TWO-STAGE ABSORBING MARKOV CHAIN FOR SALIENT OBJECT DETECTION

Qing Zhang⋆, *Desi Luo*⋆, *Wenju Li*⋆, *Yanjiao Shi*⋆, *Jiajun Lin*†

⋆Shanghai Institute of Technology, Shanghai, 201418, China
† East China University of Science and Technology, Shanghai, 200237, China

ABSTRACT

We propose a simple but effective approach to detect salient objects by exploring both patch-level and object-level cues under the framework of absorbing Markov chain. Saliency detection is carried out in a two-stage scheme. In the first stage, we conduct random walk on absorbing Markov chain with coarsely selected background seeds in the boundary. The result is integrated with a objectness map which is generated by finding potential object candidates to boost the saliency detection for object completeness. And in the second stage, we use the refined background seeds computed by the first stage as absorbing nodes for the absorbing Markov chain to obtain the final saliency map. Experimental results on four publicly available datasets demonstrate the robustness and efficiency of our proposed approach against 8 state-of-the-art methods in terms of five performance criterions.

Index Terms— Salient object detection, absorbing Markov chain, background prior, objectness measure, superpixels

1. INTRODUCTION

Salient object detection is commonly interpreted as a process that estimates the probability of each pixel belonging to salient objects and accurately segments the complete salient object from a scene. It has attracted more and more attention in the recent years and served as an important reprocessing procedure for a wide range of image processing and computer vision and applications, such as content-based image retrieval[1], image resizing[2], object tracking[3] and scene classification[4].

Most recent salient object detection works essentially follow the contrast prior, either local or global. As the evaluation in [5] shows that the local contrast methods tend to highlight the high frequency component like object boundaries rather than the entire salient object. The contrast using features from the whole image alleviates the phenomenon caused by local contrast methods that the boundaries of object can be well identified whereas the interior is attenuated. More and more complementary priors are investigated to improve the detection performance, such as center prior, background prior and focusness prior.

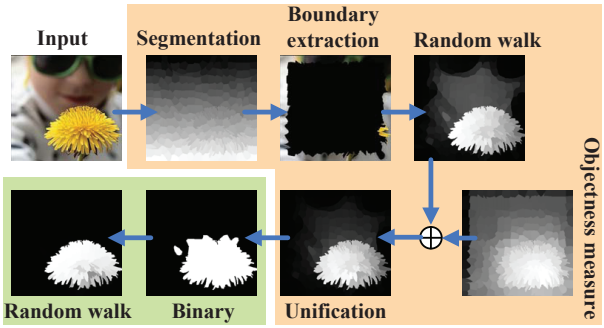


Fig. 1. The diagram of our proposed algorithm.

A simple but effective approach for salient object detection is to combine background pior and then to apply various strategies to form the saliency map. For example, it formulates saliency detection as a random walk problem in absorbing Markov chain taking all the boundary superpixels around the image borders as background absorbing nodes[6]. Graph-based manifold ranking approach exploits the background prior by using the nodes on each side of image[7]. However, the background selection is important and would affect the performance of saliency detection[8, 9]. It is too idealistic for most existing methods to simply use the whole background in the boundary because sometimes salient object touches the boundary. Though many methods improve the background prior[10], they are still insufficient for uniformly highlighting the salient region and suppressing the background. Besides, the majority of existing saliency detection algorithms are patch-level. It ignores a fact that the target object in salient object detection must be an object other than non-object. Although patch-level information performs well on retaining image details, cutting image into pieces has flaws in uniformly highlighting the whole object.

In this paper, we explore to couple patch-level and object-level cues into a unified framework to improve the performance of detecting salient object in a scene. The overview of our workflow is shown in Fig.1. In the first stage, we ex-

This work was partially supported by the National Natural Science Foundation of China (No.61401281) and Natural Science Foundation of Shanghai (No.14ZR1440700).

ploit the background prior inspired by [9]. With the coarse background seeds, we compute saliency detection as a random walk problem in the absorbing Markov chain. And then, we utilize a objectness cue to boost the saliency detection for object completeness. The two maps are then integrated. In the second stage, we apply binary segmentation on the resulted saliency map from the first stage, and take the labeled background seeds as absorbing nodes. The final saliency is computed based the refined background sees in the absorbing markov chain. The rest of our paper is organized as follows. Section 2 describes the details of our proposed method. The experimental results are given in Section 3 and we draw conclusions in Section 4.

2. PRINCIPLE OF ABSORBING MARKOV CHAIN

2.1. Graph construction

We construct a single layer graph $G = (V, E)$, where V is a set of nodes and E is a set of undirected edges. Each node is a superpixel generated by the SLIC algorithm. And each node is connected to the nodes which neighbor it or share common boundaries with its neighboring nodes. In addition, we enforce that all the transient nodes around the image borders are fully connected with each other. The weight w_{ij} of the edge e_{ij} between adjacent nodes i and j is defined as $w_{ij} = exp\{-||x_i - x_j||/\sigma^2\}$, where x_i and x_j are the mean of two nodes in the CIELab color space, and σ is a constant controls the strength of the weight.

2.2. Absorbing Markov chain

We adopt absorbing Markov chain to conduct the background-based saliency detection. Considering an absorbing Markov chain with r absorbing states and t transient states, renumber all the nodes so that the transient nodes come first, and then the transition matrix P has the following canonical form:

$$\mathrm{P} = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \tag{1}$$

where $Q \in [0,1]^{t\times t}$ is the transition probability matrix between transient nodes, $R \in [0,1]^{t\times r}$ is the transition probability matrix between the transient nodes and absorbing nodes, I is an $r \times r$ identity matrix.

The element of the affinity matrix A is defined as

$$a_{ij} = \begin{cases} w_{ij} & j \in N(i) \\ 1 & i = j \\ 0 & otherwise \end{cases} \tag{2}$$

where $N(i)$ denotes the nodes connected to node i. The degree matrix that records the sum of the weights connected to each node is written as $D = diag(\sum_j a_{ij})$. Finally, the transition matrix P on the sparsely connected graph is given as

$$P{=}\mathrm{D}^{-1} \times A \tag{3}$$

Given the transition matrix P, we can easily extract the matrix Q by Eq.1, based on which the fundamental matrix N is computed $N = (I - Q)^{-1}$, where the element n_{ij} of the fundamental matrix N is the expected number of times that the chain spends in transient state j given that it starts in transient state i. So $\sum_j n_{ij}$ represents the expected number of times before absorption, and the absorbed time vector for each transient state can be computed as:

$$y = N \times c \tag{4}$$

where c is a column vector all of whose elements are 1.

3. TWO-STAGE SALIENCY DETECTION

3.1. Saliency with preliminary background nodes

We use the boundary connectivity measure similar to [9] to select the preliminary background nodes in the boundary. Boundary connectivity is to quantify how heavily a region is connected to image boundaries.

We conduct an undirected weighted graph and compute the weight $d_{app}(p, q)$ of adjacent superpixels (p, q) as the Euclidean distance between their average colors in the CIELab color space. The geodesic distance between any two superpixels d_{geo} is defined as the accumulated edge weights along their shortest path on the graph

$$d_{geo}(p, q) = \min_{p=p_1,p_2,\cdots p_n=q} \sum_{i=1}^{n-1} d_{app}(p_i, p_{i+1}) \tag{5}$$

For convenience, we define $d_{geo}(p, p) = 0$. Then boundary connectivity of each superpixel p is defined as

$$BndCon(p) = \frac{\sum_{i=1}^{N} S(p, p_i) \cdot \delta(p_i \in Bnd)}{\sqrt{\sum_{i=1}^{N} S(p, p_i)}} \tag{6}$$

where N is the number of superpixels, $\sum_{i=1}^{N} S(p, p_i) = \sum_{i=1}^{N} \exp(-d^2_{geo}(p, p_i)/(2\sigma^2_{clr}))$, $\delta(\cdot)$ is 1 for superpixels on the image boundary and 0 otherwise and $\sigma_{clr} = 10$.

A background probability w_i^{bg} is introduced, which is mapped from the boundary connectivity value of superpixel p_i. It is close to 1 when boundary connectivity is large, and close to 0 when it is small. The definition is

$$w_i^{bg} = 1 - \exp(-\frac{BndCon^2(p_i)}{2\sigma^2_{bndCon}}) \tag{7}$$

where $\sigma_{bndCon} = 1$ and i indexes superpixel node on graph. We use w_i^{bg} to select the coarse background nodes in the

boundary. For the superpixel p_i in the boundary of d-pixel width, when it is satisfied with the following inequality, it would be selected as the background nodes.

$$\frac{w_i^{bg} - (\sum_{i=1}^{K} w_i^{bg})/K}{(\sum_{i=1}^{K} w_i^{bg})/K - \min(w_1^{bg}, w_2^{bg}, ..., w_K^{bg})} < \beta \quad (8)$$

where K is the superpixel number in the boundary of d-pixel width, $d = 5$ and $\beta = 0.8$ in our experiments. Thus, the preliminary background superpixels in the boundary are obtained.

We obtain the saliency map S by computing the absorbed time y computed by Eq.4 with the selected background superpixels as absorbing nodes, that is

$$S = norm(y) \quad (9)$$

where $norm(\cdot)$ means that it is normalized to the range between 0 and 1.

3.2. Objectness map

Alexe *et al.*[11] propose a trained method based on low-level cues to estimate an objectness score for any random image window, which measures the likelihood of the window containing a complete object. In our proposed algorithm, we use this objectness algorithm to coarsely localize the position of the salient object. The parameters are set to be default as in [11].

Given the image I, we perform objectness detection with K top object candidates, denoted by $(B_1, b_1), (B_2, b_2), ..., (B_K, b_K)$ where B_K is the bounding box and b_K is the corresponding confidence score. The objectness map is obtained by computing each pixel p's objectness score as the square root of the summed squares of the scores from all the bounding boxes that covers it, weighted by a Gaussian function for smoothness:

$$O(p) = \sqrt{\sum_{k=1}^{K} b_k^2 I(p \in B_k) \exp\{-d(p, B_i)/\sigma\}} \quad (10)$$

where $K = 500$ in our experiments, $I(p \in B_k)$ is the 0-1 indicator function denoting whether p is inside the bounding box, and $d(p, B_k)$ is the normalized distance between the pixel p and the center of the bounding box B_i measured in a scaled coordinate space where the bounding box is normalized to length 1 on both axes.

The saliency map generated by the absorbing Markov chain and the objectness map are integrated to produce the first-stage saliency detection result by the following process:

$$S_1 = S \times O(p) \quad (11)$$

3.3. Saliency with refined background nodes

The saliency map of the first stage is binary segmented using an adaptive threshold, which facilitates selecting the refined background as absorbing nodes. The threshold is set as the mean saliency over the entire saliency map.

Once the absorbing nodes are given, the absorbed time vector y can be computed by Eq.4. As carried out in the first stage, it is normalized between the range of 0 and 1 to form the final saliency map S_2.

4. EXPERIMENT

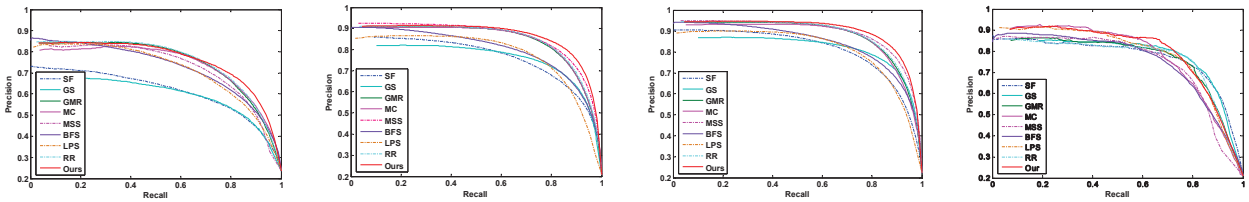
To evaluate the effectiveness and robustness of our proposed algorithm, we conduct experiments on four publicly available benchmark datasets: ECSSD[12], MSRA5K[13], MSRA10K[14] and SED2[15]. ECSSD contains 1,000 semantically meaningful but structurally complex images. MSRA5K and MSRA10K respectively contains 5,000 and 10,000 natural images, which makes it very challenging and also suitable for more comprehensive mode evaluation. Also with complex surroundings, SED2 contains 100 images with two salient object in each image. We compare our method with 8 state-of-the-art methods, including SF[16], GS[8], GMR[7], MC[6], MSS[17], BFS[10], LPS[18], and RR[19]. For comprehensive evaluation, we use five metrics including the precision-recall (PR) curve, F-measure, AUC, mean absolute error (MAE)[19] and the weighted F-measure (WF-measure) scores[20].

4.1. Quantitative evaluation

The PR curves, F-measure, WF-measure, AUC and MAE are respectively shown in Fig.2, Fig.3, Table 1 and Table2. In the PR curves (Fig.2), our method is comparable to the competitive methods. In terms of AUC score, our approach ranks second on SED2 (Table 1), and achieves the best performance on other three datasets. As reported in Table 2 and Fig.3, our model performs best in terms of MAE, F-measure and weighted F-measure on all the four datasets. These results validate our proposed approach's strong potential in handling images with complex surroundings.

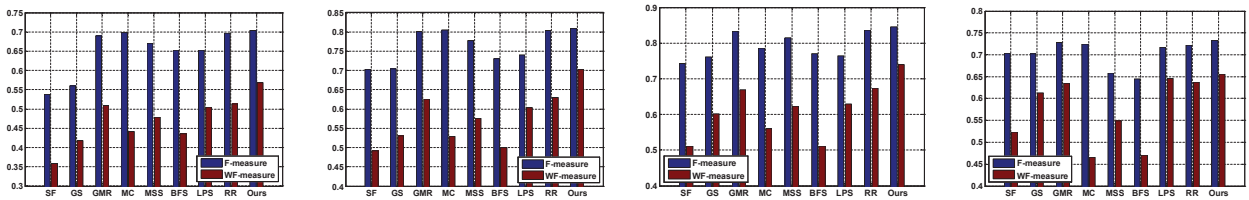
4.2. Visual comparison

Fig.3 shows some visual comparisons of best methods in the quantitative experiments. For images with simple background (see first exemplar), our model nearly accurately extracts the entire salient object and assigns uniform saliency values to all the patches within the salient object. For images with textured background (see third exemplar), the other methods incorrectly include background regions into detection result. By contrast, our model pops out all the salient object successfully. For images with complex scene (see second and last exemplars), our model locates salient objects with decent accuracy.



(a) ECSSD (b) MSRA5K (c) MSRA10K (c) SED2

Fig. 2. Quantitative comparison of saliency maps on four datasets in terms of PR curves.



(a) ECSSD (b) MSRA5K (c) MSRA10K (c) SED2

Fig. 3. Quantitative comparison of saliency maps on four datasets in terms of F-measure and weighted F-measure.

Table 1. AUC scores.(Higher is better. The top three models are highlighted in red, green and blue).

Approach	SF	GS	GMR	MC	MSS
ECSSD	0.817	0.863	0.891	0.911	0.903
MSRA5K	0.918	0.935	0.939	0.943	0.958
MSRA10K	0.905	0.946	0.944	0.951	0.964
SED2	0.876	0.917	0.870	0.886	0.866
	BFS	LPS	RR	Ours	
ECSSD	0.902	0.879	0.896	0.917	
MSRA5K	0.937	0.915	0.942	0.959	
MSRA10K	0.945	0.919	0.948	0.964	
SED2	0.889	0.877	0.871	0.897	

Table 2. MAE scores.(lower is better. The top three models are highlighted in red, green and blue).

Approach	SF	GS	GMR	MC	MSS
ECSSD	0.228	0.225	0.186	0.203	0.192
MSRA5K	0.171	0.157	0.128	0.146	0.134
MSRA10K	0.175	0.148	0.126	0.145	0.128
SED2	0.193	0.168	0.174	0.193	0.171
	BFS	LPS	RR	Ours	
ECSSD	0.206	0.185	0.184	0.167	
MSRA5K	0.167	0.134	0.126	0.108	
MSRA10K	0.174	0.139	0.124	0.110	
SED	0.200	0.159	0.173	0.159	

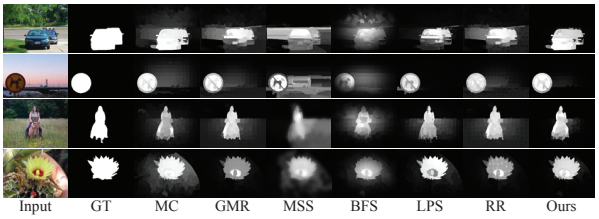


Fig. 4. Qualitative comparisons of our approach with state-of-the-art methods.

These results illustrate the effectiveness and roubustness of our proposed method.

5. CONCLUSION

In this paper, we have presented a salient object detection model that utilizes the patch-level and object-level information to obtain better discovery of salient objects. We adopt a two-stage approach with the coarse and refined background superpixels as absorbing nodes to generate the saliency maps. Experiments on four public datasets have shown that our model achieves encouraging performance compared to the state-of-the-art methods.

6. REFERENCES

[1] X. Yang, X. Qian, and Y. Xue, "Scalabel mobile image retrieval by exploring contextual saliency," *IEEE TIP*, vol. 24, pp. 1709–1721, 2015.

[2] R. Margolin, L. Zelnik-Manor, and A. Tal, "Saliency for image manipulation," *TVC*, vol. 29, pp. 381–392, 2013.

[3] J. Li, M.D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale space analysis in the frequency domain," *IEEE TPAMI*, vol. 35, pp. 996–1010, 2013.

[4] F. Zhang, B. Du, and L. Zhang, "Saliency guided unsupervised feature learning for scene classification," *IEEE TGRS*, vol. 53, pp. 2175–2184, 2015.

[5] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE TIP*, vol. 24, pp. 414–429, 2015.

[6] B. Jiang, L. Zhang, H. Lu, and M.H. Yang, "Saliency detection via absorbing markov chain," in *ICCV*. IEEE, 2013, pp. 2976–2983.

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

[8] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *ECCV*. IEEE, 2012, pp. 29–42.

[9] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *CVPR*. IEEE, 2014, pp. 2814–2821.

[10] J. Wang, H. Lu, X. Li, N. Tong, and W. Liu, "Saliency detection via background and foreground seed selection," *Neurocomputing*, vol. 152, pp. 359–368, 2015.

[11] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE TPAMI*, vol. 34, pp. 2189–2202, 2012.

[12] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *CVPR*. IEEE, 2013, pp. 1155–1162.

[13] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: a discriminant regional feature integration approach," in *CVPR*. IEEE, 2013, pp. 2083–2090.

[14] M.-M. Cheng, N.J. Mitra, X. Huang, P.H. Torr, and S. Hu, "Global contrast based salient region detection," *IEEE TPAMI*, vol. 37, pp. 569–582, 2015.

[15] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *CVPR*. IEEE, 2007, pp. 1–8.

[16] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: contrast based filtering for salient region detection," in *CVPR*. IEEE, 2012, pp. 733–740.

[17] N. Tong, H. Lu, L. Zhang, and X. Ruan, "Saliency detection with multi-scale superpixels," *IEEE Signal Proc Let*, vol. 21, pp. 1035–1039, 2014.

[18] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation salient object detection in the wild," *IEEE TIP*, vol. 24, pp. 3176–3186, 2015.

[19] C. Li, Y. Yuan, W. Cai, Y. Xia, and D.D. Feng, "Robust saliency detection via regularized random walks ranking," in *CVPR*. IEEE, 2015, pp. 2710–2717.

[20] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps," in *CVPR*. IEEE, 2014, pp. 248–255.