Vis Comput (2017) 33:1155–1167
DOI 10.1007/s00371-016-1278-0

ORIGINAL ARTICLE

Aggregating complementary boundary contrast with smoothing for salient region detection

Ruihui Li[1] · **Jianrui Cai**[2] · **Hanling Zhang**[1] · **Taihong Wang**[3]

Published online: 2 June 2016

Abstract Automatic to locate the salient regions in the images are useful for many computer vision and computer graphics tasks. However, the previous techniques prefer to give noisy and fuzzy saliency maps, which will be a crucial limitation for the performance of subsequent image processing. In this paper, we present a novel framework by aggregating various bottom-up cues and bias to enhance visual saliency detection. It can produce high-resolution, full-field saliency map which can be close to binary one and more effective in real-world applications. First, the proposed method concentrates on multiple saliency cues in a global context, such as regional contrast, spatial relationship and color histogram smoothing to produce a coarse saliency map. Second, combining complementary boundary prior with smoothing, we iteratively refine the coarse saliency map to improve the contrast between salient and non-salient regions until a close to binary saliency map is reached. Finally, we evaluate our salient region detection on two publicly available datasets with pixel accurate annotations. The experimental results show that the proposed method performs equally or better than the 12 alternative methods and retains comparable detection accuracy, even in extreme cases. Furthermore, we demonstrate that the saliency map produced by our approach can serve as a good initialization for automatic alpha matting and image retargeting.

Keywords Regional contrast · Boundary prior · Smoothing · Iterative framework · Salient object detection

1 Introduction

Visual saliency is the ability of a vision system (human or machine) to select a certain subset of visual information for further processing [14,18]. As a powerful tool for automatic detection of the most visually noticeable objects in images or videos, known as salient objects, substantial research efforts have recently been devoted to the development of visual attention by multiple disciplines. In practice, salient object detection is an important stream, which can be used as a preprocessing of many computer vision applications including image collection [8], object tracking [37], object cutout [45], image retargeting [34], image manipulation [24], etc.

In recent decades, numerous pre-attentive bottom-up saliency detection models [6,7] have been proposed for explaining visual attention. These methods [1,9,13,16,19–21,42,46] usually utilize low-level visual information such as contrast, compactness, spatial distance and structure features [35] to form saliency maps [32]. The most fundamental measure for visual saliency is the contrast computation to their surroundings. By computing pixel/region contrast with respect to local neighborhoods or the entire image, respec-

✉ Taihong Wang
hnuwth@163.com

Ruihui Li
larch18@hnu.edu.cn

Jianrui Cai
csjcai@comp.polyu.edu.uk

Hanling Zhang
jt_hlzhang@hnu.edu.cn

[1] College of Information Science and Engineering, Hunan University, Changsha 410000, China

[2] Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[3] Key Laboratory for Micro-Nano Optoelectronic Devices of Ministry of Education, and State Key Laboratory for Chemo/Biosensing and Chemometrics, Hunan University, Changsha 410000, China

tively, previous image salient object detection methods can be broadly divided into local and global categories.

The local contrast-based methods investigate the uniqueness of image regions in a small local neighborhoods. As a pioneer, Itti et al. [16] define image saliency using center-surround contrast across a difference of Gaussians (DoG) approach. Ma and Zhang [20] introduce an alternative local contrast measure using a fuzzy growth model for saliency estimation. Harel et al. [13] propose a graph-based visual saliency model to non-linearly combine the local contrast maps from different feature channels to highlight conspicuity. These methods prefer to predict human fixations on natural images and fail when the background is cluttered. In the frequency domain, Hou and Zhang [15] raise a spectral residual approach to detect saliency with the average fourier envelope and the differential spectral components, but it is insufficient to detect larger objects since the algorithm regards the large salient object as part of the scene. Frintrop et al. [11] present a method inspired by Itti's model, but they compute center-surround differences with square filters and use integral images to speed up the calculations. These methods using local computation tend to only highlight a few edges that scatter in the image, instead of highlighting uniformly the whole salient objects.

Methods modeling global properties have become popular recently as they enable the assignment of comparable saliency values across similar image regions and thus can uniformly highlight the entire object regions. Zhai and Shah [42] compute the pixel-level saliency against all other pixels in the whole image region. However, for efficiency they use only luminance information, thus ignoring distinctive clues in other channels. Achanta et al. [1] propose a frequency-tuned method that directly defines pixel saliency using the color differences from the average image color in Lab color space. The elegant method, however, when the salient objects are relatively small, is insufficient for distinguishing the salient object. Furthermore, these methods ignore spatial relationships across image parts, which will be critical for reliable and coherent saliency detection. More recently, several saliency detection approaches tend to utilize boundary connectivity or smoothing to generate the final saliency maps and demonstrate impressive results on generating high-quality saliency maps. Cheng et al. [9] present the histogram-based contrast (HC), which exploits the pixel-wise color separation to produce saliency maps, and the region-based contrast (RC), which combines the color contrast and spatial coherence. Both the methods have significant improvement by means of the color histogram smoothing, which can uniformly highlight the salient object. By exploiting image boundary as the background seeds, Wei et al. [36] define the saliency of an image region as its shortest path towards the virtual boundary nodes. Yang et al. [41] extract salient objects according to their relevance to boundary patches based graph-based manifold ranking. Owing to the boundary connectivity, these methods successfully suppress the background to a certain degree.

Saliency detection has made great progress in recent years, but there are still some issues that remain unresolved. Typically, these existing methods cannot suppress efficiently the background or just suppress background by thresholding [9] and then prefer to simultaneously highlight the salient objects and non-salient regions. Second, salient object detection isolates the object from potentially confusing background and preferentially allocates finite computational resources for subsequent image processing. Therefore, the resolution of saliency maps will be a crucial limitation of subsequent applications. As shown in Fig. 4, previous techniques prefer to give fuzzy saliency maps which are less effective in real-world applications.

Inspired by the insights and lessons from a number of previous work as well as several priors supported by psychological evidences and observations of natural images, we address the aforementioned problems in a more integral manner. In particular, we propose a coarse-to-fine measure by aggregating various bottom-up cues and bias to produce high-resolution, full-field saliency maps. First, we compute an initial prior map combined with the regional contrast and spatial relationships. The proposed method works in superpixel space instead of pixel space which can give high boundary fidelity and characterizes superpixels by the dominant color instead of mean color, which improves robustness towards variations within superpixels. Then we employ a color histogram smoothing technique to account for global context. This generates a coarse saliency map, which can be spatially coherent with discontinuities well aligned to image edges. Finally, aggregating complementary boundary contrast [36] with smoothing [9], we iteratively refine the coarse saliency map to improve the contrast between salient and non-salient regions until a close to binary saliency map is reached.

Extensively experimental results on two public datasets [1,9] show that the proposed method can produce high-resolution saliency maps and perform equally or better than the state-of-the-art methods. Some visual saliency effects of the proposed method are shown in Fig. 1. Furthermore, we assess the performances of the proposed method in two real-world applications, image retargeting and alpha matting.

Compared with the region contrast framework in [9], the proposed approach has made some modifications and improvements. First, the proposed method replaces graph-based segmentation to SLIC segmentation algorithm [2], which can give high boundary fidelity. Second, instead of calculating the mean color for each superpixel, we directly use the most frequently occurring color as the dominant color of the corresponding superpixel in contrast calculation, which can effectively reduce the artifact introduced by segmentation. Third, in color space smoothing, we quantize image by

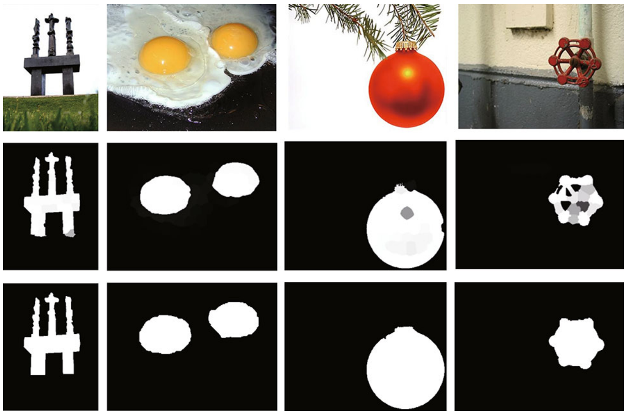


Fig. 1 Visual examples: the input original images (*top*). The high-quality saliency map computed by our proposed approach (*middle*). The ground truth at the *bottom*

a more representative color palette, which directly use the most frequency occurring color of the corresponding superpixel. Finally, in improvement of saliency map, suppressing boundary is achieved by a soft boundary score other than thresholding in [9].

The major contributions of this paper are twofold: (1) to optimize the result, we propose a novel iterative framework incorporating complementary boundary prior with smoothing, which can uniformly highlight the salient regions and simultaneously suppress the background effectively. (2) The saliency map, produced by the proposed method, has a distinct contrast between non-salient background and salient object and can be close to binary one which is more effective in real-world applications.

The rest of the paper is organized as follows: The proposed method of generating saliency map candidates is described in Sect. 2. In Sect. 3, we show the experimental results on two public datasets and analyse the performance of the proposed method qualitatively and quantitatively. In Sect. 4, we demonstrate the applicability of our saliency maps and conclude this paper with further discussion in Sect. 5.

2 Saliency model

Pixels in the same region usually have homogenous color component. Computing region-based contrast instead of pixel-wise operation enormously pulls down the computation complexity. Thus, our approach considers the super-pixel as the element of saliency estimation. Instead of using a graph-based image segmentation method [10] in Cheng et al. [9], we first over-segment the image into N super-pixels by the simple linear iterative clustering (SLIC) algorithm [2]. The measure can fit well to the boundaries between the salient objects and background regions and is more meaningful than block-level and pixel-level features. In addition, the superpixel representation facilitates preserving the better-defined object boundaries than the fixed size segmentation. Let $R = r_1, \ldots, r_N$ denote the set of superpixels.

2.1 The initial saliency map

Spatially weighted contrast measure has been shown to be effective in saliency detection [9,23]. For any superpixel r_i, we compute its saliency value by measuring its color contrast to all other superpixels in the image,

$$S_{in}(r_i) = \omega(r_i) \cdot \|c_i - c_j\| \cdot \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_p^2}\right), \tag{1}$$

where p_i and p_j are the average position whose values are normalized to [0, 1], and σ_p controls the strength of spatially weighting, c_i and c_j are the dominant color of the corresponding superpixels in the CIE Lab color space. Instead of calculating the mean color for each superpixel [40,43], we find the most frequently occurring color as the dominant element to reduce the impact of artifacts introduced by segmentation. If this color corresponds to less than five occurrences, we select an arbitrary color from the corresponding superpixel. Psychophysical studies show that human attention favors central regions of natural image [5,29] we use $\omega(r_i) = \exp(-9d_i^2)$ as a simple center bias, where d_i is the distance between the average distance in superpixel r_i and the center of the image, with pixel coordinates normalized to [0, 1]. Thus, $\omega(r_i)$ gives a high value if superpixel r_i is close to the center of the image and it gives a low value if the region is a border region away from the center.

2.2 The coarse saliency map with smoothing

As shown in Fig. 2b, although we can efficiently compute region contrast-based superpixel, this procedure may just highlight some parts of an object, leading to the whole object being indistinct. As a solution, we introduce a color space smoothing [9] to render pixels with similar colors the closer saliency values. First, we quantize the image-based histogram in RGB color space. Instead of regularly spliting the color pace [9], we generate a more representative color palette based the most frequently occurring color of the corresponding superpixel in the RGB color space, referred to Sect. 2.1, which reduces the number of colors to $N(N \ll 12^3)$. As shown in Fig. 3, we can observe that a representative palette gives a better result than the original scheme by [9], which regularly splits the R,G and B axes (leaving many bins empty, as the RGB color space is not a cube). Typically, the tomato quantized by our method is more exquisite. Second, we choose more frequently occur-

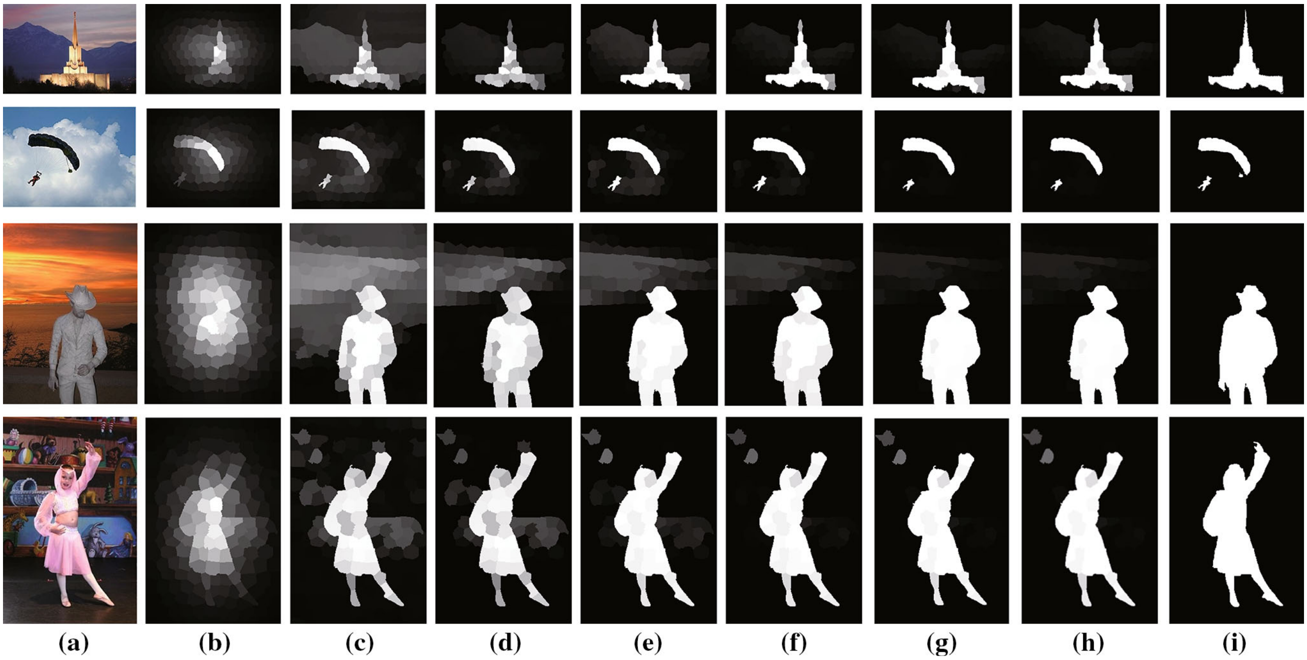


Fig. 2 Saliency computation: **a** input image, **b** the initial saliency map, **c** the coarse saliency map with smoothing, **d** and **f** refining saliency map with background prior in the first or second iteration, **e** and **g** refining saliency map with smoothing, **h** the final saliency with gaussian blur, **i** the ground truth. We get a high-quality saliency map that is comparable to human labeled ground truth

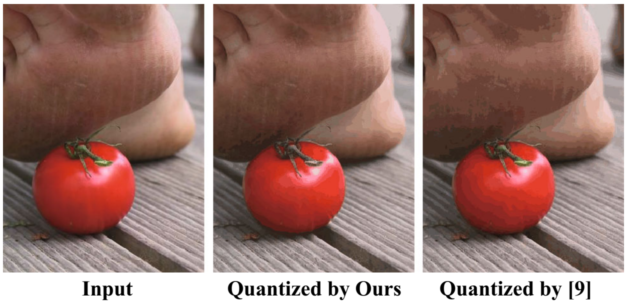


Fig. 3 Visual comparison about the quantized results between our method and [9]

ring colors in the quantized space and ensure these colors cover the colors of more than 95 % of the image pixels. We get saliency value for each color as the same color value c grouped together. Third, we replace the saliency value of each color by the weighted average of the saliency values of similar colors (measured by CIE Lab distance). We choose $m = \beta_1 \cdot N$ nearest colors to refine the saliency value of color c by

$$S_{co}(c) = \frac{1}{(m-1)T} \sum_{i=1}^{m} (T - D(c, c_i)) S_{in}(c_i), \tag{2}$$

where $T = \sum_{i=1}^{m} D(c, c_i)$ is the sum of Euclidean distances between color c and its m nearest neighbors c_i, and the normalization factor comes from $\sum_{i=1}^{m}(T - D(c, c_i)) = (m-1)T$. Finally, we reset the saliency of each superpixel as the average saliency value of its corresponding pixels and all regions will obtain the same saliency, achieving the most extreme case.

2.3 Saliency map refining with iterative framework

By comparing Fig. 2b with c, it can be found that the color histogram smoothing already achieves an evident improvement. However, as with other methods, it also cannot suppress the background effectively. In addition, since it is insufficient to distinguish the non-salient and salient regions, the smoothing procedure will highlight them both simultaneously.

To alleviate these issues, we introduce an iterative framework to further uniformly highlight the salient region and adequately suppress the background region, resulting in a close to binary saliency map. The coarse saliency map will be utilized as the S_{input} in the first iteration.

Step 1: Suppress the background For most nature images, the background always appears smoothly and homogenously. From the one-third rule in professional photography, we further observe that most photographers will not crop salient object along the view frame [36]. Taking the effort we made into consideration, we can directly define the boundary contrast of each region as its saliency value contrasts the image boundary regions, which is an abbreviated version of [36],

$$\omega_s(r_i)=\exp\left(-\frac{1}{n}\sum_{j=1}^{n}\exp\left(-\frac{|S_{input}(r_i)-S_{input}(r_j)|}{2\sigma_s^2}\right)\right), \tag{3}$$

where n is the number of superpixels on the image boundaries ($n \ll N$), $S_{input}(r_i)$ and $S_{input}(r_j)$ are the salient value of corresponding superpixels, whose values are normalized to [0,1]. We iterate the saliency map with boundary contrast and $\omega_s(r_i)$ is normalized to [0,1].

$$S_1(r_i) = S_{input}(r_i)\cdot\omega_s(r_i) \tag{4}$$

According to Eq. 4, the salient regions receive high weighting and the smooth background regions receive small weighting, so this step effectively suppresses the background. With the boundary contrast as weight, there is an obvious improvement (as shown in Fig. 2d). The boundary contrast can suppress the background to a certain degree, while it is still bumpy and noisy. In the next step, we will introduce a smoothing method-based region to further refine the saliency map.

Step 2: Highlight the salient object According to visual organization rules [17], the salient pixels are usually grouped together and most salient pixels have similar colors. Then, regions of similar color should be more likely to be assigned similar saliency values. In order to reduce noisy saliency results and uniformly highlight the salient object, referred to Sect. 2.2, we adopt a smoothing method-based region. Measured by color similarity in CIE Lab color space, we render superpixels, with similar colors the closer saliency values. We choose $m = \beta_2 \cdot N$ nearest super-pixels to refine the saliency value of super-pixel r_i by

$$S_2(r_i) = \sum_{j=1}^{m} S_1(r_j)\cdot(1-\omega_{ij}), \tag{5}$$

where $\omega_{ij} = \frac{1}{Z}\|c_i - c_j\|$ is the weight between super-pixel r_i and its *jth* nearest neighbors r_j, corresponding to color differences in CIE Lab color space because of its perceptual accuracy. $Z = \sum_{j=1}^{m}\|c_i - c_j\|$ is its normalization term that guarantees all weights summed to 1. In our experiments, we find that linearly exponent function works better than Gaussians weighting, which falls off too sharply. Through computing *step 2*, the wholly salient object becomes more uniform.

On account of the *step 1*, we have already suppressed the background effectively with a simple boundary prior, the *step 2* is more superior to highlight the salient object and further enlarges the contrast between the salient and non-salient regions. Moreover, by the weighting average, we can wipe off some noisy points to improve precision (as shown in Fig. 2e). The proposed iterative framework is simple, efficient and significantly improves the quality of the coarse saliency maps. In our experiments, we choose to iterate twice for optimizing and saving computation. The result of the *step 2* will be utilized as the input in second iteration. A visual saliency effect of each step on the saliency map is shown in Fig. 2. Experimental results show that the proposed method can produce high-quality saliency maps compared with human labeled ground truth.

3 Experiment results

To estimate the performance of the proposed method, we compare the proposed method with 12 alternative methods (IT [16], GB [13], SR [15], LC [42], HC [9], FT [1], CA [12], ZHANG [43], RC [9], SF [23], GS_SP [36] and RBD [46]). Most of these algorithm codes are available in the authors' home page. We use the standard benchmark datasets: Achanta et al. [1] dataset and MSRA10k dataset [9]. The former is widely used and relatively simple that contains 1000 images with the corresponding accurate human-labeled binary masks for salient objects. The latter contains 10,000 images with pixel-level saliency labeling and is more challenging. Figure 4 shows a visual comparison of different methods. The results indicate that our approach can achieve better performance, owing to the effect of the iterative framework.

3.1 Experiment setup

We set the number of superpixel nodes $N = 200$ in all the experiments. There are four parameters in the proposed method : $\sigma_p, \sigma_s, \beta_1, \beta_2$. Since the proposed saliency model is unsupervised, these four parameters are empirically chosen, $\sigma_p^2 = 0.2, \sigma_s^2 = 0.05, \beta_1 = \beta_2 = 0.15$, for all test images in the experiments. The sensitivity of the proposed model to the parameter settings is discussed in Sect. 3.4.

3.2 Performance comparison

Here, we evaluate our salient object detection model in terms of precision, recall, F_β, precision–recall curve (PR curve) and mean absolute error (MAE). The PR curve is produced based on the overlapping area between subjective annotation and saliency prediction. The precision value is the ratio of salient pixels correctly assigned to all the pixels of extracted regions, which reflects the accuracy of the detection algorithm. The recall value corresponds to the percentage of detected salient pixels in relation to the ground-truth number, which represents the detection consistency. The precision and recall can

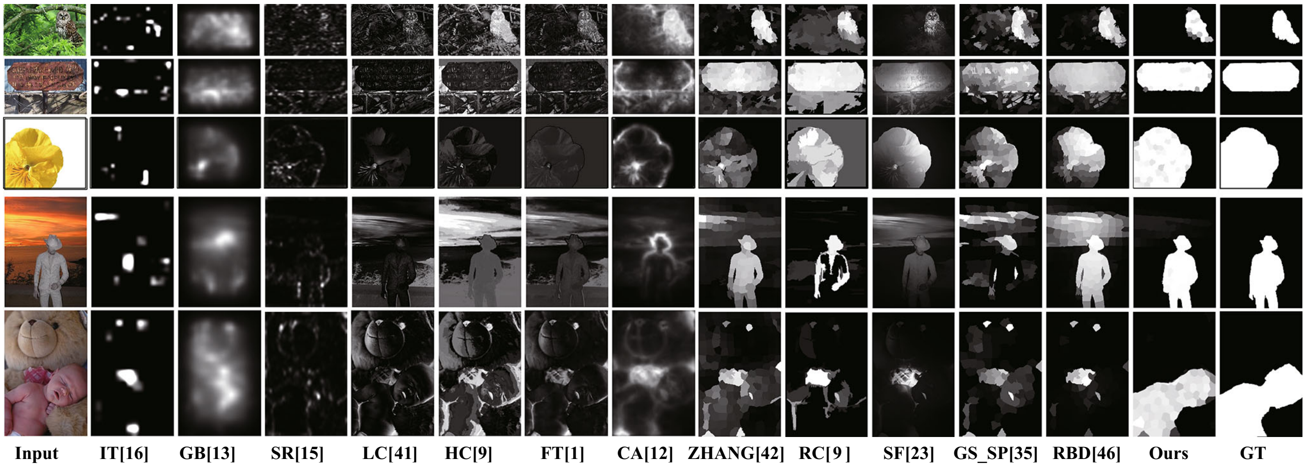


Fig. 4 Visual comparison of previous approaches to our method and ground truth(GT). As also shown in the numerical evaluation, our method consistently produces saliency maps closest to ground truth

be depicted by the PR curve on the datasets. The precision and recall rates for each image are quantified as follows:

$$Precision = \frac{|M \bigcap G|}{|M|}, Recall = \frac{|M \bigcap G|}{|G|}, \tag{6}$$

where M is the binary salient object mask generated by thresholding saliency map and G is the corresponding binary ground truth. W and H are the width and height of the saliency map.

From this definition, we can see that the binarization is the key step in the evaluation. In an ordinary way, giving a saliency map whose values are normalized to [0, 255], a set of binary images can be obtain by varying the threshold T_f from 0 to 255. As a result, precision and recall scores are computed on each fixed threshold, which are finally combined to form a PR curve to describe the model performance at different situations. The other solution to perform the binarization is the image-dependent adaptive threshold, which is defined as twice the mean value of the saliency map S,

$$T_a = \frac{2}{W \times H} \sum_{x=0}^{W} \sum_{y=0}^{H} S(x, y), \tag{7}$$

where W and H are the width and height of the saliency map in pixels, respectively, and $S(x, y)$ is the saliency value of the pixel at position (x, y). High recall rate can be achieved at the expense of reducing the precision rate and vice-versa, so it is important to evaluate both measures together. The F_β is a weighted harmonic mean between the precision and recall rate, which is the overall performance measurement.

$$F_\beta = \frac{(1 + \beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}, \tag{8}$$

where $\beta^2 = 0.3$ stresses precision rate more than recall rate, as suggested by many salient object detection works [1,9]. The resulting curves in Fig. 6a, b show that our method can perform equally or better than the others on two benchmark datasets. The performance closest to ours is the RBD [46], which proposes a more robust boundary-based measure with taking the spatial layout of image patches into consideration.

The overlap-based evaluation measures introduced above do not consider the true negative saliency assignments, i.e., the pixels correctly marked as non-salient. While the PR curve analysis is useful, it is not the only arbiter of performance and practical applications can better illustrate improvements. Previous techniques give fuzzy saliency maps which might look similar on PR curves with ours, but are less effective in real-world applications. Since the previous methods [9,23,36,46] cannot enlarge enough the contrast between the non-salient background and the salient object and its close salient background, their saliency maps always have a number of non-salient regions, which have similar saliency values as the salient object, and the selection of a uniform threshold will be unpractical. Therefore, it will be insufficient to get a distinct saliency map by thresholding or adaptive thresholding on previous methods. SaliencyCut [9] uses the computed saliency map to assist in automatic salient object segmentation. In this work, a loose threshold is used to generate the initial binary mask. Then the method iteratively uses the GrabCut segmentation method [26] to gradually refine the binary mask. Integrating our saliency maps with SaliencyCut, the salient objects can achieve a major improvement in boundary sides and faces, shown in Fig. 5. Although the RC [9] can correctly detect the salient objects, the noisy background will limit the performance of the subsequently heuristic method.

The mean absolute error (MAE) is a statistical measure that represents the difference between estimated and actual

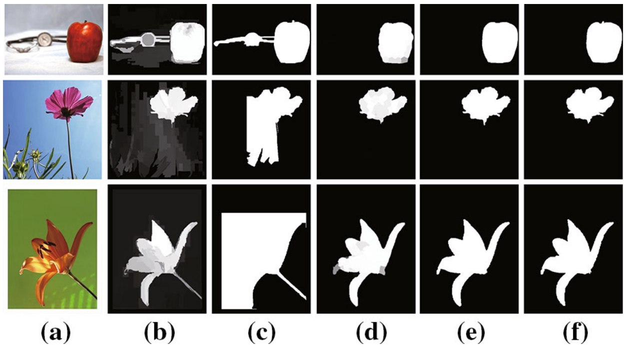


Fig. 5 Visual comparison with SaliencyCut [9]. **a** Input image. **b** Saliency map of RC [9]. **c** SaliencyCut on **b**. **d** Our Saliency map. **e** SaliencyCut on **d**. **f** Ground truth

values. In this paper, the MAE is utilized to estimate the dissimilarity between the saliency map and ground truth. The lower MAE value indicates better performance. The MAE is the average of absolute error between the continuous saliency map S and the binary ground truth G, which is defined as

$$MAE = \frac{1}{W \times H} \sum_{x=0}^{W} \sum_{y=0}^{H} |S(x, y) - G(x, y)| \tag{9}$$

Figure 6c shows that our method outperforms the other approaches in terms of the MAE measure, which provides a more balanced comparison between the saliency map and ground truth.

In addition, we also further analyze the statistical results on MSRA10K dataset. As shown in Fig. 6a, at maximum recall where $T_f = 0$, all pixels are considered to be salient, so all the methods have the same precision 0.22 and recall values 1.0 at this point. However, we further observe that there is a great difference when the value of T_f is smaller but not zero or higher close to 256. We take two more persuasive presentations ($T_f = 10$ and $T_f = 240$) into consideration. As shown in Fig. 7a, although most methods have a high recall rate, the precision rate is rather low. This is because their saliency detection methods always have a number of non-salient regions. In contrast, we perform better than the others, since that we almost efficiently suppress the gray level of background infinitely close to zero at the cost of reducing a little recall rate in Fig. 4. On the other hand, even if the threshold is close to 256 in Fig. 7b, our saliency maps also have a higher recall rate(>0.5) than the others (<0.3) with the same precision. The resulting curves show that our method can obtain a more robust performance against the others in precision and recall.

However, like most methods, our method also contains some limitation. e.g., when there are complex background. If some salient regions are pushed to the background by boundary contrast, then the subsequent smoothing and the iterative refinement will not be able to correct this error.

3.3 Human fixation dataset

While our algorithm targets salient object detection, it is also interesting to test its performance on human fixation prediction benchmarks. We use a large-scale human fixation benchmark (CAT2000) [4] for such evaluation. Some visual comparisons of the proposed method are shown in Fig. 8. To conduct a comprehensive evaluation, we use a measure of similarity to describe the spatial deviation of predicted saliency map from the actual fixation map. The similarity score (S) is a measure of how similar two distributions are. After each distribution is scaled to sum to one, the similarity is the sum of the minimum values at each point in the distributions.

Mathematically, the similarity S between two maps P and Q is

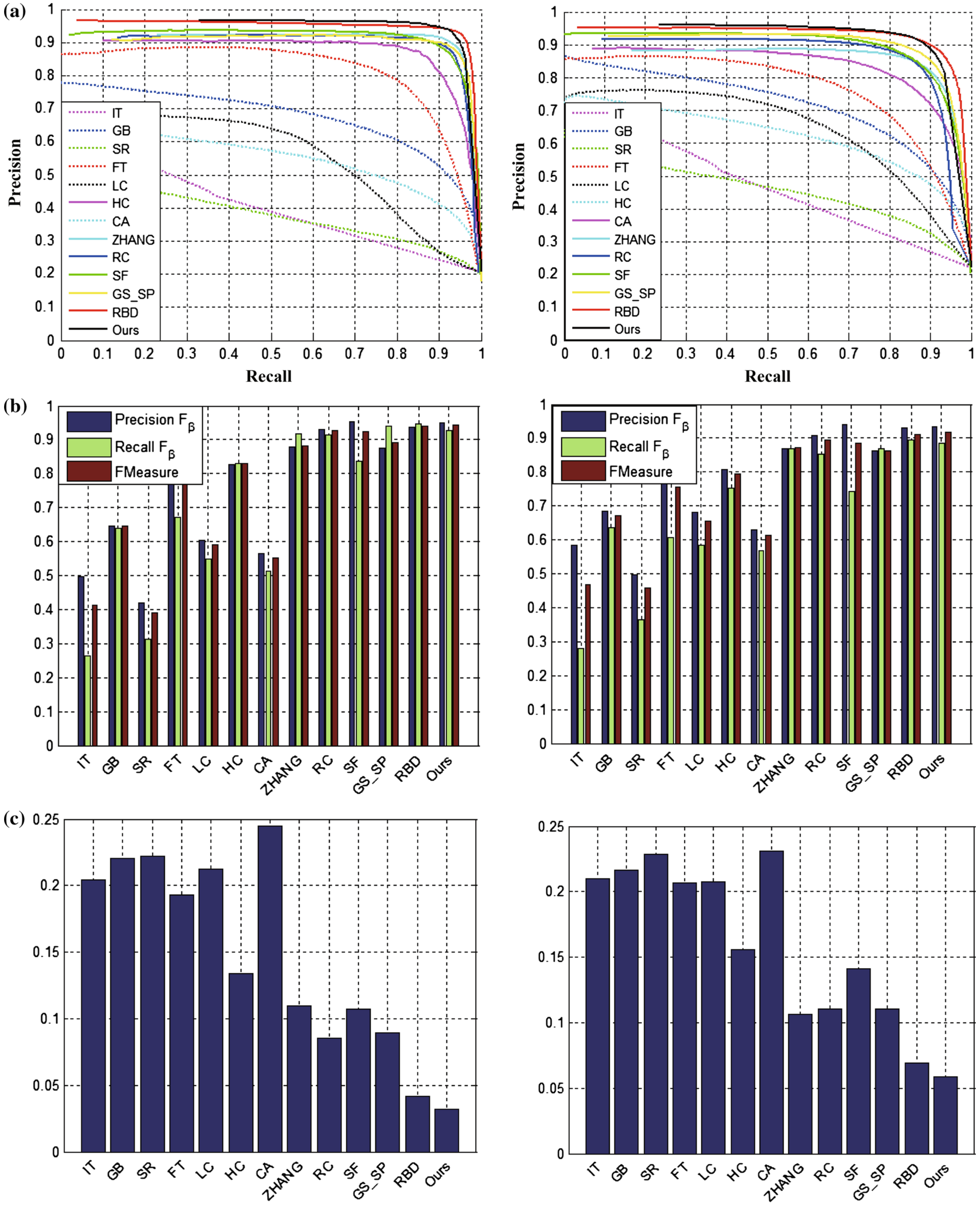
$$S = \sum_{i,j} min(P_{i,j}, Q_{i,j}), where \sum_{i,j} P_{i,j} = \sum_{i,j} Q_{i,j} = 1 \tag{10}$$

Table 1 shows that our method, although initially designed for saliency region detection, has only slightly lower performance to state-of-the-art methods [31,44] for predicting human fixation points.

3.4 Validation of the proposed model

To estimate the performance of the iterative framework, we compare the proposed approach with the coarse saliency map (CSM) and the validation framework without *step 2* (called Ours_2) by a PR curve on MSRA10K dataset. As shown in Fig. 9a, our refining method has a better effect on precision and recall rate. This is because the boundary contrast effectively enlarges the difference between salient and non-salient regions, and smoothing successfully highlights the salient objects.

We also analyze the effects of the different parameter settings. In this paper, the presented schedule utilizes superpixel method SLIC [2] to preprocess images and then detects distinctive regions. In detail, Fig. 9b gives the PR curves about the impact of different super-pixel number N on the proposed method. Considering the computational complexity and the performance of PR curves, we select superpixel number $N = 200$ for all experiments. Meanwhile, the other different quantitative results comparison has been made in Fig. 9. Furthermore, we empirically find that the proposed model is not sensitive to the choice of parameter settings. For all the data set in this paper,we use the same parameters.



(i) Achanta et al.[1] dataset **(ii) MSRA10K dataset**

Fig. 6 Statistical comparison results. **a** Precision and recall rates for all algorithms. **b** Precision, recall, and F-measure for adaptive thresholds. **c** Mean absolute error of the different saliency methods to ground truth

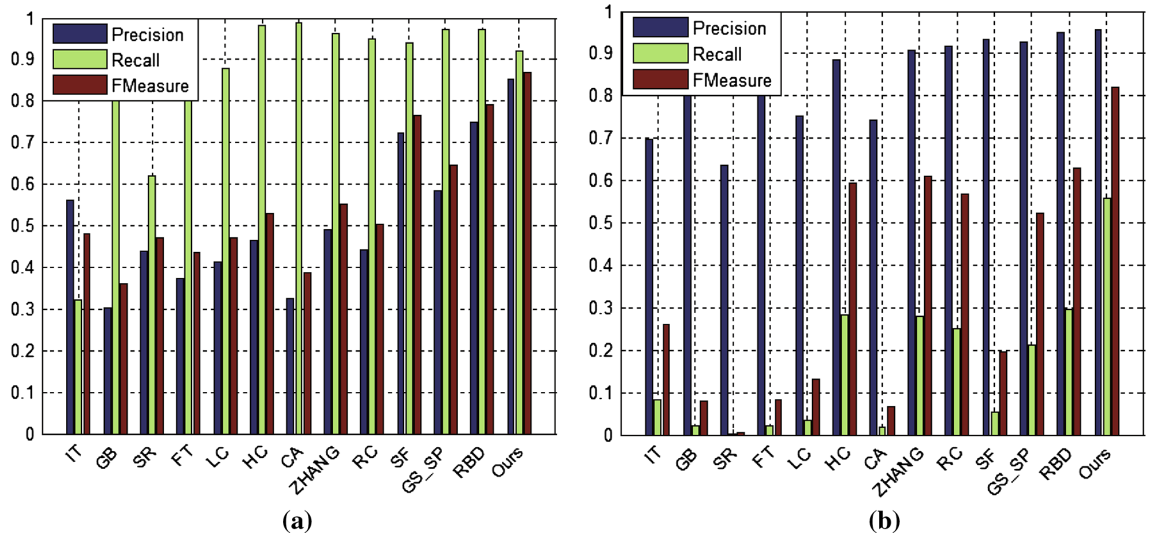


Fig. 7 Precision, recall and F-measure where $T_f = 10$ and $T_f = 240$

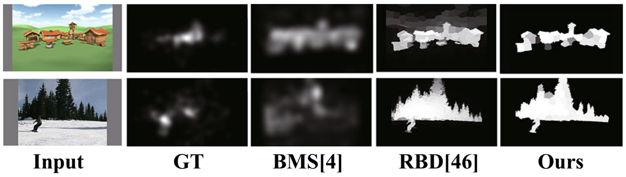


Fig. 8 Visual comparison on human fixation dataset (CAT2000) [4]

Table 1 Performance scores for models (Similarity)

Method	GVBS [13]	RBD [46]	FES [31]	BMS [44]	Ours
Score	0.5	0.51	0.57	0.61	0.56

3.5 Average run time

The average run times of all the methods on the MSRA10K dataset are presented in Table 2 based on a machine with an Intel Quad-Core i3 2.53 GHz CPU with 2GB RAM.

4 Applications

Many applicants require saliency maps as input. In this section we show via two applicants to demonstrate that our method is helpful.

4.1 Image retargeting

Image retargeting aims to adapt images to display of target sizes and different aspect ratios. Effective retargeting requires emphasizing the important content while retaining surrounding context with minimal visual distortion. Therefore, retargeting methods rely on the defining of the importance of pixels. Since it effectively highlights the meaningful objects, we believe the proposed approach will be beneficial to incorporate with image retargeting.

Seam carving [3] and image warping [27] are recent techniques of content-aware retargeting. Seam carving iteratively removes or inserts a seam passing through unimportant regions. This approach may generate jagged edges because of the removal of discontinuous seams. In contrast, image

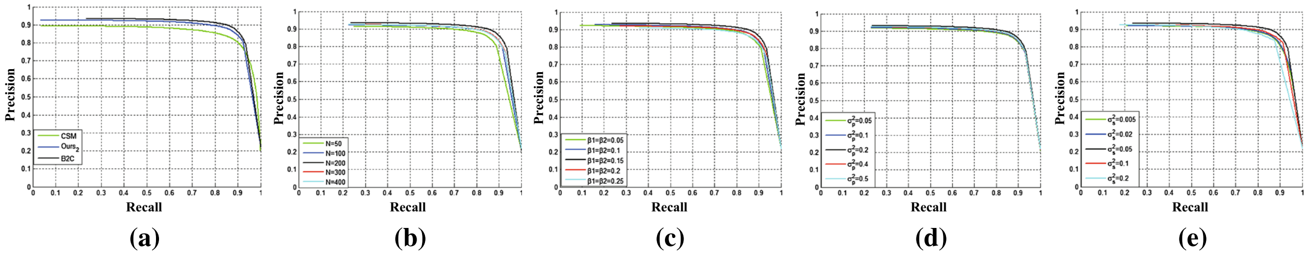


Fig. 9 Precision-recall curves for validation of the proposed model on the MSRA10K dataset. **a** Comparison of the proposed method with CSM and Ours_2. **b** The effects of changes of super-pixel number N. **c** Comparison of saliency maps smoothing with different β_1 and β_2. **d** Comparison of initial saliency maps with different σ_p. **e** Comparison of refined saliency maps with different σ_s

Table 2 Averaging running time

Method	HC	ZHANG	RC	SF	RBD	Ours
Time (s)	0.02	0.39	0.3	0.41	0.36	0.35
Code	C++	Matlab	C++	C++	Matlab	C++

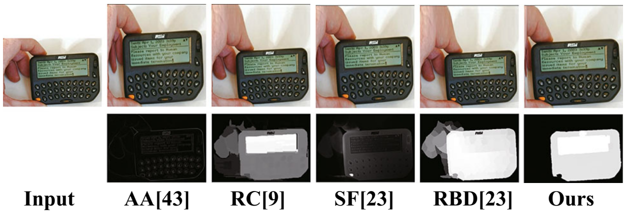


Fig. 10 Image retargeting results by enlarging 50 % height

warping offers a better possibility of producing a continuous deformation for content-aware retargeting. We run the original code of [22] based on axis-aligned (AA) deformations and compare the results with those produced after replacing their saliency map with four methods, including ours.

Figure 10 presents a representative result by retargeting to 150 % of the original height. Our saliency map guarantees that the salient object (the pager) is not distorted. The improved results can be explained by comparing the saliency maps. In the saliency maps of [9,23,46], the dominant object is detected correctly, but some non-salient regions (the fingers) exist in their saliency maps, which will occupy the extra space after retargeting. Meanwhile, the saliency values only rely on the edge gradients in [22]. Consequently, the retargeting results of [9,22,23,46] have more or less distortions against ours. On the other hand, our saliency maps differentiate between the non-salient background and the salient object and its close salient background. Both are maintained after retargeting, resulting in visual images. Further comparisons are provided in Fig. 11.

4.2 Automatic image matting

Although the results shown in Fig. 4 are visually compared well, the details of their boundaries are not as good as that of image matting techniques. This motivates us to employ alpha matting techniques to further improve the boundary of the extracted salient object.

Our running time is similar to that of RC and RBD (all methods involve segmentation). Specifically, our method spends 0.172, 0.152 and 0.026 s on initial map computation with super-pixel generation, coarse map smoothing and saliency map refining, respectively. Note that we just compare against several competitive accuracy methods or those similar to ours.

Alpha matting aims at softly and accurately extracting the foreground from an image and user-specified trimap, which indicates the known foreground/background and the unknown pixels are often required. With the salient object extracted by our method, the trimap can be automatically created with a uniform bandwidth (set by the user) through eroding and dilating the binary mask of the extracted object. Once the trimaps are obtained, any standard matting methods can be adopted to estimate the matte, and a finer boundary of

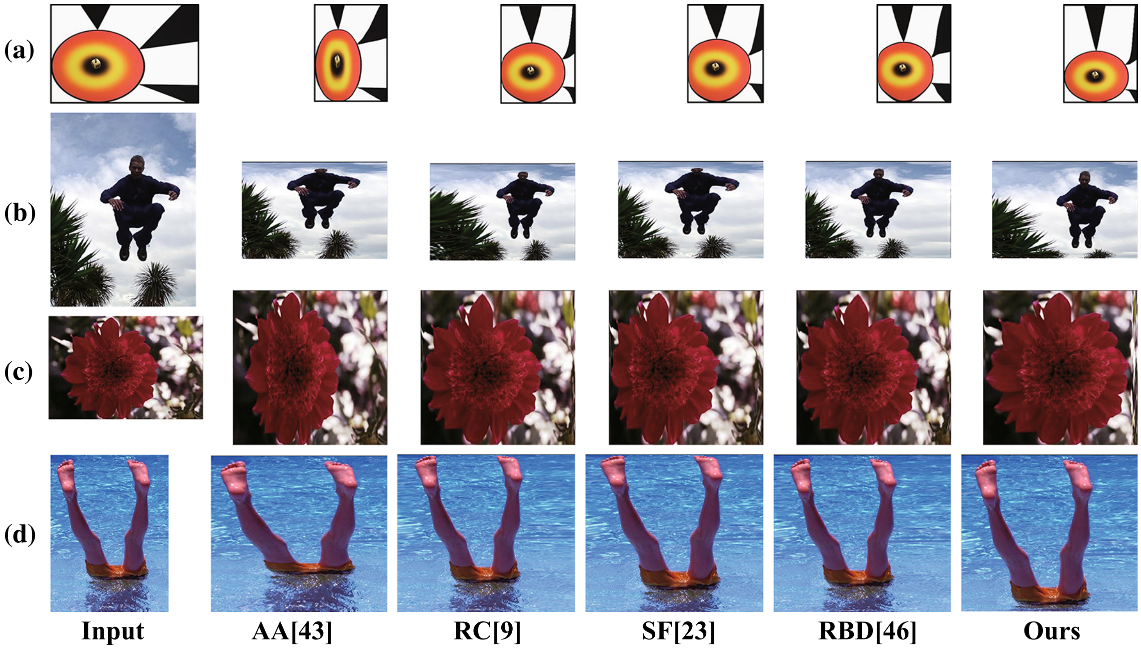


Fig. 11 Image retargeting results. **b** and **c** reduce or enlarge 50 % height. **a** and **d** reduce or enlarge 50 % width

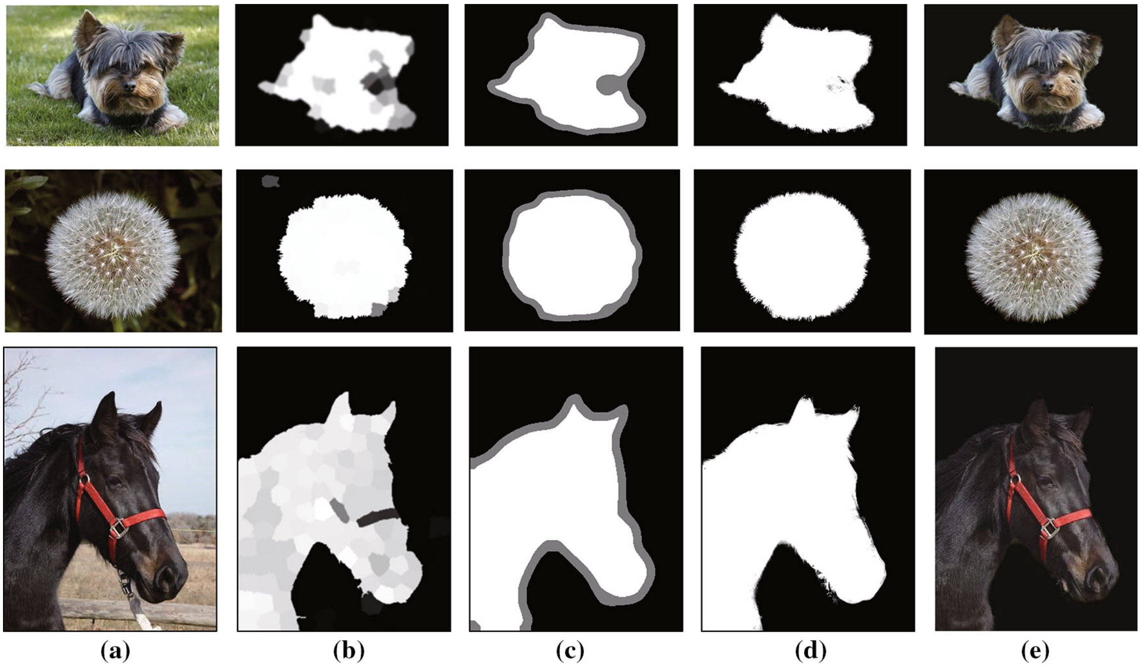


Fig. 12 Matting results on MSRA10k dataset [9]. **a** Input image. **b** Our saliency map. **c** Trimap created from **b**. **d** Matting result tested on **c** by the comprehensive sampling sets [28]. **e** Compositing result with a constant background using **d**

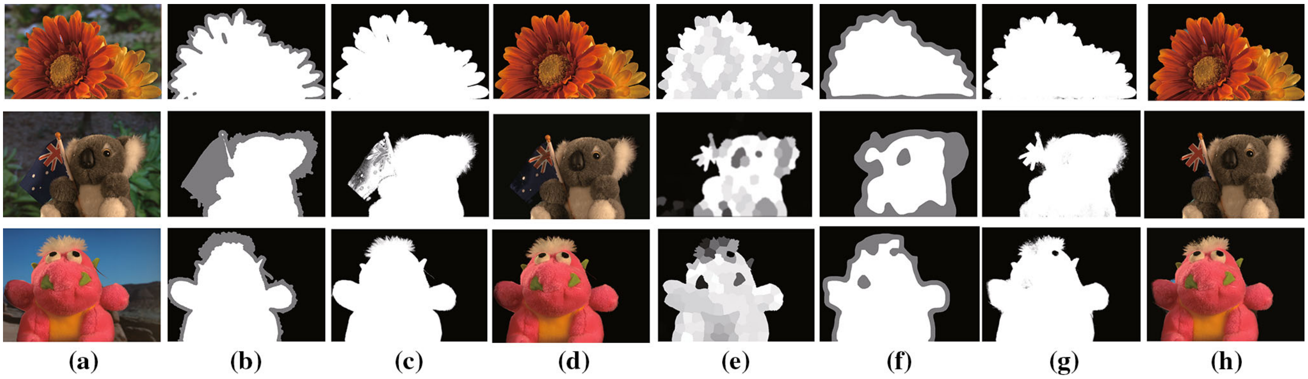


Fig. 13 Comparison with different trimaps. **a** Input image from [25]. **b** Finest trimap from [25]. **c** Matting result tested on **b** by the comprehensive sampling sets [28]. **d** Compositing result with a constant background using **c**. **e** Our saliency map. **f** Trimap created from **e**. **g** Matting result tested on **f** by the comprehensive sampling sets [28]. **h** Compositing result with a constant background using **g**

the salient object can be obtained. Here we use the comprehensive sampling sets proposed by [28] as the alpha matting system.

Figure 12 shows several matting results on images from the MSRA10k dataset [9]. The results show that our method can be seamlessly fitted into the automatic image matting system [28] as an intelligent frontend. To further evaluate effectiveness of the trimaps created by our method, we compare final boundaries when the image matting system [28] works with different trimaps from our method and from a benchmark dataset [25]. Figure 13 presents some comparison results tested on images from [25]. It can be observed that the matting results obtained using the trimaps created by our method compare favorably with the results derived using the finest trimaps, as the saliency object produced by our method is close enough to the ground truth.

5 Conclusions

In this paper, we present a simple and efficient object-level saliency detection model, which can produce high-resolution, full-field saliency maps. The proposed method aggregates multiple saliency cues and priors. The saliency confidence is further jointly modeled with a unified iterative framework combined boundary contrast and smoothing, which can be complementary to each other to provide more informative evidences for extracting complete salient objects. The iterative framework refines the coarse saliency map to improve the contrast between salient and non-salient regions until a close to binary saliency map is reached, which is more practical in real-world applications. The experimental results on public datasets demonstrate that the proposed approach can obtain a more robust performance and consistently outperforms the existing saliency detection methods in terms of precision and recall rate, even in extreme cases. In addition, we evaluate the contribution of the proposed method on image retargeting and automatic image matting.

In the future, we believe that incorporating more sophisticated techniques, such as hierarchical structure [38] and multi-scale analysis [32], will be helpful to improve our saliency detection.

References

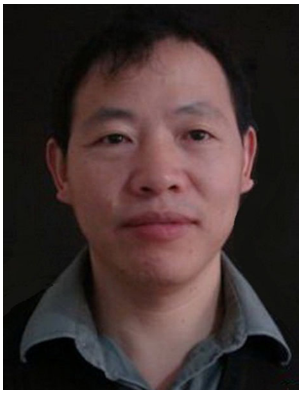
1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1597–1604 (2009)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
3. Avidan, S., Shamir, A.: Seam carving for content-aware image resizing. ACM Trans, Graph (2007)
4. Borji, A., Itti, L.: CAT2000: A Large Scale fixation dataset for boosting saliency research. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Workshop on Future of Datasets (2015)
5. Borji, A., Itti, L.: State-of-the-art in visual attention modeling. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 185–207 (2013)
6. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: a benchmark. In: European Conference on Computer Vision (2012)
7. Borji, A., Cheng, M., Huaizu, J., Jia, L.: Salient object detection: a benchmark. IEEE Trans. Image Process. **24**(12), 5706–5722 (2015)
8. Cheng, M., Mitra, N.J., Huang, X., Hu, S.: SalientShape: group saliency in image collections. Vis. Comput. **30**(4), 443–453 (2014)
9. Cheng, M., Mitra, N.J., Huang, X., Torr, P.H.S., Hu, S.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)
10. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. Int. J. Comput. Vis. (IJCV) **59**(2), 167–181 (2004)
11. Frintrop, S., Klodt, M., Rome, E.: A real-time visual attention system using integral images. In: International Conference on Computer Vision Systems (2007)
12. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 1915–1926 (2012)
13. Harel, J., Koch, C., Perona, P.: Graph based visual saliency. In: Advances in Neural Information Processing Systems, pp. 545–552 (2006)
14. Hayhoe, M., Ballard, D.: Eye movements in natural behavior. Trends Cogn Sci **9**, 188–194 (2005)
15. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
16. Itti, L., Koch, C., Niebur, E.: A model of saliency based visual attention for rapid scene analysis. In: IEEE Trans. Pattern Anal. Mach. Intell. pp. 1254–1259 (1998)
17. Koch, C., Poggio, T.: Predicting the visual world: silence is golden. Nat. Neurosci. **2**, 9–10 (1999)
18. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurbiol. **4**, 219–227 (1985)
19. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N.: X, T., H.Y., S.: Learning to detect a salient object. IEEE Trans. Pattern Anal. Mach. Intell. **33**(2), 353–367 (2011)
20. Ma, Y.F., Zhang, H.J.: Contrast based image attention analysis by using fuzzy growing. In: ACM Multimedia, pp. 374–381 (2003)
21. Ma, Y., Zhang, H.: Contrast-based image attention analysis by using fuzzy growing. ACM Multimedia (2003)
22. Panozzo, D., Weber, O., Sorkine, O.: Robust image retargeting via axis-aligned deformation. Eurographics. **31**(2), 229–236 (2012)
23. Perazzi, F., Krahenbuhl, P., Pritch, Y., et al.: Saliency filters: contrast based filtering for salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
24. Ran, M., Zelnik-Manor, L., Tal, A.: Saliency for image manipulation. Vis. Comput. **29**(5), 381–392 (2013)
25. Rhemann,C., Rother, C., Wang, J,. Gelautz, M., Kohli, P., Rott, P.: A perceptually motivated online benchmark for image matting. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
26. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: interactive foreground extraction using iterated graph cuts. ACM TOG **23**(3), 309314 (2004)
27. Rubinstein, M., Shamir, A., Avidan, S.: Improved seam carving for video retargeting. ACM Trans. Graph. **27**, 15–19 (2008)
28. Shahrian, E., Rajan, D., Price, B., Cohen, S.: Improving image matting using comprehensive sampling sets. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
29. Tatler, B.: The central fixation bias in scene viewing: selecting an optimal viewing position ndependently of motor biases and image feature distributions. J. Vis. **7**(14), 1–17 (2007)
30. Tavakoli, HR., Rahtu, E., Heikkil, J.: Fast and efficient saliency detection using sparse sampling and Kernel density estimation. In: Scandinavian Conference on Image Analysis (2011)
31. Tong, N., Lu, H., Zhang, L., Xiang, R.: Saliency Detection with Multi-Scale Superpixels. IEEE Signal Process. Lett. **21**(9), 1035–1039 (2014)
32. Treisman, A., Gelade, G.: A feature-integration theory of attention. Cogn. Psychol. **12**, 97–136 (1980)
33. Wang, D., Li, G., Jia, W., Luo, X.: Saliency-driven scaling optimization for image retargeting. Vis. Comput. **27**(9), 853–860 (2011)
34. Wang, K., Lin, L., Lu, J., Li, C., Shi, K.: PISA: pixelwise image Saliency by aggregating complementary appearance contrast measures with edge-preserving coherence. IEEE Trans. Image Process. **24**(10), 3019–3032 (2015)
35. Wei, Y.C., Wen, F., Zhu, W.J., Sun, J.: Geodesic saliency using background priors. In: European Conference on Computer Vision (2012)
36. Wu, H., Li, G., Luo, X.: Weighted attentional blocks for probabilistic object tracking. Vis. Comput. **30**(2), 229–243 (2014)
37. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. IEEE Conf. Comput. Vis. Pattern Recognit. **9**(4), 1155–1162 (2013)
38. Yang, J.,Yang, M.: Top-down visual saliency via joint crf and dictionary learning. In: IEEE Conference on Computer Vision and Pattern Recognition. pp, 2296–2303 (2012)
39. Yang, C., Zhang, L.H., Lu, H.C.: Graph-regularized saliency detection with convex-hull-based center prior. IEEE Signal Process. Lett. **20**(7), 637–640 (2013)

40. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.: Saliency detection via graph-based manifold ranking. In: IEEE Conference on Computer Vision and Pattern Recognition. pp, 3166–3173 (2013)
41. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: ACM Multimedia, pp. 815–824 (2006)
42. Zhang, H., Xu, M., Zhuo, L., Havyarimana, V.: A novel optimization framework for salient object detection. Vis. Comput. **32**(1), 1–11 (2014)
43. Zhang, J., Sclaroff, S.: Exploiting surroundedness for saliency detection: a Boolean Map approach. IEEE Trans. Pattern Anal. Mach. Intell (2015). doi:10.1109/TPAMI.2015.2473844
44. Zhong, G., Liu, R., Cao, J., Su, Z.: A generalized nonlocal mean framework with object-level cues for saliency detection. Vis. Comput. (2015). doi:10.1007/s00371-015-1077-z
45. Zhu, H., Cai, J., Zheng, J., Wu, J., Magnenat Thalmann, N.: Salient object cutout using Google images. In: IEEE International Symposium on Circuits and Systems. pp, 19–23 (2013)
46. Zhu, W., Liang, S., Wei, Y., Sun, J., Cottrell, G.: Saliency optimization from robust background detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)

Ruihui Li received his B.S degree in Computer Science and Technology from Hunan University, Changsha, China, and he is an M.S. candidate student in Information Science and Engineering at Hunan University, Changsha, China. His current research interests include saliency detection and objective tracking.

Jianrui Cai received his B.S and M.S degrees from Hunan University, Changsha, China, in 2012 and 2015, respectively. He is now a Ph.D. Candidate with The Hong Kong Polytechnic University. His research interests include machine learning and computer vision.

Hanling Zhang is an associated Professor at the Department of Information Sciences and Engineering at Hunan University since 2003. He has a Ph.D. in Signal and Information Process from Northwestern Polytechnical University, China,2003. From 2007 till 2008 he was a visiting professor at University of western Ontario, Canada. His research interests include machine learning, image processing.



Taihong Wang is a Professor at the Department of Physics and Microelectronics Science at Hunan University. His research interests include micro-nano processing, super sensitive detection and the internet of things.