Background Prior-based Salient Object Detection via Adaptive Figure-Ground Classification

Jingbo Zhou[1*], Jiyou Zhai[1,2], Yongfeng Ren[1] and Ali Lu[1]
[1]School of Computer Engineering, Nanjing Institute of Technology, Nanjing, 211167, P. R. China
[e-mail: jbzhou2013@aliyun.com]
[2]College of Computer and Information, Hohai University, Nanjing, 211100, P. R. China
*Corresponding author: Jingbo Zhou

Received November 21, 2016; revised July 24, 2017; revised October 21, 2017; accepted November 23, 2017; published March 31, 2018

Abstract

In this paper, a novel background prior-based salient object detection framework is proposed to deal with images those are more complicated. We take the superpixels located in four borders into consideration and exploit a mechanism based on image boundary information to remove the foreground noises, which are used to form the background prior. Afterward, an initial foreground prior is obtained by selecting superpixels that are the most dissimilar to the background prior. To determine the regions of foreground and background based on the prior of them, a threshold is needed in this process. According to a fixed threshold, the remaining superpixels are iteratively assigned based on their proximity to the foreground or background prior. As the threshold changes, different foreground priors generate multiple different partitions that are assigned a likelihood of being foreground. Last, all segments are combined into a saliency map based on the idea of similarity voting. Experiments on five benchmark databases demonstrate the proposed method performs well when it compares with the state-of-the-art methods in terms of accuracy and robustness.

Keywords: Saliency Object Detection, Soft-Label Partition, Similarity Voting, Adaptive Figure-Ground Classification

This work is sponsored by the Scientific Research Foundation of Nanjing Institute of Technology (No.YKJ201722, No. YKJ201740), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 14KJB520006), the National Natural Science Foundation of China (Grant No. 61403188).

http://doi.org/10.3837/tiis.2018.03.016 **ISSN : 1976-7277**

1. Introduction

Human saliency is usually referred as local contrast [1]. It typically originates from contrasts between an item and its surroundings, such as differences in color, texture, shape, etc. This mechanism measures intrinsically salient stimuli to the vision system that primarily attracts human attention in the early stage of visual exposure to an input image [2]. Intermediate and higher visual processes may automatically judge the importance of different regions of the image, and conduct detailed processes only on the "salient objects" that mostly related to the current tasks, while neglecting the remaining "background" regions [3]. The detection of such salient objects in the image is of significant importance, as it directs the limited computational resources to faster solutions in the subsequent image processing and analysis [4]. There are many applications for salient object detection, for example, visual tracking [5], object retargeting [6-7], image categorization [8] and image segmentation[9], and so forth.

From the perspective of information processing mechanisms, existing saliency estimation algorithms can be broadly categorized as either bottom-up approaches [10-30] or top-down approaches [31-35]. Bottom-up methods are usually based on low-level visual information, and are more effective in detecting fine details rather than global shape information. In contrast, top-down saliency models are able to detect objects of certain sizes and categories based on more representative features from training samples. Since the bottom-up strategy of saliency detection is pre-attentive and data-driven and therefore has been widely applied. It is usually fast to execute and easy to adapt to various cases compared with top-down approaches. In this paper, we focus on bottom-up salient object detection.

All bottom-up saliency methods rely on some prior knowledge about salient objects and backgrounds, such as contrast, compactness, etc. Different saliency methods characterize the prior knowledge from different perspectives. The most widely utilized assumption is that appearance contrasts between objects and their surrounding regions are high. This is called contrast prior and is used in almost all saliency methods. As a pioneer, Itti et al. [1] extract center-surround contrast at multiple spatial scales to find the prominent region. Bruce et al. [10] exploit Shannons self-information measure in local context to compute saliency. However, the local contrast does not consider the global influence and only stands out at object boundaries. Region contrast based methods [4,11] first segment the image and then compute the global contrast of those segments as saliency, which can usually highlight the entire object. Fourier spectrum analysis is used to detect visual saliency [12-13]. Recently, Perazzi et al. [14] unify the contrast and saliency computation into a single high-dimensional Gaussian filtering framework.

Besides contrast prior, several recent approaches [15-17] exploit boundary prior [18], i.e., image boundary regions are mostly backgrounds, to enhance saliency computation. Such methods achieve state-of-the-art results, suggesting that boundary prior is effective. For example, Wei et al. [16] exploit background priors and geodesic distance for saliency detection. Yang et al. [17] cast saliency detection into a graph-based ranking problem, which performs label propagation on a sparsely connected graph to characterize the overall differences between salient object and background. However, we observe two drawbacks. The first is they simply treat all image boundaries as background. This is fragile and may fail even when the object only slightly touches the boundary. The second is their usage of boundary prior is mostly heuristic. It is unclear how it should be integrated with other cues for saliency computation.

In this paper, we propose a background prior based salient object detection method (BPS for short), which exploits an adaptive figure-ground classification method [18] to classify the object from background, to make the salient object pop-out automatically in given image. The block diagram of proposed algorithm is shown in **Fig. 1**. The new framework can be divided into four steps. First, an initial segmentation, i.e., SLIC [19], is required to partition the image into homogeneous regions for measuring saliency. Second, to generate background prior, we take the superpixels located in four borders into consideration and exploit a mechanism based on image boundary information to remove the foreground noises. These superpixels are used to form the background prior. Then, an initial foreground prior is obtained by selecting superpixel that is the most dissimilar to the background prior. To classify the superpixel into the foreground and the background, a threshold is used. According to the threshold, the remaining superpixels are iteratively assigned based on their proximity to the foreground or background prior, with the foreground prior being updated with new superpixels. As the threshold changes, different foreground priors generate multiple soft-label partitions that are not explicitly assigned a foreground or background label, but instead assigned a likelihood of being foreground, based on foreground likelihood of preceding labeled patches. Last, based on the idea of similarity voting, all soft-label partitions are combined into a saliency map.

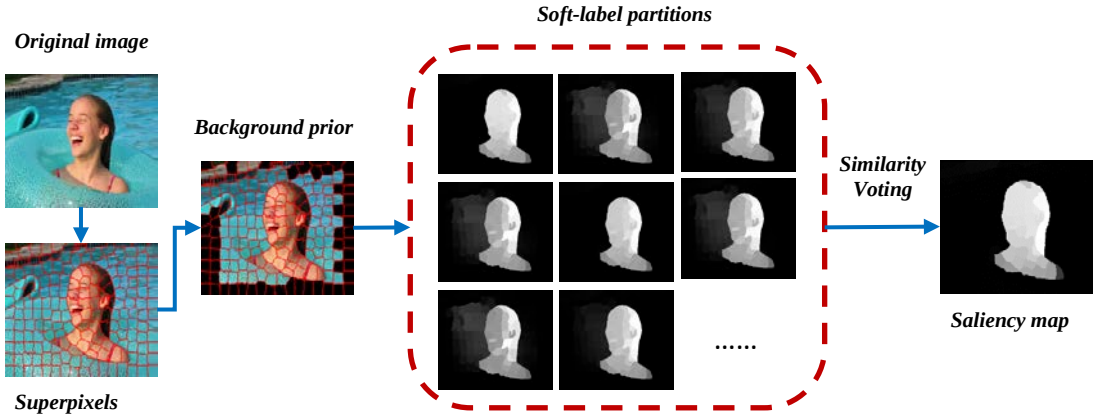


Fig. 1. Main steps of the proposed approach

In summary, the contributions of this paper include: (1) A novel salient object detection algorithm is proposed that is based on background prior and adaptive figure-ground classification. Exploiting figure-ground classification, the superpixels of the given image can be classified into background and foreground, which fulfill the goal of salient object detection effectively. (2) A method of background prior selection is designed for salient object detection in the proposed algorithm. Unlike conventional methods that use a problematic boundary as the prior of the background in saliency estimation, BPS algorithm optimizes the boundary influences by locating and eliminating erroneous boundaries before the saliency detection.

The remainder of the paper is organized as follows: Section 2 reviews related work that is related to our approach. We demonstrate framework of our saliency detection method in detail in Section 3. Then, we demonstrate our experimental results based on five public image datasets and compare the results with other state-of-art saliency detection methods in Section 4. The final section concludes the paper by summarizing our findings.

2. Related Work

Saliency estimation methods can be explored from different perspectives. Regarding the early problem of fixation prediction, Itti et al. [1] propose a well-known saliency model which is implemented based on the biological attention mechanisms and feature integration theories. In this model, elementary features, e.g., color and luminance computed from different scales, are integrated using a center-surround operator to generate the saliency map, in which visually salient points are highlighted, as the prediction of fixations. After that, a number of fixation prediction models are proposed (e.g., [12,20]). A comprehensive survey on the fixation prediction models can be found in [21].

Concerning the salient object detection problem, which is the focus of this paper, in [22] it is defined as a binary segmentation problem for application to object recognition. Since then, plenty of saliency models are proposed for detecting salient objects in images based on various theories and principles, such as information theory [23], graph theory [17,24-25], statistical modeling [26-28], low-rank matrix recovery [29-30], partial differential equations [31], and machine learning [32-34]. Moreover, a variety of priors are explored to achieve a higher performance of salient object detection, e.g., center prior [35], boundary connectivity prior [16,36], focusness prior [37,38], objectness prior [38,39] and background prior [17,24,31,40].

Some early works use the so called center prior to bias the image center region with higher saliency. Usually, center prior is realized as a gaussian fall-off map. It is either directly combined with other cues as weights [33,41,26], or used as a feature in learning-based methods [15]. This makes strict assumptions about the object size and location in the image.

However, center prior alone is not very effective. The other most commonly used cue is based on the assumption that most photographers do not crop the salient object along the view frame. Hence the image boundary forms the background. However boundary prior is fragile and it's prone to fail even when the object is slightly touching the background. Considering the connectivity of regions in the background, Wei et al. [16] define each region's saliency value as the shortest-path distance towards the boundary. In [15], the contrast against image boundary is used as a new regional feature vector to characterize the background. In [17], Yang et al. compute the saliency of image regions according to their relevance to boundary patches via manifold ranking. Recently, Zhu et al. [36] propose a boundary connectivity measure that utilizes both contrast prior and boundary prior. Foreground and Background weights obtained are then combined using an optimization framework.

Objectness proposal generation methods propose small number of windows which are likely to contain the object in an image thereby reducing search space for classifiers. Alexe et al. [42] propose an objectness measure that combines several image cues measuring an object's characteristics in a Bayesian framework. Zhang et al. [43] propose cascaded ranking SVM to generate an ordered set of proposals. Cheng et al. [44] proposes a binarized version of normed gradient features (BING) which can be tested using few atomic operations to generate Objectness proposals. Jiang et al [37] integrate Objectness with Uniqueness and Focusness to obtain saliency maps. However these maps are not smooth and it is difficult to attribute these results to specific algorithm properties [14].

It is also worth noting that there are some previous works involving forground-background classification in saliency detection [45,46]. In [45], Gopalakrishnan et al. exploit the hitting time on the fully connected graph and the sparsely connected graph to find the most salient seeds, based on which some background seeds are determined again. They then use the difference of the hitting times to the two kinds of seeds to compute the saliency for each node. However, the hitting time based saliency measure prefers to highlight the global rare regions

and does not suppress the backgrounds very well, thereby decreasing the overall saliency of objects. BL [46], which exploits both weak and strong bootstrap learning models, integrates multiscale saliency maps to improve the detection performance. In weak bootstrap learning, they compute a weak contrast-based saliency map based on superpixels of an input image. In strong bootstrap learning, a strong classifier based on Multiple Kernel Boosting is learned to measure saliency where three feature descriptors are extracted and four kernels are used to exploit rich feature representations. However, it makes the algorithm cannot suppress the noise in the background and preserve the object boundary well. In contrast, BPS focuses on the background prior and exploit the dissimilarities between the background prior and the rest superpixels to generate multiple soft-label partitions. To form saliency map from soft-label partitions, BPS uses the idea of similarity voting. From this point of view, BPS is similar to the multiscale saliency detection, for one, [46].

3. Proposed Salient Object Detection

In this section, we describe BPS algorithm which is based on adaptive figure-ground classification and background prior propagation in detail. Since BPS is mainly divided into four steps, this section introduces BPS by four points as following: 1) Initial background prior, which present a mechanism based on image boundary information to remove the foreground noises and select background seeds from the border superpixels; 2) Similarity measure between patches that defines a suitable dissimilarity measure between two patches, and between a patch and a region; 3) Soft-label partitions section describes assigning each image patch a likelihood (soft-label) of belonging to the foreground category. In final subsection, we exploit the methods aforementioned to form the new salient object detection framework.

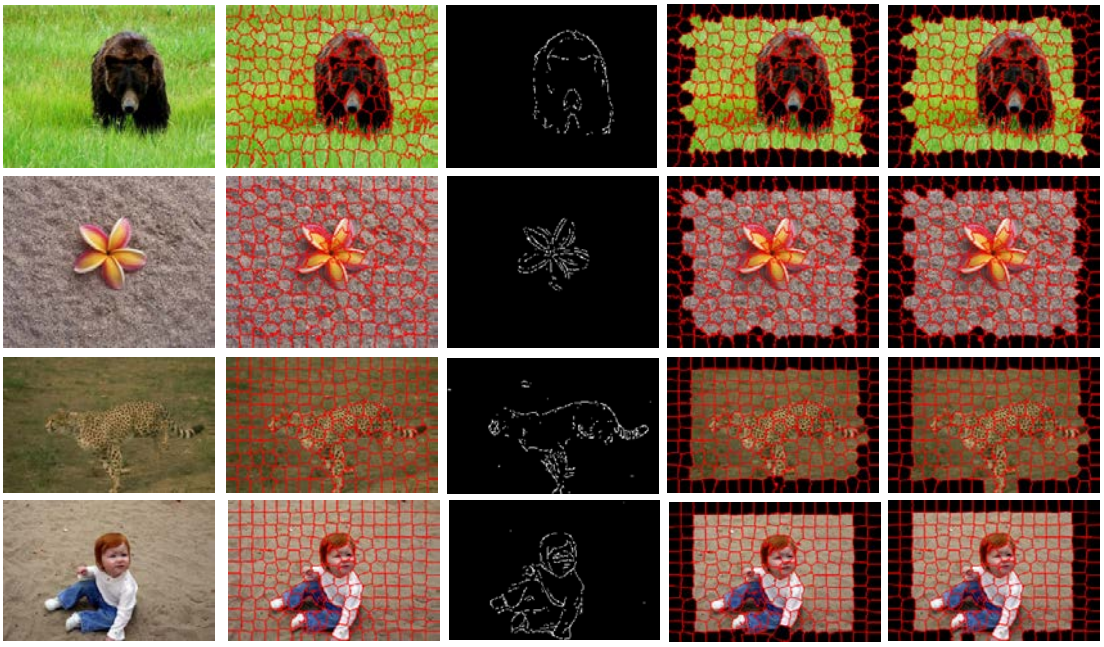
3.1 Initial Background Prior

Due to the advantages in information transfer and computational efficiency, BPS first segments the input image into superpixels for salient object detection. Since the background priors are so important for the new framework, in this subsection, we first discuss how to obtain the priors in given image.

In the conventional problems of background-based salient object detection, the background priors are manually labeled with the ground truth. Recently, the cues in bottom-up saliency detection have gained much popularity [34]. Different proposals include measures of pixel contrast [33], various implementations [4,33] of the discriminant center-surround saliency principle of [47] on the graph-based saliency model of [48]. Although several seed mechanisms have been proposed in the literature, they tend to be heuristic in nature, e.g. selecting the superpixels that most differ from those along the image border. They simply treat all image boundaries as background. Nevertheless, some of them maybe incorrect that there may be some foreground noises in the border regions, leading to negative effects on saliency detection. Since they have insignificant effects on the final results, a mechanism based on image boundary information is proposed to remove the foreground noises and selected background seeds from the border superpixels.

Next, some examples are given to illustrate the selection of background prior, which are shown in **Fig. 2**. As an input image is over-segmented into 200 superpixels as shown in **Fig. 2**(b), the superpixels are selected as the border set whose centroids locate within a certain number of pixels to the image borders. Since the most distinct boundary of an image is likely to be the contour between the object and background, we can roughly remove the image

superpixels with strong boundaries (see **Fig. 2**(c)), which are regarded as the foreground noises, out of the border set.



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

Fig. 2. Illustration of the background prior, (a) input image, (b) superpixels, (c) PB map, (d) border set (the black regions along image borders), (e) background prior

We first adopt the probability of boundary (PB) [49] to detect image boundary (see **Fig. 2**(c)). The boundary feature of the $i-th$ superpixel is calculated by the average PB value of pixels along the edge contour of superpixel i, as follows:

$$PB_i = \frac{1}{|B_i|}\sum_{I\in B_i} I^{pb} \tag{1}$$

where B_i is the edge pixel set of superpixel i and $|B_i|$ denotes its cardinality. The PB value of pixel I is denoted by I^{pb}. Since the superpixel with large boundary feature is more likely to be the object, we remove the superpixels whose boundary features are larger than the adaptive gray threshold derived by [50]. Then the remaining superpixels in the border set are selected as background priors, containing more stable and reliable background information. As shown in **Fig. 2** (e), the selected background seeds (e) have less foreground noise than the border set (see **Fig. 2** (d)).

3.2 Superpixels Similarity Measure

In the next stage of BPS algorithm, superpixels are gradually assigned likelihood labels, based on their similarities to the background priors. Hence, we must first define a suitable

dissimilarity measure between two superpixels, and between a superpixel and a region (a group of superpixels).

To measure the dissimilarity between two superpixels, the conventional methods [15-17] adopted a Gaussian function to calculate the weights, which measure the difference of the mean color between two superpixels. However, it omit the texture information of the superpixel. In natural images, two superpixels with the same average color exactly, have different textures completely. To exploit the texture information in saliency detection framework, we model each pixel as a 5D feature vector in a joint color-spatial feature space, i.e.,

$$f_{x,y} = (L(x,y), a(x,y), b(x,y), x, y) \tag{2}$$

where (x, y) are the 2D pixel coordinates and $(L(x,y), a(x,y), b(x,y))$ are the corresponding pixel values in the Lab color space. We use the Lab space because it is better modeled by a normal distribution in comparison to RGB [51]. Then, each superpixel p_i can be represented as a multivariate normal distribution $N(\mu_i, \Sigma_i)$ in the 5D feature space, where the mean vector μ_i and the covariance matrix Σ_i are estimated from the superpixel. All superpixels are eroded with a 3×3 structural element to avoid border effects.

The Kullback-Leibler divergence (KLD) can be used to measure dissimilarity between two distributions, but is not symmetric [52]. Here, we use the minimum KLD between two superpixels as our dissimilarity measure, i.e.,

$$D(p_1, p_2) = \min(KL(p_1, p_2), KL(p_2, p_1)) \tag{3}$$

where superpixels p_1 and p_2 are represented by two Gaussians, with distributions $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$, and the KLD between two d-dimensional Gaussians[52] is

$$KL(p_1, p_2) = \frac{1}{2}\left[(\mu_1 - \mu_2)^T \Sigma_2^{-1}(\mu_1 - \mu_2) + Tr(\Sigma_2^{-1}\Sigma_1) + \log\frac{|\Sigma_2|}{|\Sigma_1|} - d\right] \tag{4}$$

Note that Eq. (3) is a symmetrized version of the KLD in (4), and has an intuitive interpretation that two superpixels are similar if either of them can be well described by the other. With this dissimilarity the background holes can be reliably identified as similar to the background.

A region in an image (e.g., the background) is represented as a set of superpixels, $R = \{p_{r_1}, \cdots, p_{r_k}\}$, where $\{r_k\}$ are the indices of the superpixels forming the region. Using the dissimilarity between superpixels in Equation (2), we define the dissimilarity between a superpixels p and a region R as the minimum dissimilarity between the superpixel p and any superpixel in R,

$$D(p, R) = \min_{r \in R} D(p, r) \tag{5}$$

We define the dissimilarity between two regions R_1 and R_2 as the minimum dissimilarity between their superpixels,

$$D(R_1, R_2) = \min_{r \in R_1, p \in R_2} D(r, p) = \min_{r \in R_1} D(r, R_2) \tag{6}$$

Fig. 3. Two examples with both background and foreground have very different distributions

Note that both background and foreground can be multimodal, which is shown in **Fig. 3**. That is, superpixels in one region (e.g., background) may have very different distributions (e.g., tree and grass in the left example in **Fig. 3**). Therefore, for the superpixel-region dissimilarity, we use the minimum dissimilarity so as to match the superpixel to the most similar part in the region. Likewise, the minimum dissimilarity measure between two regions implies that they are similar if they have superpixels in common (e.g., both contain grass in the right example in **Fig. 3**). In the context of salient object detection, using alternatives such as median dissimilarity or max-min dissimilarity may not work well due to the regions being multi-modal.

3.3 Soft-Label Segmentations

With the superpixel distances defined in Equation (3) we next describe our foreground extraction algorithm. Under the assumption that the background priors provides sufficient background statistics, we first initialize the background priors, and then gradually compute a soft label segmentation. Generally, the objective is to assign each image superpixel p_i a likelihood (soft-label) of belonging to the foreground category, denoted by $L(p_i)$.

The partitioning process proceeds as follows. First, all patches p_i overlapping with the background mask form the initial background prior B, and are given zero likelihood,

$$L(p_i) = 0, \ \ \forall p_i \in B \tag{7}$$

Next, the initial foreground region F_0 is formed using the set of patches that are sufficiently far from B,

$$F_0 = \{p_i | D(p_i, B) > D_t\} \tag{8}$$

where D_t is a foreground threshold whose value will be discussed at the end of the subsection. The foreground likelihood of these initial foreground patches is set to 1,

$$L(p_i) = 1, \ \forall p_i \in F_0 \tag{9}$$

The remaining unlabeled patches are progressively labeled with patches furthest from the background considered first, i.e., in descending order based on their distances from the background prior B, $D(p_i, B)$. Let Θ be the set of currently labeled patches. For each patch p_i under consideration, a local conditional probability with respect to any labeled patch $p_j \in \Theta$ is computed by comparing the distances from p_i to the background prior B and p_j using the softmax (logistic) function,

$$l(p_i | p_j) = \frac{e^{-D(p_i, p_j)}}{e^{-D(p_i, p_j)} + e^{-D(p_i, B)}} \tag{10}$$

Since the feature space represents both color and location, Eq. (10) will give high likelihood when the two patches are both visually similar and spatially close together, while also being dissimilar to B. The overall likelihood of patch p_i being foreground is estimated by calculating the maximum likelihood score over all preceding patches,

$$L(p_i) = \max_{p_j \in \Theta} L(p_j) l(p_i | p_j) \tag{11}$$

Eq. (11) considers both the conditional probability of the current patch being foreground given the labeled patch, and the probability of the labeled patch also being foreground. Note that these patches are not explicitly assigned a foreground or background label, but instead assigned a likelihood of being foreground, based on foreground likelihood of preceding labeled patches. After all unlabeled patches are processed with (11), a likelihood L is defined for every patch, resulting in a soft-labeling of foreground regions in the image.

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

Fig. 4. Different initial foreground region by selecting different thresholds, (a) original image, (b) D_t=10, (c) D_t=25, (d) D_t=35, (e) D_t=45

We now turn our attention to the threshold D_t that determines the initial foreground region F_0. The choice of threshold is important since it may lead to different foreground seeds and hence different soft-label partitions (as shown in **Fig. 4**). Rather than select a single threshold, we instead consider multiple thresholds, i.e., multiple foreground initializations, and produce various candidate soft-label partitions for consideration. In practice, we use all

thresholds D_t between the lower and upper bounds, $D_l = 5$ and $D_u = 50$. This interval allows a large enough set of initial foreground priors but excludes unnecessary initializations. Since there are a finite number of possible $D(p_i, B)$ values (one for each image patch), we only need to try a finite number of thresholds. In particular, we sort all values of $D(p_i, B)$ within the interval $[D_l, D_u]$ in ascending order and use the midpoints between two successive values as the set of thresholds. Running the soft-label partitioning method for each threshold, we obtain a large set of soft-label partitions. The size of the set depends on the number of patches in the image. Simple images will have few patches. On the contrary, cluttered images will have more patches and obtain a larger set of soft-label partitions.

3.4 Salient Object Detection

Based on the soft-label partitions aforementioned, the proposed algorithm next builds a foreground probability map by fusing all soft-label partitions. The fusion is based on the idea of similarity voting. That is, partitions sharing more similarities are given higher influence. Denote F_i as the $i-th$ soft-label partition from the previous stage, and F_i^m as the likelihood value of the $m-th$ pixel in F_i, where pixels take the likelihoods of their corresponding patches. Then, the similarity between two soft-label partitions F_i and F_j is defined as

$$d(F_i, F_j) = \frac{\sum_{m=1}^{M} \left| F_i^m - F_j^m \right|}{\sum_{m=1}^{M} sign(F_i^m + F_j^m)} \tag{12}$$

where M is the total number of pixels, and $sign(x)$ is 0 when $x = 0$ and 1 when $x > 0$. We then construct a symmetric affinity matrix A with entries

$$A(i, j) = \exp(-d(F_i, F_j)^2 / 2\sigma^2) \tag{13}$$

where σ^2 is the variance of the pairwise distances between all partitions $\{F_i\}$ [53]. Finally, a real-valued probability map is calculated as the weighted sum of the soft-label partitions $\{F_i\}$,

$$S_{map} = \sum_i w_i^2 F_i \tag{14}$$

The weight vector w is determined using the following constrained optimization problem,

$$\max w^T A w, s.t. \ \|w\|^2 = 1 \tag{15}$$

Eq. (15) is a standard Rayleigh quotient problem [54], and the optimal w is given by the top eigenvector of A. Intuitively, the weights found by Eq. (15) are higher for partitions sharing more similarities. In short, the probability map is computed as the weighted sum of all soft-label partitions, where larger weights are given to more similar partitions. This

corresponds to a similarity voting process leading to a better probability map, compared to [18]. Since the probability map indicates the probability of the pixel belongs to the target, S_{map} is defined as the final saliency map in the BPS algorithm.

4. Experimental Results and Analysis

In this section, the performance of BPS algorithm is evaluated over five datasets that are widely used in previous works, e.g. [24,46]. Next, we describe the datasets shortly, discuss the parameter in our algorithm and report both quantitative and qualitative comparisons of BPS approach with state-of-the-art approaches in detail. To save the space, the method is compared with several prior ones, including LRMR [29], RC [4], GS [16], MR [17], MC [24], LPS [55], MS [56], wCtr [36], SCA [57], GP [58], BL [46] and GL [59]. **Table 1** summarizes the details of these algorithms. To evaluate these methods aforementioned, the results are originating from either provided by authors or ran their implementations based on the available codes or software. BPS algorithm is implemented in Matlab 2010a. All experiments are conducted on a PC with an Intel Core i3-3240 3.4 GHz CPU and 4 GB RAM. In the experiment, all input images are over-segmented into 200 superpixels. For the threshold in **Eq.(8)**, we fix it as described in Section 3.3.

Table 1. Summary of existing methods compared in this paper. "Code": "M" = Matlab, "C" = C/C++

Model	Year	Code	Prior	Dataset(s)
LRMR [29]	2012	M+C	Location + Semantic + Color	MSRA
RC [4]	2015	C++	Center	MSRA
GS [16]	2013	M	Boundary	MSRA, Berkeley
MR [17]	2013	M	Boundary	MSRA, DUT-OMRON
MC [24]	2013	M+C	Boundary	MSRA,ECSSD, DUT-OMRON
MS [56]	2014	M+C	Center	MSRA
wCtr [36]	2014	M	Boundary connectivity	MSRA, SED
SCA [57]	2015	M+C	Boundary	MSRA, ECSSD, DUT-OMRON
GP [58]	2015	M+C	Center	MSRA, ECSSD
BL [46]	2015	M+C	Center	MSRA, Berkeley, SED, DUT-OMRON
LPS [55]	2015	M+C	Boundary + Objectness	MSRA, ECSSD
GL [59]	2015	M+C	Boundary + Center + Objectness	MSRA

4.1 Datasets and Evaluation Measures

1) Datasets: Experiments are performed on five different image sets which are generated from five publicly available saliency object detection databases (all with human-marked binary mask for salient regions as ground truth), i.e., MSRA [33], ECSSD [41], DUT-OMRON [17], SED [60] and Berkeley [61]. MSRA [33] includes 5000 images, originally containing labeled rectangles from nine users drawing a bounding box around what they consider the most salient object. There is a large variation among images including natural scenes, animals, indoor, out-door, etc. We use the salient object (contour) as binary masks. To represent natural image situations, Yan et al. [41] extended their CSSD dataset in to the larger ECSSD dataset, which contains 1000 images. It includes many semantically meaningful but structurally complex images. Ground truth masks were produced by five subjects. DUT-OMRON [17] which consists of 5,168 images carefully labeled by five users, have one or more salient objects and

relatively complex background. Compared with the MSRA dataset, images in the DUT-OMRON dataset are more difficult, thus more challenging, and provide more space of improvement for saliency detection. To test methods in the image with multiple objects, we also conduct experiments on the SED saliency benchmark database with 200 natural images. SED [60] contains two subsets: SED1 that has 100 images containing only one salient object and SED2 that has 100 images containing two salient objects. Pixel-wise ground truth annotation for the salient objects in both SED1 and SED2 are provided. The last image set used in this part is the even more challenge Berkeley saliency object database, which is based on the well-known 300 images in Berkeley segmentation database [61]. Those images usually contain multiple foreground objects of different sizes and positions in the image. Furthermore, the appearance of foregrounds and backgrounds are also more complex. Please see **Table 2** for details of these datasets.

Table 2. Summary of datasets in the comparison

Name	Scale	Type	GT
MSRA	5000	Single object, simple background	Binary pixel-wise mask
DUT-OMRON	5168	Single or multiple object(s), simple or complex	
ECSSD	1000	Single object, complex scene	
SED	200	Single or multiple object(s), simple background	
Berkeley	300	Single or multiple object(s), complex scene	

2) Fixed Threshold: In the first evaluation measure, we use binary masks, which obtained by directly thresholding a saliency map S_{map} using threshold from the range [0,…,255], to calculate the precision and recall rate. Precision corresponds to the percentage of salient pixels correctly assigned, while recall corresponds to the fraction of detected salient pixels in relation to the number of salient pixels in ground truth maps [4]. Here, the precision P and recall R values are calculated as:

$$P = \frac{|SF \cap GF|}{|SF|} \text{ and } R = \frac{|SF \cap GF|}{|GF|} \tag{16}$$

where GF is the ground truth map, $|\cdot|$ denotes the sum area of masks. SF is the binary mask obtained by directly thresholding a saliency map using threshold.

3) Adaptive Threshold: In the second evaluation measure, we employ the saliency-map-dependent threshold proposed by [16] and define it as proportional to the mean saliency of a map:

$$\tau_\alpha = \frac{2}{H * W} \sum_{x=1}^{H} \sum_{y=1}^{W} S_{map}(x, y) \tag{17}$$

where W and H are the width and height of the saliency map in pixels, respectively. If the saliency value of a pixel is larger than threshold τ_α, it is considered as the part of salient object. In many applications, high precision and high recall are both required. Then a weighted harmonic mean measure between precision and recall, i.e., F-measure, is introduced by

$$F = \frac{(1+\beta^2) \times P \times R}{\beta^2 \times P + R} \tag{18}$$

where we use $\beta^2 = 0.3$ as that in [4] to weight precision more than recall. As can be seen later, one method cannot have in all the highest precision, recall and F-measure as the former two are mutually exclusive and the F-measure is a complementary metric to balance them.

4) Mean Absolute Error: In the third evaluation measure, we introduce the Mean Absolute Error (MAE) between the continuous saliency map S_{map} and the binary mask of ground truth GT:

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left\| S_{map}(x, y) - GF(x, y) \right\| \tag{19}$$

where W and H are the width and height of the saliency map in pixels, respectively. The metric takes the true negative saliency assignments into account whereas the precision and recall favor the successfully assigned saliency to the salient pixels [42]. Moreover, the quality of the weighted continuous saliency maps may be of higher importance than the binary masks in some cases [14].

4.2 Quantitative Comparisons

1) Mean Absolute Error: The MAE estimates the approximation degree between the saliency map and the ground truth map, and it is normalized into [0, 1]. **Fig. 5** shows the MAE metric of BPS algorithm and other methods on MSRA [33], ECSSD [41], DUT-OMRON [17], SED [60] and Berkeley [61]. Considering the recent and well-performed methods, such as MS [56], wCtr [36], SCA [57], GP [58], BL [46], LPS [55] and GL [59], BPS achieves the lowest error of 0.1116, 0.1711, 0.1386, 0.1466 (SED1) and 0.1162 (SED2), and 0.2254 on the corresponding datasets.

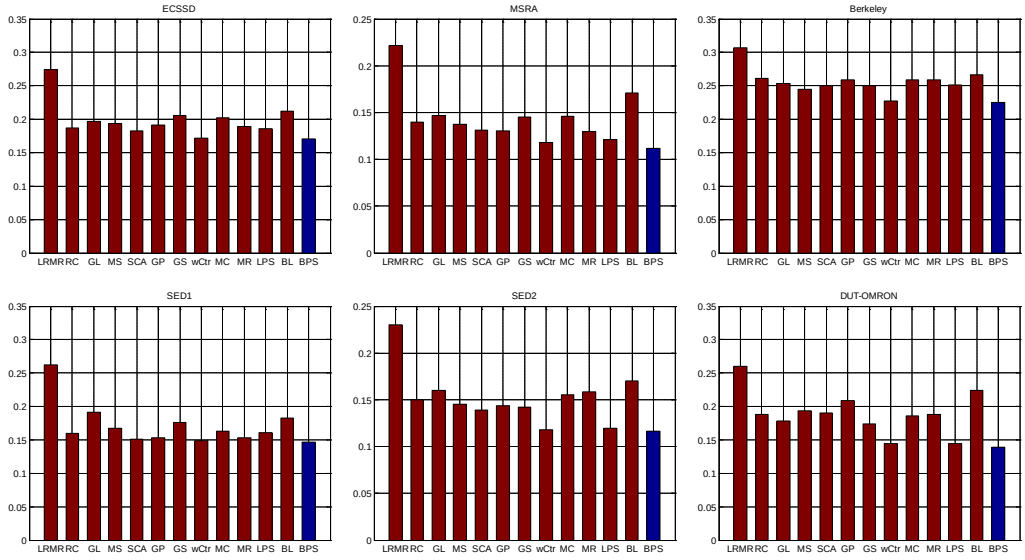


Fig. 5. MAE metric on five image datasets

It can be seen that LRMR [59] has the highest MAE scores in five image datasets. In addition, the MAE values of BL is slightly lower than LRMR, but higher than other methods. The rest algorithms, except wCtr,have similar MAE scores in these image databases. The MAE values of wCtr [36] are slightly higher than or the same as BPS algorithm. According to the definition of the MAE, it provides a direct way of measuring how close a saliency map is to the ground truth. BPS algorithm achieves the lowest MAE scores on the four corresponding datasets, which indicates that the resultant maps are closest to ground truth.

2) Evaluation on DUT-OMRON: **Fig. 6** displays the P-R curves and F-measure on DUT-OMRON benchmark. On one hand, BPS method achieves the highest precision rate covering most ranges of the recall while other models, such as wCtr [36], SCA [57] and MS [56], have similar performance competing BPS and yet lower precision at specific ranges of recall. On the other hand, the highest precision and F-measure score of 0.6474 and 0.6158 is accomplished by BPS outperforming other 12 methods. Note that due to the cluttered background, many models treat the regions near the salient object as salient area, such as GS and wCtr, their model has the highest recall value in **Fig. 6**; however, it is more important to have a high value of precision or F-measure in the saliency community.

3) Evaluation on MSRA and ECSSD: **Figs. 7-8** reports the performance comparison on these two datasets. For the ECSSD benchmark, compared with most methods, BPS achieves the best curve performance as well as the highest F-measure. Specifically, BPS has the higher precision of 0.7767, lowest MAE error and similar F-measure. In addition, since GL [59] achieves better precision score of 0.7821, their recall is lower than most methods, including BPS algorithm.

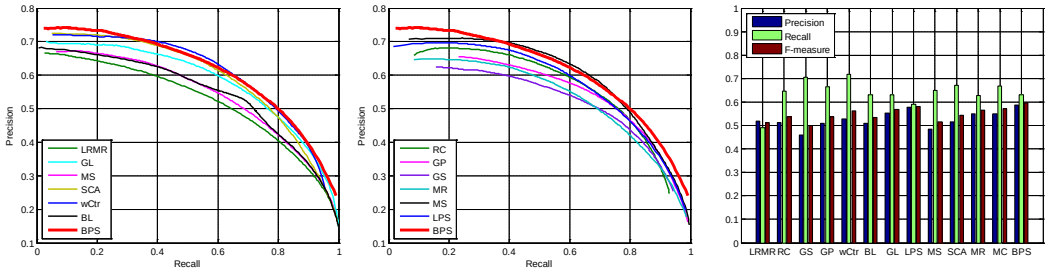


Fig. 6. Quantitative comparisons of saliency maps produced by different approaches on DUT-OMRON dataset

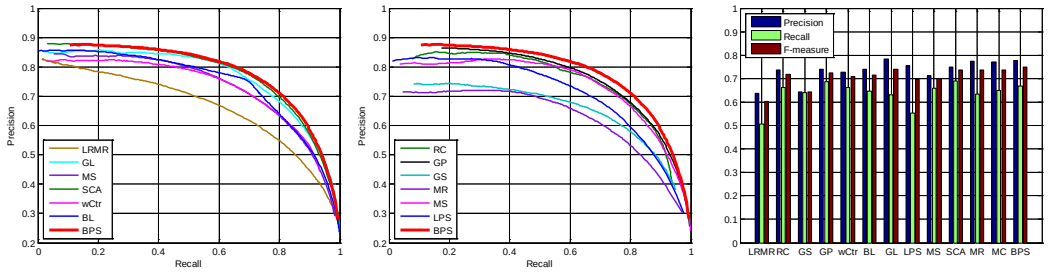


Fig. 7. Quantitative comparisons of saliency maps produced by different approaches on ECSSD dataset

For the MSRA database, we plot the precision-recall curves and F-measure in **Fig. 8** to compare our method with twelve state-of-the-art approaches. Though the images in MSRA database contain different kinds of content, the background of each image is usually large while the foreground is relatively compact with the same color. Most of the state-of-the-art

algorithms have similar performance on this image dataset. Compared with other methods, BPS model achieves the best curve performance spanning most ranges of recall as well as the highest precision and F-measure of 0.8523, 0.8344, respectively. From **Fig. 8**, the precision, recall and F-measure of BPS approach are comparable to the currently state-of-the-art methods, which indicate our method could generate reasonable saliency maps.

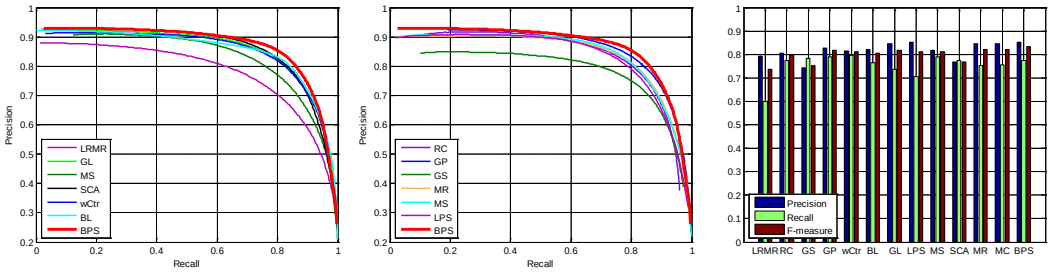


Fig. 8. Quantitative comparisons of saliency maps produced by different approaches on MSRA dataset

4) Evaluation on SED and Berkeley datasets: We also conducted the comparisons on the more challenging Berkeley database. All the comparison results, including P-R curve and weighted F-measure, are shown in **Fig. 9**. As can be seen, the performance of BPS outperforms those of other state-of-the-art algorithms in terms of all metrics. Specifically, the precision values overall P-R curve of BPS are almost higher than that of all other state-of-the-art approaches as well. From **Fig. 9**, it is observed that BPS approach can achieve the highest weighted F-measure score. It is also worth noting that because a large number of images in the Berkeley dataset contain complicated content and multiple salient objects, many excellent approaches, such as GL [59], SCA [57], GP [58], BL [46] and wCtr [36], cannot work effectively in this dataset though they can achieve promising performance on the MSRA and ECSSD datasets. On the contrary, BPS has the capability to yield consistently satisfactory results on both of the datasets, especially on the more challenging Berkeley dataset.

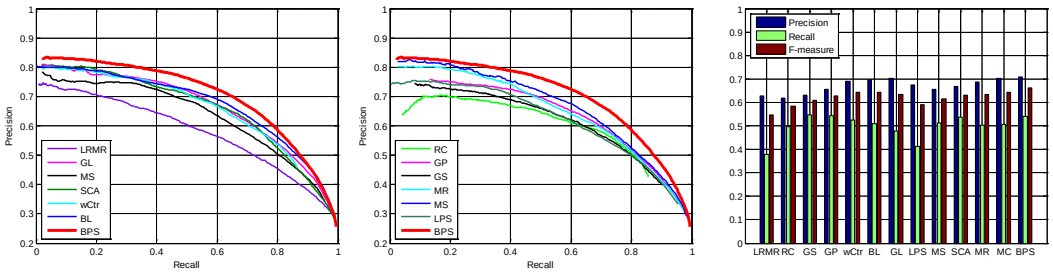


Fig. 9. Quantitative comparisons of saliency maps produced by different approaches on Berkeley dataset

The proposed approach was also tested on the SED database, another challenging dataset. We compare the proposed approach with the state-of-the-art methods in this group of experiments. All the comparison results, including P-R curve and the weighted F-measure, are shown in **Fig. 10**. As can be seen, BPS performs similar to other state-of-the-art algorithms in SED1 database. More encouragingly, compared with other state-of-the-art algorithms, BPS achieves the highest weighted F-measure value compared with other state-of-the-art methods. Similar to the Berkeley dataset, the SED2 dataset also contains a large number of images with complicated content and multiple salient objects. More encouragingly, compared with other

state-of-the-art algorithms, BPS has achieved a higher precision values along almost the whole P-R curve. Note that the wCtr [36] model has a competitive high value of precision in the recall range from 0.7 to 0.8, which means they have a strong capability to suppress the image background in this interval. In addition, BPS achieves the highest recall and F-measure, which indicates that it tends to highlight the entire salient objects and has more capability to handle tough scenarios.

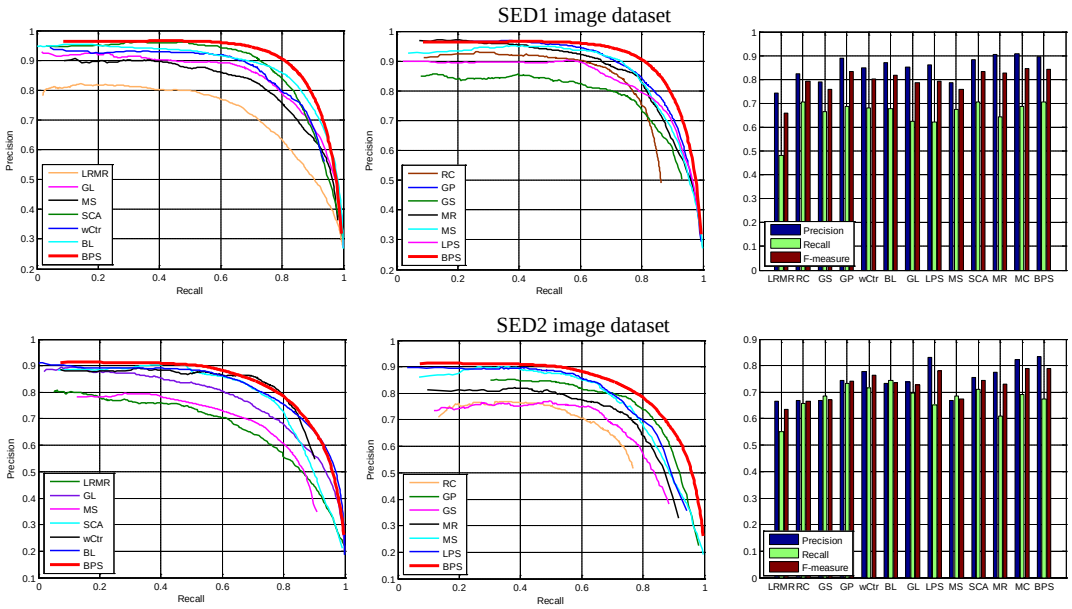


Fig. 10. Quantitative comparisons of saliency maps produced by different approaches on SED dataset

5) Execution Time: **Table 3** shows the average execution time of processing one image in the MSRA dataset. Experiments are conducted and timed on a PC with an Intel Core i3-3240 3.4 GHz CPU and 4 GB RAM, running Matlab 2010a. From **Table 3**, BL [46] spends much more time to weak and strong bootstrap learning for each image. In addition, LPS and GL take much longer than BPS because the calculation of the objectness measure [55] is time consuming. By exploiting boundary prior, the time complexity of BPS is similar to that of MR and wCtr. In contrast, those methods that directly utilize the objectness measure for each single image (LPS [55], GL [59]) have suffered from poor efficiency as well as inferior P-R curves.

Table 3. Running time analysis of different methods on MSRA database (time/s)

Methods	**LRMR** [29]	**RC** [4]	**GS** [16]	**GP** [58]	**wCtr** [36]	**BL** [46]	**GL** [59]
Time(s)	47.28	**0.16**	**0.25**	0.99	**0.28**	52.14	5.4
Methods	**LPS** [55]	**SCA** [57]	**MS** [56]	**MR** [17]	**MC** [24]	**BPS**	/
Time(s)	1.96	0.78	14.68	**0.26**	0.37	**0.28**	/

Note that some methods such as RC [4] and GS [16] have faster efficiency than BPS; we believe that using C++MEX implementation can substantially improve the computational efficiency.

4.3 Qualitative Comparisons

Several natural images with complex background are shown in **Fig. 11** for visual comparison

of our method w.r.t. the most recent state-of-the-arts. For single-object images, BPS accurately extracts the entire salient object with few scattered patches, and assigns nearly uniform saliency values to all patches within the salient objects. For images with multiple objects, some methods (e.g., LRMR [29], LPS [55] and BL [46]) miss detecting parts of the objects, while some (e.g., GS [16] and GP [58]) incorrectly include background regions into detection results. By contrast, BPS pops out all the salient objects successfully. For the images with complex scenes, most methods fail to identify the salient objects, while BPS locates them with decent accuracy. Finally, for the images whose foreground and background share similar appearance, BPS often separates the salient objects from the background. However, by fusing all soft-label partitions, parts of background near the salient object are detected as the object in this case. In general, these results illustrate the robustness of the BPS algorithm since it almost successfully highlights the salient object.

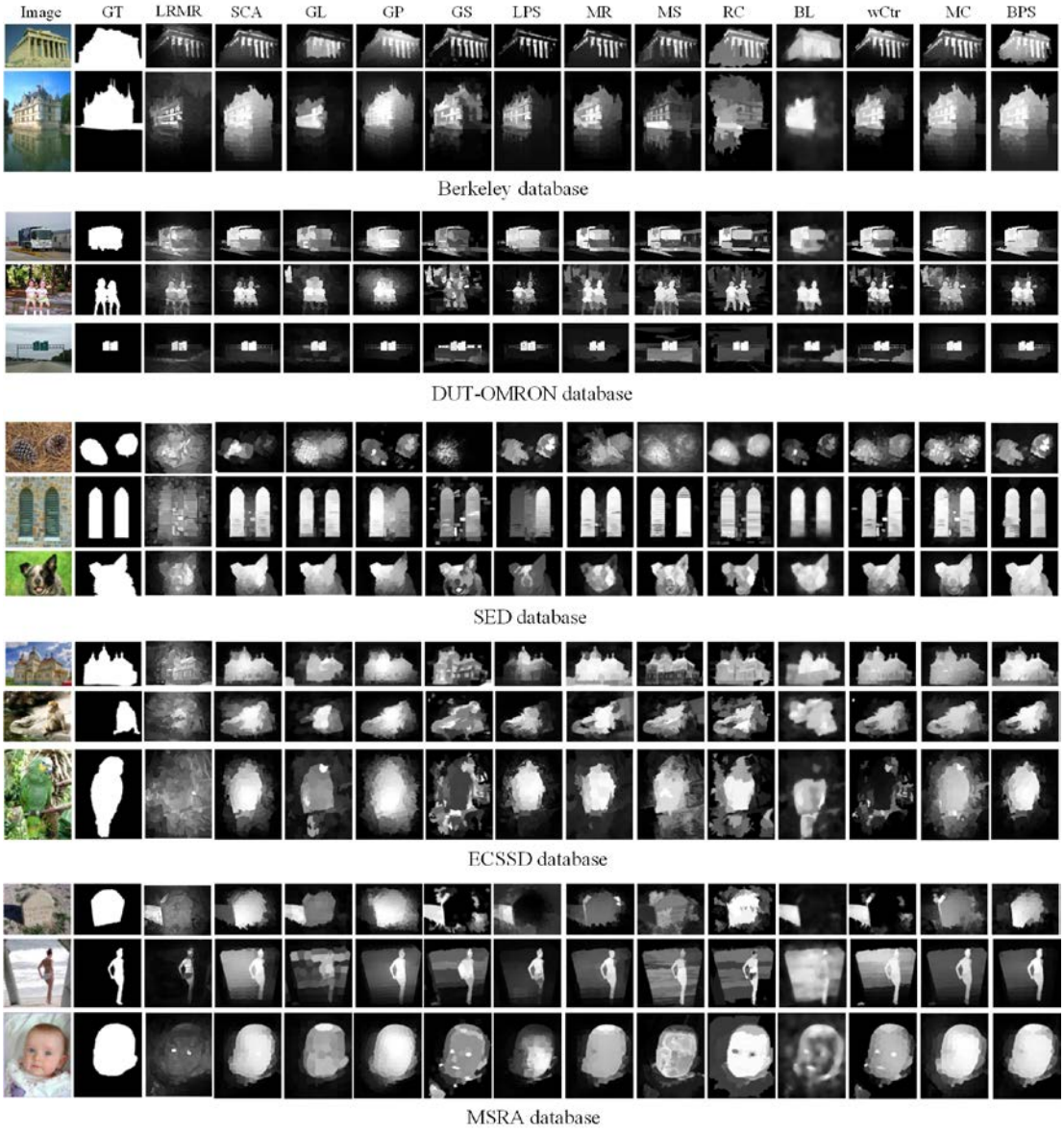
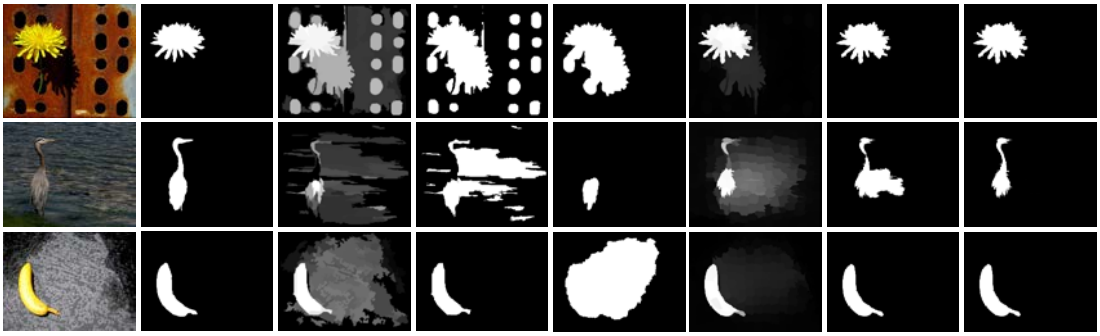


Fig. 11. Visual comparisons of the saliency maps by different methods



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 12. Demonstration of black-and-white salient region generated by BPS and RCC, (a) original image, (b) ground truth, (c) the results of RC, (d) the binarized results of (c), (e) the results of RCC; (f) the results of BPS, (g) the binarized results of (f), (h) black-and-white salient region generated by BPS

It is also worth pointing out that our approach performs well when the object touches the image border, e.g. the last row of SED database in **Fig. 11**, even though it violates the pseudo-background assumption. MR [17], which the first stage is based on the pseudo-background assumption, cannot label the saliency seeds correctly when the object touches the image border. The similar case also find in MC algorithm. Since MC [24] exploits the boundary prior to determine the absorbing nodes, the small salient object touching image boundaries may be incorrectly suppressed. In addition, according to the computation of the absorbed time, a node with sharp contrast to its surroundings often has abnormally large absorbed time, which results that most parts of object even the whole object are suppressed.

Table 4. Comparison of black-and-white salient regions by BPS and RCC on MSRA database

Methods	Precision	Recall	F-measure
RC+Fixed threshold	0.8045	0.7737	0.7972
BPS+Fixed threshold	**0.8523**	**0.7742**	**0.8344**
RC+GrabCut(RCC)	0.8762	0.8054	0.8587
BPS+GrabCut	**0.8914**	**0.8127**	**0.8719**

In [4,62], Cheng et al. present the histogram-based contrast (HC), which exploits the pixel-wise color separation to produce saliency maps. In addition, RC (Region-based Contrast), which is an improvement of HC that takes spatial distances into account at the cost of reduced computational efficiency, can handle complex foreground and background with different details, as shown in **Fig. 11**. By relying solely on the color of pixels/regions that is much different from the dominant one, RC [4] often mistakenly focus on distinct background colors, e.g., the holes in the background are also detected as salient objects in **Fig. 12**. In order to detect exact black-and-white salient objects, the saliency map obtained by RC can be binarized for the initialization of classical GrabCut [9] by using a fixed threshold defined in **Eq. (17)**. For image pixels with saliency value bigger than τ_α, the largest connected region is considered as initial candidate region of the most dominate salient object. Once initialized, GrabCut is ran iteratively to improve the saliency result (denoted as RCC [62]). To compare with RCC, we carry out the image binary by using a fixed threshold τ_α and also use the GrabCut method to detect exact black-and-white salient objects. The results are shown in **Fig. 12** (h). In **Fig. 12** (c) and (f), we can see that the saliency map generated by BPS is more accurate than that of RC. Note that the binarized results of BPS are closer to the ground truth.

It makes the GrabCut method possible to get accurate final results with only a few iterations. Simultaneously, we evaluate the results of RCC and ours in MSRA dataset and list them in **Table 4**. From **Table 4**, the value of precision and F-measure of the proposed algorithm is higher than RCC algorithm. It shows from a side view that BPS algorithm is better than RC.

For other state-of-the-art approaches in **Fig. 11**, it can be seen that while GS [16] which based on boundary priors also detects the regions in background fails when objects touch the image boundary to quite some extent, or when connectivity assumptions are invalid in the presence of complex backgrounds or textured scenes. LRMR [29], which integrates the high-level priors, is focus on the center and the warm color of image. It can be seen that the salient objects with warm colors such as red and yellow are more pronounced. BL [46], which exploits both weak and strong bootstrap learning models, integrate multi-scale saliency maps to improve the detection performance. However, it makes the algorithm cannot suppress the noise in the background and preserve the object boundary well. MS [56] detects the salient object by multi-scale analysis on superpixels. Unlike BL, the results of MS reserve the boundary of salient object better. However, it cannot reserve the object as a whole. For example, in the first line of SED database, the fruits cannot detect as the salient region simultaneously.

5. Conclusion and Future Work

We propose a bottom-up method to detect salient regions in images based on adaptive figure-ground classification. We remove the foreground noises for the background prior by taking the superpixels located in four borders into consideration. The initial foreground prior is obtained by selecting superpixels that are the most dissimilar to the background prior. According to a group of threshold, foreground priors generate multiple soft-label partitions that are not explicitly assigned a foreground or background label. We combine all soft-label partitions into a saliency map based on the idea of similarity voting. Both qualitative and quantitative comparisons show that the proposed approach performs slightly better than several recently state-of-the-art algorithms. Our future work will focus on high-level knowledge, which could be beneficial for handling cases that are more challenging and other kinds of saliency cues or priors to be embedded into our framework.

References

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

[2] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature reviews neuroscience*, vol. 2, pp.194-203, 2001.

[3] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," *In Proc. of ECCV*, pp. 414-429, 2012. Article (CrossRef Link)

[4] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.409-416, 2011. Article (CrossRef Link)

[5] V. Mahadevan and N. Vasconcelos, "Saliency-based discriminant tracking," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.1007-1013, 2009. Article (CrossRef Link)

[6] Y. Ding, J. Xiao, and J. Yu, "Importance filtering for image retargeting," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.89-96, 2011. Article (CrossRef Link)

[7] J. Sun and H. Ling, "Scale and object aware image retargeting for thumbnail browsing," in *Proc. of 13th IEEE International Conference on Computer Vision (ICCV)*, pp.1511-1518, 2011. Article (CrossRef Link)
[8] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.29, no.2, pp.300-312, 2007. Article (CrossRef Link)
[9] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *In ACM Transactions on Graphics (TOG)*, vo.23, pp.309-314, 2004. Article (CrossRef Link)
[10] N. Bruce and J. Tsotsos, "Saliency based on information maximization," *Advances in Neural Information Processing Systems*, pp.155-162, 2006.
[11] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proc. of BMVC*, 2011. Article (CrossRef Link)
[12] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.1-8, 2007. Article (CrossRef Link)
[13] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1-8, 2008. Article (CrossRef Link)
[14] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 733-740, 2012. Article (CrossRef Link)
[15] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2083-2090, 2013. Article (CrossRef Link)
[16] Y. Wei, F.Wen,W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. of ECCV*, 2012. Article (CrossRef Link)
[17] C. Yang, L. Zhang, H. Lu, X. Ruan and M. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.3166-3173, 2013. Article (CrossRef Link)
[18] Y. Chen, A. Chan, "Enhanced figure-ground classification with background prior propagation," *IEEE Transactions on Image Processing*, vol.24, no.3, pp.873-885, 2015. Article (CrossRef Link)
[19] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, Nov., 2012. Article (CrossRef Link)
[20] Y. Lin, Y. Y. Tang, B. Fang, Z. Shang, Y. Huang, and S. Wang, "A visual-attention model using earth mover's distance-based saliency measurement and nonlinear feature combination," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.35, no.2, pp.314-328, 2013. Article (CrossRef Link)
[21] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.35, no.1, pp.185-207, 2013. Article (CrossRef Link)
[22] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol.19, no.9, pp.1395-1407, 2006. Article (CrossRef Link)
[23] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2368-2375, 2010. Article (CrossRef Link)
[24] B. Jiang, L. Zhang, H. Lu, C. Yang, and M. Yang, "Saliency detection via absorbing markov chain," in *Proc. of 14th IEEE International Conference on Computer Vision (ICCV)*, pp.1665-1672, 2013. Article (CrossRef Link)
[25] X. Li, Y. Li, C. Shen, A. Dick, and A. Hengel, "Contextual hypergraph modeling for salient object detection," in *Proc. of 13th IEEE International Conference on Computer Vision (ICCV)*, pp.3328-3335, 2013. Article (CrossRef Link)

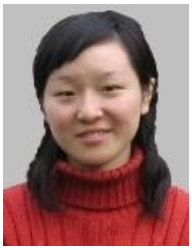
[26] M. Cheng, J. Warrell, W. Lin, et al, "Efficient salient region detection with soft image abstraction," in *Proc. of 13th IEEE International Conference on Computer Vision (ICCV)*, pp.1529-1536, 2013. Article (CrossRef Link)
[27] C. Scharfenberger, A. Wong, K. Fergani, J. Zelek, and D. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.979-986, 2013. Article (CrossRef Link)
[28] Y. Xie, H. Lu, and M. Yang, "Bayesian saliency via low and mid-level cues," *IEEE Transactions on Image Processing*, vol.34, no.11, pp.1689-1698, 2013. Article (CrossRef Link)
[29] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.853-860, 2012. Article (CrossRef Link)
[30] W. Zou, K. Kpalma, Z. Liu, and J. Ronsin, "Segmentation driven low-rank matrix recovery for saliency detection," in *Proc. of BMVC*, 2013. Article (CrossRef Link)
[31] R. Liu, J. Cao, Z. Lin, and S. Shan, "Adaptive partial differential equation learning for visual saliency detection," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.3866-3873, 2014. Article (CrossRef Link)
[32] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.883-890, 2014. Article (CrossRef Link)
[33] T. Liu, Z. Yuan, J. Sun, et.al, "Learning to detect a salient object," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.33, no.2, pp.353-367, 2011. Article (CrossRef Link)
[34] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.2790-2797, 2014. Article (CrossRef Link)
[35] Z. Jiang and L. Davis, "Submodular salient region detection," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.2043-2050, 2013. Article (CrossRef Link)
[36] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.2814-2821, 2014. Article (CrossRef Link)
[37] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by UFO: Uniqueness, Focusness and Objectness," in *Proc. of 13th IEEE International Conference on Computer Vision (ICCV)*, pp.1976-1983, 2013. Article (CrossRef Link)
[38] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.1605-1616, 2014. Article (CrossRef Link)
[39] K. Chang, T. Liu, H. Chen and S. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. of 12th IEEE International Conference on Computer Vision (ICCV)*, pp.914-921, 2011. Article (CrossRef Link)
[40] X. Li, H. Lu, L. Zhang, X. Ruan, and M.Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. of 13th IEEE International Conference on Computer Vision (ICCV)*, pp.2976-2983, 2013. Article (CrossRef Link)
[41] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.1155-1162, 2013. Article (CrossRef Link)
[42] B. Alexe, T. Deselaers, V.Ferrari, "Measuring the objectness of image windows," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.34, no.11, pp.2189-2202, 2012. Article (CrossRef Link)
[43] Z. Zhang, J. Warrell, P. Torr, "Proposal generation for object detection using cascaded ranking SVMs," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.1497-1504, 2011. Article (CrossRef Link)
[44] M. Cheng, Z. Zhang, W. Lin, et al, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.3286-3293, 2014. Article (CrossRef Link)

[45] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3232-3242, 2010. Article (CrossRef Link)
[46] N. Tong, H. Lu, L. Zhang, X. Ruan., M. Yang, "Salient object detection via bootstrap learning," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1884-1892, 2015. Article (CrossRef Link)
[47] D.Gao, V.Mahadevan, and N.Vasconcelos, "On the plausibility of the discriminant center-surround hypothesis for visual saliency," *Journal of Vision*, vol.8, no.7, pp.1-18, 2008. Article (CrossRef Link)
[48] J. Harel, C. Koch, P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, pp.545-552, 2006.
[49] D. Martin, C. Fowlkes, J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.26, no.5, pp.530-549, 2004. Article (CrossRef Link)
[50] N. OTSU, "A threshold selection method from gray level histograms," *IEEE Transactions on Systems Man & Cybernetics*, vol.9, no.1, pp.62-66, 1979. Article (CrossRef Link)
[51] S. Rao, H. Mobahi, A.Yang, S. Sastry, and Y. Ma, "Natural image segmentation with adaptive texture and boundary encoding," *In Proc. ACCV*, pp.135-146, 2009. Article (CrossRef Link)
[52] T. M. Cover and J. A. Thomas, "Elements of Information Theory (Telecommunications)," New York, NY, USA: Wiley, 1991.
[53] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp.888-905, Aug. 2000. Article (CrossRef Link)
[54] R. Horn and C. Johnson, "Matrix Analysis," Cambridge, U.K.: Cambridge Univ. Press, 1985.
[55] H. Li, H. Lu, Z. Lin, et al., "Inner and inter label propagation: salient object detection in the wild," *IEEE Transactions on Image Processing*, vol. 24, no.10, pp.3176-3186, 2015. Article (CrossRef Link)
[56] N. Tong, H. Lu, L. Zhang, X. Ruan, "Saliency detection with multi-scale superpixels," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp.1035-1039, 2014. Article (CrossRef Link)
[57] Qin Y, Lu H, Xu Y, et al, "Saliency detection via cellular automata," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.110-119, 2015. Article (CrossRef Link)
[58] P. Jiang, N. Vasconcelos, J. Peng, "Generic promotion of diffusion-based salient object detection," in *Proc. of 14th IEEE International Conference on Computer Vision (ICCV)*, pp.217-225, 2015. Article (CrossRef Link)
[59] N. Tong, H. Lu, Y. Zhang, et al, "Salient object detection via global and local cues," *Pattern Recognition*, vol.48, no.10, pp.3258-3267, 2015. Article (CrossRef Link)
[60] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp.315-327, 2011. Article (CrossRef Link)
[61] D. Martin, C. Fowlkes, D. Tal, et al, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. of 7th IEEE International Conference on Computer Vision (ICCV)*, pp.416-423, 2001. Article (CrossRef Link)
[62] M. Cheng, N. Mitra, X. Huang, P. Torr, et al, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp.569-582, 2015. Article (CrossRef Link)

Jingbo Zhou received his PhD degree in control science and engineering from Nanjing University of Science and Technology (NUST) in 2013. He is a lecturer at Nanjing Institute of Technology. His research interests include pattern recognition, image processing, data clustering, etc.

Jiyou Zhai is a doctoral student majoring in College of Computer and Information at Hohai University from September 2012 to present. He is a lecturer at Nanjing Institute of Technology. His research interests include pattern recognition, image processing, etc.

Yongfeng Ren received his PhD degree in College of Computer and Information from Hohai University in 2016. He is a lecturer at Nanjing Institute of Technology. His research interests include pattern recognition, image processing, etc.



Ali Lu received her PhD degree in control science and engineering from Nanjing University of Science and Technology in 2012. She is a lecturer at Nanjing Institute of Technology. Her research interests include computer vision, pattern recognition, etc.