CrossMark

Salient region detection via unit boundary distribution and energy optimization

Hong Li[1] · Enhua Wu[1,2] · Wen Wu[1]

Abstract Due to recent rapid development of computer vision applications such as object recognition and image segmentation, it has become increasingly important to generate reliable saliency maps to uniformly highlight the desired salient object. In this paper, we present a novel bottom-up salient region detection method by exploiting contrast prior and the relationship between the salient region detection and graph based semi-supervised learning problem. First, we compute a preliminary initial saliency map by a newly proposed technique named unit boundary distribution and several refinement schemes. Second, after obtaining the indication map generated via a double threshold operation on the initial saliency map, we model the final saliency inference problem as a graph based semi-supervised learning approach by solving a energy minimization problem. Both quantitative and qualitative evaluations on three widely used datasets demonstrate the superiority of the proposed method to other twenty-one state-of-the-art methods.

Keywords Salient region detection · Unit boundary distribution · Global contrast · Local contrast · Energy minimization

✉ Hong Li
  yb27441@umac.mo; valhongli@gmail.com

  Enhua Wu
  ehwu@umac.mo

  Wen Wu
  wenwu@umac.mo

[1] Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau 999078, China

[2] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100864, China

⑩ Springer

1 Introduction

Salient region detection mainly aims to detect and uniformly emphasize the most important objects in a scene. It has attracted mass attention and accomplished big progress during the past two decades for its wide range of applications such as object aware image retargeting [11, 33], image categorization [31], image and video compression [15], to name a few.

Generally, we can categorize existing works into top-down or bottom-up methods. Top-down models are task-driven and usually need high-level knowledge. On the other hand bottom-up methods are usually based on low-level visual features like intensity, pattern, or orientation from pixels or regions. In this work, we only focus on bottom-up salient region detection models.

For automatic bottom-up models, the most widely-used principles are contrast prior and background prior. Contrast prior assumes that the appearance contrast between salient objects and background regions are high. For a specific region, its contrast is computed as the sum of differences between it and its local neighboring or the entire image regions respectively. This region will be considered as salient if the computed contrast is high. This assumption is very intuitive and easy to realize. So it has been widely applied to numerous models [2, 5, 10, 14, 16, 17, 24, 27, 36, 42], implicitly or explicitly.

Previous contrast prior based models can be categorized as local methods [5, 14, 16, 17, 24, 36] or global [2, 10, 27, 42] methods according to the extent of context where the contrast is evaluated. Although contrast prior has enjoyed remarkable success, they still have various limitations. The most typical one is that they tend to detect the boundaries of the salient object instead of highlighting the entire object uniformly. As we know, the purpose of salient region detection is to detect uniform object regions because most applications usually require entire object regions instead of boundaries, such as in [33]. On the other hand, it is difficult to get the entire object regions by using the generated unclosed boundaries. So it is insufficient to evaluate saliency only using contrast prior.

Recently, to tackle above shortcoming, background prior has been widely adopted to evaluate saliency. It is based on an observation that most photographers usually will not crop salient object along the view frame. That is to say, four image borders (top, right, bottom and left) are mostly consisted of background regions. Based on this prior, a coarse map is obtained by propagating the background information from those background regions to other regions. Then an initial saliency map is generated by computing the complement of the coarse map. The final saliency map can be computed by using this initial saliency map. The first influential background prior based model is proposed by Wei et al. [37]. They investigate saliency from a different perspective: modeling the background instead of the object. From then on, many works built upon this prior have been proposed [22, 23, 28, 34, 43, 45, 46] and they all have a better performance than previous models which merely consider the contrast prior. This suggests that the background prior is effective. However, all most all these models simply treat the whole image boundary as background and then get the background inference which can be used to further generate final foreground saliency. This simple treatment may fail when the desired salient object touches the image border.

In this paper, we simulate salient region detection according to above-mentioned priors. Firstly, we introduce a novel contrast based background model: unit boundary distribution. This measurement can effectively exploit the intrinsic relationship between contrast prior and background prior to model more robust backgrounds. The saliency map is generated by solving an energy minimization problem. The main pipeline of the proposed algorithm is shown in Fig. 1. And the main contributions of this paper are as follows:
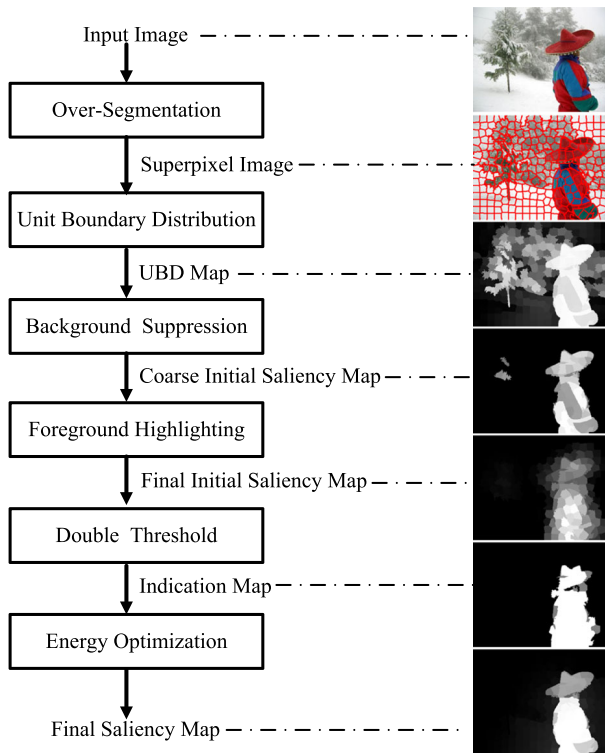
Fig. 1 The framework of our proposed salient region detection method

- A novel technique named unit boundary distribution which can exploit the background prior and contrast prior more effectively.
- A more accurate initial saliency map generation scheme which is built upon unit boundary distribution and several techniques.
- A novel algorithm which combines contrast prior, background prior and energy minimization to effectively detect the desired salient region.

The rest of the paper is organized as follows. First, related works are summarized in Section 2. In Section 3, we first present the details of generation of initial saliency map in Section 3.1 and then present our final saliency map generation scheme in Section 3.2. Experimental results and analysis are give in Section 4. Finally, conclusions and future work are given in Section 5.

2 Related works

During the past two decades, numerous bottom-up saliency models have been proposed to detect the salient region in an image. A very comprehensive survey of these previous models can be found in [6–8]. Our work is based on two priors, i.e., contrast prior and background prior. So we only review several most influential works based on these two priors.

One of the earliest local contrast based models is proposed by Itti et al. [17]. They employ DoG (Difference of Gaussian) technique to model multi-scale information of features including color, intensity and orientation. Then they generate the saliency map by computing the center-surround differences according to the multi-scale information. Harel et al. [16] further extend this idea by using a graph-based approach to non-linearly combine these different feature channels. Later, Goferman et al. [14] simultaneously combine local low-level clues and visual organization rules to highlight the salient region along with their context. These local contrast based algorithms tend to generate higher saliency scores near edges instead of uniformly highlighting the smooth object interior.

Viewed from another perspective, global contrast based models evaluate saliency by exploiting the contrast relationships over whole image. Achanta et al. [2] propose a frequency-tuned method to gain consistent results by utilizing the difference of the average image color. Perazzi et al. [27] exploit the variance of spatial distribution of each color and show that high-dimensional Gaussian filters can be used to measure saliency. These global contrast based methods cannot distinguish salient regions from background regions when they have similar colors.

Background prior is proposed to complement the contrast prior. It is based on a different view point: exploiting the feature distribution of background. Wei et al. [37] find that the distance of a pair of background regions is shorter than that of a region from the salient object and a region from the background. They employ both background prior and geodesic distance to evaluate saliency. Later, Yang et al. [43] treat all four image sides as background and utilize graph-based manifold ranking to generate the final saliency map. Jiang et al. [19] also treat four image sides as background and regard these pixels on image borders as absorbing nodes. Then saliency is measured according to absorbed time in a Markov chain. More recently, Zhu et al. [46] propose a new background measurement named boundary connectivity and achieve the final saliency via an energy minimization technique. Sun et al. [34] treat left and top image borders as background cues and employ Markov absorption probability on a sparse 2-ring graph to estimate saliency. These models achieve better performance than previous contrast prior based models. However, they only apply this prior in a straightforward manner. This will make the model fail when the salient object touches the image border.

3 Proposed model

Contrast based salient object detection usually consists of two main components: contrast evaluation and saliency inference. Our model also consists of two main steps: initial saliency map evaluation and final saliency map generation. In Section 3.1, we give the details of evaluation of initial saliency map. The final saliency map generation will be presented in Section 3.2.

To reduce the computational cost, we employ superpixel as the basic processing unit to represent the image. There are many edge-preserving models for generating these superpixels [3, 13, 30, 39]. Here, we employ SLIC [3] to achieve this goal for its high efficiency. Given an input image I, we over-segment it into N (e.g., $N = 300$) regions $\{t_1, t_2, ..., t_N\}$. For each region (also known as superpixel), we use the average feature value of pixels belonged to this region to represent it. In this work, we utilize $CIE\ Lab$ color space to evaluate saliency as this color space has been shown to be effective in saliency detection [2, 8]. Therefore, for each superpixel t_i, $F_i = \{F_i^L, F_i^a, F_i^b\}$ is a feature vector, denoting L, a and b feature of superpixel t_i, respectively.

3.1 Initial saliency map via unit boundary distribution and contrast based refinement

3.1.1 Unit boundary distribution

Contrast prior is usually employed as either global or local perspective. We employ both of them in our model. Firstly, the global contrast is utilized to generate the coarse initial saliency based on our unit boundary distribution. Then, local contrast is employed to generate the fine initial saliency map. Based on above notations, the global contrast is denoted as follows:

$$GC_i = \sum_{j=1}^{N} \|F_i - F_j\|^2 \tag{1}$$

where i denotes current superpixel t_i, N is the number of superpixels.

To better exploit the background information, we compute the boundary contrast according to different image boundaries. Take top image boundary as example, , we define the top boundary contrast as:

$$BC_i^t = \sum_{j=1}^{n_t} \|F_i - F_j\|^2 \tag{2}$$

here, n_t represents the superpixels along top image boundary. Similarily, We define other three boundary contrast. The proposed unit boundary distribution is defined as:

$$UBD_i^t = \frac{\sum_{j=1}^{n_t} \|F_i - F_j\|^2}{\sum_{k=1}^{N} \|F_i - F_k\|^2} \tag{3}$$

we use different subscripts (j and k) for clearer expression. Finally, the overall unit boundary distribution is computed via

$$UBD_i = Normalize(UBD_i^t + UBD_i^r + UBD_i^b + UBD_i^l) \tag{4}$$

here r, b and l denotes right, bottom and left image boundary, respectively.

Figure 2 is an illustration of proposed unit boundary distribution technique. To present the computation process of (3) and (4) more clearly, we only use a small number of superpixels in (a) and (b). (a) shows the boundary contrast of current superpixel and green arrows denote all the involved boundary superpixels. (b) shows the global contrast and blue arrows denote all the involved superpixels. (c) is the corresponding top, right, bottom and left boundary contrast map, respectively. (d) and (e) are the boundary contrast map and global contrast map. The final unit boundary distribution map is given in (f). It can be seen that although some background regions are highlighted, the desired foreground region is extracted uniformly. Next, we will present the schemes used to tackle above shortcoming.

#### 3.1.2 Refined final initial saliency map

Extensive experiments have shown that global contrast based models usually generate undesired high saliency values for some non-salient regions. Figure 3b gives an illustration of this situation. We can see that some background regions also have high saliency values. As shown in second row of Fig. 3b, it should also be noted that foreground region may be wrongly suppressed when it has similar color distribution to background regions.
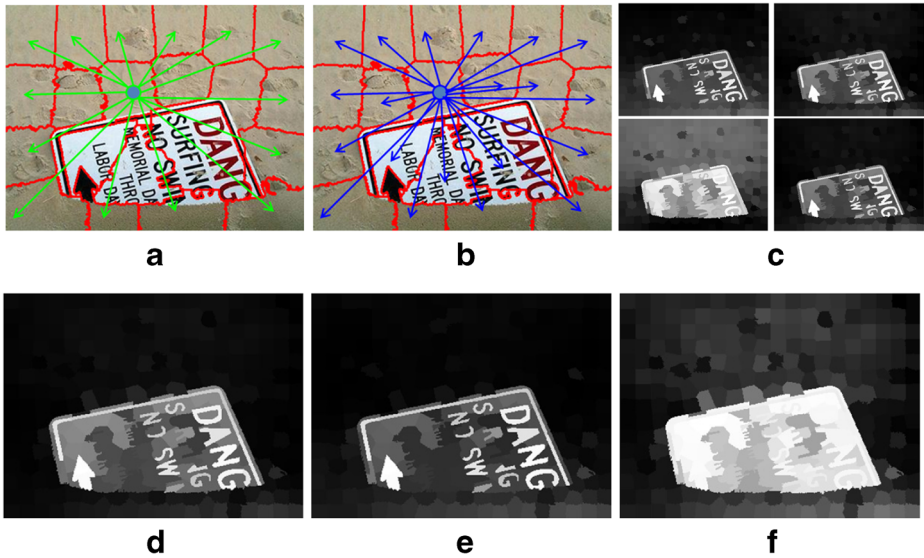
Fig. 2 Unit boundary distribution. **a** boundary contrast. **b** global contrast **c** *top*, *right*, *bottom* and *left* boundary contrast map **d** Boundary contrast map **e** Global contrast map. **f** Unit boundary distribution

To overcome these two shortcomings, we propose to utilize local contrast to refine the coarse initial saliency maps. Firstly, we tackle the problem of highlighting non-salient regions. Then comes the problem of wrongly suppressing the foreground regions.

Background Suppression To suppress the non-salient regions, i.e., the background regions, two techniques are proposed: contrast weights and adaptive selection. We find that
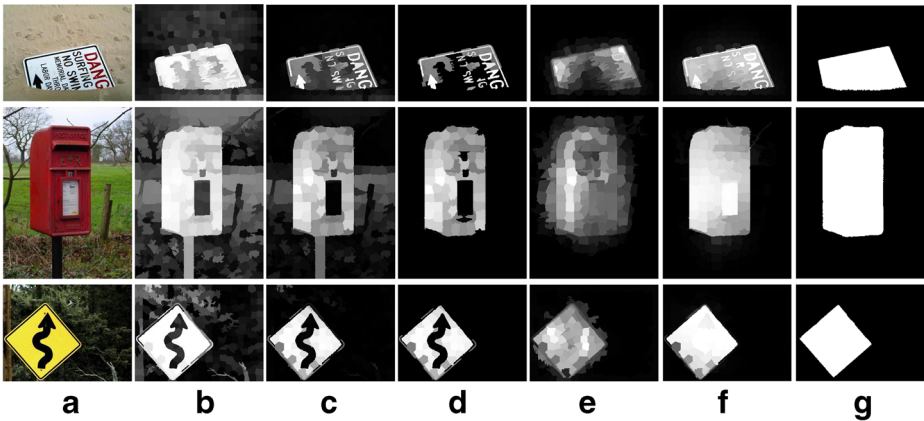
Fig. 3 Illustration of local contrast enhancement. **a** are input images. **b** are the unit boundary distribution maps. **c** are the *top* and *left* image sides weighted maps (BS). **d** are the final background suppressed saliency maps ($\widetilde{BS}$). **e** local contrast enhanced initial saliency maps (LC). **f** are final saliency maps. **g** are ground truth images

salient regions rarely touch the top and left image border. Based on this observation, the coarse background suppressed saliency map is defined as

$$BS_i = UBD_i * \sum_{j=1}^{n_t+n_l} \|F_i - F_j\|^2 \tag{5}$$

where n_t and n_l denotes superpixels located on top and left image side respectively. We can see from Fig. 3c that most non-salient regions are eliminated by this process. For the images with more complex backgrounds, there still remain some background regions. To remove these redundant regions, an adaptive selection scheme is defined as

$$\widetilde{BS}_i = \begin{cases} BS_i & , if \ BS_i > \tau \\ max(0, \ BS_i - \tau) & , else \end{cases} \tag{6}$$

where max is to choose the maximum between 0 and $BS_i - \tau$, τ is defined as $\tau = 0.2 * (max(BS) - mean(BS)) + mean(BS)$, BS is a vector obtained via (5), max and $mean$ denotes maximum and mean value of a vector respectively. Figure 3d shows the final background suppressed saliency map $\widetilde{BS}$. We can see that the results are much cleaner than that of Fig. 3c. It should be noted that the pillar in second row has been removed because it is different from the red box. We can see from the last row that the salient region should not include this pillar.

Foreground highlighting Although background suppression can remove undesired background regions, it sometimes may also suppress salient regions. So we employ local contrast to highlight the wrongly suppressed salient regions, i.e., the foreground regions. To prevent Local contrast based models from highlighting undesired background regions, we use coarse saliency map obtained via (6) to suppress the wrongly emphasized background regions. The final initial saliency map based on local contrast is defined as

$$LC_i = \frac{\sum_{j \in N_i} \widetilde{BS}_j * A_j}{\sum_{j \in N_i} A_j} * (\max_{j \in N_i} \|F_i - F_j\|^2 - \min_{j \in N_i} \|F_i - F_j\|^2) \tag{7}$$

where N_i denotes the neighboring nodes of superpixel i, A_j denotes the area of superpixel j, in this work we regard it as the pixel number of superpixel j. Figure 3e shows the results of final initial saliency maps. We can see that, especially in second and third row, the wrongly suppressed foreground regions are highlighted. This proves the effectiveness of our proposed scheme.

3.2 Saliency detection via energy optimization

When we get the initial saliency map, the very core problem is to generate the final saliency map according to the initial saliency map. The final initial saliency map obtained via (7) is a good prior distribution for salient region detection. Based on this prior knowledge, we model the final saliency detection problem as a graph based semi-supervised learning problem. It consists of four components: formation of initial query, construction of affinity graph, energy minimization and label querying.

Initial queries by double threshold We model the final saliency detection as a two class inference problem: background and foreground detection. Given initial saliency map LC, the initial queries are defined as

$$q_i = \begin{cases} 1 & , if \ LC_i \geq (\Theta_{FG} * \widehat{LC}) \\ -1 & , if \ LC_i \leq (\Theta_{BG} * \widehat{LC}) \\ 0 & , else \end{cases} \tag{8}$$

where $\widehat{LC}$ denotes the mean value of initial saliency map LC, Θ_{FG} and Θ_{BG} are two parameters which are used to defined the determinate foreground and background labels, respectively. These two parameters are empirically chosen, $\Theta_{FG} = 2$ and $\Theta_{BG} = 1$, for all the experiments. Then the problem is to classify the data points which are labeled as 0 into either -1 (background) or 1 (foreground). Figure 4d shows that this double threshold can effectively separate the indeterminate regions from determinate regions.

Affinity graph Graph is usually defined as $G = (V, E, W)$, where V, E and W denotes graph nodes, edge connection and edge weights respectively. It mainly consists of two step: graph structure modeling and graph edge weights formation. Graph structure is usually modeled as $k - nn$ and edge weights are formed using Gaussian kernel: $\omega_{ij} = exp(-d^2/\sigma^2)$. However, the $k - nn$ graph only considers the feature distribution.

We over-segment each input image into homogenous regions and regard each region as a node in the graph G. For graph structure, to take local smoothness constraint into consideration, we construct the graph as a k-ring sparse graph: each node is not only connected to
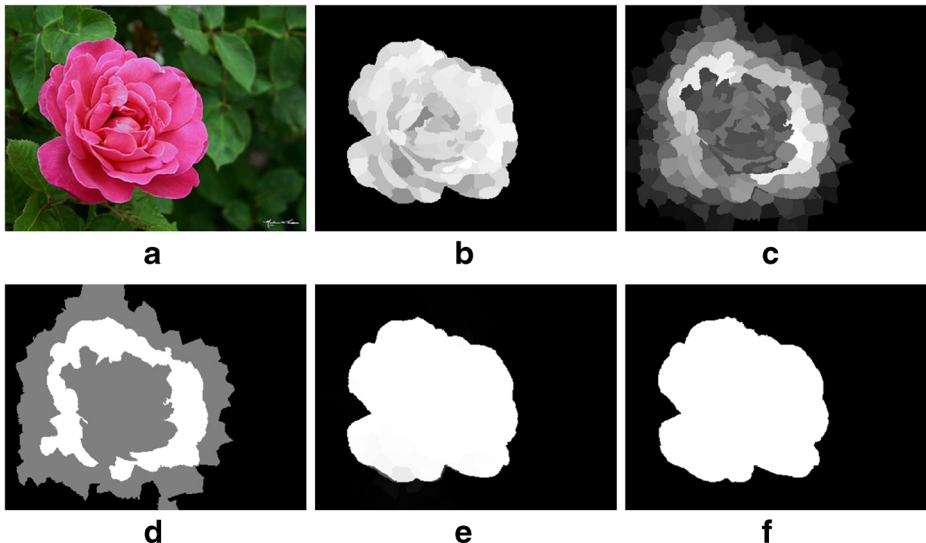
Fig. 4 Illustration of saliency inference. **a** Input image. **b** Coarse initial saliency map obtained via (6). **c** Final initial saliency map obtained via (7). **d** Indication map obtained via (8), *black* regions are determinate background regions, white regions are determinate foreground regions, *gray* regions are regions to be inferenced. **e** Inferenced final saliency map. **f** Ground truth

its direct neighboring nodes, but also connected to its k-layer neighboring nodes. For graph edge weights, they are defined as

$$\omega_{ij} = \begin{cases} exp(\frac{-\|F_i - F_j\|^2}{\sigma}), if \ j \in N_i \\ 0, \qquad\qquad\quad else \end{cases} \tag{9}$$

where N_i denotes all the nodes have connection with node i (k-ring connection), $\sigma = 0.1$ is used to control the weight strength. This graph modeling scheme is illustrated in Fig. 5. It shows the cases of 1-ring sparse graph and 2-ring sparse graph. (a) and (c) are the graph edge connection illustration. We plot the graph edge weights matrix in (b) and (d). In this work, we employ 2-ring sparse graph as our graph.

Energy minimization The energy minimization model is defined as

$$\arg \min_x \sum_{i=1}^n \|d_i x_i - \sum_{j \in N_i} \omega_{ij} x_j\|_2^2 + \sum_{i=1}^n \|x_i - q\|_2^2 \tag{10}$$
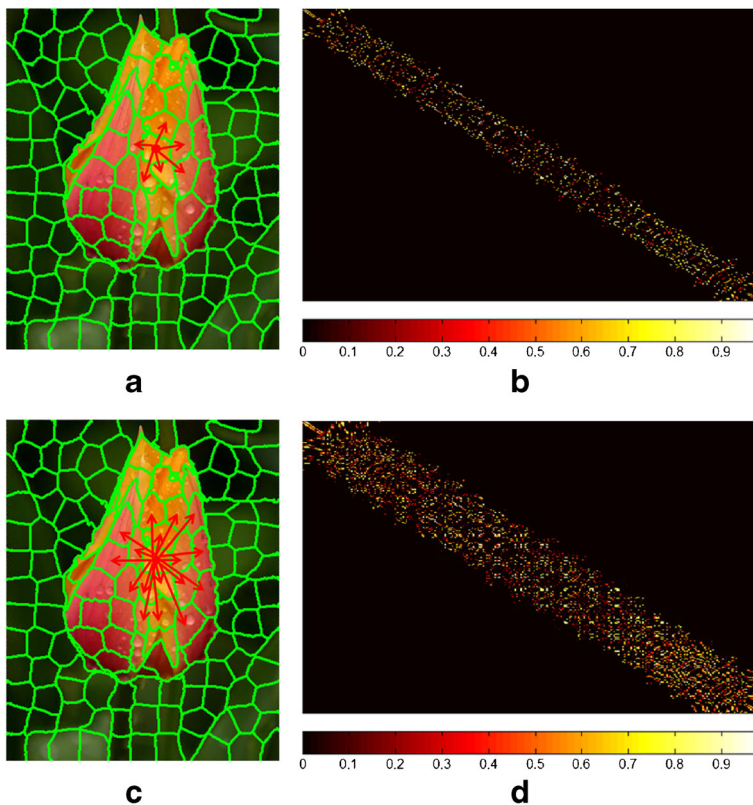
Fig. 5 Illustration of k-ring graph construction. **a** edge connection of 1-ring graph. **b** graph edge weights of **(a)**. **c** edge connection of 2-ring graph. **d** graph edge weights of **(c)**

where q denotes initial queries obtained via (8), N_i denotes all the connected nodes of node i, $d_i = \sum_{j=1}^{n} \omega_{ij}$, n is the number of graph nodes, i.e., the number of superpixels. This energy minimization problem can be easily solved as:

$$(I_q + L^2) * x = q \tag{11}$$

where I_q is a diagonal matrix and is defined as

$$I_q(i, i) = \begin{cases} 1, if \ i \ \in l \\ 0, else \end{cases} \tag{12}$$

where l is the indexes of determinate foreground and background defined in (8), $L = D - W$ is graph Laplacian matrix, W is graph edge weights matrix, D is a diagonal matrix where $D_{ii} = d_i$. This energy minimization model is motivated by the work of [32, 40], in which they use a similar energy optimization scheme to tackle the surface processing problem in geometry processing.

Label inference After solving the energy minimization problem (10), the solution vector x is the propagated saliency score. The final label of node i is defined as

$$S_i = \begin{cases} -1 \ , if \ x_i \leq -1 \\ 1 \ , if \ x_i \geq \ 1 \\ x_i \ , else \end{cases} \tag{13}$$

The determinate foreground and background nodes are denoted as 1 and -1, respectively. The solution vector x stands for the propagated saliency value. Equation (13) is employed to make sure the saliency value stay in range. The label vector S is normalized to [0, 1] to get the final saliency value. Figure 4 shows an example of saliency inference.

4 Experimental results and analysis

In this section, we make extensive quantitative and qualitative evaluations of our model against several state-of-the-art models on four widely used datasets.

4.1 Datasets and compared models

Datasets ASD [2] is also known as MSRA-1000 and consists of 1000 images with accurate binary human-labeled masks. Although it has various images, the foreground is actually obvious among the simple and structured background. It is the mostly widely used dataset.

SOD [26] contains 300 images with complex objects and scenes. Some image contains two or more objects. It is more challenging.

SED1 [4] has 100 images with one salient object. Pixel-wise groundtruth annotations for the salient object are provided.

ECSSD [41] has 1000 images with varied patterns in both background and foreground regions. It contains many semantically meaningful but structurally complex images. It represents more general cases that natural images fall into.

Compared models We compare our model with twenty-one state-of-the-art salient object detection models on above four widely used datasets. The compared models are: IT [17], GB [16], CA [14], FT [2], SF [27], GS [37], GMR [43], MAP [34], MC [19], HS [41], BM [38], CB [18], CHM [21], FES [35], HDCT [20], LRMR [29], MSS [1], PCA [25],SVO

[9], SWD [12], LGH [44]. All the compared saliency maps of these twenty-one models are generated by using the source code released by the authors of corresponding paper. The parameters of each model are set to optimal according to the paper for a fair comparison.

4.2 Quantitative evaluation

To evaluate the saliency detection performance quantitatively, we use three commonly used metrics including the PR (precision-recall) curve, F-measure and MAE (mean absolute error). Precision is defined as the ratio of correctly detected salient pixels number with
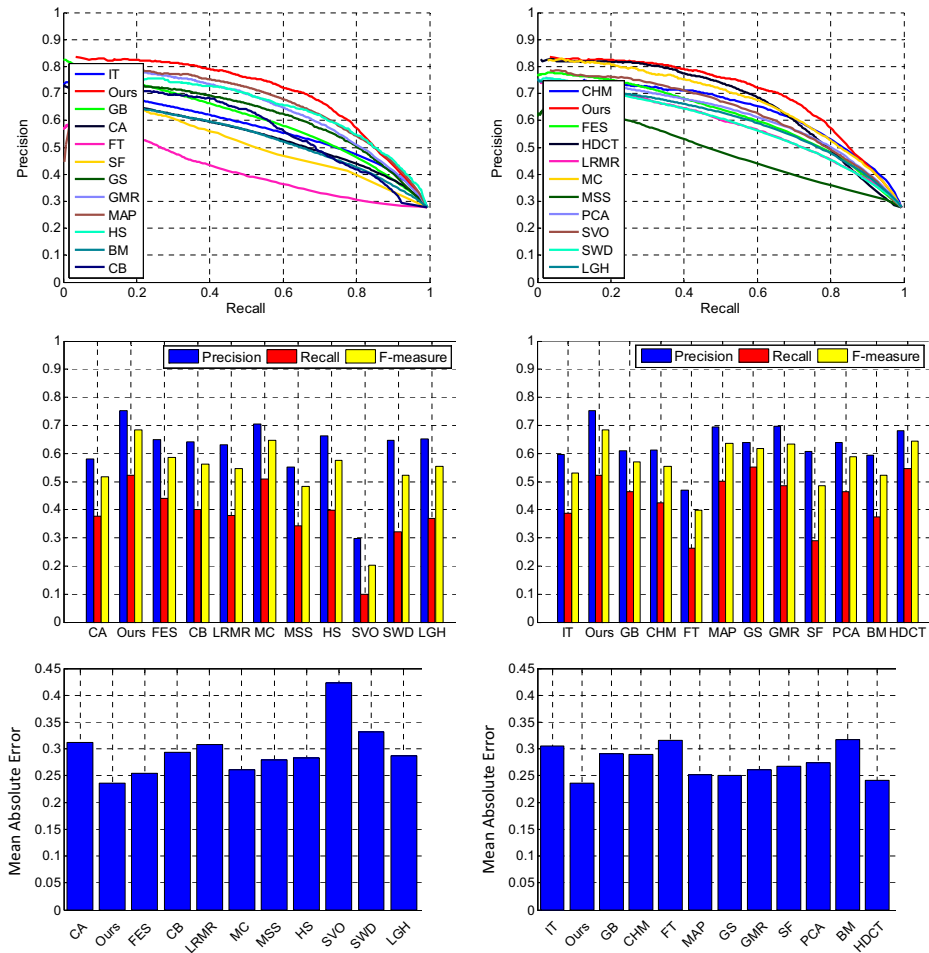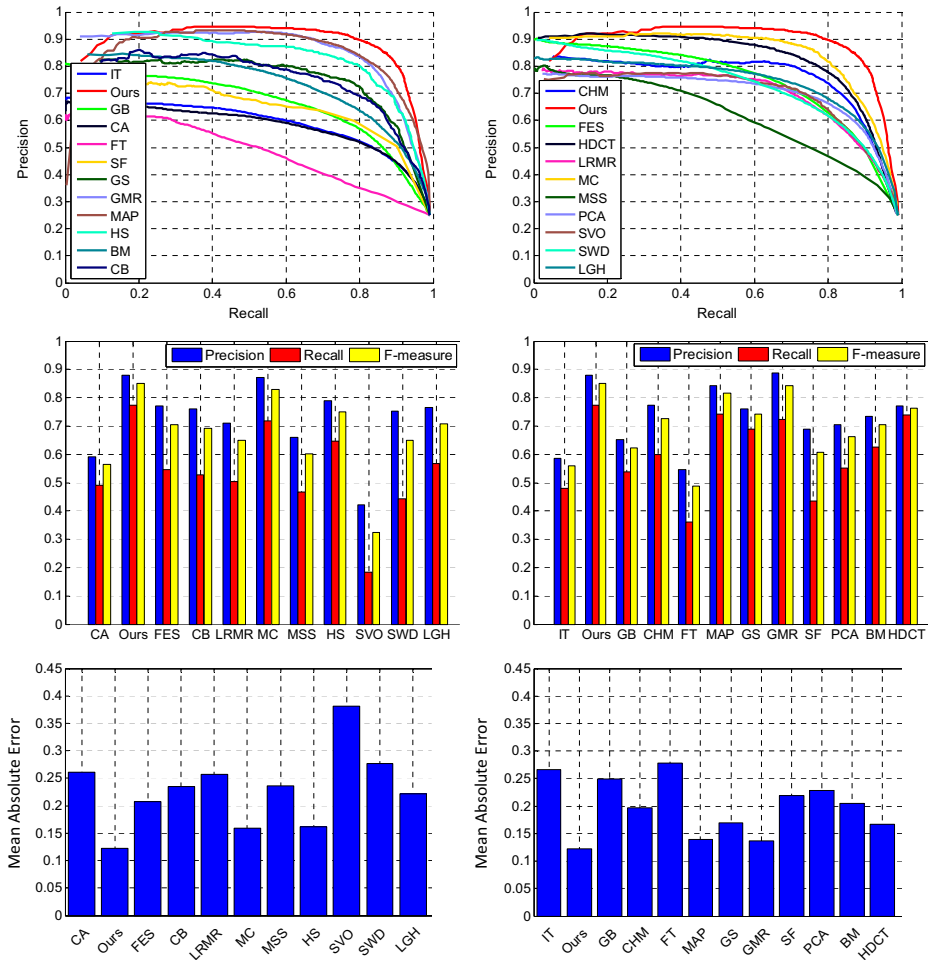

Fig. 6 *First row*: precision-recall curves of different methods. *Second row*: precision, recall and F-measure using an adaptive threshold. *Last row*: MAE. All experiments are carried out on ECSSD dataset. The proposed method performs well for all these metrics

respect to all salient pixels number. Recall is defined as the ratio of correctly detected salient pixel number with respect to ground truth number.

Given the saliency map, the binarized saliency map is generated using threshold value from 0 to 255. The precision and recall at each value of the threshold are computed via above definition. We plot the precision-recall curve using generated precision-recall pairs. The average precision-recall curve is obtained by combining the results from all the images of each dataset. The F-measure is defined as

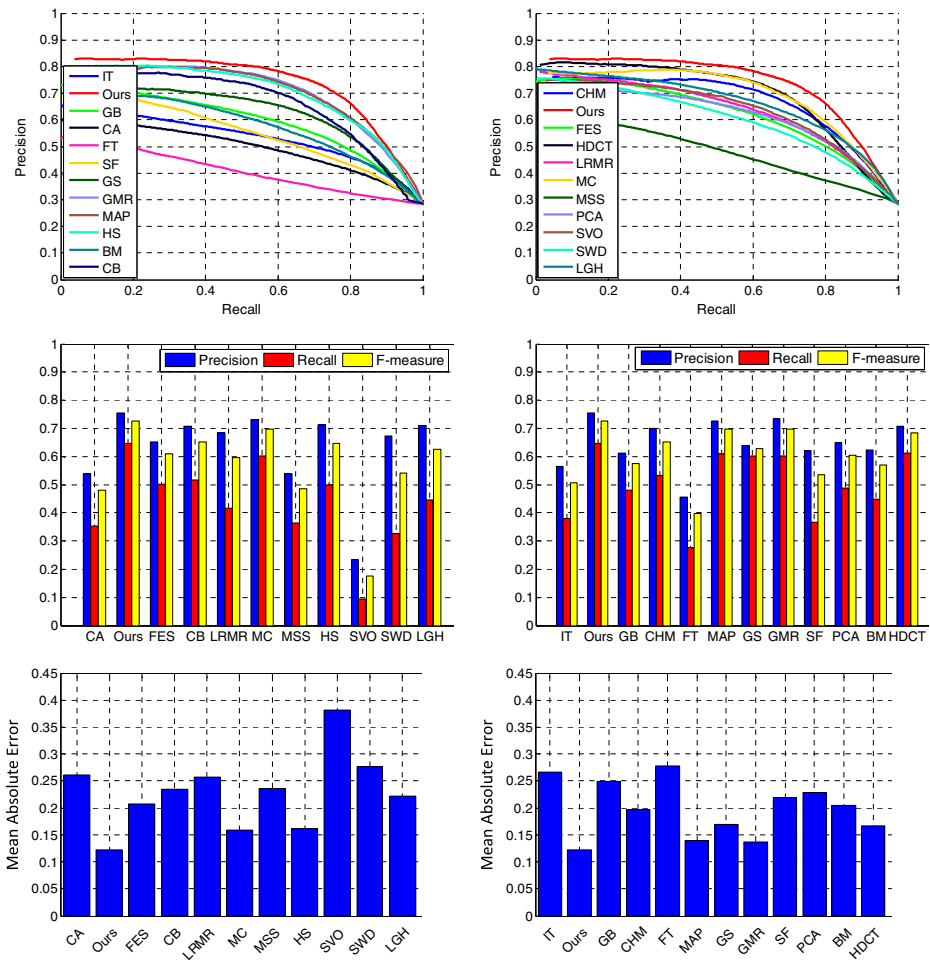$$F_{\beta^2} = \frac{(1 + \beta^2) * Precision * Recall}{\beta^2 * Precision + Recall}$$

Fig. 7 *First row*: precision-recall curves of different methods. *Second row*: precision, recall and F-measure using an adaptive threshold. *Last row*: MAE. All experiments are carried out on ASD dataset. The proposed method performs very well

It jointly considers recall and precision. To compute F_{β^2}, we set $\beta^2 = 0.3$ according to [2], and apply adaptive threshold σ_a to the saliency map before computing F_{β^2}, σ_a is defined as

$$\sigma_a = \frac{2}{W * H} \sum_{i=1}^{W} \sum_{j=1}^{H} S_{ij}$$

where W and H denote the width and height of the saliency map S, respectively. For salient region detection evaluation, MAE (Mean Absolute Error) is better than PR curves because the PR curves are limited in that they only consider whether the object saliency is higher than
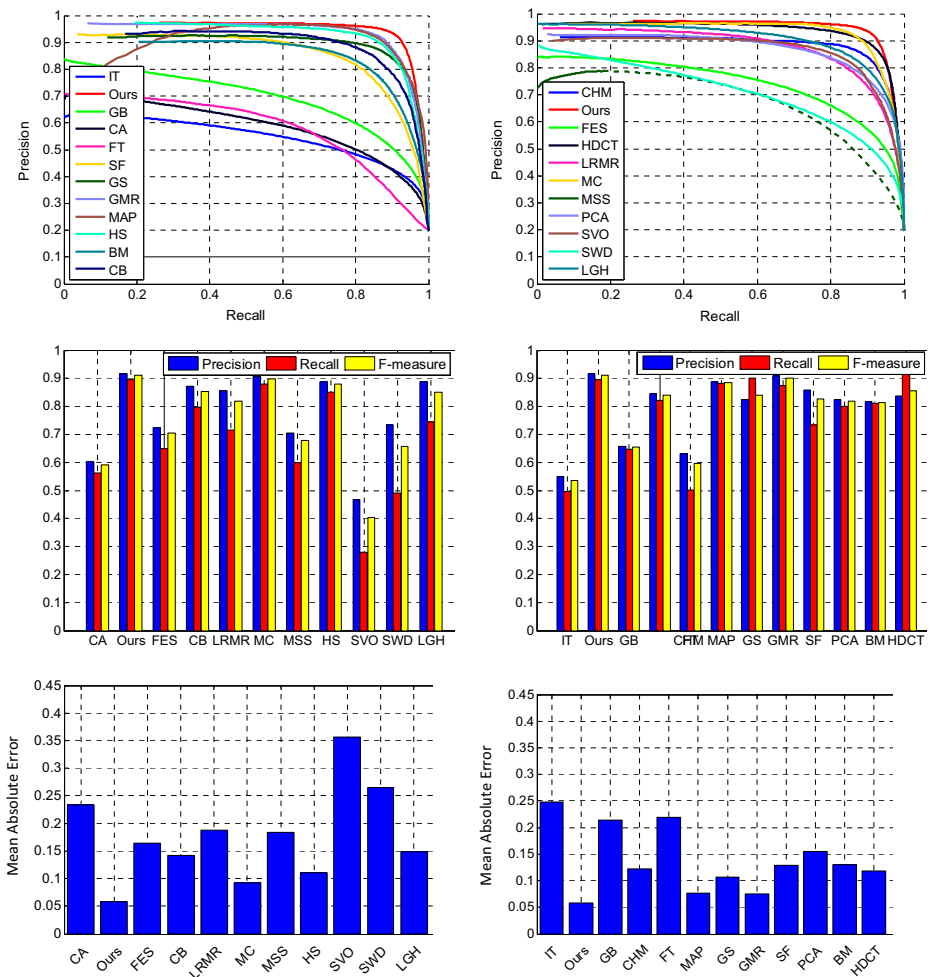

Fig. 8 *First row*: precision-recall curves of different methods. *Second row*: precision, recall and F-measure using an adaptive threshold. *Last row*: MAE. All experiments are carried out on SOD dataset. The proposed method performs well in all these metrics

the background saliency. MAE is employed to evaluate the dissimilarity between saliency map S and ground truth G. It is defined as

$$MAE = \frac{1}{W * H} \sum_{i=1}^{W} \sum_{j=1}^{H} \|S_{ij} - G_{ij}\|^2$$

Therefore, MAE is the average per-pixel difference between the pixel-wise annotation and the computed saliency map. It directly measures how close a saliency map is to the ground truth and is more meaningful and complementary to PR curves.

Quantitative comparisons of our model against other twenty-one models on four datasets are shown in Figs. 6, 7, 8 and 9. In each figure, first, second and last rows show the PR curve, F-measure and MAEs of all models on four datasets, respectively.

Fig. 9 *First row*: precision-recall curves of different methods. *Second row*: precision, recall and F-measure using an adaptive threshold. *Last row*: MAE. All experiments are carried out on SED1 dataset. The proposed method performs very well

As we can see, our model consistently outperforms others on all four data sets in terms of these three metrics. Specifically, the PR curve of proposed method outperforms PR curves of all other methods on dataset ECSSD, SOD and SED1. On ASD dataset, our model is among the best models. Benefiting from our proposed background suppression and foreground highlighting, our model can generate more clean saliency map. Therefore, our model can achieve higher precision and recall. For F-measure, our model gets the best performance on ECSSD, SOD and SED1. For the F-measure on ASD dataset, the difference between our model and others is not clear. To present F-measure more clearly, we present all the values in Table 1. Our model has best performance. Finally, for MAE, our model has the smallest value on all these four datasets and this indicates that our saliency maps are closest to the ground truth maps.

4.3 Qualitative evaluation

For qualitative evaluation, the results of applying the various models to representative images are shown in Fig. 10. We note that the proposed algorithm uniformly highlights the salient regions and preserves finer object boundaries than all other methods.

In first example, i.e., cup images in top three rows, the saliency maps of our model, HDCT [20] and SVO [9] models can all detect whole object. However, our saliency map is much cleaner than others. Especially in the background regions. This good performance benefits from the background suppression step in our model. From seventh row to ninth row, the red flowers image has textured background regions. Only our model can detect the whole salient object with few background noise. In last two examples, i.e., images shown in last six rows, both the salient objects have similar color distribution with background regions. Therefore, all the saliency maps generated by other models will be greatly influenced by these background regions. In fact, our model will be affected by these regions too. However, with the help of background suppression and foreground highlighting of our model, the initial saliency map will have as less background regions as possible and the difference between background and foreground gets bigger. The final saliency map after propagating the initial saliency to other regions will be more accurate than that of other models.

4.4 Efficiency

To demonstrate the efficiency of our model, we show the average running time of different models in Table 2. In column code, M, M+C and EXE denotes MATLAB, MATLAB with

Table 1 F-measure of ASD dataset

Model	F-measure	Model	F-measure	Model	F-measure
IT [17]	0.5356	MAP [34]	0.8856	LRMR [29]	0.8191
GB [16]	0.6548	HS [41]	0.8789	MC [19]	0.8994
CA [14]	0.5916	BM [38]	0.8140	MSS [1]	0.6784
FT [2]	0.5957	CB [18]	0.8533	PCA [25]	0.8177
SF [27]	0.8264	CHM [21]	0.8400	SVO [9]	0.4037
GS [37]	0.8409	FES [35]	0.7043	SWD [12]	0.6584
GMR [43]	0.9012	HDCT [20]	0.8554	LGH [44]	0.8502
Ours	0.9122				
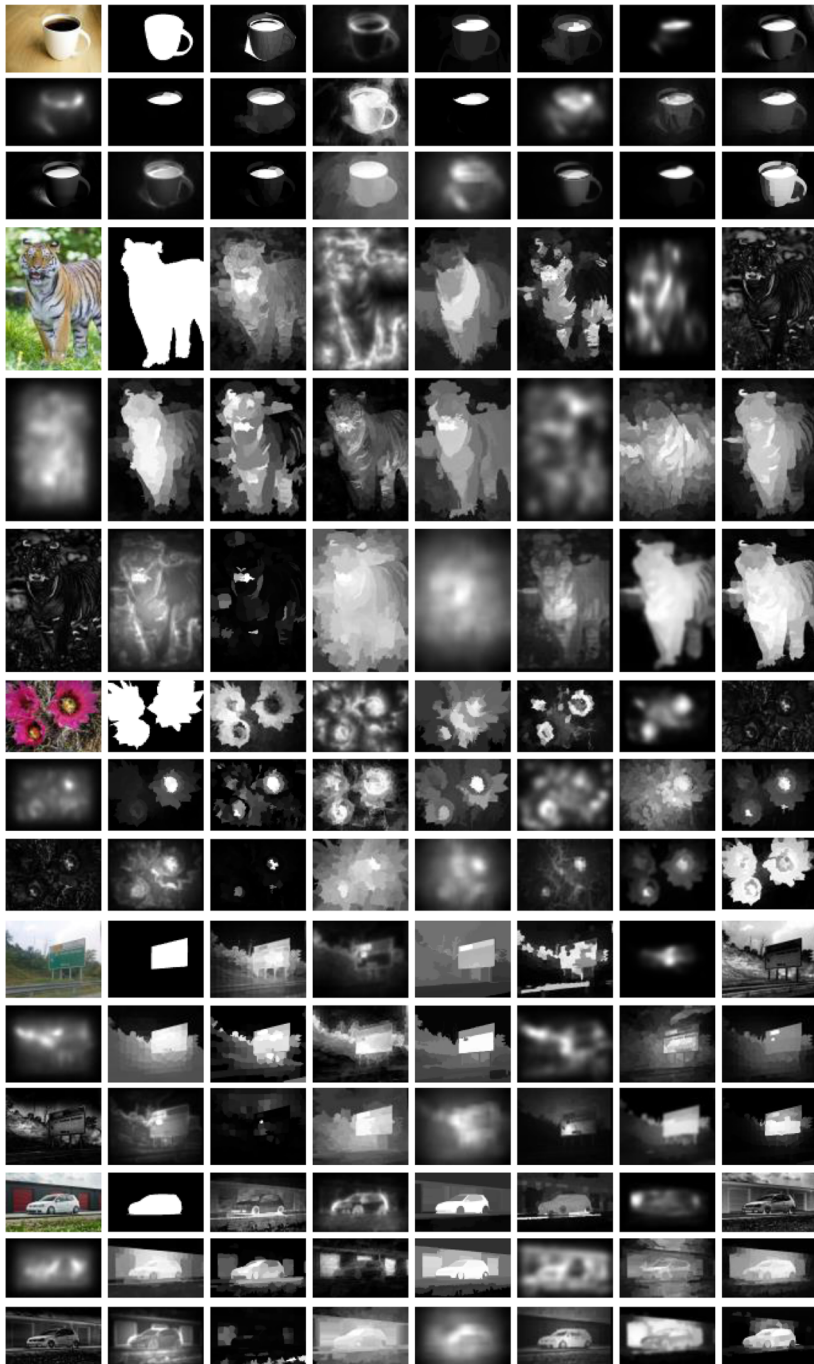
Fig. 10 Visual comparison of proposed model and twenty-one other models. From top to bottom and *left* to *right* are input image, ground truth, and saliency maps of BM [38], CA [14], CB [18], CHM [21], FES [35], FT [2], GB [16], GMR [43], GS [37], HDCT [20], HS [41], IT [17], LRMR [29], MC [19], MSS [1], PCA [25], SF [27], SVO [9], SWD [12] LGH [44], MAP [34] and Ours

Table 2 Comparison of average running time (seconds per image)

Model	Time(s)	Code	Model	Time(s)	Code	Model	Time(s)	Code
IT	0.25	M	MAP	0.08	M+C	LRMR	13.9	M+C
GB	0.63	M	HS	0.31	EXE	MC	0.17	M+C
CA	41.4	M+C	BM	2.16	M+C	MSS	3.07	M
FT	0.13	M	CB	1.94	M+C	PCA	3.69	M+C
SF	0.30	M+C	CHM	2.97	M+C	SVO	23.49	M+C
GS	0.22	M+C	FES	0.20	M	SWD	0.28	M
GMR	1.61	M+C	HDCT	8.15	M+C	LGH	1.81	M
Ours	0.9	M						

C/C++ and executable program, respectively. The experiments are carried out on ECSSD dataset with a typical 400×300 image using a PC with an Intel i7 CPU of 3.2GHz and 16GB memory. It should be noted that our model is implemented by using MATLAB without optimization. Therefore, the computational complexity of our model is comparable to that of other models. The main reason for this low computational cost is that we employ superpixels as our basic processing unit, not pixels. This will greatly reduce the computational cost. Given a 400×300 image, we segment it into 300 superpixels. The overall running time of our model is 0.9s. The time used for solving energy minimization problem (10) is only 0.008s. Another reason is that the computation of contrast, background suppression and foreground highlighting is carried out in a vector form. As we all know, MATLAB has big advantage in vector and matrix operation. So these operations are computed very fast.

4.5 Failed cases

Though our proposed model achieves good results in most cases, it still have some limitations. Firstly, the final saliency map will be greatly influenced by the unit boundary
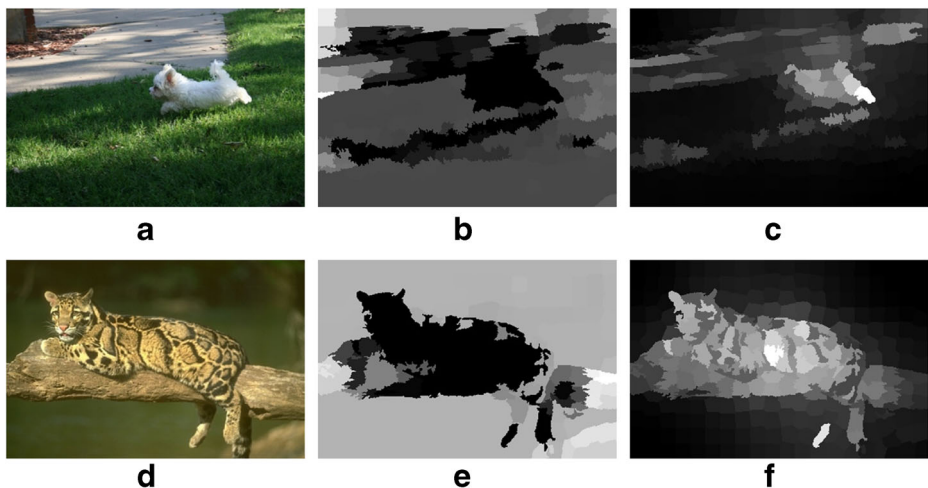
Fig. 11 Failed cases of our model. **a, d** Input images. **b, e** Unit boundary distribution. **c, f** Our saliency map

distribution. As shown in Fig. 11b and c, when the unit boundary distribution map has many undesired regions, the final saliency map will be inaccurate. The road has different values of unit boundary distribution. When the background suppression is employed, some background regions will still remain. This will affect the final saliency detection. Secondly, the saliency map will contain background regions when the object and background regions have similar color distribution. As shown in Fig. 11e and f, although the unit boundary distribution is good enough, the final saliency map is not satisfying. This may be caused by the feature used in our work. We only employ color feature to evaluate saliency. When the salient object and background regions have similar color distribution, the detection result may be not good enough.

5 Conclusions and future works

In this paper, we have presented a novel model for salient region detection. Based on the global and local contrast prior, we propose unit boundary distribution to model background distribution. Then, we use contrast between other regions and top, left image side to suppress background regions. A coarse initial saliency map is then generated by highlighting foreground regions. By background suppression and foreground highlighting, the initial saliency map has good estimation of location where the salient object is. Finally, we model the final saliency evaluation as a graph based semi-supervised learning problem via solving an energy minimization problem. We evaluate our model on four widely used datasets and demonstrate promising results with comparisons to other twenty-one state-of-the-art models.

The failed cases presented in Section 4.5 motivate our future works. Firstly, we will exploit more techniques to model more reliable background distribution because our model is based on it. Secondly, more cues should be considered into our model, e.g., texture, patterns. Further more, high-level knowledge may be incorporated to produce more accurate saliency maps.

References

1. Achanta R, Susstrunk S (2010) Saliency detection using maximum symmetric surround. In: 2010 17th IEEE international conference on Image processing (ICIP), pp 2653–2656
2. Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency-tuned salient region detection. In: 2009 IEEE conference on Computer vision and pattern recognition (CVPR), pp 1597–1604
3. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2012) Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(11):2274–2282
4. Alpert S, Galun M, Basri R, Brandt A (2007) Image segmentation by probabilistic bottom-up aggregation and cue integration. In: 2007 IEEE Conference on computer vision and pattern recognition, pp 1–8
5. Arya R, Singh N, Agrawal RK (2015) A novel hybrid approach for salient object detection using local and global saliency in frequency domain. Multimedia Tools and Applications:1–21
6. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(1):185–207

7. Borji A, Cheng M-M, Jiang H, Li J (2014) Salient object detection: a survey. ArXiv e-prints
8. Borji A, Cheng M, Jiang H, Li J (2015) Salient object detection: a benchmark. IEEE Trans Image Process 24(12):5706–5722
9. Chang K-Y, Liu T-L, Chen H-T, Lai S-H (2011) Fusing generic objectness and visual saliency for salient object detection. In: 2011 IEEE international conference on computer vision (ICCV), pp 914–921
10. Cheng M, Mitra NJ, Huang X, Torr PHS, Hu S (2015) Global contrast based salient region detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(3):569–582
11. Ding Y, Xiao J, Jingyi Y (2011) Importance filtering for image retargeting. In: 2011 IEEE conference on Computer vision and pattern recognition (CVPR), pp 89–96
12. Duan L, Wu C, Miao J, Qing L, Fu Y (2011) Visual saliency detection by spatially weighted dissimilarity. In: 2011 IEEE conference on computer vision and pattern recognition (CVPR), pp 473–480
13. Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. Int J Comput Vis 59(2):167–181
14. Goferman S, Zelnik-Manor L, Tal A (2010) Context-aware saliency detection. In: 2010 IEEE conference on Computer vision and pattern recognition (CVPR), pp 2376–2383
15. Guo C, Zhang L (2010) A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19(1):185–198
16. Harel J, Koch C, Perona P (2006) Graph-based visual saliency. In: Neural information processing systems, pp 545–552
17. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11):1254–1259
18. Jiang H, Wang J, Yuan Z, Liu T, Zheng N (2011) Automatic salient object segmentation based on context and shape prior. In: Proceedings of the british machine vision conference, pp 110.1–110.12
19. Jiang B, Zhang L, Lu H, Yang C, Yang M-H (2013) Saliency detection via absorbing markov chain. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp 1665–1672
20. Kim J, Han D, Tai Y-W, Kim J (2014) Salient region detection via high-dimensional color transform. In: 2014 IEEE conference on computer vision and pattern recognition (CVPR), pp 883–890
21. Li X, Li Y, Shen C, Dick A, van den Hengel A (2013) Contextual hypergraph modeling for salient object detection. In: 2013 IEEE international conference on Computer vision (ICCV), pp 3328–3335
22. Li Y, Keren F, Zhou L, Qiao Y, Yang J, Li B (2014) Saliency detection based on extended boundary prior with foci of attention. In: 2014 IEEE international conference on Acoustics, speech and signal processing (ICASSP), pp 2798–2802
23. Li C, Yuan Y, Cai W, Xia Y, Feng DD (2015) Robust saliency detection via regularized random walks ranking. In: 2015 IEEE conference on Computer vision and pattern recognition (CVPR), pp 2710–2717
24. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H-Y (2011) Learning to detect a salient object. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(2):353–367
25. Margolin R, Tal A, Zelnik-Manor L (2013) What makes a patch distinct? In: 2013 IEEE conference on computer vision and pattern recognition (CVPR), pp 1139–1146
26. Movahedi V, Elder JH (2010) Design and perceptual validation of performance measures for salient object segmentation. In: 2010 IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW), pp 49–56
27. Perazzi F, Krahenbuhl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE conference on Computer vision and pattern recognition (CVPR), pp 733–740
28. Roy S, Das S (2014) Saliency detection in images using graph-based rarity, spatial compactness and background prior. In: 2014 international conference on Computer vision theory and applications (VISAPP), vol 1, pp 523–530
29. Shen X, Wu Y (2012) A unified approach to salient object detection via low rank matrix recovery. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR), pp 853–860
30. Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8):888–905
31. Siagian C, Itti L (2007) Rapid biologically-inspired scene classification using features shared with visual attention. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2):300–312
32. Sorkine O (2006) Differential representations for mesh processing. Comput Graph Forum 25(4):789–807
33. Sun J, Ling H (2011) Scale and object aware image retargeting for thumbnail browsing. In: 2011 IEEE international conference on Computer vision (ICCV), pp 1511–1518
34. Sun J, Lu H, Liu X (2015) Saliency region detection based on markov absorption probabilities. IEEE Trans Ind Appl 24(5):1639–1649

35. Tavakoli HR, Rahtu E, Heikkil J (2011) Fast and efficient saliency detection using sparse sampling and kernel density estimation. In: Heyden A, Kahl F (eds) Image Analysis, volume 6688 of lecture notes in computer science. Springer, Berlin, pp 666–675
36. Valenti R, Sebe N, Gevers T (2009) Image saliency by isocentric curvedness and color. In: 2009 IEEE 12th international conference on Computer vision, pp 2185–2192
37. Wei Y, Wen F, Zhu W, Sun J (2012) Geodesic saliency using background priors. In: Computer vision ECCV 2012, volume 7574 of lecture notes in computer science, pp 29–42
38. Xie Y, Lu H (2011) Visual saliency detection based on bayesian model. In: 2011 18th IEEE international conference on Image processing (ICIP), pp 645–648
39. Xie Y, Huang C, Xu L (2014) Semantic superpixel extraction via a discriminative sparse representation. Multimedia Tools and Applications 73(3):1247–1268
40. Xu K, Zhang H, Cohen-Or D, Xiong Y (2009) Technical section: dynamic harmonic fields for surface processing. Comput Graph 33(3):391–398
41. Yan Q, Xu L, Shi J, Jia J (2013) Hierarchical saliency detection. In: 2013 IEEE conference on Computer vision and pattern recognition (CVPR), pp 1155–1162
42. Yang B, Xu D (2014) Color boosted visual saliency detection and its application to image classification. Multimedia Tools and Applications 69(3):877–896
43. Yang C, Zhang L, Lu H, Ruan X, Yang M-H (2013) Saliency detection via graph-based manifold ranking. In: 2013 IEEE conference on Computer vision and pattern recognition (CVPR), pp 3166–3173
44. Yeh H-H, Liu K-H, Chen C-S (2014) Salient object detection via local saliency estimation and global homogeneity refinement. Pattern Recogn 47(4):1740–1750
45. Zhang H, Xu M (2015) Saliency detection based on boundary feature and smoothness energy function. Optik - International Journal for Light and Electron Optics 126(1):81–86
46. Zhu W, Liang S, Wei Y, Sun J (2014) Saliency optimization from robust background detection. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2814–2821

Hong Li is currently a third year PhD candidate at Macau University (Macao,China), Computer Graphics laboratory. His researches focus on image processing, semi-supervised learning, computer vision and so on. Contact him at yb27441@umac.mo.

Enhua Wu is a professor of the State Key Laboratory of Computer Science at the Chinese Academy of Sciences Institute of Software and the University of Macau. His interests include VR, realistic-image synthesis, and data visualization. He received his PhD in computer science from University of Manchester.

Wen Wu received her PhD degree in computer science and engineering from The Chinese University of Hong Kong. She is currently an assistant professor in Department of Computer and Information Science at University of Macau. Her research interests include medical simulation, virtual reality and physically-based animation.