SALIENT OBJECT DETECTION VIA DEFORMED SMOOTHNESS CONSTRAINT

Xiyin Wu [a,b], Xiaodi Ma [a,b], Jinxia Zhang [c], Andong Wang [a,b] and Zhong Jin [a,b,]*

[a]School of Computer Science & Engineering, Nanjing University of Science & Technology, China
[b]Key Lab of Intelligent Perception & Systems for High-Dimen. Info. of Ministry of Education, China
[c]School of Automation, Southeast University, China

ABSTRACT

In recent years, various graph-based salient object detection methods have been successfully proposed. Since existing methods may miss some object regions with low contrast to background, a novel propagation model via deformed smoothness constraint is proposed to address this problem. By regularizing nodes and their neighbors locally, the deformed smoothness constraint is able to prevent erroneous label propagation. Thus, the object regions with low contrast to background can be emerged. Besides, the deformed smoothness constraint is further utilized in a map refinement model, which can suppress the background noises in label propagation result. Experiments on three public datasets show that the proposed method outperforms eleven state-of-the-art salient object detection methods.

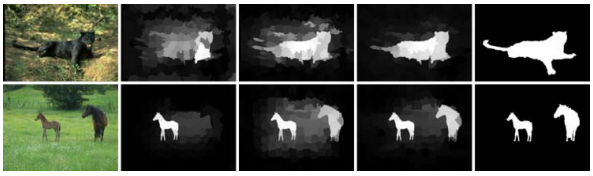
Index Terms— Salient object detection, deformed smoothness constraint, label propagation, map refinement

1. INTRODUCTION

Salient object detection aims to identify important objects in a scene. It can be viewed as a preprocessing step to reduce the amount of computations in numerous applications, such as image and video compression [1], image retrieval [2], visual tracking [3] and video retargeting [4].

The graph-based salient object detection methods have attracted much attention recently for their simplicity and efficiency [5, 6]. To conduct salient object detection, an image is partitioned into superpixels and described as a graph, where superpixels are represented as nodes and connected to their adjacent nodes by weighted edges. Saliency information is diffused over the graph by seeds and a propagation model [7]. The propagation model is generally based on cluster assumption and smoothness assumption [8]. The former means that nodes in the same cluster are likely to have the same label. The latter describes that nodes on the same manifold structure are likely to have the same label. It has been widely observed that a solution under the smoothness assumption can achieve higher accuracy [9].

Although the performances of graph-based methods have achieved state-of-the-art results, most of them still have two common limitations: 1) Most label propagation models based on smoothness assumption cannot handle the situation of missing object regions which have low contrast to background (see Fig. 1(b)). These models usually adopt a standard smoothness [10] to constrain the labels of every pair of nodes based on their similarities. Since the standard smoothness evaluates the smoothness globally, it may miss the local information of missing regions. Thus, the nodes belong to missing regions cannot be propagated reliably. 2) Background noises may occur in the label propagation result. They need to be suppressed to improve the accuracy of result.



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

Fig. 1. Examples of saliency maps. (a) Input images, (b) label propagation based on MR, (c) label propagation based on proposed DSMR, (d) map refinement of (c), (e) Ground truth.

In this paper, a novel method is proposed to address the problems of graph-based methods. A deformed smoothness (DS) constraint is used in the label propagation model. Different from the standard smoothness, DS evaluates the smoothness of nodes with their neighbors locally. It can prevent erroneous label propagation of the nodes which have low contrast to background. Besides, a map refinement process is added by utilizing the label propagation result and a visual cue called objectness. This process can suppress most background noises. The contributions of this paper are threefold: 1) A new propagation model called DS-based manifold ranking (DSMR) is proposed. 2) DSMR is applied in salient object detection to generate a coarse map. 3) The coarse map is further refined by a new model to obtain the refined map, which can make the results more accurate.

2. METHODOLOGY

The diagram of the proposed method is shown in Fig. 2. Firstly, an image is segmented into superpixels and represented as a graph. Secondly, a coarse map is generated

*Corresponding author: zhongjin@njust.edu.com. This work is supported by National Natural Science Foundation of China (Nos. 61602244, 61702262, 61602444, 91420201, 61472187, 61703100).

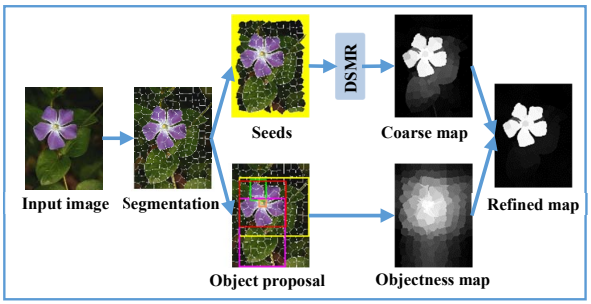


Fig. 2. The diagram of the proposed method.

via background seeds and the proposed propagation model (DSMR). Moreover, an objectness map is constructed by object proposal. Lastly, a refined map is generated by integrating the information of coarse map and objectness map. All the details of the proposed method are introduced in the following subsections.

2.1 Graph representation

An image I is segmented into n superpixels using SLIC algorithm [11]. Then an undirected weighted graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ is constructed, where $\mathcal{V}=\{v_1,\ldots,v_n\}$ represents the set of nodes corresponding to superpixels and $\mathcal{E}$ represents the set of edges. The edges $\mathcal{E}=\mathcal{E}_1\bigcup\mathcal{E}_2\bigcup\mathcal{E}_3$ are added with three rules:

$$\begin{aligned}&\text{Rule 1}:\mathcal{E}_1=\{(v_i,v_j)\mid v_j\in\mathcal{N}_i^s\vee(\exists v_k\in\mathcal{N}_i^s,v_j\in\mathcal{N}_k^s)\}\\&\text{Rule 2}:\mathcal{E}_2=\{(v_i,v_j)\mid v_i,v_j\in\mathcal{B}\}\\&\text{Rule 3}:\mathcal{E}_3=\{(v_i,v_j)\mid v_j\in\mathcal{N}_i^c\}\end{aligned}\tag{1}$$

where $\mathcal{B}$ is a set of boundary nodes at the image border. $\mathcal{N}_i^s$ denotes the immediate spatial neighbors of v_i. $\mathcal{N}_i^c$ is the cluster of v_i in Lab color space computed by APC [12]. Rule 1 links v_i with its geometric neighbors [13]. Since the background regions are usually close to the image boundaries, the nodes close to boundaries are considered to be proximal. The arbitrary boundary nodes are connected with each other based on Rule 2 [13]. Rule 3 connects v_i with its color similar nodes.

The edge weight matrix $\mathbf{W}=(w_{ij})_{n\times n}$ is a symmetrical matrix representing similarity between graph nodes. The weight entry w_{ij} between v_i and v_j is defined as follows:

$$w_{ij}=\exp\left(-\frac{\left\|\mathbf{c}_i-\mathbf{c}_j\right\|+\left\|\sin(\pi\mid\mathbf{d}_i-\mathbf{d}_j\mid)\right\|}{\delta^2}\right)\tag{2}$$

where $\sin(\cdot)$ computes the sine function of a vector element-wisely. δ is a constant controlling the strength of the weight. In Eq. (2), the first term is the Euclidean distance between v_i and v_j in the Lab color space, while the second term is geometrical distance computed in sine space [14].

2.2 Label propagation based on deformed smoothness

The label propagation model has different forms [15], in which the manifold ranking (MR) has been proved to achieve good performance in terms of effectiveness and efficiency [10]. The goal of MR is to compute a rank vector $\mathbf{f}=(f_i)_n$ with respect to an indication vector $\mathbf{z}=(z_i)_n$, where $z_i=1$ if v_i is a seed, and $z_i=0$ otherwise. It can be formulated as:

$$\begin{aligned}&\min_{\mathbf{f}}\left\{S(\mathbf{f})+\mu_1R(\mathbf{f})\right\}\\&=\frac{1}{2}\left\{\mathbf{f}^{\mathrm{T}}(\mathbf{D}-\mathbf{W})\mathbf{f}+\mu_1(\mathbf{f}-\mathbf{D}^{-1}\mathbf{z})^{\mathrm{T}}\mathbf{D}(\mathbf{f}-\mathbf{D}^{-1}\mathbf{z})\right\}\end{aligned}\tag{3}$$

where $\mathbf{D}=\mathrm{diag}(d_{11},\ldots,d_{nn})$ is a degree matrix with $d_{ii}=\sum_j w_{ij}$. μ_1 is a parameter controlling the balance of the standard smoothness constraint $S(\mathbf{f})$ and the label fitness constraint $R(\mathbf{f})$.

However, MR may miss some salient object regions since the nodes of these regions have low contrast to background, as shown in Fig. 1(b). The local smoothness $S^L(\mathbf{f})$ is introduced to solve this issue:

$$S^L(\mathbf{f})=\mathbf{f}^{\mathrm{T}}(\mathbf{I}-\frac{\mathbf{D}}{v})\mathbf{f}\tag{4}$$

where $v=\sum_i d_{ii}$ is the volume of graph $\mathcal{G}$. Different from $S(\mathbf{f})$ which evaluates the smoothness between pairs of nodes over the entire image, $S^L(\mathbf{f})$ considers the label smoothness of nodes with their neighbors as a whole [16]. Therefore, the missing regions with large d_{ii} (strong connection with their neighbors) acquire a confident soft ranking value and vice versa. Combining the standard and local smoothness, the DS constraint is formulated as:

$$S^D(\mathbf{f})=S(\mathbf{f})+\mu_2S^L(\mathbf{f})=\mathbf{f}^{\mathrm{T}}[\mathbf{D}-\mathbf{W}+\mu_2(\mathbf{I}-\frac{\mathbf{D}}{v})]\mathbf{f}\tag{5}$$

where μ_2 is a non-negative parameter to balance the weights of two smoothness constraints.

Replacing the standard smoothness constraint with the DS constraint, a label propagation model called DSMR is proposed as follows:

$$\begin{aligned}&\min_{\mathbf{f}}\left\{S^D(\mathbf{f})+\mu_1R(\mathbf{f})\right\}\\&=\frac{1}{2}\left\{\mathbf{f}^{\mathrm{T}}[\mathbf{D}-\mathbf{W}+\mu_2(\mathbf{I}-\frac{\mathbf{D}}{v})]\mathbf{f}+\mu_1(\mathbf{f}-\mathbf{D}^{-1}\mathbf{z})^{\mathrm{T}}\mathbf{D}(\mathbf{f}-\mathbf{D}^{-1}\mathbf{z})\right\}\end{aligned}\tag{6}$$

The first term of Eq. (6) is the DS constraint, which indicates that the ranking values of two nearby nodes should not change too much. Meanwhile, it considers local smoothness which regularizes nodes and their nearby nodes locally. The second label fitness constraint guarantees the ranking values of seeds should not differ too much from their initial values. By setting the derivate of Eq. (6) to be zero, the optimal solution can be written as:

$$\mathbf{f}=[\mathbf{D}-\alpha\mathbf{W}+\beta(\mathbf{I}-\frac{\mathbf{D}}{v})]^{-1}\mathbf{z}\tag{7}$$

where $\alpha=1/(1+\mu_1)$ and $\beta=\mu_2/(1+\mu_1)$.

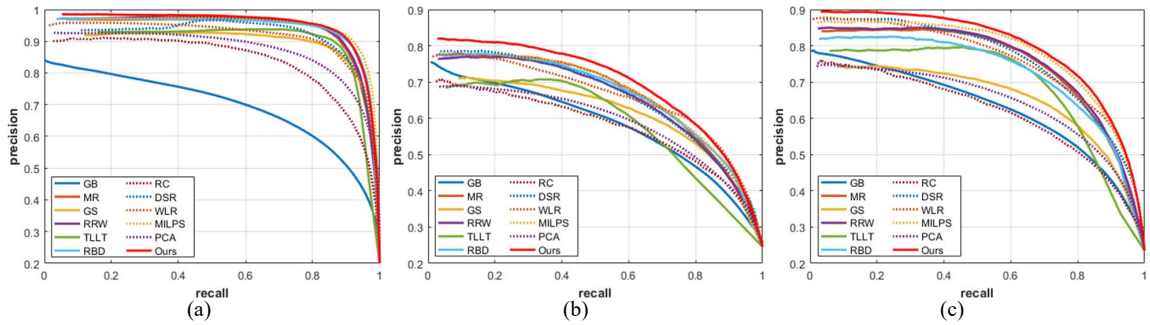


Fig. 3. Quantitative comparison of PR curves on three datasets. (a) ASD, (b) PASCAL-S, (c) ECSSD. (Please view in color)

The separation/combination (SC) strategy [17] is used to generate a coarse map in this paper. That is, four propagation results $\mathbf{f}(\mathrm{T})$, $\mathbf{f}(\mathrm{D})$, $\mathbf{f}(\mathrm{L})$ and $\mathbf{f}(\mathrm{R})$ are obtained by taking the top, down, left and right boundary nodes as seeds, respectively. Then the four propagation results are combined to generate the coarse map $\mathbf{M}^c = (m_i^c)_n$:

$$m_i^c = \prod_{k\in\{\mathrm{T,D,L,R}\}} \left[1-\bar{\mathbf{f}}_i(k)\right] \tag{8}$$

where $\bar{\mathbf{f}}$ denotes the normalized vector of $\mathbf{f}$. As shown in Fig. 1(c), DSMR could help to detect the missing regions with low contrast to background.

2.3 Map refinement based on deformed smoothness

Although most salient regions can be highlighted in $\mathbf{M}^c$, the coarse map usually suffers from background noises (see Fig. 1(c)). Therefore, other information should be combined to further improve the performance of $\mathbf{M}^c$.

Objectness [18] is a useful information to locate the object regions and exclude the background regions. It can generate a set of bounding boxes or regions, which are likely to conclude the objects of image. Edge box [19] generates object bounding box proposals from edges. In our paper, the edge box is used to obtain the objectness map $\mathbf{M}^o = (m_i^o)_n$ for each node:

$$m_i^o = \sum_{j=1}^{B} R_j \cdot \delta(v_i \in \Omega_j) \tag{9}$$

where R_j is the score of bounding box Ω_j and $\delta(\cdot)$ is an indicator function denoting whether v_i is inside the bounding box. Generally, the edge box outputs thousands of bounding boxes and some of them cannot locate the objects well. The reasonable object proposals are selected as [20] and the number of object proposals is B.

Most of the previous works combine information of maps using weighted summation or multiplication [6], which is heuristic and hard for generalization. In this paper, a map refinement model is proposed to combine the information of $\mathbf{M}^o$ and $\mathbf{M}^c$. The map refinement model can generate the refined result $\mathbf{g}$ by solving the following optimization problem:

$$\min_{\mathbf{g}} \frac{1}{2}\left\{\mathbf{g}^{\mathrm{T}}[\mathbf{D}^c - \mathbf{W}^c + \mu_2(\mathbf{I} - \frac{\mathbf{D}^c}{v^c})]\mathbf{g} + \left\|\mathbf{g}-\mathbf{M}^c\right\|^2 + \mathbf{g}^{\mathrm{T}}\mathbf{D}^o\mathbf{g}\right\} \tag{10}$$

where $\mathbf{D}^o = (d_{ii}^o)_{n\times n}$ represents a diagonal matrix with $d_{ii}^o = \mathrm{diag}(\exp(-m_i^o))$. $\mathbf{W}^c = (w_{ij}^c)_{n\times n}$ is a new weight matrix computed by $\mathbf{M}^c$:

$$w_{ij}^c = \exp\left(-\frac{\left\|m_i^c - m_j^c\right\|}{\delta^2}\right) \tag{11}$$

$\mathbf{D}^c$ and v^c are the degree matrix and the volume of $\mathbf{M}^c$ respectively. The three terms of Eq. (10) define costs from different constraints. The first one is the DS constraint which encourages continuous saliency values. The second one is a fitness constraint, which means the refined result $\mathbf{g}$ should not change too much from the coarse map $\mathbf{M}^c$. The third one is a regularization constraint constructed by the objectness map. This term can suppress the background regions which may not belong to object proposals, and highlight the regions probably to be a part of objects. The optimal solution of Eq. (10) is expressed as:

$$\mathbf{g} = [\mathbf{D}^c - \mathbf{W}^c + \mu_2(\mathbf{I} - \frac{\mathbf{D}^c}{v^c}) + \mathbf{D}^o]^{-1}\mathbf{M}^c \tag{12}$$

The elements of $\mathbf{g}$ are normalized to $[0,1]$ and assigned to the corresponding superpixels to generate the refined map $\mathbf{M}^r$. As shown in Fig. 1(d), by combining the information of coarse map and objectness map, the refined map has a better result in highlighting the objects and restraining the background noises.

3. EXPERIMENTAL RESULTS

The salient object detection performance of the proposed method is validated on three standard datasets: ASD [21], PASCAL-S [22] and ECSSD [23]. ASD is a widely used dataset which contains 1,000 images. Most images in ASD have only one object and their background are relatively simple. PASCAL-S contains 850 images with multiple objects. ECSSD is a challenging dataset which contains 1,000 images with semantically meaningful but structurally complex scenes.

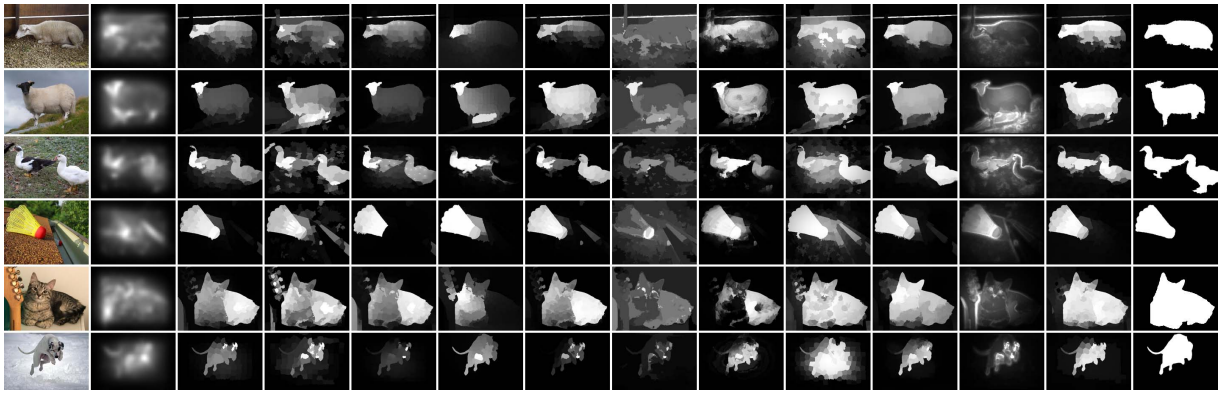


Fig. 4. Visual comparisons of saliency maps. GT is short for Ground Truth.

Two metrics are used to evaluate the performance: precision-recall (PR) curve and F-measure. The saliency map is segmented with varying thresholds, and compared the binary map with ground truth mask to compute the values of PR curve. To evaluate the quality of a saliency map comprehensively, the F-measure considers as a weighted harmonic mean of precision and recall [24].

3.1 Comparison with state-of-the-art methods

The compared eleven state-of-the-art methods are: GB [7], MR [17], GS [25], RRW [13], TLLT [5], RBD [6], RC [26], DSR [27], WLR [24], MIL [20] and PCA [28]. Fig. 3 and Table 1 show the PR curves and F-measure on three benchmark datasets respectively. The proposed method outperforms all other methods both on relatively simpler ASD dataset, and on more challenging PASCAL-S and ECSSD datasets. Moreover, the PR curves show that the proposed method has a higher minimum precision value compared with other methods. That means the regions belong to objects can be better highlighted in saliency maps.

Visual comparisons of saliency maps are shown in Fig. 4. The saliency maps generated by our method are more consistent with the ground truth. Our method not only tackles the situation that the salient objects have low contrast to background, but also suppresses the background noises preferably.

3.2 Evaluation of the effectiveness of proposed method

Experiments are conducted to verify the effectiveness of the proposed method on ECSSD dataset. The effectiveness of the proposed DSMR model and the map refinement model are evaluated. The DS constraint and the refinement process are removed from our method respectively, thus three situations are obtained: method without DS constraint and refinement process, method without DS constraint, method without refinement process. Fig. 5 demonstrates that both DS constraint and refinement process make significant contributions to our method.

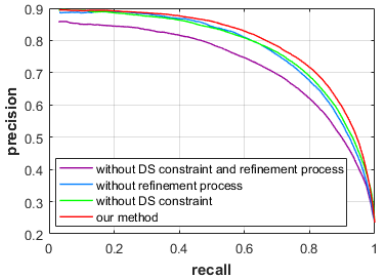


Fig. 5. Evaluation of the effectiveness of our method.

Table 1. Quantitative comparison of F-measure.

Models	ASD	Pascal-S	ECSSD	Average
GB	0.6237	0.4963	0.5497	0.5565
GS	0.8280	0.5588	0.6082	0.6650
MR	0.8962	0.6023	0.6888	0.7291
RRW	0.8966	0.5870	0.6967	0.7268
TLLT	0.8806	0.5325	0.6658	0.6930
RBD	0.8835	0.596	0.6762	0.7186
RC	0.6803	0.4009	0.4546	0.5119
DSR	0.8559	0.5845	0.6901	0.7102
WLR	0.8116	0.5699	0.6423	0.6746
PCA	0.7968	0.5254	0.5777	0.6333
MIL	0.8975	0.6002	0.7132	0.7370
Ours	**0.9042**	**0.6053**	**0.7185**	**0.7427**

4. CONCLUSION

In this paper, a label propagation model called DSMR is proposed for salient object detection. While the MR model may not handle the situation of missing regions with low contrast to background, the DSMR model is proposed to prevent erroneous label propagation by suppressing the ranking values of nodes belonging to missing regions. Moreover, a map refinement model is developed by taking advantages of objectness and DS constraint. Experimental results show that the proposed method outperforms eleven state-of-the-art saliency detection methods. The future work will focus on analyzing the potential of DSMR to build better label propagation model.

5. REFERENCES

[1] C. Guo, and L. Zhang, "A novel multi resolution spatiotemporal saliency detection model and its applications in image and video compression," IEEE TIP, vol. 19, no. 1, pp. 185–198, 2010.

[2] L. Li, S. Jiang, Z. Zha, Z. Wu, and Q. Huang, "Partial duplicate image retrieval via saliency-guided visually matching," IEEE Multimedia, vol. 20, no. 3, pp. 13–23, 2013.

[3] L. Zhang, M.-H. Tong, T.-K. Marks, H. Shan, and G.-W. Cottrell, "SUN: A bayesian framework for saliency using natural statistics," J. Vis, vol. 8, no. 7, pp. 1–20, 2008.

[4] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," ACM TOG, vol. 27, no. 3, pp. 443–453, 2008.

[5] C. Gong, D. Tao, W. Liu, S.-J. Maybank, M. Fang, K. Fu, and J. Yang, "Saliency propagation from simple to difficult," in CVPR, pp. 2531–2539, 2015.

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

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in NIPS, pp. 545–552, 2007.

[8] O. Chapelle, B. Schölkopf, and A. Zien, "Semi-supervised learning," IEEE TNN, vol. 20, no. 3, pp. 542–542, 2009.

[9] X. Zhu, and A.-B. Goldberg, "Introduction to semi-supervised learning," SLAIML, vol. 3, no. 1, pp. 1–130, 2009.

[10] L. Zhang, C. Yang, H. Lu, X. Ruan, and M.-H. Yang, "Ranking saliency," IEEE TPAMI, vol. 39, no. 9, pp. 1892–1904, 2017.

[11] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE TPAMI, vol. 34, no. 11, pp. 2274–2282, 2012.

[12] B.-J. Frey, and D. Dueck, "Clustering by passing messages between data points," Science, vol. 315, no. 5814, pp. 972–976, 2007.

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

[14] J. Zhang, K.-A. Ehinger, H. Wei, K. Zhang, and J. Yang, "A novel graph-based optimization framework for salient object detection," Pattern Recognition, vol. 64, pp. 39–50, 2017.

[15] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in CVPR, pp. 2790–2797, 2014.

[16] C. Gong, T. Liu, D. Tao, K. Fu, E. Tu, and J. Yang, "Deformed graph laplacian for semisupervised learning," IEEE TNNLS, vol. 26, no. 10, pp. 2261–2274, 2015.

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

[18] X. Cheng, J. Lu, J. Feng, J. Yuan, and J. Zhou. "Scene recognition with objectness". PR, vol. 74, pp. 474–487, 2018.

[19] C.-L. Zitnick, and P. Dollár, "Edge boxes: Locating object proposals from edges," in ECCV, pp. 391–405, 2014.

[20] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan, "Salient object detection via multiple instance learning," IEEE TIP, vol. 26, no. 4, pp. 1911–1922, 2017.

[21] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in CVPR, pp. 1597–1604, 2009.

[22] Y. Li, X. Hou, C. Koch, J.-M. Rehg and A.-L. Yuille, "The secrets of salient object segmentation," in CVPR, pp. 280–287, 2014.

[23] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended CSSD," IEEE TPAMI, vol. 38, no. 4, pp. 717–729, 2016.

[24] C. Tang, P. Wang, C. Zhang, and W. Li, "Salient object detection via weighted low rank matrix recovery," IEEE Signal Processing Letters, vol. 24, no. 4, pp. 490–494, 2017.

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

[26] M.-M. Cheng, G.-X. Zhang, N.-J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in CVPR, pp. 409–416, 2011.

[27] H. Lu, X. Li, L. Zhang, X. Ruan, and M.-H. Yang, "Dense and sparse reconstruction error based saliency descriptor," IEEE TIP, vol. 25, no. 4, pp. 1592–1603, 2016.

[28] Y. Li, X. Hou, C. Koch, J.-M. Rehg and A.-L. Yuille, "The secrets of salient object segmentation," in CVPR, pp. 280–287, 2014.