Postprint: Salient Object Detection Based on Divergence-Shape Guidance and Optimization Function

Authors: Liang Lixiang, Xia Chenxing, Shengwen Wang, Zhang Hanling

Date: 2018-05-24T00:00:00+00:00

Abstract

To accurately perform salient object detection, an effective saliency detection framework based on divergence-shape guidance and an optimization function is proposed. First, a discriminative similarity measure is proposed by considering color, spatial position, and edge information; subsequently, divergence prior is utilized to eliminate foreground noise from image boundaries to obtain a background set, which is then combined with the similarity measure to compute a background-based saliency map. To improve detection quality, shape completeness is proposed and the corresponding shape-complete saliency map is generated by statistically computing the expected number of times regions are activated in hierarchical space. Finally, an optimization function is used to optimize the fused result of the two saliency maps to obtain the final result. Experimental validation on public datasets ASD, DUT-OMRON, and ECSSD demonstrates that the proposed method can accurately and effectively detect salient objects located at arbitrary positions within images.

Full Text

Abstract

To accurately detect salient objects, this paper proposes an effective framework for saliency detection based on scatter-shape guidance and an optimization function. First, by considering color, spatial position, and edge information, we introduce a discriminative similarity metric. Next, we utilize scatter prior to eliminate foreground noise from image boundaries and obtain a background set, which is then combined with the similarity metric to compute a background-based saliency map. To improve detection quality, we propose a shape completeness cue and generate a corresponding shape completeness saliency map by statistically measuring the expected number of times regions are activated in

hierarchical space. Finally, an optimization function is employed to refine the fused result of the two saliency maps to obtain the final result. Experimental validation on the public datasets ASD, DUT-OMRON, and ECSSD demonstrates that the proposed method can accurately and effectively detect salient objects located anywhere in the image.

Keywords: saliency detection; scatter-shape guidance; optimization function; similarity metric; hierarchical space

0 Introduction

Saliency detection aims to simulate the human visual system to extract the most prominent regions in an image. In recent years, saliency detection has attracted increasing attention and been widely applied in computer vision tasks including image segmentation [**?**], object recognition [**?**], video compression [**?**], and image retrieval [**?**]. From a psychological perspective, saliency detection can be categorized into bottom-up (data-driven) and top-down (task-driven) approaches. The former primarily uses low-level features such as color, position, and texture to detect salient regions, while the latter requires learning specific object visual information in a supervised manner to form saliency maps.

Since there currently lacks a unified definition of salient objects, most methods rely on effective heuristic priors based on foreground or background features. Among these, boundary prior is the most widely used criterion [**?**, **?**], which assumes that regions located on image boundaries have a high probability of being background. Although many studies have demonstrated that image boundaries are likely to be background, salient objects may still touch the image boundaries. Directly treating all boundary regions as background often introduces pathological information into the selected background, leading to errors. To address this issue, numerous improved methods have been proposed. Li et al. [**?**] suggested discarding the boundary edge with the maximum distinctiveness among the four edges and using the remaining three as background. Wang et al. [**?**] treated superpixels with strong edge strength in the boundary set as foreground noise to obtain a reliable background. Li et al. [**?**] computed color differences for each pixel in the boundary set, sorted them in descending order, and empirically removed the top 30% ranked pixels from the boundary set. While these processing mechanisms effectively improve results, they are limited to computing distinctiveness within the boundary set, yielding insufficiently reliable and robust outcomes.

Although boundary prior works well for simple scenes, it is clearly inadequate for complex scenes. Consequently, many researchers have focused on building more discriminative similarity metrics based on high-level features to increase the difference between foreground and background. Lee et al. [**?**] proposed concatenating encoded low-level distance maps with high-level features to compute saliency maps. Zhao et al. [**?**] introduced a multi-context deep learning framework for saliency detection. Wang et al. [**?**] fused local and global features under

deep networks for salient object detection. While these methods demonstrate good performance on datasets, they require collecting large amounts of manually labeled images and incur substantial overhead in building learning frameworks, severely limiting their applicability. Meanwhile, some researchers have proposed various propagation models to improve saliency detection results. Li et al. [**?**] introduced a normalized random walk model, while Jiang et al. [**?**] utilized Markov absorption probabilities on image graph models for saliency detection. However, when background structures are complex or the difference between background and foreground is small while the similarity matrix lacks strong discriminative power, these methods may incorrectly highlight background regions, producing chaotic results.

To address these problems, this paper proposes an effective framework for salient object detection based on scatter-shape guidance and an optimization function. First, edge guidance is employed for region segmentation, and a discriminative similarity metric matrix is constructed through a new feature space. Unlike previous methods for extracting robust backgrounds, we consider the entire image from a global perspective, using scatter prior to eliminate foreground noise in image boundaries combined with boundary distinctiveness prior to obtain a corresponding background-based saliency map. To ensure the integrity of salient object boundaries and that all internal region elements are highlighted, we propose utilizing shape completeness cues and generating a shape completeness saliency map by statistically counting the expected number of times regions are activated in hierarchical segmentation space. Finally, a propagation algorithm fusing mid-level features and seed selection is used to optimize the fused result of the two saliency maps to obtain the final saliency result.

1 Saliency Detection Based on Scatter-Shape Guidance and Optimization Function

We first adopt the method from [**?**] to compute the probability of boundary (PB), then obtain superpixels by thresholding the ultrametric contour map (UCM) derived from PB. Here, we define the initial segmentation as $P_0 = \{R_i\}$ with the number of segmentation regions being K_0.

1.1 Scatter-Guided Background

In an image, salient regions typically exhibit similar colors and compact spatial distribution, while background regions often show diverse appearances and loose distribution. Therefore, we introduce scatter prior to eliminate foreground noise from the image boundary set (BS). We treat the entire image as our research object to compute the scatter of each superpixel:

$$Div(i) = \text{normalize}\left(\sum_{j=1}^{N} w_{ij} \cdot |s_j - \rho_i|\right) \quad (1)$$

where s_j represents the coordinates of superpixel j, ρ_i denotes the weighted average position of superpixel i, N represents the number of superpixels (here equal to K_0), normalize(x) denotes normalization of x, and w_{ij} represents the edge weight. To effectively distinguish between background and foreground, we compute w_{ij} using a joint matrix that simultaneously considers color differences, spatial distances, and edge information. The main considerations are:

a) According to the cognitive characteristics of color similarity, image regions with similar colors often belong to the same category.

b) Based on spatial proximity, adjacent regions may have the same label in space.

c) In some cases, using edge maps can better highlight contours between foreground and background than equations (2) and (3).

Based on the above analysis, we express w_{ij} as:

$$\rho_i = \text{normalize}\left(\sum_{j=1}^{N} w_{ij} \cdot s_j\right) \quad (2)$$

$$w_{ij} = \exp\left(-\frac{d_c(i,j) + d_s(i,j) + d_e(i,j)}{2\sigma_w^2}\right) \quad (3)$$

where σ_w controls the intensity of distance between a pair of nodes, $d_c(i,j)$ represents the Euclidean distance between superpixels i and j in LAB color space, $d_s(i,j)$ denotes their spatial Euclidean distance, and $d_e(i,j)$ is defined as the sum of the shortest path weights between them, i.e.:

$$d_e(i,j) = \min_{u_1,\ldots,u_{k-1}} \sum_{m=1}^{k-1} e(u_m, u_{m+1}) \quad (4)$$

Here, $e(i,j)$ represents the edge strength between two superpixels in the UCM.

Since different boundaries have different probabilities of contacting salient objects, we adopt a scatter-information-based approach that sets different thresholds for different boundaries to eliminate foreground noise from the four boundaries, thereby obtaining a robust background set (BG). Based on boundary prior, we can obtain a saliency map by computing the distinctiveness from BG. However, background superpixels may only be similar to part of BG rather than all of it. Therefore, computing the distinctiveness between a superpixel and all superpixels in BG as its saliency value may not work well in complex scenes. Consequently, we define the saliency value of superpixel i as its distinctiveness from the k nodes in BG with the most similar scatter:

$$S_B(i) = \sum_{j \in BG} w_{ij} \quad (5)$$

where $j \in BG$ and $k = \text{Num}(BG)/10$, with $\text{Num}(BG)$ representing the number of superpixels in BG. We then normalize S_B to obtain the background-based saliency map. To further increase the difference between background and foreground, we process S_B using a logistic function:

$$S_B^+ = f(S_B) = \frac{1}{1 + \exp(-a(S_B - b))} \quad (6)$$

where a controls the weight of difference and b is the threshold that determines whether saliency values are enhanced or suppressed. We set $a = 10$ and $b = 0.7$. [Figure 1: see original paper] shows the effects of S_B and S_B^+. It can be observed that S_B highlights salient regions, while S_B^+ further increases the distinction between background and foreground, enhancing the foreground and suppressing the background.

1.2 Shape Completeness Saliency Map

While the background-based saliency map tends to highlight objects, the shape completeness saliency map better suppresses background noise. To achieve complementary effects, we fuse the background-based saliency map S_B^+ with the shape completeness saliency map S_C^+:

$$S_I = f(S_B^+) \cdot f(S_C^+) \quad (9)$$

Considering that salient objects have well-defined closed boundaries [**?**], this section introduces shape completeness to improve saliency detection quality.

By applying different thresholds $\xi \in [0, \xi_N]$ to the UCM, we obtain a hierarchical segmentation $\{P_\xi\}$. To determine whether a region in $P_\xi = \{R_i\}$ forms a complete closed shape, we propose measuring the expected number of times the region is activated across the entire hierarchical segmentation space. Using this approach, the more frequently similar regions are activated, the higher their probability of being salient. The indicator map for threshold ξ can be defined as:

$$Ind_\xi(x) = \begin{cases} s & \text{if } x \in R, R \in P_\xi^{\text{in}} \\ 0 & \text{if } x \in R, R \in P_\xi^{\text{out}} \end{cases} \quad (7)$$

where $P_\xi^{\text{in}} = \{R_i | R_i \cap BG = \emptyset\}$ denotes the set of interior regions, $P_\xi^{\text{out}} = P_\xi \setminus P_\xi^{\text{in}}$, and s represents the weight of each pixel, taken as S_B^+. A similar method was proposed in [**?**], but our approach differs in several aspects: First, equation (7) uses the robust BG, whereas [**?**] directly selects the boundary set BS composed of four edges. This mechanism in [**?**] would treat any region connected to boundaries as background, potentially producing poor results when salient objects touch boundaries (see the third row in Figure 1: see original

paper). Our improved mechanism can successfully detect target regions even when salient objects connect to boundaries. Additionally, hierarchical segmentation results depend on threshold selection; inappropriate thresholds may cause over-segmentation or under-segmentation. This could lead to background being included in P_ξ^{in} during under-segmentation or foreground being included in P_ξ^{out} during over-segmentation. Therefore, we introduce saliency weights to assign different probabilities to each pixel: for $x \in R$ where $R \in P_\xi^{\text{out}}$, we set $Ind_\xi(x) = s$ instead of 0 [**?**].

Based on the obtained indicator map, we generate a contour completeness map by computing its expectation over the entire hierarchical segmentation space:

$$Q(x) = \int Ind_\xi(x) \cdot p(\xi) d\xi \quad (8)$$

where ξ follows a uniform distribution with probability density function $p(\xi)$. Similarly, we enhance the contrast between background and foreground in the obtained S_C using equation (6), with the result denoted as S_C^+. Figure 2: see original paper-(f) show the results of S_C and S_C^+. Clearly, our method performs better when objects appear at boundary connections.

In most cases, the background-based saliency map S_B^+ with robust BG works well, but this is not true for some complex scenes. As shown in the second example of Figure 1: see original paper, relying solely on boundary prior may incorrectly highlight background regions. The fused result shown in Figure 2: see original paper effectively suppresses background and highlights foreground. However, we note that small non-salient objects (marked by red ellipses in Figure 2: see original paper) are still detected. To eliminate background noise, we propose a propagation algorithm based on seed selection mechanism to smooth saliency values and suppress background. Given a query y_i, manifold ranking solves the following optimization problem [**?**]:

$$f^* = \arg\min_f \frac{1}{2} \left(\sum_{i,j=1}^{N} w_{ij} \left\| \frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}} \right\|^2 + \mu \sum_{i=1}^{N} \|f_i - y_i\|^2 \right) \quad (10)$$

where w_{ij} is an element of the affinity matrix W representing similarity between two superpixels, $D = \text{diag}\{d_{11}, \ldots, d_{nn}\}$ is the degree matrix with $d_{ii} = \sum_j w_{ij}$. However, incorrect seed nodes can lead to incorrect propagation results, especially when background and foreground are similar. To reduce propagation errors, we propose a propagation algorithm that fuses mid-level features with a seed selection mechanism.

We use the mid-level clustering algorithm proposed in [**?**] to merge superpixels. Superpixels with similar structures may have similar saliency values. Therefore, we define a mid-level-feature-based similarity matrix A as:

$$p_{ij} = \begin{cases} w_{ij} & \text{if } i \text{ and } j \text{ belong to the same region} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where $NB(i)$ denotes the neighbors of node i (including directly and indirectly connected nodes). We define $Z = \text{diag}\{\theta_1, \ldots, \theta_n\}$, where $\theta_i = 1$ if node i is a seed node and 0 otherwise. Based on this definition, the loss function of the propagation algorithm can be written as:

$$f^* = \arg\min_f \frac{1}{2} \left(\sum_{i,j=1}^{N} p_{ij} \left\| \frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}} \right\|^2 + \mu \sum_{i=1}^{N} \theta_i \|f_i - y_i\|^2 \right) \quad (13)$$

Equation (13) can be rewritten as:

$$f^* = (D - \alpha P)^{-1} ZY \quad (14)$$

where $D = \text{diag}\{d_{11}, \ldots, d_{nn}\}$ with $d_{ii} = \sum_j p_{ij}$, and Y represents the query matrix corresponding to determined foreground. The seed selection matrix Z can be obtained by three-level thresholding of S_I:

$$\theta_i = \begin{cases} 1 & \text{if } S_I(i) \geq \text{th} \\ 0 & \text{if } S_I(i) < \text{th} \end{cases} \quad (15)$$

Here, $\text{th} = \text{mean}(S_I)$. We obtain the ranking vector f^* using equation (14) and normalize it to $(0, 1)$ to obtain the final saliency map:

$$S_{\text{final}}(i) = f^*(i), \quad i = 1, \ldots, N \quad (16)$$

Examples in Figure 2: see original paper demonstrate the effectiveness of our proposed method, showing that foreground objects are highlighted more accurately while background is further suppressed.

2 Experiments

To validate our algorithm' s performance, we conducted evaluations on three datasets: ASD [?], DUT-OMRON [?], and ECSSD [?]. We compared our method with nine representative and related state-of-the-art methods: SF [?], MR [?], BFS [?], LPS [?], MB [?], RR [?], CGV [?], SRD [?], and HCC [?]. For fair comparison, results of other algorithms were obtained by running the source code provided by the authors.

2.1 Qualitative Analysis

[Figure 3: see original paper] shows visual quality comparisons between our algorithm and other methods. The results demonstrate that our algorithm generates saliency maps closer to ground truth. For images containing single objects (first two rows), our method perfectly highlights salient targets while effectively suppressing background noise. When images have complex background structures, our algorithm still achieves good results with minimal background noise. For example, in rows 3 and 4, our saliency map can uniformly highlight foreground objects, whereas other algorithms fail to extract salient objects from complex backgrounds. Moreover, when salient targets and background have similar appearances, our algorithm can still accurately detect salient regions, while other methods either fail to identify salient targets or incorrectly over-highlight background regions, as shown in rows 5 and 6. These results prove the robustness and effectiveness of our method in highlighting salient targets and suppressing background regions.

2.2 Quantitative Analysis

For performance comparison, we first evaluate our method using standard Precision-Recall (PR) curves. The PR curves are obtained by binarizing saliency maps with fixed thresholds ranging from [0,255] to generate 256 binary maps, which are then compared with ground truth. Figure 4: see original paper shows the PR curve results, with ASD, DUT, and ECSSD displayed from top to bottom.

Since precision and recall often influence each other, we adopt F-measure for comprehensive evaluation. F-measure is the weighted harmonic mean of precision (P) and recall (R): $F_\beta = (1+\beta^2)P \cdot R/(\beta^2 P + R)$, with balance factor $\beta^2 = 0.3$ [**?**]. Additionally, these metrics focus primarily on the probability of correctly detecting salient pixels while ignoring the impact of correctly assigning non-salient pixels, thus not fully measuring saliency quality. Therefore, we also introduce Mean Absolute Error (MAE) for performance evaluation: MAE = mean($|S - G|$), where S represents the saliency map and G denotes ground truth. Figure 4: see original paper and 4(c) show F-measure and MAE results, with corresponding data listed in .

As shown in [Figure 4: see original paper], our method outperforms other algorithms across all evaluation metrics and datasets, demonstrating its overall superiority. Taking the challenging ECSSD dataset as an example, compared to the second-best method, our approach improves precision, recall, and F-measure by 5.44%, 2.72%, and 1.5% respectively, while reducing MAE by 12.8%. Furthermore, some methods achieve good precision at the cost of low recall (e.g., MR, SRD), leading to imbalance between precision and recall. In contrast, our algorithm achieves the best F-measure across all test datasets. We also observe that our method obtains more advantageous precision values in high recall ranges, indicating its strong capability in suppressing background, which can be

attributed to the joint use of robust boundary background, shape completeness, and robust propagation algorithms.

2.3 Performance Evaluation of Components

Our algorithm involves four main components: background-based saliency map, shape completeness saliency map, fusion mechanism, and seed selection propagation mechanism. [Figure 5: see original paper] shows PR curves testing each component on the ASD dataset. The green and purple curves (see electronic version) represent whether the boundary set undergoes foreground noise processing. Clearly, foreground noise processing improves the curve of our background distinctiveness map. The cyan and pink curves represent whether the shape completeness map adopts our proposed improved mechanism. The cyan curve significantly outperforms the pink curve, well demonstrating the effectiveness of our shape completeness improvement. The blue and red curves represent the fused result and the final result after optimization, respectively. The results show that the optimized fusion curve achieves significant improvement, strongly proving the contribution of our optimization mechanism.

2.4 Running Time Comparison

shows the average running time of our algorithm compared with other methods on the ASD dataset, implemented using MATLAB R2014b. All running times were computed on a 64-bit Windows system with an Intel Core i5-4460 3.20 GHz CPU and 8 GB RAM. As shown in , our algorithm is faster than BFS, LPS, CGV, and HCC. Although it is slightly slower than SF, MR, RR, and SRD, our method' s performance far exceeds these approaches across all evaluation metrics. Since our algorithm is implemented in MATLAB, rewriting it in C++ would yield further speed improvements.

3 Conclusion

This paper proposes a saliency detection algorithm based on scatter-shape guidance and an optimization function. To accurately extract a robust background, we introduce a scatter prior mechanism and obtain a corresponding background-based saliency map by computing distinctiveness from boundary backgrounds. We then propose a shape completeness mechanism and generate a shape completeness saliency map by combining robust boundary nodes with the background-based saliency map. Finally, to further eliminate background and highlight foreground, we propose a propagation algorithm based on mid-level features and seed selection mechanism to optimize the fused result of the two saliency maps.

Future work will consider using deep learning methods to extract higher-level features for improving saliency detection accuracy. We also plan to apply saliency detection to other tasks such as object recognition and image retrieval to improve computational efficiency and accuracy.

References

[1] Huang Juan, Mei Zhechuan, Huang Xiaoming. Color image segmentation based on integration of regional combination and graph cuts [J]. Computer Engineering and Applications, 2016, 52 (17): 225-228.

[2] Gu Suhang, Ma Zhenghua, Lyu Jidong. Recognition method of apple target based on significant contour [J]. Application Research of Computers, 2017, 34 (8): 2551-2556.

[3] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system: an overview [J]. IEEE Trans on Consumer Electronics, 2000, 46 (4): 1103-1127.

[4] Yang Yang, Yang Linjun, Wu Gangshan, et al. Image relevance prediction using query-context bag-of-object retrieval model [J]. IEEE Trans on Multimedia, 2014, 16 (6): 1700-1712.

[5] Wang Jiaojiao, Liu Zhengyi, Li Hui. Feature fusing and objectness enhanced approach of saliency detection [J]. Computer Engineering and Applications, 2017, 53 (2): 195-200.

[6] Xia Chenxing, Zhang Hanling. Saliency detection combining multi-layer integration algorithm with background prior and energy function [C]// Proc of Pacific Rim Conference on Multimedia. 2016: 11-21.

[7] Li Changyang, Yuan Yuchen, Cai Weidong, et al. Robust saliency detection via regularized random walks ranking [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2710-2717.

[8] Wang Jianpeng, Lu Huchuan, Li Xiaohui, et al. Saliency detection via background and foreground seed selection [J]. Neurocomputing, 2015, 152 (C): 359-368.

[9] Li H, Lu H, Lin Z, et al. Inner and inter label propagation: salient object detection in the wild [J]. IEEE Trans on Image Processing, 2015, 24 (10): 3176-3186.

[10] Lee G, Tai Yuwing W, Kim J. Deep saliency with encoded low level distance map and high level features [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 660-668.

[11] Zhao Rui, Ouyang Wanli, Li Hongsheng, et al. Saliency detection by multi-context deep learning [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1265-1274.

[12] Wang Lijun, Lu Huchuan, Ruan Xiang, et al. Deep networks for saliency detection via local estimation and global search [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3183-3192.

[13] Sun Jingang, Lu Huchuan, Liu Xiuping. Saliency region detection based on Markov absorption probabilities [J]. IEEE Trans Image Process, 2015, 24 (5):

1639-1649.

[14] Dollár P, Zitnick C L. Structured forests for fast edge detection [C]// Proc of International Conference on Computer Vision. 2013: 1841-1848.

[15] Jiang Huaizu, Wang Jingdong, Yuan Zejian, et al. Automatic salient object segmentation based on context and shape prior [C]// Proc of British Machine Vision Conference. 2011.

[16] Liu Qin, Hong Xiaopeng, Zou Beiji, et al. Hierarchical contour closure-based holistic salient object detection [J]. IEEE Trans on Image Processing, 2017, 26 (9): 4537-4552.

[17] Yang Chuan, Zhang Lihe, Lu Huchuan, et al. Saliency detection via graph-based manifold ranking [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3166-3173.

[18] Kim T H, Lee K M, Sang U L. Learning full pairwise affinities for spectral segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010: 2101-2108.

[19] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2009: 1597-1604.

[20] Yan Qiong, Xu Li, Shi Jianping, et al. Hierarchical saliency detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1155-1162.

[21] Perazzi F, Krähenbühl P, Pritch Y, et al. Saliency filters: contrast based filtering for salient region detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2012: 733-740.

[22] Zhang Jianming, Sclaroff S, Lin Zhe, et al. Minimum barrier salient object detection at 80 FPS [C]// Proc of International Conference on Computer Vision. 2015: 1404-1412.

[23] Yang Kaifu, Li Hui, Li Chaoyi, et al. A unified framework for salient structure detection by contour-guided visual search [J]. IEEE Trans on Image Processing, 2016, 25 (8): 3475-3488.

[24] Zhou Li, Yang Zhaohui, Yuan Qing, et al. Salient region detection via integrating diffusion-based compactness and local contrast [J]. IEEE Trans on Image Processing, 2015, 24 (11): 3308-3320.

Note: Figure translations are in progress. See original paper for figures.