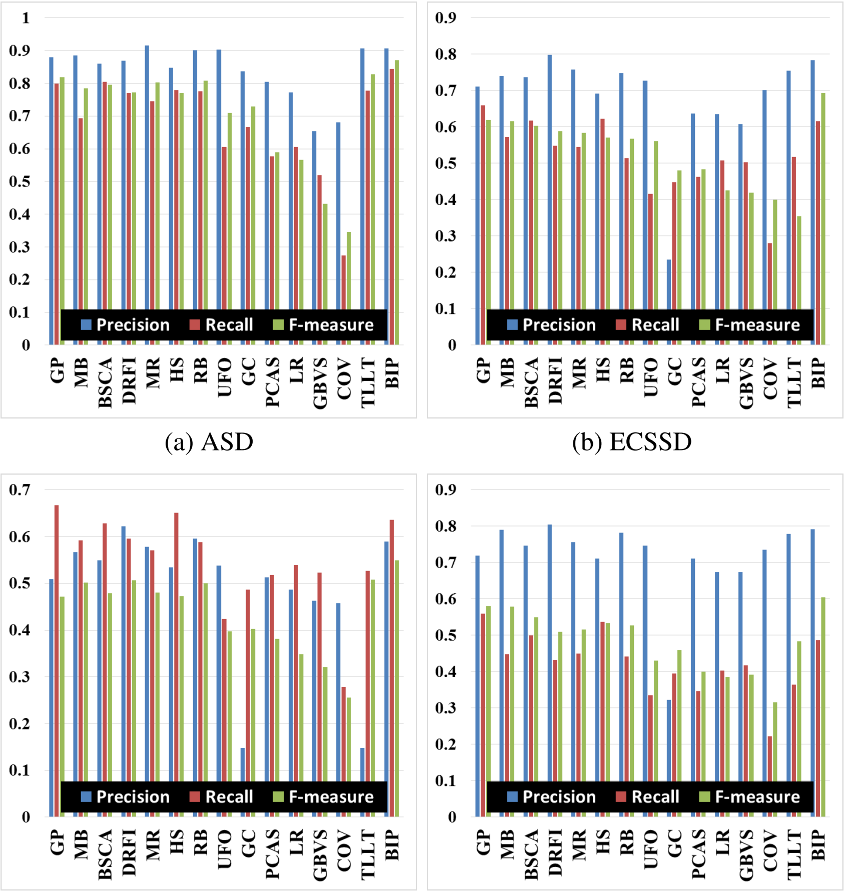
Model	ECSSD	ASD	OMRON	PASCAL
GP	0.191	0.083	0.209	0.223
MB	0.174	0.091	0.157	0.196
BSCA	0.183	0.086	0.191	0.214
DRFI	0.170	0.091	0.150	0.201
MR	0.186	0.076	0.187	0.224
HS	0.228	0.111	0.227	0.251
RB	0.171	0.066	0.144	0.197
UFO	0.203	0.111	0.170	0.226
GC	0.235	0.111	0.170	0.264
PCAS	0.247	0.156	0.207	0.240
LR	0.274	0.189	0.260	0.274
GBVS	0.263	0.215	0.240	0.261
COV	0.220	0.189	0.175	0.236
TLLT	0.172	0.064	0.144	0.209
BIP	**0.137**	**0.050**	**0.130**	**0.190**

Table 4: MAE scores of fourteen state-of-the-art saliency models and the proposed BIP model on ASD, ECSSD, DUT-OMRON and PASCAL-S datasets. The best results for each measurement over different datasets are in bold, while the second best ones are underlined.

sion, recall and F-measure of each saliency model over ECSSD, ASD, DUT-OMRON and PASCAL-S datasets respectively, as shown in Figure 6. From the four bar charts, it is obvious that our BIP model maintains both high precision and recall compared to the other saliency models, and thus results in best performance in F-measure.

3.3. Propagation Efficiency

We use the foreground seeds and background seeds obtained by BIP model and perform difficulty-based propagation in bi-directional way and single directional way respectively on ECSSD dataset. For single directional propagation, the average iteration number of the foreground propagation is 6 while the average iteration number of the background propagation is 7. However, bi-directional propagation only needs averagely T_2=4 iterations, which is much fewer than single directional method either for foreground propagation or background propagation. Moreover, the average computational time of bi-directional propagation is 0.023s per image, while that of the single directional propagation is 0.030s (foreground propagation) and



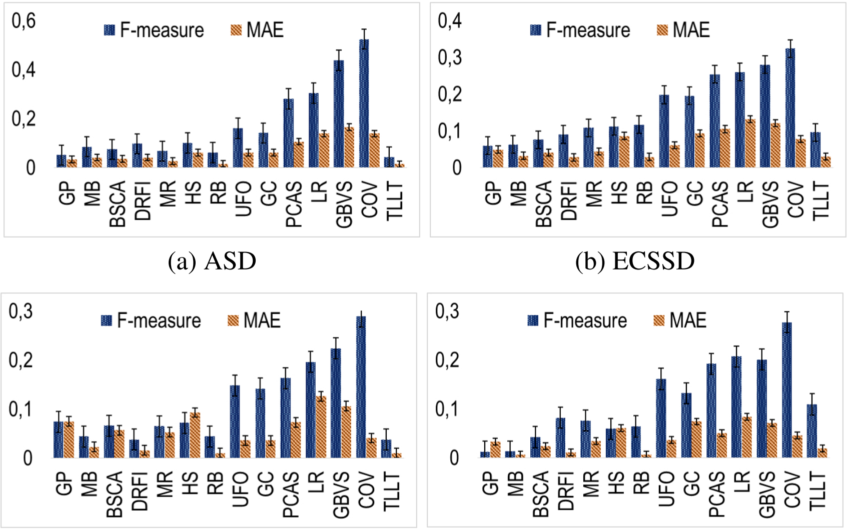
(a) ASD

(b) ECSSD

(c) DUT-OMRON

(d) PASCAL

Figure 6: The average precision, recall, and F-measure of the fourteen state-of-the-art saliency models and the proposed BIP model on ASD, ECSSD, DUT-OMRON and PASCAL-S datasets.



(a) ASD

(b) ECSSD

(c) DUT-OMRON

(d) PASCAL-S

Figure 7: Performance enhancement by comparing the proposed BIP model to the fourteen state-of-the-art saliency models (increase in F-measure and decrease in MAE) on ASD, ECSSD, DUT-OMRON and PASCAL-S datasets.

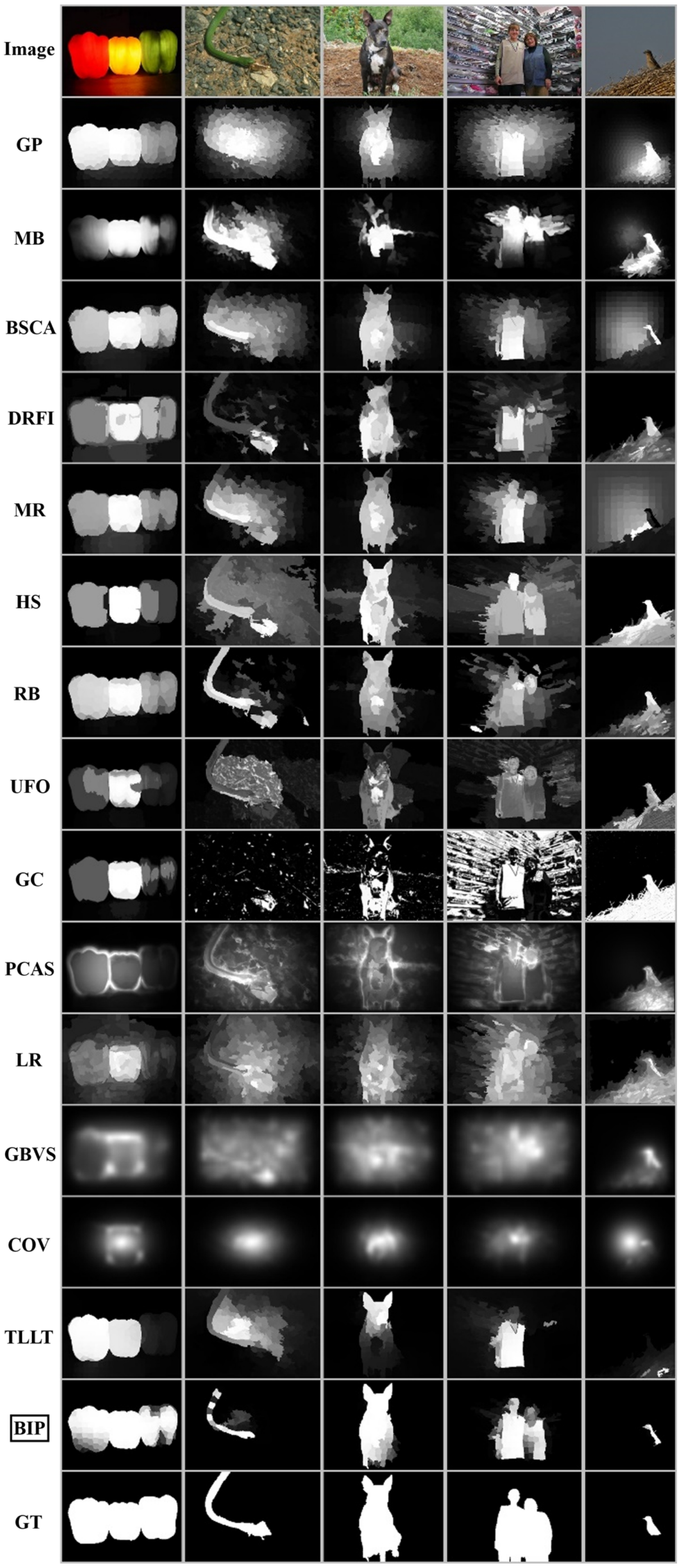


Figure 8: Examples of the results of the fourteen state-of-the-art saliency models and the proposed BIP model. The original images, results of fourteen saliency models and BIP model, and ground truth (GT) are sequentially presented.

0.039s (background propagation) for each image. Thus, bidirectional propagation outperforms single directional propagation in both iteration numbers and computational time. Therefore, the BIP model achieves high propagation efficiency.

4. Conclusion

We propose a bi-directional propagation model that performs both foreground propagation and background propagation in every individual iteration with a difficulty-based rule. The difficulty-based rule evaluates the difficulties of each unlabeled superpixel to the labeled foreground set and the labeled background set respectively by its distinctness to the neighborhood and its connectivity to the two unlabeled sets accordingly. The proposed model outperforms fourteen the state-of-the-art saliency models on four challenging datasets, and largely enhances the propagation efficiency compared to single directional propagation models.

Acknowledgments

This work was supported by Academy of Finland, Tekes Fidipro program (Grant No. 1849/31/2015) and Tekes project (Grant No. 3116/31/2017), Infotech, National Natural Science Foundation of China (No. 61772419). The authors wish to acknowledge CSC IT Center for Science, Finland, for generous computational resources. Xiaopeng Hong is partly supported by the Natural Science Foundation of China under the contract No.61572205. Xin Liu is partly supported by the Natural Science Foundation of China under the contract No.61601362. Besides, we appreciate all the code and results from the corresponding authors, especially the code of TLLT model from Mr. Chen Gong and the code of BSCA model from Prof. Huchuan Lu.

References

[1] H. Tang, C. Chen, X. Pei, Visual saliency detection via sparse residual and outlier detection, IEEE Signal Processing Letters 23 (12) (2016) 1736–1740.

[2] C. Tang, J. Wu, C. Zhang, P. Wang, W. Li, Salient object detection via weighted low rank matrix recovery, IEEE Signal Processing Letters.

[3] Z. Liu, W. Zou, L. Li, L. Shen, O. Le Meur, Co-saliency detection based on hierarchical segmentation, IEEE Signal Processing Letters 21 (1) (2014) 88–92.

[4] J. Li, Y. Tian, L. Duan, T. Huang, Estimating visual saliency through single image optimization, IEEE Signal Processing Letters 20 (9) (2013) 845–848.

[5] T. Judd, K. Ehinger, F. Durand, A. Torralba, Learning to predict where humans look, in: IEEE Proc. ICCV, 2009, pp. 2106–2113.

[6] N. Tong, H. Lu, X. Ruan, M.-H. Yang, Salient object detection via bootstrap learning, in: IEEE Proc. CVPR, 2015, pp. 1884–1892.

[7] J. Yang, M.-H. Yang, Top-down visual saliency via joint crf and dictionary learning, in: IEEE Proc. CVPR, 2012, pp. 2296–2303.

[8] M. Jiang, J. Xu, Q. Zhao, Saliency in crowd, in: Proc. ECCV, Springer, 2014, pp. 17–32.

[9] C. Shen, Q. Zhao, Webpage saliency, in: Proc. ECCV, 2014, pp. 33–46.

[10] N. Liu, J. Han, D. Zhang, S. Wen, T. Liu, Predicting eye fixations using convolutional neural networks, in: IEEE Proc. CVPR, 2015, pp. 362–370.

[11] R. Zhao, W. Ouyang, H. Li, X. Wang, Saliency detection by multi-context deep learning, in: IEEE Proc. CVPR, 2015, pp. 1265–1274.

[12] D. Zhang, J. Han, C. Li, J. Wang, Co-saliency detection via looking deep and wide, in: IEEE Proc. CVPR, 2015, pp. 2994–3002.

[13] C. Gong, D. Tao, W. Liu, S. J. Maybank, M. Fang, K. Fu, J. Yang, Saliency propagation from simple to difficult, in: IEEE Proc. CVPR, 2015, pp. 2531–2539.

[14] L. Wang, H. Lu, X. Ruan, M.-H. Yang, Deep networks for saliency detection via local estimation and global search, in: IEEE Proc. CVPR, 2015, pp. 3183–3192.

[15] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, TPAMI 20 (11) (1998) 1254–1259.

[16] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: NIPS, 2006, pp. 545–552.

[17] D. Walther, C. Koch, Modeling attention to salient proto-objects, Neural Networks 19 (9) (2006) 1395–1407.

[18] N. Bruce, J. Tsotsos, Saliency based on information maximization, in: NIPS, 2005, pp. 155–162.

[19] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in: IEEE Proc. CVPR, 2007, pp. 1–8.

[20] E. Erdem, A. Erdem, Visual saliency estimation by nonlinearly integrating features using region covariances, Journal of Vision 13 (4) (2013) 11.

[21] R. Margolin, A. Tal, L. Zelnik-Manor, What makes a patch distinct?, in: IEEE Proc. CVPR, 2013, pp. 1139–1146.

[22] Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, in: IEEE Proc. CVPR, 2015, pp. 110–119.

[23] G. Li, Y. Yu, Deep contrast learning for salient object detection, IEEE Proc. CVPR.

[24] A. Volokitin, M. Gygli, X. Boix, Predicting when saliency maps are accurate and eye fixations consistent, IEEE Proc. CVPR.

[25] L. Wang, L. Wang, H. Lu, P. Zhang, X. Ruan, Saliency detection with recurrent fully convolutional networks, in: Proc. ECCV, Springer, 2016, pp. 825–841.

[26] E. Rahtu, J. Kannala, M. Salo, Heikkil Segmenting salient objects from images and videos.

[27] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, TIP 19 (1) (2010) 185–198.

[28] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, M. Cohen, Gaze-based interaction for semi-automatic photo cropping, in: Proc. CHI, ACM, 2006, pp. 771–780.

[29] Y. Xu, X. Hong, Q. He, G. Zhao, M. Pietikinen, A task-driven eye tracking dataset for visual attention analysis, in: Proc. ACIVS, 2015, pp. 637–648.

[30] Q. Liu, X. Hong, B. Zou, J. Chen, Z. Chen, G. Zhao, Hierarchical contour closure based holistic salient object detection, TIP.

[31] Z. Chen, H. Wang, L. Zhang, Y. Yan, H.-Y. M. Liao, Visual saliency detection based on homology similarity and an experimental evaluation, Journal of Visual Communication and Image Representation 40 (2016) 251–264.

[32] L. Xu, L. Zeng, H. Duan, An effective vector model for global-contrast-based saliency detection, JVCI 30 (2015) 64–74.

[33] A. Hati, S. Chaudhuri, R. Velmurugan, An image texture insensitive method for saliency detection, Journal of Visual Communication and Image Representation 43 (2017) 212–226.

[34] B. Zou, Q. Liu, Z. Chen, H. Fu, C. Zhu, Surroundedness based multiscale saliency detection, Journal of Visual Communication and Image Representation 33 (2015) 378–388.

[35] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, TIP 19 (12) (2010) 3232–3242.

[36] B. Jiang, L. Zhang, H. Lu, C. Yang, M.-H. Yang, Saliency detection via absorbing markov chain, in: Proc. ICCV, 2013, pp. 1665–1672.

[37] Z. Ren, Y. Hu, L.-T. Chia, D. Rajan, Improved saliency detection based on superpixel clustering and saliency propagation, in: Proc. Multimedia, ACM, 2010, pp. 1099–1102.

[38] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: IEEE Proc. CVPR, 2013, pp. 3166–3173.

[39] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: Proc. ECCV, Springer, 2012, pp. 29–42.

[40] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: IEEE Proc. CVPR, 2014, pp. 2814–2821.

[41] X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, in: IEEE Proc. ICCV, 2013, pp. 2976–2983.

[42] P. Jiang, H. Ling, J. Yu, J. Peng, Salient region detection by ufo: Uniqueness, focusness and objectness, in: IEEE Proc. ICCV, 2013, pp. 1976–1983.

[43] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: IEEE Proc. CVPR, 2013, pp. 1155–1162.

[44] Y. Li, X. Hou, C. Koch, J. M. Rehg, A. L. Yuille, The secrets of salient

object segmentation, in: IEEE Proc. CVPR, 2014, pp. 280–287.

[45] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: IEEE Proc. CVPR, 2009, pp. 1597–1604.

[46] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, TPAMI 33 (2) (2011) 353–367.

[47] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results, http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html.

[48] M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, S. Hu, Global contrast based salient region detection, TPAMI 37 (3) (2015) 569–582.

[49] P. Jiang, N. Vasconcelos, J. Peng, Generic promotion of diffusion-based salient object detection, in: IEEE Proc. ICCV, 2015, pp. 217–225.

[50] X. Shen, Y. Wu, A unified approach to salient object detection via low rank matrix recovery, in: IEEE Proc. CVPR, 2012, pp. 853–860.

[51] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, R. Mech, Minimum barrier salient object detection at 80 fps, in: Proc. ICCV, 2015, pp. 1404–1412.

[52] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, Slic superpixels compared to state-of-the-art superpixel methods, TPAMI 34 (11) (2012) 2274–2282.