EXEMPLAR BASED IMAGE SALIENT OBJECT DETECTION

Zezheng Wang[1] *Rui Huang*[1,†] *Liang Wan*[2] *Wei Feng*[1]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] School of Computer Software, Tianjin University, Tianjin, China
{zzwang, ruihuang, lwan, wfeng}@tju.edu.cn

ABSTRACT

Saliency detection is an important problem. Researchers in this area mainly focus on advanced models to achieve high performance on benchmark datasets with a large number of labeled images. However, most conventional saliency detection methods only use these benchmark datasets for saliency evaluation. We argue that we can use these valuable labeled data to generate precise saliency results. In this paper, we propose to exploit these labeled data by retrieving labeled images with similar foreground to a query image as exemplars. Then we learn to generate precise saliency from these exemplars. We conduct extensive experiments on four benchmark datasets, and we compare our method with eleven state-of-the-arts. Experimental results show the promising performance improvements of our method over compared methods.

Index Terms— Saliency detection, exemplar-based, data-driven, multiscale superpixel

1. INTRODUCTION

Saliency detection, a considerable topic in computer vision, aims to detect conspicuous foreground objects from images or videos. This detection has been widely used for segmentation, recognition, compression and inpainting [1, 2, 3, 4, 5, 6].

Recently, numerous saliency detection models have achieved reasonable detection results on the existing labeled benchmark datasets (MSRA5000 [7], MSRA1000 [8], ECSSD [9], and PASCAL-S [10]). For example, the state-of-the-art saliency detection methods [11, 7, 12, 13, 14, 15] mainly pay much attention on developing novel saliency detection models. Note that these methods are based on well-designed models to achieve high detection performance and use these benchmark datasets for saliency evaluation. These datasets contain valuable label for precise saliency detection. How to exploit the valuable labeled datasets is an important issue.

Generally, the labeled images of benchmark datasets can be exploited in two ways. One way is to train classifiers on specifical dataset. A notable work [16] uses random forest regressor [17] to generate initial salient regions based on 2000 images of MSRA5000 [7]. Note, the classifiers of this kind of methods are often trained on specific dataset, which lacks the adaptive ability on wild images. Recently, deep learning is introduced in saliency detection. The representative work [18] trains a deep saliency network with $9,500$ training images ($6,000$ images from MSRA10K [19] and $3,500$ images from DUT-OMRON [13]). Success of this kind of saliency detectors is based on the complex deep models and large amount of training data. Another way is to detect class-specific top-down salient objects. For example, [20, 21, 22] train classifier with certain type of objects (i.e., cars, people, dogs) to obtain class-specific saliency, which constrains the use to general saliency detection.

Existing benchmark datasets contain valuable labeled images, which are beneficial for precise detection of salient objects. In this paper, we investigate the use of labeled images of these datasets. Specifically, we use randomly sampled 4,000 images of MSRA5000 [7] as exemplar set. For an image, we first search similar images by foreground region $\mathbf{B}^{\mathrm{C}}$. Afterward, we use the background of the input image as negative samples and the foreground of searched images as positive samples to generate initial saliency $\mathbf{B}^{\mathrm{R}}$. Both salient foregrounds $\mathbf{B}^{\mathrm{C}}$ and $\mathbf{B}^{\mathrm{R}}$ are refined in global image propagation, and Bayesian integration is subsequently performed to generate final single-scale saliency map $\mathbf{S}^{\mathrm{B}}$. We use our method with multiscale superpixel to obtain robust saliency detection. A similar idea has been proposed in [23], which uses unlabeled images retrieved from internet that must first generate the corresponding saliency detection by additional saliency detectors. Our major contribution is a method to exploit the valuable labeled images for saliency detection. Our method achieves superior performance compared with eleven state-of-the-arts.

2. SALIENCY DETECTION MODEL

The main steps of our saliency detector are shown in Fig. 1.

2.1. Candidate images selection

We randomly select 4000 images from MSRA5000 [7] as image pool, represented as $\mathbb{D} = \{I_1, I_2, ..., I_N\}, N = 4000$.

† is the corresponding author. This work is supported by the National Natural Science Foundation of China (NSFC 61671325, 61572354).

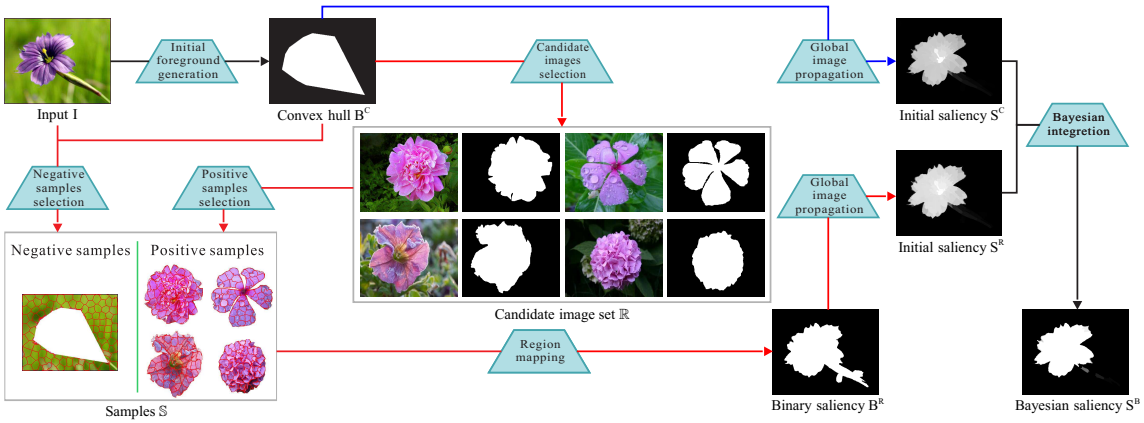

Fig. 1. Illustration of our single-scale saliency detection model. For an image I, we first get the convex hull $\mathbf{B}^{\mathrm{C}}$. And we utilize $\mathbf{B}^{\mathrm{C}}$ as the initial foreground to retrieve candidate image set $\mathbb{R}$. Then we extract exemplars to generate training set $\mathbb{S}$ to train a SVM model for region mapping, and obtain binary saliency map $\mathbf{B}^{\mathrm{R}}$. We deal $\mathbf{B}^{\mathrm{C}}$ and $\mathbf{B}^{\mathrm{R}}$ with global propagation to smooth the saliency detection. Finally, we use Bayesian integration to obtain final single-scale saliency detection.

When detecting saliency for a query image I, we first obtain the initial foreground by using convex hull method [24]. Then we extract the foreground features of I to search similar images from $\mathbb{D}$. Specifically, we use 4 types of global features , including spatial pyramid, gist, tiny image, and color histogram, to describe the foreground. For each feature type, the similarity is calculated by Chi-square distance. According to each type of feature, we gradually reduce the candidate image members. In our selection, we use 100, 80, 70, 60. All global features of image pool are calculated off time to facilitate our online application. Finally, We obtain a candidate image set $\mathbb{R} = \{I_1, I_2, ..., I_M\}, M = 60$ in this step.

2.2. Regional features extraction

We label the query image I with the foreground of the images in the candidate image set $\mathbb{R}$ and corresponding ground truth $\mathbb{G} = \{G_1, G_2, ..., G_M\}$. However, assigning pixel-level labels may be inefficient. Instead, we use superpixel-level labels and feature extraction, which are sufficient for the information expression (we use SLIC [25] here). And we use 17 types of regional features for superpixel representation, i.e., mask of superpixel shape over its bounding box, bounding box width/height relative to image, superpixel area relative to image, mask of superpixel shape over the image, top height of bounding box relative to image height, texton histogram, dilated texton histogram, SIFT histogram, dilated SIFT histogram, left/right/top/bottom boundary SIFT histogram, RGB color mean , RGB color std. dev, RGB color histogram, RGB dilated hist, color thumbnail, masked color thumbnail, and grayscale gist over superpixel bounding box [26].

2.3. Salient region mapping

2.3.1. *Samples selection*

Given a candidate image set $\mathbb{R}$, we aim to find "good" positive (i.e., foreground) and negative (i.e., background) exemplars, which can effectively distinguish salient foreground and complex background of the query image. The direct and effective formation of selected samples $\mathbb{S}$ is that positive exemplars are obtained from the foreground regions of candidate image set $\mathbb{R}$ and the negative exemplars are obtained from the query image itself. In this paper, we use complementary regions of the convex hull [24] to get negative exemplars. The binary map of the convex hull is denoted as $\mathbf{B}^{\mathrm{C}}$.

2.3.2. *Region mapping*

The number of positive exemplars is larger than that of negative exemplars, thereby inducing computation burdens. In our experiment, we find that randomly sampling one-fifth of the positive exemplars can accelerate training process and harm little of the performance. We employ Support Vector Machine (SVM) [27] with radial bias kernel function, as SVM is a light classifier and presents good classification ability. The salient regions of the query image are mapped by the trained SVM classifier to generate binary salient indication map $\mathbf{B}^{\mathrm{R}}$.

2.4. Global image propagation

Currently, two binary salient foreground cues (i.e., $\mathbf{B}^{\mathrm{C}}$ and $\mathbf{B}^{\mathrm{R}}$) are obtained. However, $\mathbf{B}^{\mathrm{C}}$ provides an initial foreground regions which are often contaminated by background regions. Furthermore, $\mathbf{B}^{\mathrm{R}}$ may have incomplete foreground and noised background regions. To address these problems, we propose the use of a graphical model to revalue the saliency probability of each superpixel by considering the global and local relations of the graph. In [13], Yang *et al.* proposed an effective graph-based model that minimizes

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \frac{1}{2}\Big(\sum_{i,j=1} w_{ij}\Big|\Big|\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\Big|\Big|^2 + \mu\sum_{i=1}||f_i - y_i||^2\Big), \tag{1}$$

where $\mathbf{y} = [y_1, ..., y_n]$ is the initial salient value and μ is a balanced parameter; additionally, the closed form solution is

$$\mathbf{f}^* = (\mathbf{D} - \alpha\mathbf{W})^{-1}\mathbf{y}, \tag{2}$$

where $\mathbf{W} = [w_{ij}]_{n\times n}$ is superpixels' affinity matrix, $\mathbf{D} = diag\{d_{11}, ..., d_{nn}\}$ is degree matrix computed by $d_{ii} = \sum_j w_{ij}$, and $\alpha = 1/(1+\mu)$. To further improve saliency detection, we use a quadratic optimization

$$\begin{aligned}\mathbf{s}^* = \arg\min_{\mathbf{s}}(&\sum_i w_i^{fg}||s_i||^2 + \sum_i w_i^{bg}||1 - s_i||^2 \\ &+ \sum_{i,j\in\mathcal{N}(i)} w_{ij}||s_i - s_j||^2),\end{aligned} \tag{3}$$

to get optimal results, where s_i is the objective saliency score of superpixel i, w_i^{bg} is background probability [28], $w_i^{fg} = f_i$ in $\mathbf{f}^*$ is the foreground probability, $\mathcal{N}(i)$ is the neighborhood of i. In this paper, we use the foreground regions that identified by $\mathbf{B}^{\mathrm{C}}$ and $\mathbf{B}^{\mathrm{R}}$ as query to generate two saliency maps, namely, $\mathbf{S}^{\mathrm{C}}$ and $\mathbf{S}^{\mathrm{R}}$.

2.5. Optimal Bayes integretion

As mentioned in Sec. 2.4, the saliency map $\mathbf{S}^{\mathrm{C}}$ and $\mathbf{S}^{\mathrm{R}}$ are complementary to each other. Fused saliency map $\mathbf{S}^{\mathrm{B}}$ can be computed as follows:

$$\mathbf{S}^{\mathrm{B}}(i) = p(F^{\mathrm{R}}|\mathbf{S}^{\mathrm{C}}(i)) + p(F^{\mathrm{C}}|\mathbf{S}^{\mathrm{R}}(i)), \tag{4}$$

where $p(F^{\mathrm{R}}|\mathbf{S}^{\mathrm{C}}(i))$ treats $\mathbf{S}^{\mathrm{R}}$ as the prior and uses $\mathbf{S}^{\mathrm{C}}$ to compute the likelihood:

$$p(F^{\mathrm{R}}|\mathbf{S}^{\mathrm{C}}(i)) = \frac{\mathbf{S}^{\mathrm{R}}(i)p(\mathbf{S}^{\mathrm{C}}(i)|F^{\mathrm{R}})}{\mathbf{S}^{\mathrm{R}}(i)p(\mathbf{S}^{\mathrm{C}}(i)|F^{\mathrm{R}}) + (1-\mathbf{S}^{\mathrm{R}}(i))p(\mathbf{S}^{\mathrm{C}}(i)|B^{\mathrm{R}})}, \tag{5}$$

where we threshold $\mathbf{S}^{\mathrm{R}}$ by its mean saliency value and obtain its foreground and background described by F^{R} and B^{R}. In each region, we compute the likelihoods by comparing $\mathbf{S}^{\mathrm{R}}$ and $\mathbf{S}^{\mathrm{C}}$ in terms of the foreground and background bins at superpixel i:

$$p(\mathbf{S}^{\mathrm{C}}(i)|F^{\mathrm{R}}) = \frac{N_{bF(\mathbf{S}^{\mathrm{C}}(i))}}{N_{F^{\mathrm{R}}}}, p(\mathbf{S}^{\mathrm{C}}(i)|B^{\mathrm{R}}) = \frac{N_{bB(\mathbf{S}^{\mathrm{C}}(i)))}}{N_{B^{\mathrm{R}}}}, \tag{6}$$

where $N_{F^{\mathrm{R}}}$ denotes the number of superpixels in the foreground F^{R} and $N_{bF(\mathbf{S}^{\mathrm{C}}(i))}$ is the number of superpixels whose features fall into the foreground bin $bF(\mathbf{S}^{\mathrm{C}}(i))$ which contains feature $\mathbf{S}^{\mathrm{C}}(i)$, while $N_{B^{\mathrm{R}}}$ and $N_{bB(\mathbf{S}^{\mathrm{C}}(i))}$ are denoted likewise in the background bins. $p(F^{\mathrm{C}}|\mathbf{S}^{\mathrm{R}}(i))$ is also calculated in the similar way.

2.6. Multiscale saliency fusion

Scale is an important problem in salient object detection. Large superpixel scale can capture large objects, and fine superpixel scale is good at detecting small objects. Recent work [14, 15, 9], show that one can achieve robust results by fusing multiscale saliency results. To obtain robust saliency detection, we average the saliency detection results as follows:

$$\mathbf{S}^{final} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{S}^{B}(l), \tag{7}$$

where $\mathbf{S}^{B}(l)$ is the l-th integral saliency map generated in Eq.(5). Note that larger l means larger superpixel number.

In our experiment, we implement the superpixel numbers of 50, 100, 200, 300 and 400, which show good performance. Specifically, we ultilize a fixed scale of 300 in global image propagation.

3. EXPERIMENT

3.1. Setting

Baselines. We compare our saliency detection method with eleven state-of-the-art models including FT [8], GS [29], HS [9], RA [30], SF [31], DSR [15], GRSD [24], HDCT [16], MR [13], GB [12], and wCtr [28]. We name our exemplar based image saliency detection method as **EBIS**.

Benchmark. Four benchmark datasets, including MSRA5000 [7], MSRA1000 [8], ECSSD [9], and PASCAL-S [10] are used in our experiments. We randomly obtain $4,000$ images from MSRA5000 [7] to form our image pool for similar image searching. The remaining $1,000$ images are used to evaluate all saliency detectors.

Metric. Average Precision, Recall, F-measure and Mean absolute error (MAE) are used for quantitative comparison.

3.2. Results

As shown in Fig. 3, our model generates considerably accurate saliency detections that have uniformly highlighted foreground and well-suppressed background. In the second row of Fig. 3, the salient object is a man with dark shadow. Accordingly, the man is more salient than his shadow. However, almost all compared saliency detectors fail to suppress the shadow. By contrast, our method successfully suppresses the shadow and shows precise salient object detection. We can also observe from other images that our method preserves finer and integrity object boundaries.

Fig. 2 shows the PR cures of our method and 11 state-of-the-art saliency detectors on 4 benchmark datasets. According to the observations on Fig. 2, 1) All saliency detection methods exhibit different performances on four benchmark datasets. Compared with other datasets, most of saliency detection methods achieve good performance on MSRA1000 [8]; 2) On ECSSD [9], none of the curves can stride over line of precision at 0.9, which verifies the complexity of ECSSD; 3) Our method has stabler saliency detection results. Considering the different thresholds, we have the shortest PR-curves; additionally, when recalls are higher than certain

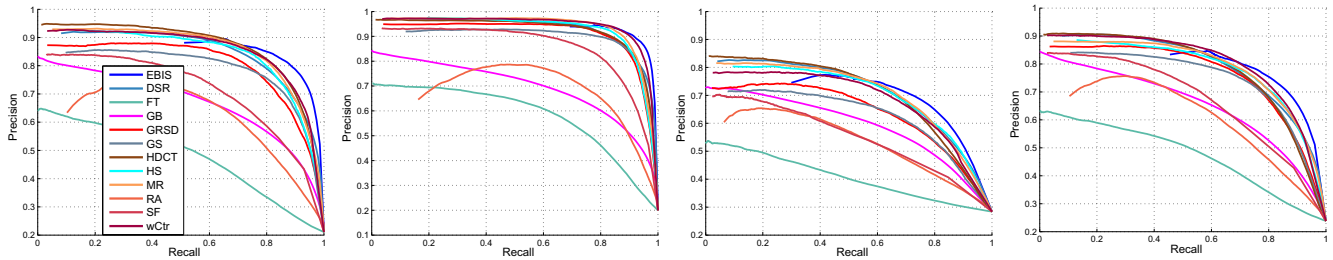

Fig. 2. Comparison of PR curves of different saliency models on 4 benchmark datasets (from left to right): MSRA5000 [7], MSRA1000 [8], ECSSD [9], and PASCAL-S [10].

Table 1. Comparison of F-measure and MAE of different saliency models on 4 benchmark datasets. We use red, blue and green colors to denote the first, second and third ranked methods, respectively.

Datasets	Criteria	EBIS	DSR	FT	GB	GRSD	GS	HDCT	HS	MR	RA	SF	wCtr
MSRA5000	F-measure	0.8479	0.8170	0.5198	0.6277	0.7170	0.7415	0.8009	0.7669	0.8211	0.5370	0.7098	0.8201
	MAE	0.0897	0.1216	0.2514	0.2292	0.2149	0.1458	0.1439	0.1653	0.1281	0.3230	0.1677	0.1111
MSRA1000	F-measure	0.9152	0.8784	0.5935	0.6405	0.7941	0.8128	0.8329	0.8408	0.8976	0.5561	0.8244	0.8885
	MAE	0.0474	0.0796	0.2195	0.2146	0.1607	0.1069	0.1147	0.1107	0.0752	0.3080	0.1285	0.0652
ECSSD	F-measure	0.7243	0.7189	0.4407	0.5793	0.6113	0.6259	0.7030	0.6700	0.7063	0.5133	0.5477	0.6950
	MAE	0.2078	0.2253	0.3283	0.3050	0.3163	0.2542	0.2483	0.2682	0.2357	0.3673	0.2669	0.2242
PASCAL-S	F-measure	0.7959	0.7819	0.5138	0.6298	0.7032	0.7174	0.7556	0.7324	0.7716	0.5480	0.6902	0.7823
	MAE	0.1196	0.1326	0.2504	0.2326	0.2282	0.1614	0.1587	0.1912	0.1566	0.3257	0.1611	0.1280

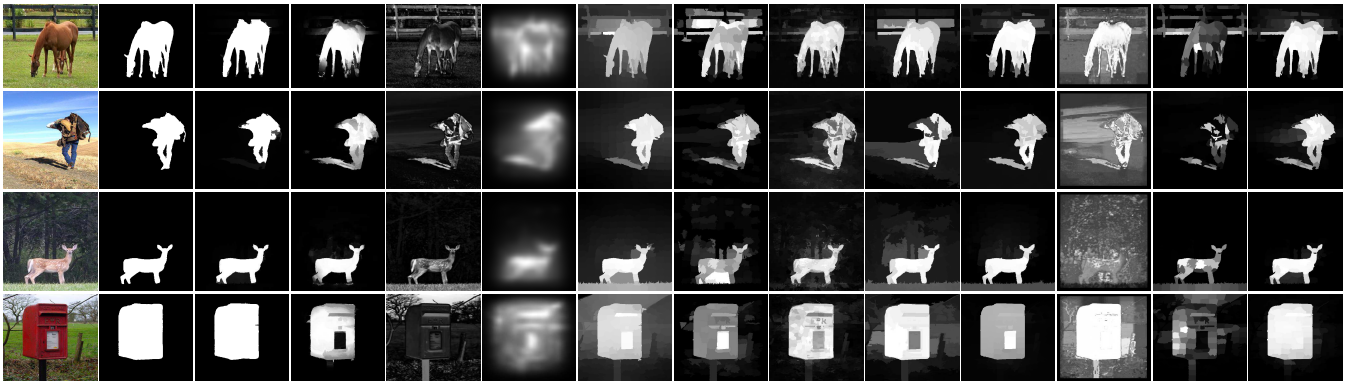

Fig. 3. Some examples of saliency detection results of different saliency detection methods on 4 benchmark datasets.

values (0.7 on MSRA5000 [7]), our precision is higher than that of all compared methods. To verify that our method is superior, we further compare F-measure and MAE in Table 1.

We use three colors to denote the first, second and third ranked saliency detection methods. As shown in the Table 1, our method consistently outperforms compared methods on 4 benchmark datasets. The wCtr [28] ranks second except F-measure on ECSSD [9]. Most of the saliency detection methods have better performances on MSRA5000 [7] and MSRA1000 [8] than that on ECSSD [9], and PASCAL-S [10]. In detail, on MSRA5000 [7], our MAE is the only one that is lower than 0.1 and presents an improvement of 19.3% relative to the second ranked wCtr [28]. On MSRA1000 [8], only our F-measure reaches 0.915, which is higher than that of second ranked MR [13] at approximately 1.9%. On complex datasets ECSSD [9], and PASCAL-S [10], the performance of all saliency detection methods slightly decreases. Nevertheless, on these two datasets, our method still obtains the most remarkable saliency detections. We obtain more improvement on MAE evaluations than that of F-measure. The conclusions of Table 1 are consistent with the results in Fig. 2 and Fig. 3, thereby demonstrating the superiority of our method.

4. CONCLUSION

In this paper, we propose an exemplar based saliency detection method by exploiting valuable labelled images from existing saliency detection benchmark datasets. We elaborately design a progressive way to select candidate images and yield an effective saliency propagation method to provide saliency from exemplars. The multiscale strategy is also used in our proposed method to generate robust saliency detection. Our method achieves better saliency detection performance on 4 benchmark datasets compared with 11 state-of-the-arts. In the future, we plan to analyze the effects of scale variations and extend our method for co-saliency detection.

5. REFERENCES

[1] Z. Liu, R. Shi, L. Q. Shen, and Y. Z. Xue, "Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut," *IEEE TMM*, vol. 14, no. 4, pp. 1275–1289, 2012.

[2] A. Browet, P. A. Absil, and P. V. Dooren, "Contour detection and hierarchical image segmentation," *IEEE TPAMI*, vol. 33, no. 5, pp. 898–916, 2011.

[3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition," in *CVPR*, 2004.

[4] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE TIP*, vol. 13, no. 10, pp. 1304–18, 2004.

[5] J. Y. Wu and Q. Q. Ruan, "Object removal by cross isophotes exemplar-based inpainting," in *ICPR*, 2006.

[6] H. A. Roy and V. Jayakrishna, "2d image reconstruction after removal of detected salient regions using exemplar-based image inpainting," in *AISC*, 2014.

[7] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *IEEE TPAMI*, vol. 33, no. 2, pp. 353–67, 2007.

[8] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009.

[9] Q. Yan, L. Xu, J. P. Shi, and J. Y. Jia, "Hierarchical saliency detection," *ACM*, vol. 9, no. 4, pp. 1155–1162, 2013.

[10] Y. Li, X. D. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in *CVPR*, 2014.

[11] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. M. Hu, "Salient object detection and segmentation," *IEEE TPAMI*, vol. 37, no. 3, pp. 1, 2011.

[12] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *NIPS*, 2010.

[13] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, "Saliency detection via graph-based manifold ranking," in *CVPR*, 2013.

[14] R. Huang, W. Feng, and J. Z. Sun, "Saliency and co-saliency detection by low-rank multiscale fusion," in *ICME*, 2015.

[15] 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.

[16] J. W. Kim, D. Y. Han, Y. W. Tai, and J. M. Kim, "Salient region detection via high-dimensional color transform and local spatial support," *IEEE TIP*, vol. 25, no. 1, pp. 1–1, 2015.

[17] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] N. Liu and J. W. Han, "Dhsnet: Deep hierarchical saliency network for salient object detection," in *CVPR*, 2016.

[19] A. Borji, M. M. Cheng, H. Z. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE TIP*, vol. 24, no. 12, pp. 5706–5722, 2015.

[20] J. M. Yang and M. H. Yang, "Top-down visual saliency via joint crf and dictionary learning," in *CVPR*, 2012.

[21] J. Zhu, Y. Y. Qiu, R. Zhang, J. Huang, and W. J. Zhang, "Top-down saliency detection via contextual pooling," *SPS*, vol. 74, no. 1, pp. 33–46, 2014.

[22] S. F. He, R. W.H.Lau, and Q. X. Yang, "Exemplar-driven top-down saliency detection via deep association," in *CVPR*, 2016.

[23] L. W. Ye, Z. Liu, X. F. Zhou, and L. Q. Shen, "Saliency detection via similar image retrieval," *IEEE SPL*, vol. 23, no. 6, pp. 1, 2016.

[24] C. Yang, L. H. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE SPL*, vol. 20, no. 7, pp. 637–640, 2013.

[25] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "Slic superpixels," *EPFL*, p. 15, 2010.

[26] J. Tighe and S. Lazebnik, "Superparsing: Scalable nonparametric image parsing with superpixels," in *ECCV*, 2010.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[28] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *CVPR*, 2014.

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

[30] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *ECCV*, 2010.

[31] P. Krahenbuhl, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*, 2012.