

# Salient Object Detection Based on the Proto-objects and Background Prior

Gaopeng Zhao<sup>\*</sup>, Mingfeng Yin and Liang He

Nanjing University of Science and Technology, China

**Abstract:** Image salient object detection is useful for applications like image retrieval, object recognition etc. Aiming to obtain the pixel-level saliency map, most traditional saliency models ignore the object-level information and become inappropriate in complex scenes. A new salient object detection method is proposed in this paper by incorporating the notion of object and background prior directly into the saliency measurement. Firstly the collection of proto-objects is obtained by adapting selective search method. Secondly for each proto-object, we measure its saliency score by computing the distance of histogram feature between the proto-object and the prior background region, then the saliency scores are ranked, and the top-ranked proto-object is considered as the salient object. The publicly MSRA dataset is adapted for experiment evaluation, compared with several other state-of-the-art methods, the proposed method produces superior performance. Experimental results show that the proposed method is effective.

**Keywords:** Background prior, proto-object, salient object detection.

## 1. INTRODUCTION

Human vision has the selective attention characteristics which can help them in focusing on the significant region of the scene quickly and accurately. However how to use the selective attention mechanism to quickly and efficiently extract the salient object information from the image is still a difficult task for computer vision. The task of salient object detection is to identify the visual attentive area in an image. Recently it has been used in many computer vision applications including image segmentation, object recognition, image retrieval [1-3].

A detailed comparison is presented in [4] for the latest saliency model. A lot of research focuses on the bottom-up models. The idea can be described as the saliency map is calculated by measuring the saliency value for each pixel by using the lower visual feature, then the salient object is obtained by segmenting the saliency map. The contrast visual feature is usually adapted, which is computed according to the lower visual features, including color, edge, gradients, spatial frequencies, histograms, multi-scale descriptors and combinations thereof.

The purpose of salient object detection is to identify the position and size of the salient object. The drawback of the above model is that the result depends on the quality of the saliency map and the image segmentation algorithm. The salient map obtained by simple pixel feature is difficult to be effectively applied to the salient object detection in complex scenes because such saliency map does not take into account the information about objects of an image. A simple example is presented in Fig. (1). From the Fig. (1), we can see that it is difficult to locate the salient object from the saliency map.

Considering that the human visual system pre-segments the complex scene into proto-objects, and then human attention is focused on the most salient proto-object [7]. In this paper, we aim to directly obtain the salient object rather than saliency map and propose a new salient object detection method by incorporating the notion of an object directly in the saliency measurements. The novelty of the paper is that: (1) the saliency computation is based on object level rather than single pixel. (2) The saliency scores are computed based on the proto-objects which are extracted from the source image, and the salient objects are obtained by ranking the saliency scores. (3) The background prior is introduced to calculate the saliency score. We demonstrate that the proposed method achieves state-of-the-art performance on the standard dataset.

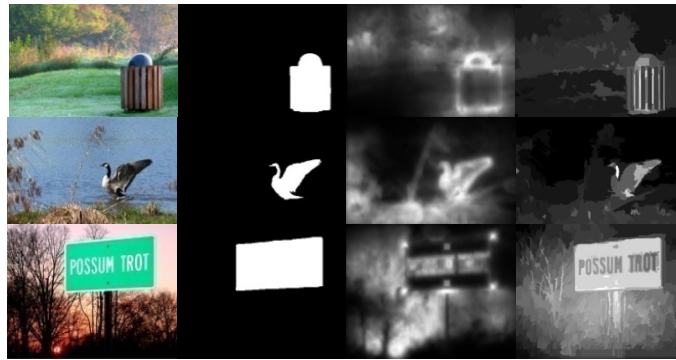
The paper is organized as follows. We provide a brief review of the recent start-of-the-art saliency detection methods and a description of their strengths and shortcomings in Section 2. The proposed method is illustrated in Section 3. In Section 4, we describe the experiments and the analysis. Finally, we conclude this work in Section 5.

## 2. RELATED WORK

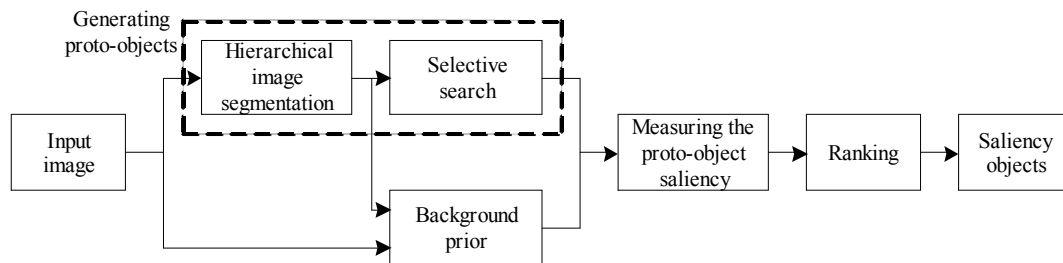
Most of the existing saliency detection methods can be categorized as top-down computational model and bottom-up computational model. We focus on relevant literature targeting the bottom-up saliency detection. It is common to follow a spatial-based attention theory by calculating a pixel-based saliency map with the task of salient object detection. According to the computational scheme and the mathematical theory, these methods can be broadly classified into several categories, including spectral methods, image contrast methods, center-surround methods etc.

A comprehensive survey of salient object detection methods can be found from [1-4]. Recently, some latest

<sup>\*</sup>Address correspondence to this author at the Nanjing University of Science and Technology in China; Tel: 86-025-84315172; Fax: 86-025-84315172; E-mail: [zhaogaopeng@sina.com](mailto:zhaogaopeng@sina.com)



**Fig. (1).** Saliency detection results. From left to right, each column respectively represents the input images, the ground truth, the saliency map detected by Goferman's algorithm [5], the saliency map detected by Cheng's algorithm [6]. Input images are taken from MSRA dataset.



**Fig. (2).** The diagram of the proposed method.

methods produce superior performance. Li *et al.* [8] combine the global information from frequency domain analysis and local information from spatial domain analysis to generate the saliency map. Goferman *et al.* [5] propose context-aware saliency, which detects the important parts of the scene rather than the traditional salient object. However, the method tends to acquire a high saliency value near the edge and does not highlight the entire salient region. Jiang *et al.* [9] computes the saliency by using the context and shape prior, but the shape of object is difficult to obtain in complex scene. The two methods [5, 9] adopt the context information to improve the saliency effect. Cheng *et al.* [6] employ the color histogram feature and region contrast but it is difficult to handle cluttered and textured background. Achanta *et al.* [10] determines the saliency of a pixel by using the maximum symmetric surrounds and Vikram *et al.* [11] adopt the center-surround method to estimate saliency by generating random regions of an image. The above three methods [6, 10, 11] can generate full resolution saliency map since they do not downscale the input image to a low resolution, however, they calculate saliency map of each pixel independently like many other methods, so the correlation between pixels is not considered properly.

Additionally, there are several works directly checking if an image window contains an object. Alexe *et al.* [12] propose the generic objectness measure which is defined by combining several image cues in a Bayesian framework to quantify the possibility that a window contains an object. Siva *et al.* [13] propose an unsupervised learning approach to object saliency. The two methods can effectively obtain the object location, however they require a lot of images to learn the parameters.

Despite so many methods have been proposed and contributed to the specific task, the start-of-the-art performance

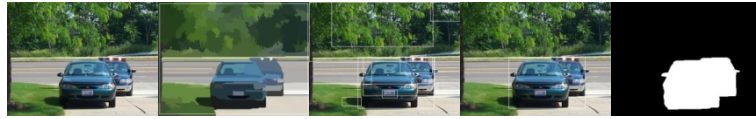
is not satisfying. As motivated before, we propose a saliency object detection method. We extract proto-objects directly from the source image by feature grouping, which are feasible candidates for salient objects. To do this, we follow the strategy introduced by [14], which propose the selective search method to find candidates for objects locations. We observe that the background often presents local or global appearance and connectivity with each of four image boundaries and foreground presents appearance of coherence and consistency. Thus, we can obtain the background prior by four image boundary regions which are labeled through segmentation. We also observe that the distance between a pair of background regions is shorter than that of the salient object region and the background region. So we measure the saliency of the proto-objects by computing the distance of histogram feature between the proto-object and the prior background region. The proto-object with the maximal saliency value is selected as the most salient object. Experimental results using several datasets show that the proposed method performs efficiently.

### 3. PROPOSED METHOD

The overview of the proposed method is presented in Fig. (2). We assume that the human attention focus on proto-objects, and the image boundary regions are considered as background prior, we model the saliency on object level and get the saliency detection results by ranking. In the remaining of this section we provide more details of each part of our method.

#### 3.1. Generating Proto-objects

We adapt the selective search method [14] to obtain a collection of candidate proto-objects. Inspired by bottom-up segmentation and grouping strategy, a hierarchical image



**Fig. (3).** An example of the proposed method. From left to right, each column respectively represents the input image, the background prior regions under the white line, part of the proto-objects represented by white rectangular, the final salient object detected by the proposed method, the ground truth labeled by human.

segmentation algorithm [15] is firstly used to create the initial regions, because this method can provide a set of small regions which ideally do not span multiple objects and can capture objects in all scales. Then a grouping procedure works on the basis of initial regions. A greedy algorithm is used to iteratively group regions together. First, the similarities between all neighboring regions are calculated. The two most similar regions are grouped together, and new similarities are calculated between the resulting region and its neighbors. The process of grouping the most similar regions is repeated. Then object locations are extracted from all regions.

The similarity function is defined as (1):

$$s(r_i, r_j) = s_{\text{colour}}(r_i, r_j) + s_{\text{texture}}(r_i, r_j) + s_{\text{size}}(r_i, r_j) + s_{\text{fill}}(r_i, r_j) \quad (1)$$

where  $s(r_i, r_j)$  represents the similarity value between region  $r_i$  and  $r_j$ .  $s_{\text{colour}}(r_i, r_j)$  measures color similarity which is computed by color histogram.  $s_{\text{texture}}(r_i, r_j)$  measures texture similarity which is computed by SIFT-like feature and texture histogram.  $s_{\text{size}}(r_i, r_j)$  is defined as the fraction of the image that  $r_i$  and  $r_j$  jointly occupy.  $s_{\text{fill}}(r_i, r_j)$  measures how well region  $r_i$  and  $r_j$  fit into each other.

The hierarchical image segmentation is described in detail in [15] and the selective search method is described in detail in [14]. We adapt the algorithms to generate proto-objects, where the parameters are set to the default values.

The features of the proto-object are computed as the color histogram in Lab color space.

### 3.2. Background Prior

There exist a few algorithms attempting to make use of the characteristics of the background in saliency detection [16]. It has been observed that the pixels lie in the border region of the image often belong to the background. Therefore, the background prior is extracted by the 15-pixel wide narrow border region of the image. Obviously the fixed width method is subjective. Based on the assumption, we present an objective method for obtaining more accurate background prior than a fixed width.

We adapt the hierarchical image segmentation algorithm [15] to generate the background prior region in the paper. The image is segmented into different regions in all scales. So we define the regions lie in the border of the image in the finest scale as the background prior, as illustrated in Fig. (3).

The features of the background prior are computed as the color histogram in Lab color space.

### 3.3. Measuring the Object Saliency and Ranking

We measure the saliency score for each proto-object instead of computing the saliency map of the image. Saliency map is a two-dimensional image which describes the

pixel's saliency degree of the scene. However the saliency value for each proto-object is a single value which directly represents the saliency degree in object level. The larger the saliency value the more salient the proto-object.

We compute the saliency score by feature distance between the proto-object and the background prior, which are described in section 3.1 and section 3.2. Usually it is considered that the color of the salient object is scarcely in the overall image, the color distribution of the salient object is far different from the background. Therefore we adapt the color histogram in Lab color space and  $\chi^2$ -distance, as depicted in Equation 2.

$$\chi^2(h_o, h_b) = \sum_{i=1}^m \frac{2(h_{oi} - h_{bi})^2}{h_{oi} + h_{bi}} \quad (2)$$

where,  $h_o$  represents the normalized color histogram of the proto-object.  $h_b$  represents the normalized color histogram of the background prior.  $m$  represents the number of the histogram bins.

For the set of proto-objects  $R = r_1, r_2, \dots, r_n$ , we calculate the corresponding object saliency scores  $A = a_1, a_2, \dots, a_n$  where  $n$  is the number of the proto-objects.

We rank the saliency scores  $A$  in descending order. For the single salient object detection, the top-ranked proto-object is considered as the salient object. For the multiple salient object detection, several top-ranked proto-objects are considered as the salient object, the number of salient object can be defined in advance for special application or a certain threshold can be selected in advance, all proto-objects with saliency score above the threshold are considered as the salient objects.

## 4. EXPERIMENTS

### 4.1. Setup

We test our method on the MSRA dataset, where the task is to detect a bounding box around the most salient object in an image. ASD dataset [10] is a subset of MSRA with manually produced ground annotations. As our method detects the most salient object, we report a tight bounding box around it. We evaluate the performance using the measures used in [5, 6, 9-11], including the precision, recall and F-measure.

We also compare state-of-the-art bottom-up saliency methods with the proposed method, including FS [8], CA [5], CBSal [9], RC [6], MSSS [10], RCSS [11]. These methods are implemented with the author's original codes. Our method is implemented with matlab codes, and it is abbreviated as ObjSal.

The parameters settings of our method include: the number of histogram bin  $m$  is set to  $12 \times 12 \times 12$ , and the parameters of the selective search are set to the default values in [14].



Fig. (4). Visual comparison of the part of results. Column 1~5: an image and its corresponding results; Row 1~9: input image, ground truth, FS, CA, CBSal, RC, MSSS, RCSS, ObjSal.

4.2. Qualitative Comparison

We provide the visual comparison of different methods in Fig. (4). We report the results of the compared methods as saliency map, and report our detection result as bounding box on input image.

It can be observed in Fig. (4) that the performance of each method is quite different. For example, MSSS method and RCSS method are less effective than the others. CA method highlights the edges of the salient object. CBSal can get better results than the above method. The proposed method ObjSal can accurately detect the salient object. In the salient object detection task, the detection results of the compared methods depend on the quality of the saliency map and the image segmentation algorithm. If the saliency map is poor, it is difficult to obtain the accurate locations of the salient object. However the proposed method can directly obtain the salient object.

4.3. Quantitative Comparison

The performance evaluation Precision, Recall and F-measure are computed as in (3)~ (5).

$$P = \frac{\sum(B \times A)}{\sum B} \tag{3}$$

$$P = \frac{\sum(B \times A)}{\sum A} \tag{4}$$

$$F = \frac{(1 + \beta^2) \times P \times R}{(\beta^2 \times P + R)} \tag{5}$$

where P is the precision value, R is the recall value, F is the F-measure value. B represents the segment result of the saliency map by threshold, A is the ground truth by manual annotation,  $\beta$  2. is set to 0.3 which indicates the F-measure value has more emphasis on accuracy. The larger the P value and F value, the better the effect of the method.

The adaptive threshold T is adapted to segment the saliency map, which is computed as (6).

$$T = \frac{2}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} S(x, y) \tag{6}$$

where W is the image width, H is the image height, S represents the saliency map.

By computing the P value, R value, and F value for every saliency maps in the results obtained by above method, the final values are given by averaging the corresponding values. The results are shown in Table 1. It is worth pointing out that the result of our method is represented by bounding box, so we compute the quantitative evaluation results with bounding box, rather than saliency map.



**Table 1. Comparison of quantitative evaluation results.**

	CA	FS	MSSS	RC	RCSS	CBSal	ObjSal
P	0.7122	0.8036	0.5007	0.7721	0.7241	<b>0.8630</b>	<b>0.9347</b>
R	<b>0.9517</b>	0.8294	<b>0.9727</b>	0.7731	0.7701	0.8334	0.8426
F	0.7402	0.7872	0.5449	0.7187	0.7104	<b>0.8361</b>	<b>0.9117</b>

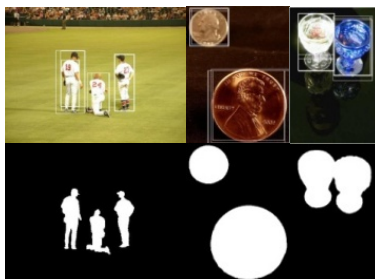
**Table 2. The average computing time for an image.**

	CA	FS	MSSS	RCSS	CBSal	ObjSal
Time (Second)	59.5	2.4	0.5	1.6	2.5	5.6

From the quantitative evaluation in Table 1, we can see that precision value and F-measure value obtained by the proposed method are larger than the other methods. It shows that the proposed method is more effective and it is consistent with the visual effect in Fig. (4).

#### 4.4. Multiple Salient Objects

The proposed method is not limited to the detection of a single salient object. It can select any number of the most salient objects. Depending on the special application, the number can be defined in advance. If the number is not known, all proto-objects with saliency above a certain threshold can be selected, or we can select specific number of the top-ranked proto-objects as salient objects. In Fig. (5), the bounding boxes around 10 most salient proto-objects are shown for several images. Particularly we can remove the near-duplicate proto-objects with a specific overlap threshold and get the most salient objects.



**Fig. (5).** Examples of multiple salient objects (the first row: the ten top-ranked proto-objects as the detection result of the proposed method; the second row: ground truth).

#### 4.5. Discussion

We evaluate the runtime of the proposed method by computing the average processing time in the ASD dataset. The CPU is dual-core 2.93GHz and the RAM is 3.0G.

The RC method is C++ executable codes. The others are programmed with matlab codes. So we evaluate the runtime of the proposed method with these methods except RC method in the Matlab environment. The more hierarchical segmentations are used, the more proto-objects are generated, therefore it needs a trade-off between accuracy and computational efficiency. In our experiment the default parameters are used. The results are shown in Table 2. The

proposed method is time-consuming due to the selective search step, which takes ninety percent of the total time.

## 5. CONCLUSION AND FUTURE WORK

Differently from the traditional saliency models, we propose a novel method for salient object detection application based on the object-level visual attention. The concepts of proto-object and background prior are taken into account for assessing proto-object saliency. And we compute the saliency score for each proto-object. The top-ranked proto-objects are considered as the most salient objects. The proposed method can directly locate the salient objects, and achieves superior performance on a well-known benchmark. In the future, we plan to improve the method for specific application such as ground object detection, object recognition and so on.

## CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

This Project is supported by the National Natural Science Foundation of China (Grant No. 61203266).

## REFERENCES

- [1] A. Toet, "Computational versus psychophysical bottom-up image saliency: a comparative evaluation study," *IEEE Trans. PAMI*, vol. 33, no. 11, pp. 2131-2146, 2011.
- [2] R. N. Ma, X. P. Tu, J. D. Ding, J. Y. Yang, "To evaluate saliency map towards popping out visual objects," *Acta Auto. Sin.*, vol. 5, pp. 870-876, 2012.
- [3] A. Borji, L. Itti, "State-of-the-art in visual attention modeling" *IEEE Trans. PAMI*, vol. 35, no. 1, pp. 185-207, 2013.
- [4] A. Borji, D. N. Sihite, L. Itti, "Salient object detection: a benchmark," In: Proceedings of ECCV, pp. 414-429, 2012.
- [5] S. Goferman, L. Z. Manor, A. Tal, "Context-aware saliency detection," *IEEE Conf. CVPR*, pp. 2376-2383, 2010.
- [6] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, S. M. Hu, "Global contrast based salient region Detection," In: *IEEE Conf. CVPR*, pp. 409-416, 2011.
- [7] D. Walther, C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, pp. 1395-1407, 2006.
- [8] J. Li, M. Levine, X. J. An, H. He, "Saliency detection based on frequency and spatial domain analyses," In: *Proc. BMVC*, vol. 86, pp. 1-86, 2011.

- [9] H. Z. Jiang, J. D. Wang, Z. J. Yuan, T. Liu, N. Zheng, S. Li, "Automatic salient object segmentation based on context and shape prior," In: Proc. BMVC, vol. 110, pp. 1-110, 2012.
- [10] R. Achanta, S. Ssstrunk, "Saliency detection using maximum symmetric surround," In: IEEE Conf. Image Process., pp. 2653-2656, 2010.
- [11] T. N. Vikram, M. Tscherepanow, B. Wrede, "A saliency map based on sampling an image into random rectangular regions of interest", *P.R.*, vol. 45, no. 9, pp.3114-3124, 2011.
- [12] B. Alexe, T. Deselaers, V. Ferrari, "Measuring the objectness of image windows", *IEEE Trans. PAMI*, vol. 34, no. 11, pp. 2189-2202, 2012.
- [13] P. Siva, C. Russell, T. Xiang, L. Agapito, "Looking beyond the image: unsupervised learning for object saliency and detection", In: IEEE Conf. CVPR, pp. 3238-3245, 2013.
- [14] J. R. R. Uijlings, K. E. A. Van de Sande, T. Gevers, A.W.M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vision*, vol. 104, no. 2, pp. 154-171, 2013.
- [15] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. PAMI*, vol. 33, no. 5, pp. 898-916, 2010.
- [16] H. Z. Jiang, J. D. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, "Salient object detection: a discriminative regional feature integration approach," In: IEEE Conf. CVPR, pp. 2083-2090, 2013.

---

Received: September 22, 2014

Revised: November 30, 2014

Accepted: December 02, 2014

© Zhao et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.