Journal of Visual Communication and Image Representation 53 (2018) 113–121

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

# Saliency detection via bi-directional propagation☆

Yingyue Xu, Xiaopeng Hong, Xin Liu, Guoying Zhao*

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

ARTICLE INFO

*Keywords:*
Saliency detection
Bi-directional
Propagation

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.

## 1. Introduction

Recently, saliency detection [1–25], 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], and human behaviour analysis [29]. 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]), 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 [10–12].

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 [15–19,31,20–22,32], making different assumptions, for example, boundary prior assumption [22,33,34].

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 [35–38], 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.

There are two types of propagation: foreground propagation and background propagation. Foreground propagation (Fig. 1.1) selects the most salient values from a coarse 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(Fig. 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 [35–37], but recent saliency models [13,38] firstly compute a coarse saliency map by background propagation to obtain

☆ This paper has been recommended for acceptance by Zicheng Liu.
* Corresponding author.
*E-mail addresses:* Yingyue.Xu@oulu.fi (Y. Xu), Xiaopeng.Hong@oulu.fi (X. Hong), Xin.Liu@oulu.fi (X. Liu), Guoying.Zhao@oulu.fi (G. Zhao).

https://doi.org/10.1016/j.jvcir.2018.02.015
Received 11 April 2017; Received in revised form 5 January 2018; Accepted 26 February 2018
Available online 14 March 2018
1047-3203/ 

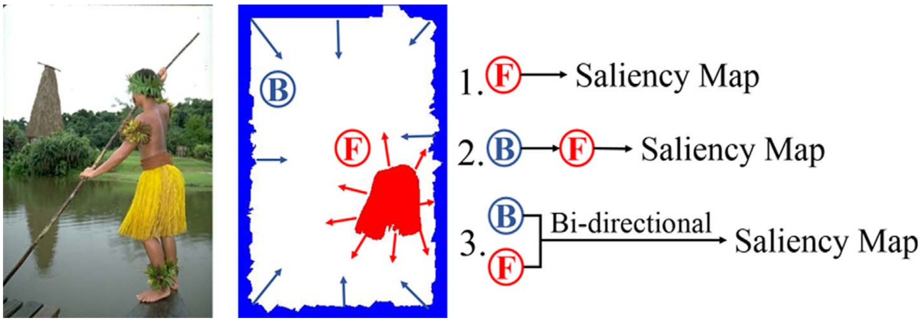


**Fig. 1.** There are two directions in which the propagation methods spread the selected seeds to the whole image, including foreground propagation Ⓕ from the most salient seeds and background propagation Ⓑ from the most unsalient seeds. Most previous propagation methods are single directional with 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.

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(Fig. 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 Fig. 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 $N_k, k = 1,...,K$. 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 (see Fig. 3):

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

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 $\mu_n$ and $\mu_m$ on CIELab-XY features, where $\mu_n = [\varphi_n^T; x_n; y_n]^T$, $x_n$ and $y_n$ are the

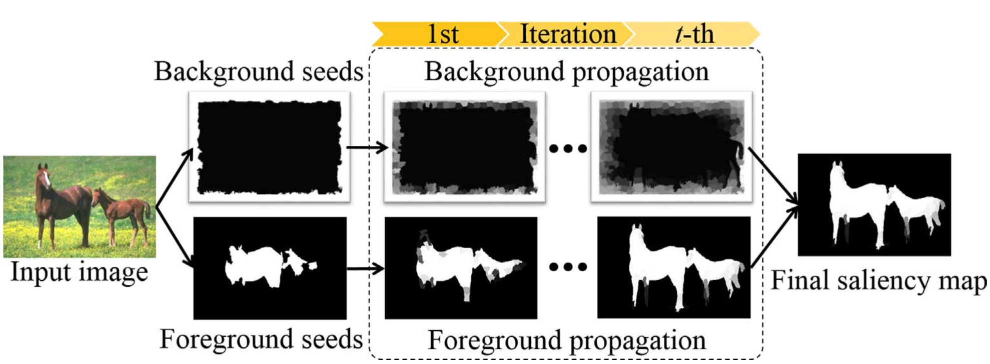


**Fig. 2.** Framework of the proposed BIP model. Given an input image, both foreground seeds and background seeds are chosen as two initial labeled sets. In each iteration, the unlabeled superpixels are evaluated according to their difficulties to the labeled foreground set and the labeled background 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.