Hierarchical Contour Closure based Holistic Salient Object Detection

Qing Liu, Xiaopeng Hong, Beiji Zou, Jie Chen, Zailiang Chen, and Guoying Zhao, *Senior Member, IEEE*

***Abstract*—Most existing salient object detection methods compute the saliency for pixels, patches or superpixels by contrast. Such fine-grained contrast based salient object detection methods are stuck with saliency attenuation of the salient object and saliency overestimation of the background when the image is complicated. To better compute the saliency for complicated images, we propose a hierarchical contour closure based holistic salient object detection method, in which two saliency cues, i.e., closure completeness and closure reliability are thoroughly exploited. The former pops out the holistic homogeneous regions bounded by completely closed outer contours, and the latter highlights the holistic homogeneous regions bounded by averagely highly reliable outer contours. Accordingly, we propose two computational schemes to compute the corresponding saliency maps in a hierarchical segmentation space. Finally, we propose a framework to combine the two saliency maps, obtaining the final saliency map. Experimental results on three publicly available datasets show that even each single saliency map is able to reach the state-of-the-art performance. Furthermore, our framework which combines two saliency maps outperforms the state of the arts. Additionally, we show that the proposed framework can be easily used to extend existing methods and further improve their performances substantially.**

***Index Terms*—salient object detection, contour closure, closure completeness, closure reliability.**

I. Introduction

SALIENT object detection is to compute a saliency map in which the object attracting the humans' attention most is uniformly highlighted via mimicking the human visual system. It has raised a lot of attention because of its wide applications in image segmentation [1], compression [2] and object discovery [3], etc.

Recently, a number of methods for salient object detection are proposed. By answering the question "what the salient object should look like", the *contrast prior* is exploited. It assumes that there are apparent differences in appearance between the regions from a salient object and background regions. To implement the *contrast prior*, numerous strategies such as [4] [5] [6] [7] [8] [9] [10] [11] are proposed. Simultaneously, by answering the opposite question "what the background should look like", the *background prior* is explored. It assumes that the regions on the image borders are background and the entire background is smoothly connected. Thus, the dissimilarity to the boundary regions in turn reflects the saliency of a region. To implement the *background prior*, strategies such as [12] [13] [14] [15] [16] [17] are proposed. For either contrast prior or background prior, the essence is the contrast between the salient object and the background.

However, simply using contrast prior is insufficient. There still exists two challenges: (1) saliency attenuation of the salient object: when the salient object is in non-uniform appearance, as exemplified in the first example in Fig. 1, previous methods tend to highlight the salient object non-uniformly rather than the holistic salient object uniformly. (2) Saliency overestimation of the background: when the image background is cluttered, as the second example shows in Fig. 1, existing methods always erroneously highlight the non-salient structures with high contrast to others, and generate a chaos saliency map.

To address these issues, we tackle the salient object detection from a different direction. In this paper, inspired by the psychology finding that contour closure does contribute to the object-based attention [18] [19] [20], we propose a computational method to explore two cues about contour closure for salient object detection.

The first cue is *closure completeness*, indicating whether a homogeneous region is bounded by closed outer contours. From the view of human perceiving object, it has been demonstrated that targets with closed outer contours (e.g. R_1 in Fig. 2) attract humans' attention most [18] [21] [22]. From the view of photographing, because salient objects are always photographed in focus and located according to the *rule of thirds*, the salient objects have sharp and closed outer contours. Thus, extracting the object bounded by completely closed outer contours powerfully facilitates the holistic salient object detection.

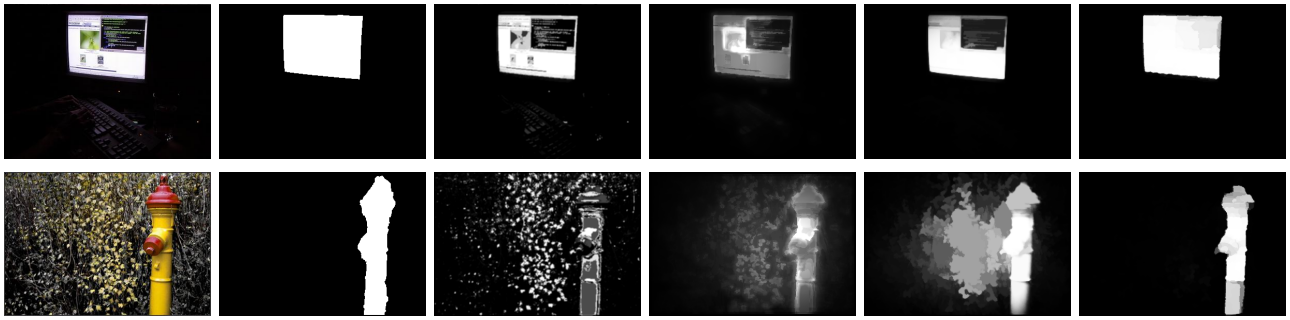
The second one is *closure reliability*. It is defined by the average reliability of the contours partitioning the object and its surroundings, characterising the closure extent of the contours bounding the object. Usually, the closure reliability of a salient object is high due to its strong contrast to its surroundings. For example, in Fig. 3, the salient object R_1 bounded by contours with high reliability attracts more attention than the background region R_2 bounded by contours with low reliability. Thus, we naturally exploit the *closure reliability* for saliency detection.

Albeit beautiful in concept, practically measuring the *closure completeness* and *closure reliability* for the object is difficult due to the absence of the exact location and shape

Q. Liu was with the School of Information Science and Engineering, Central South University, Changsha, Hunan, 410083, China, and Center for Machine Vision and Signal Analysis, P.O.Box 4500 FI-90014, University of Oulu, Finland. (emails: qing.liu.411@gmail.com)

B. Zou and Z. Chen were with the School of Information Science and Engineering, Central South University, Changsha, Hunan, 410083 China. (emails: bjzou@csu.edu.cn, xxxyczl@163.com)

X. Hong, J. Chen and G. Zhao were with Center for Machine Vision and Signal Analysis, P.O.Box 4500 FI-90014, University of Oulu, Finland. (emails: {xhong, jiechen, gyzhao}@ee.oulu.fi)



(a) Input images (b) GT (c) GC [10] (d) PCAS [11] (e) MB+ [15] (f) ours

Fig. 1: Examples of challenging images for salient object detection. (a) Input images. (b) Ground truth images (GT). (c∼f) Saliency maps generated by GC [10], PCAS [11], MB+ [15] and ours.

about the object. To solve this problem, we propose to compute the *closure completeness* and *closure reliability* for homogeneous regions in a hierarchical segmentation space. For *closure completeness*, we measure the completeness of a region by the expectation of times for which the region is bounded by completely closed outer contours over the hierarchical space. In this way, the homogeneous regions bounded by uniformly reliable outer contours are highlighted. For *closure reliability*, we define it as the length weighted reliability of the region's bounding contours at an optimal hierarchical level automatically obtained by searching the whole hierarchical space. In this way, homogeneous regions bounded by averagely high reliable outer contours are highlighted. To take the advantages of these two cues, we fuse them into a final saliency map.

We test our method on three widely used datasets and compare it with sixteen state-of-the-art methods. The experimental results show that the proposed method outperforms the previous methods on three datasets consistently. To validate the flexibility of the proposed method, by combining the two cues, we extend the previous background prior based methods, and substantially improve their performances, which in turn demonstrates the effectiveness of the cues we explore.

The contributions of this paper are summarised as follows:

- We highlight the importance of contour closure for salient object detection and explore two saliency cues about closure, i.e., *closure completeness* and *closure reliability*, which are ignored in the previous salient object detection methods.
- We propose to measure the two saliency cues in a hierarchical segmentation space, which treat the homogeneous region as a whole such that the entire salient homogeneous region can be highlighted uniformly.
- We propose a framework to combine the closure completeness and the closure reliability, which can be easily used to extend the previous background prior based saliency methods and further improve their performances.

The rest of this paper is organised as follows. Section II reviews the state-of-the-art saliency detection methods. Section III provides a detailed description of the proposed saliency detection framework. In Section IV, experimental results are reported. The following is the conclusion of this paper and our future work.

(a) Binary contour map (b) Salient object

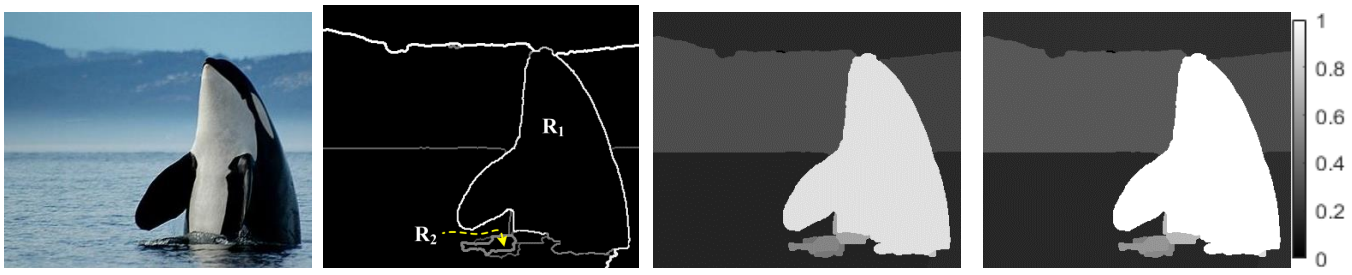
Fig. 2: Illustration for the closure completeness: regions with completely closed contours attract more attention than those with opened contours. The salient region R_1 is bounded by closed contour in red in (a). The background region R_2 reaches the image borders, resulting two gaps along the image borders, i.e., the dotted lines in magenta. Its outer contours are opened.

II. Related Works

During the past years, numerous methods about salient object detection have been proposed and tremendous progress has been made. Also their success paved the way to tackle harder problems such as saliency attenuation of the salient object and saliency overestimation of the background in cluttered images. According to the disciplines that these methods used, we broadly divide them into three classes:

(1) *Biologically inspired methods.* Itti et al. [4] measured the saliency via the centre-surround contrast of multi-scale and multi-channel features. After that, variants for contrast were proposed, such as the discriminant centre-surround contrast [23], centre-surround divergence of feature statistic [24], histogram-based contrast [7], to name a few. Additionally, several methods used deep learning architecture to learn the contrast, such as [25] and [26]. Though these methods achieved a big success, they tended to highlight the boundaries of the salient object rather than the whole salient object in the image, especially in the cluttered scene images.

(2) *Background prior based methods.* Wei et al. [12] proposed to use the background prior for saliency detection. They validated that most photographers follow the rules of photographic composition and do not crop salient object along the view frame. In other words, regions on the image



(a) Input image (b) A segmentation (c) Closure reliability map (d) Normalised closure reliability map

Fig. 3: Closure reliability for salient object detection. (a) Input image. (b) A segmentation associated with the contour reliability. (c) Closure reliability map. (d) Normalised closure reliability map. Regions such as R_1 with strong contours are more salient than those such as R_2 with weak contours.

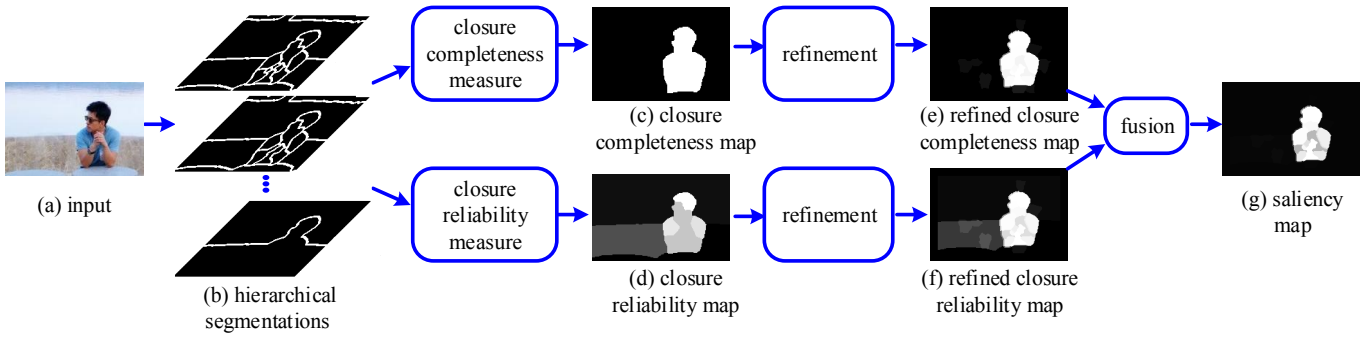


Fig. 4: The proposed framework for saliency detection[1]. In this framework, two schemes for closure completeness and closure reliability respectively are proposed to generate the closure completeness map and closure reliability map. Then they are further refined using background prior. Finally, we combine the refined maps and yield the final saliency map. In this framework, the refinement is highly flexible. Existing background prior based methods after combining with the closure based saliency cues can achieve substantially better performances than before.

border are always background. Inspired by such a rule, they proposed a geodesic saliency model. Thereafter, numerous saliency detection models leveraging the background prior were proposed. For example, in [13], the superpixels on the image border were regarded as the background queries to generate a coarse saliency map. In [14] [27] [28] [29] [30] [31] [32] [33], background prior played a similar role, though various strategies for the saliency computation were adopted. Background prior based methods achieved excellent performances in uniformly highlighting whole salient objects when the image is simple. But they are limited when the image is cluttered.

(3) *Psychology inspired methods.* According to the suggestions from psychologists that figures are more likely to be attended than the background, psychology inspired methods formulated the saliency detection problem as a figure-ground separation problem in which the configural figure-ground principles such as convexity, surroundedness were explored as saliency cues. For example, Lu et al. [34] proposed a convexity context based saliency model while Zhang et al. [35] explored the surroundedness cue and presented a Boolean map based saliency model.

In terms of the granularity level, region-based saliency detection becomes more and more popular. In [36], Jiang et al. proposed to estimate the saliency for the segmented regions in a segmentation, in which region-level focusness and objectness were estimated based on the pixel-level focusness and objectness. In [8], Jiang et al. computed a set of saliency maps according to the multi-level segmentations, then fused them into a final saliency map by the learnt linear fusor. Li et al. [37] proposed to compute the saliency for each object candidate generated by an object proposal method, then generated the salient object segmentation by averaging the top-K segments at pixel level. The saliency of each object candidate was obtained by the learnt random forest fed by the shape features and fixation distribution features. Our method is different from these methods in three aspects: (1) The saliency cues our method proposes are contour closure based, which has not been elaborated before; (2) In our method, two of the properties about the contour closure are designed to estimate the saliency of the regions directly while the saliency cues such as the objectness and focussness in [36] and the fixation map in [37] are first computed in pixel-level, then the region-level saliency cues are formulated based on the pixel-level ones; (3) We generate the closure completeness based saliency map and closure reliability based saliency map in the whole hierarchical segmentation space while the hierarchical structure is ignored in [36] [8] [37].

[1]The input image is from http://www.weibo.com/musiclijian?is_all=1

One thing worth to mention here is that contour closure has been used for object proposal [38] [39], contour completion [40], perceptual grouping [41] [42]. However, few previous salient object detection methods use contour closure, even though it plays an important role in priming low-level saliency. In addition to its applications in computer vision, another related work is about how to compute the contour closure. Stahl et al. [41] directly found the closed contours from line segments via minimising the closure cost with regard to boundary gap while Levinshtein et al. [42] found the closed contours from superpixel boundaries. Our method differs from these methods in two aspects: (1) In contour closure computation [41] [42], the number of salient structures should be given beforehand. However, in salient object detection, the number of the salient objects in images is uncertain. Thus we cannot directly apply closure computation methods like [41] [42] to salient object detection. (2) Contour closure computation [41] [42] aim to find closed and single object boundaries while our goal is to design reasonable closure measures for salient object detection.

III. Contour Closure Based Saliency Detection

The proposed framework is illustrated in Fig. 4, in which two closure based maps are generated for the salient object detection in the hierarchical segmentation space. One is the closure completeness map which integrates the information over all the levels whether the outer contours of regions are completely closed or not (see Sec. III-A). The other is the closure reliability map in which the closure reliability of each region is computed at an automatically selected detailed level (see Sec. III-B). Then we further refine these two saliency maps and fuse them into a final saliency map (see Sec. III-C).

As aforementioned, directly computing the closure completeness and closure reliability for the entire object are challenging due to the absence of the exact location and shape about the object. To ensure the computational unit is a holistic object, we propose to measure these two cues in a hierarchical segmentation space. On one hand, hierarchy is one of the main characteristics of human perceptual organisation [43]. On the other hand, an image can be represented by a hierarchical tree of regions easily. As the basis of the proposed method, the formulation of the hierarchical segmentation is first provided.

A segmentation K of an image I, whose domain is $\Omega \subset \mathbb{R}^2$, can be defined by a finite set of rectifiable Jordan curves, i.e., the contours of K. The regions $\{R_i\}$ of K, defining a partition $P = \{R_i\}$ of Ω, are connected pixels of $\Omega \setminus K$. Given an initial partition P_0 of Ω and a level parameter ξ, a hierarchical segmentation [43] [44] *[45]* can be expressed by a sequence of partitions $\{P_\xi\}$ via assigning each partition P_ξ to the couple (P_0, ξ) satisfying:

$$\xi \geq \xi' \Rightarrow P_\xi \sqsupseteq P_{\xi'}, \tag{1}$$

$$P_\xi = P_0, \forall\ \xi \leq 0, \tag{2}$$

$$\exists\ \xi'' \in \mathbb{R}^+ :\ P_\xi = \{\Omega\}, \forall\ \xi \geq \xi''. \tag{3}$$

Let $\xi_N = \inf(\xi'')$, then the effective range of levels is $[0, \xi_N]$. $\sqsupseteq$ denotes the partial order of partitions, i.e., any region in partition P_ξ is a disjoint union of regions in partition $P_{\xi'}$, which enforcing a hierarchical structure to the family $\mathcal{H} = \{R_i \subseteq \Omega | \exists\ \xi :\ R_i \in P_\xi\}$. In this paper, we adopt [46] to obtain the hierarchical segmentation $\{P_\xi\}$ where $\xi \in [0, \xi_N]$.

A. Closure Completeness for Saliency Detection

In this section, closure completeness is explored to highlight the pixels in a salient region which is bounded by completely closed outer contours. In the hierarchical segmentation space, we activate the regions with closed outer contours level by level. Then we measure the closure completeness by the expectation of times for which the regions are activated in the whole hierarchical segmentation space. In this way, for more times the holistic homogeneous regions are activated, the more salient they are.

To describe whether the outer contours of regions are completely closed or not in segmentation $P_\xi = \{R_i\}, \xi \in [0, \xi_N]$ from a hierarchical segmentation space, we define an indicator map Q at level ξ by:

$$Q_\xi(x) = \begin{cases} 1 & \text{if } x \in R\ ,\ R \in P_\xi^{in}, \\ 0 & \text{if } x \in R\ ,\ R \in P_\xi^{border}, \end{cases} \tag{4}$$

where $P_\xi^{in} = \{R_i \mid R_i \cap B = \phi\}$ is the set of inner regions, $P_\xi^{border} = P_\xi \setminus P_\xi^{in}$ is the set of border regions, and B is the set of pixels on the image border.

In Q, inner regions with completely outer contours are activated whereas regions with opened outer contours are suppressed. For any two indicator maps Q_{ξ_1} and Q_{ξ_2} where $\xi_1 < \xi_2$, if a pixel x is activated in Q_{ξ_2}, then it is activated in Q_{ξ_1} (i.e., $Q_{\xi_2}(x) = 1 \Rightarrow Q_{\xi_1}(x) = 1$). In this way, as ξ increases from 0 to ξ_N, the background regions are gradually peeled off from the image border towards the image centre. Concomitantly, the salient objects pop out gradually. By computing the expectation of the indicator maps over the whole hierarchical segmentation space, we obtain the closure completeness based saliency map $S_\mathsf{C} \in [0, 1]$:

$$S_\mathsf{C}(x) = \int Q_\xi(x) \cdot p(\xi) d\xi, \tag{5}$$

where ξ obeys a uniform distribution with the probability density function $p(\xi)$. In S_C, high saliency value is assigned to the region which is activated in most of the indicator maps $\{Q_\xi\}$. Inversely, small saliency value is assigned to the region which is suppressed in most of the indicator maps $\{Q_\xi\}$.

Fig. 5 gives an example of the computation of S_C. For the given input image Fig. 5(a), we uniformly sample ξ from $0.1 \times \xi_N$ to $0.9 \times \xi_N$ with a step of $0.1 \times \xi_N$, and generate a sequence of segmentations and corresponding indicator maps. In the indicator maps, regions on the image borders are suppressed due to the gaps along the image borders. For example, background region R_i in Fig. 5(b1) is assigned 0 in Fig. 5(c1) due to the gap (marked by dotted line) along the image border. Inversely, the salient region R_j is activated in Fig. 5(c1) for its completely closed outer contour. By computing the expectation of the times for which the regions are activated, we obtain the closure completeness map, i.e., Fig. 5(d). In Fig. 5(d), the more times the regions are activated in indicator maps, the more salient they are.

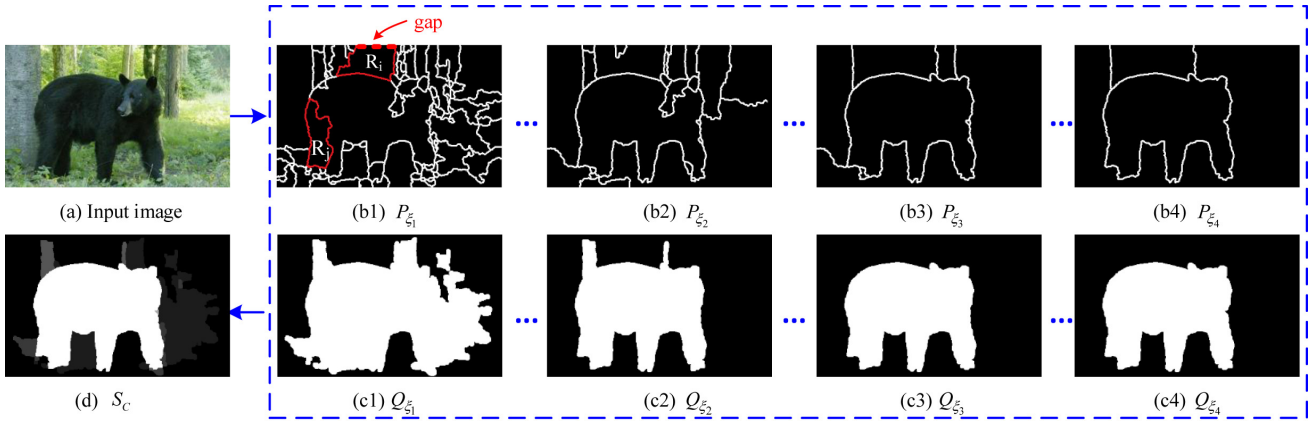


Fig. 5: An example for closure completeness map generation. ξ is uniformly sampled. For concise illustration, only four segmentations and the corresponding indicator maps at levels $\xi_1 = 0.1 \times \xi_N, \xi_2 = 0.3 \times \xi_N, \xi_3 = 0.5 \times \xi_N$ and $\xi_4 = 0.9 \times \xi_N$ are shown.

B. Closure Reliability for Saliency Detection

The closure extent of salient objects are higher than that of the background regions. In this section, we define the closure reliability by:

$$s(R) = s_r(R) \cdot s_b(R). \tag{6}$$

The first item $s_r(R)$ on the right side of Eq. 6 is the length weighted contour reliability of the region's bounding contour lines:

$$s_r(R) = \frac{\sum_{j=1}^{J} W(\kappa_j) \cdot L(\kappa_j)}{\sum_{j=1}^{J} L(\kappa_j) + |R \cap B|}, \tag{7}$$

where J is the number of R's surrounding regions; κ_j is the contour line partitioning R and its j^{th} surrounding region; $W(\kappa_j)$ is the contour reliability of κ_j determined by the dissimilarity between the region R and its surrounding region; $L(\kappa_j)$ is the length of κ_j; B is the pixel set on image borders; and $|R \cap B|$ is the number of pixels located on image borders in the region R. Note that though several weighting strategies on contrast have been proposed for decades, they mainly focus on (1) the spatial distance of the two regions [7] [9] [10] [32] [5] [29], (2) the area of the compared region [7] [8] [5], (3) the distance to the image centre [7], (4) the boundary connectivity [29], and (5) a constant weight [33] [47] [12] [28] [13] [36] [16] [30] [15] [14]. Unlike them, $s_r(R)$ considers both the dissimilarity to its surrounding regions and the surrounded extent by its surrounding regions. It uses the length of the shared contour line as weight. To the best of our knowledge, the weight we use has not been elaborated for saliency detection. Thereby, our contour length weighted contour reliability is novel to saliency detection.

The second item $s_b(R)$ on the right side of Eq. 6 is a penalty item for regions on the image border:

$$s_b(R) = \exp\{-\frac{|R \cap B|}{\sum_{j=1}^{J} L(\kappa_j) + |R \cap B|} - \alpha \cdot \mathrm{sgn}(|R \cap B|)\}. \tag{8}$$

In Eq. 8, $\mathrm{sgn}(\cdot)$ is a sign function. It equals to 1 if $|R \cap B| > 0$ whereas it is 0 if $|R \cap B| = 0$. α a positive value to penalise the saliency of regions on the image border. $s_b(R)$ is determined by the ratio of the number of R's pixels on image borders to the total length of R's outer contours. The larger the ratio is, the more severe the penalty is. For an inner region, s_b is 1, indicating that no punishment is performed on it. Otherwise, for a region on the image border, penalty is imposed on it.

However, there are two problems we have to solve to compute the closure reliability map:

(1) What should we choose? Compute the level based closure reliability map in each segmentation from the hierarchical segmentation space, then average them to obtain the closure reliability map, like what we do for the computation of closure completeness map? Or compute the closure reliability map in an optimal segmentation P^* in which any two adjacent regions are heterogeneous to each other.

(2) If we compute the closure reliability map in P^*, then how to get such an optimal segmentation P^*?

By averaging or computing in an optimal segmentation?

In the hierarchical segmentation space $\{P_\xi\}, \xi \in [0, \xi_N]$, P_ξ is always over-segmented when ξ is small while P_ξ is always under-segmented when ξ is large. In an over-segmentation, a homogeneous salient region is decomposed into small regions. For example, the region b_2 in Fig. 6(c) is decomposed into several small regions in Fig. 6(b). Due to the low reliability of the contour line between two homogeneous small regions in the object, their closure reliabilities are smaller than that computed when taking the salient object as a whole. In an under-segmentation, the salient regions together with background regions are merged into one region. For example, the salient region b_2 and background region a_1 in Fig. 6(c) are grouped into one region in Fig. 6(d). Due to the low reliabilities of the contour lines bounding the background regions, the closure reliabilities of the salient regions are smaller than those computed when only taking the salient object as a whole. What's more, because of the high reliabilities of the contour lines bounding the salient regions, the closure reliabilities of the background regions are larger than those computed when only taking the background as the input. Thus computing the closure reliability map by averaging the closure reliability over

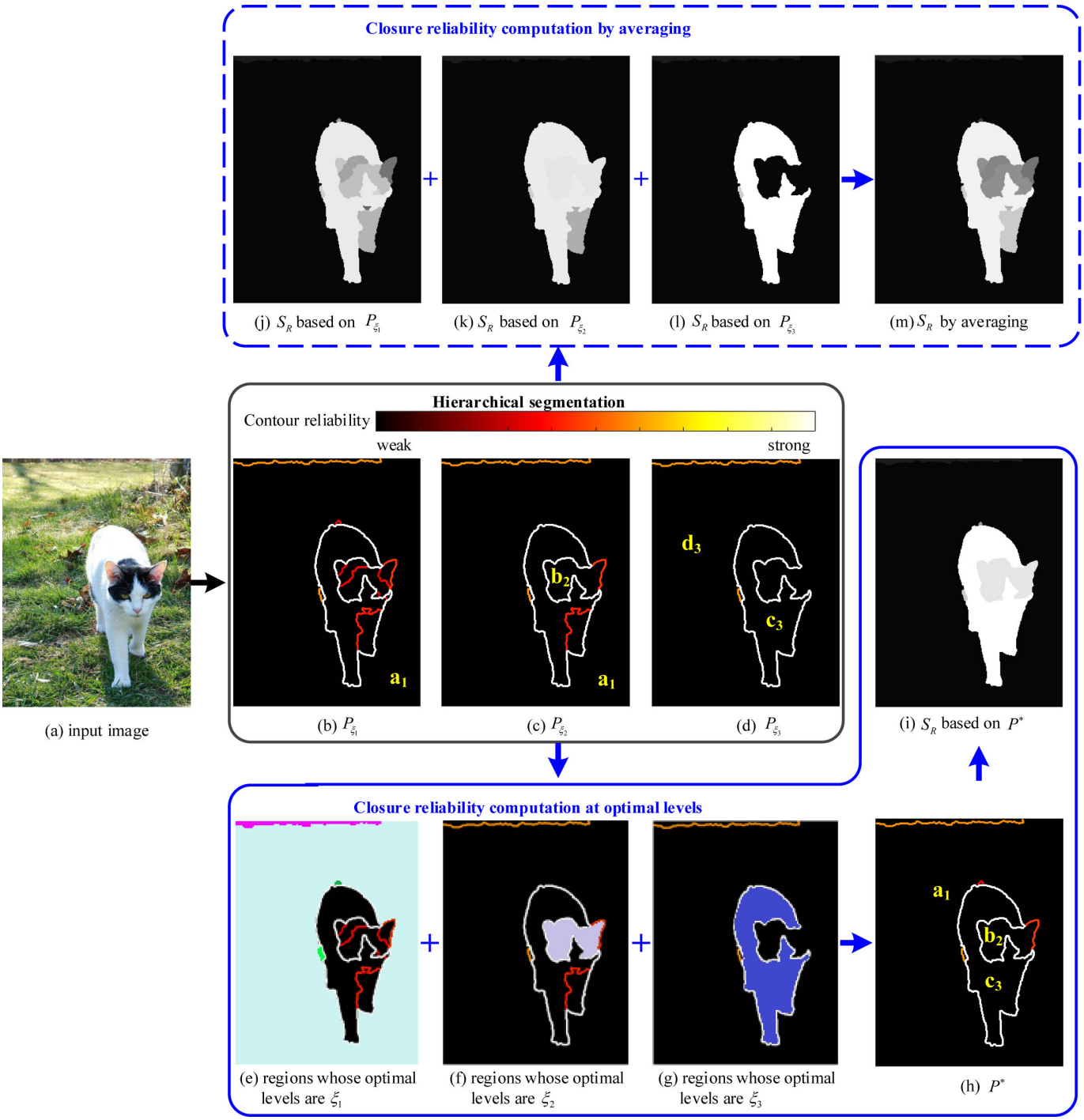


Fig. 6: An example for the computation of closure reliability map. We discard the highly over-segmentations and and under-segmentations as computing closure reliability on them do not make sense. From the middle levels from $\xi_1 = 0.3 \times \xi_N$ to $\xi_3 = 0.5 \times \xi_N$ with a step of $0.1 \times \xi_N$, we uniformly sample three segmentations P_{ξ_1}, P_{ξ_2}, and P_{ξ_3}, i.e., (b)-(d). We colour the contour lines in (b)-(d) to characterise their reliabilities. We use Eq. 9 to search for the optimal level for each pixel. Pixels whose optimal levels are ξ_1 and belonging to the same region in P_{ξ_1} are marked in one colour in (e). Similarly, pixels whose optimal levels are ξ_2/ξ_3 and belonging to the same region in P_{ξ_2}/P_{ξ_3} are marked in one colour in (f)/(g). By grouping the pixels with the same optimal level ξ^* and belonging to the same region in P_{ξ^*}, we produce the optimal segmentation P^*, i.e., (h). Then we compute the closure reliability for each region in P^*, and obtain the closure reliability map (i). In (i), the saliency of the salient regions such as b_2 and c_3 are uniformly highlighted while the saliency of the background regions such as a_1 are suppressed. Comparing to (m) which is obtained by averaging the closure reliability over the hierarchical segmentation space, the saliency of the pixels in the cat in (i) are much more uniform(better view in colour).

the whole hierarchical segmentation is not a good choice. Therefore, we choose to generate an optimal segmentation P^* based on the hierarchical segmentation space such that any two adjacent regions in P^* are heterogeneous to each other. Then we compute the closure reliability map in P^*.

Optimal segmentation generation

In the following context, we provide a solution to generate an optimal segmentation P^* based on the hierarchical segmentation.

Hierarchical segmentation does not guarantee that there is a ξ^* such that all the regions in P_{ξ^*} are heterogeneous to each other. For example in Fig. 6, the optimal level to compute the closure reliability for pixels in b_2 is ξ_2 while the optimal level to compute the closure reliability for pixels in c_3 is ξ_3. The optimal level for the rest pixels is ξ_1. To obtain an segmentation in which any two adjacent regions are heterogeneous to each other, for each pixel x in the image I, we first search for an optimal level for it from the whole hierarchical segmentation space $\{P_\xi\}, \xi \in [0, \xi_N]$ by:

$$\xi^*(x) = \underset{\xi \in [0,\xi_N]}{\arg\max}\ s(R),\ s.t.\ x \in R,\ R \in P_\xi, \tag{9}$$

Then by grouping the pixels with the same optimal level ξ^* and belonging to the same region in P_{ξ^*}, we generate the optimal segmentation P^*.

Closure reliability map computation in P^*

By computing the closure reliability for each region R in P^* using Eq. 6, we obtain the closure reliability map $S_{\mathtt{R}} \in [0, 1]$ by:

$$S_{\mathtt{R}}(x) = s(R), \quad s.t. \quad \forall R \in P^*, x \in R. \tag{10}$$

The pipeline to compute the closure reliability map is illustrated in Fig. 6.

C. Saliency Refinement and Combination

Though salient regions are highlighted in $S_{\mathtt{C}}$ and $S_{\mathtt{R}}$, in complicated images, small non-salient objects, such as the regions marked in yellow ellipse in Fig. 7, are also highlighted due to their contours with non-ignorable reliability. To further reduce the saliency of the background regions, we refine $S_{\mathtt{C}}$ and $S_{\mathtt{R}}$ using background priors.

Taking the contour completeness map as an example, Fig. 8 shows the pipeline for the refinement. First, we choose the regions with low saliency values in $S_{\mathtt{C}}$ and large distance to the image centre as the background queries. Then similar to the geodesic saliency model [12], we obtain a dissimilarity map $S_{B_{\mathtt{C}}}$ by computing the minimun geodesic distance to the background queries. Compared with the geodesic saliency model [12] by directly taking background queries from the image border, our strategy for background query selection are more moderate. By combining the contour completeness map and dissimilarity map, we obtain the refined closure completeness map:

$$\hat{S}_{\mathtt{C}} = \exp(S_{\mathtt{C}}) \cdot \exp(S_{B_{\mathtt{C}}}). \tag{11}$$

In a similar way, we compute a dissimilarity map $S_{B_{\mathtt{R}}}$ according to the background queries chosen from $S_{\mathtt{R}}$, and obtain a refined closure reliability map $\hat{S}_{\mathtt{R}}$ by:

$$\hat{S}_{\mathtt{R}} = \exp(S_{\mathtt{R}}) \cdot \exp(S_{B_{\mathtt{R}}}). \tag{12}$$

Fig. 7 gives an example of the two saliency maps before refinement and after refinement. It is obvious that the background regions in the refined maps are much clearer than that in the maps before refinement. Note that any background prior based method using the proposed background query selection strategy and combining the closure completeness map and closure reliability map in this way can achieve an improved performance, which is validated in Section IV.

$\hat{S}_{\mathtt{C}}$ highlights the salient objects in the inner of the image and misses the salient objects on image borders. $\hat{S}_{\mathtt{R}}$ highlights regions with strong contour reliability no matter where they are. To fully take the advantages of the two saliency maps, we combine them into one saliency map S by:

$$S = f(\exp(\hat{S}_{\mathtt{C}}) \cdot \exp(\hat{S}_{\mathtt{R}})), \tag{13}$$

where f is a Logistic function for enhancement, defined by:

$$f(x) = \frac{1}{(1 + \exp(-\gamma(x - \tau)))}, \tag{14}$$

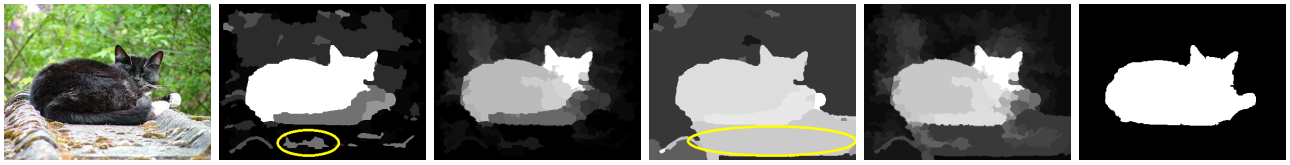
where γ is a predefined parameter to control the level of contrast. τ is a bias such that saliency larger than τ is enhanced while saliency less than τ is suppressed. Additionally, to handle the different details of the salient object, we compute the saliency map at three resolutions, i.e., 100%, 75% and 50% of the input image size. Then we rescale them into the original size and combine them into one map.

Fig. 9 shows two examples to illustrate the effectiveness of fusion and the complement between the closure completeness map $\hat{S}_{\mathtt{C}}$ and closure reliability map $\hat{S}_{\mathtt{R}}$. In the first example, the salient object in Fig. 9(a1) is located on the image border and not fully highlighted in $\hat{S}_{\mathtt{C}}$, i.e., Fig. 9(c1), but it stands out uniformly in $\hat{S}_{\mathtt{R}}$, i.e., Fig. 9(d1). Therefore, we can highlight the salient object via combining those two maps, as Fig. 9(e1) shows. In the second example, the background regions in Fig. 9(a2) have strong contour reliability such that they are falsely labelled as salient regions in $\hat{S}_{\mathtt{R}}$, i.e., Fig. 9(d2). However, since they are located on the image border, they are assigned low saliency values in $\hat{S}_{\mathtt{C}}$, i.e., Fig. 9(c2). Therefore, we can suppress the saliency of the background regions via Eq. 13, as Fig. 9(e2) shows. After fusion, the performance is further improved.

IV. Experiments

A. Experimental setup

In our implementation, we adopt [46] to generate a hierarchical segmentation space and the corresponding contour reliability map for a given image for its high efficiency. All the saliency maps are the multiscale ones. When computing the closure completeness map $S_{\mathtt{C}}$, we uniformly sample segmentation from the base level to the upper level with a step $0.1 \times \xi_N$. When computing the closure reliability map $S_{\mathtt{R}}$, we leave out the highly over-segmentations and under-segmentations, and find the optimal level for each region from middle levels from $0.3 \times \xi_N$ to $0.5 \times \xi_N$ with a step of $0.1 \times \xi_N$. α in Eq. 8 is set to 0.5 and γ and τ in Eq. 14 are set to 10 and 0.7 times the mean value of the input map respectively. These parameters are set empirically and fixed in the following experiments. Besides,



(a) Input image (b) S_{C} (c) $\hat{S}_{\mathrm{C}}$ (d) S_{R} (e) $\hat{S}_{\mathrm{R}}$ (f) GT

Fig. 7: Illustration for the effectiveness of refinement. (a) Input image. (b) Closure completeness before refinement. (c) Closure completeness after refinement. (d) Closure reliability before refinement. *(e)* Closure reliability after refinement. *(f)* Ground truth. The saliency of background regions such as those marked by yellow ellipses in (b) and (d) are further reduced after refinement (better view in color).

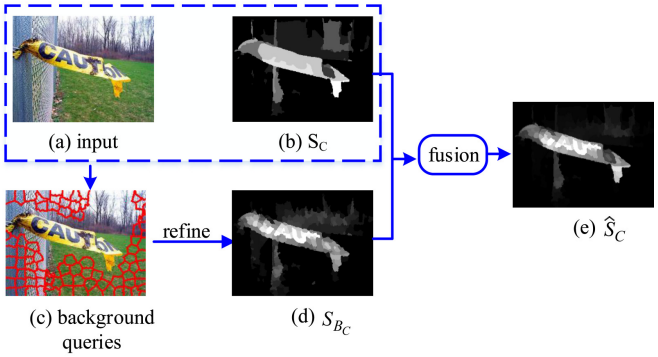


Fig. 8: Pipeline for refinement using the background prior.

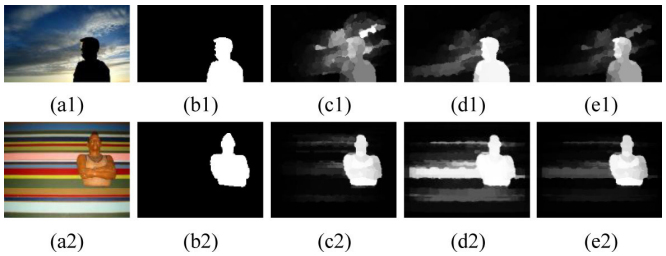


Fig. 9: Illustration for the effectiveness of fusion and the complement between the closure completeness map $\hat{S}_{\mathrm{C}}$ and closure reliability map $\hat{S}_{\mathrm{R}}$. (a1)&(a2) are the input images. (b1)&(b2) are the ground truth images. (c1)&(c2) are the closure completeness maps $\hat{S}_{\mathrm{C}}$ for (a1)&(a2) respectively. (d1)&(d2) are the closure reliability maps $\hat{S}_{\mathrm{R}}$. (e1)&(e2) are the final saliency maps.

the results of our methods, i.e., the closure completeness based saliency map and its refined one, the closure reliability based saliency map and its refine one are the ones which combine the saliency maps from the three scales.

B. Evaluation Metrics

Precision-Recall (PR) Curve and F-measure

We use the widely used metrics PR curve to evaluate the overall performance of a method. Precision measures the fraction of the salient pixels correctly assigned while the recall measures the fraction of ground truth salient pixels which are labelled as salient pixels. For each saliency map, binary maps are obtained by binarizing the saliency map with fixed thresholds ranging from 0 to 255. On a dataset, then the averaging precisions and recalls for different thresholds are computed to plot the PR curve. As well, F-measure, the harmonic mean of the precision and recall, is adopted:

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 Precision + Recall}. \tag{15}$$

Similar to [48] [49], the $Precision$ and $Recall$ are obtained by adaptively thresholding the saliency maps, and the weight β^2 is set to 0.3.

Receiver Operating Characteristics (ROC) Curve

In addition to PR curve, we also report the false positive rate and true positive rate. The false positive rate measures the fraction of true background pixels which are erroneously labelled as salient pixels while the true positive rate measures the fraction of true salient pixels which are labelled as salient pixels. For each threshold from 0 to 255, we binarize each saliency map to compute the mean false positive rates and the mean true positive rates, then plot the ROC curve. The Area Under ROC Curve (AUC) scores of the compared methods are also reported.

Mean Absolute Error (MAE)

As complementary to PR curves and ROC curves, we also use the MAE for quantitative analysis, which considers the saliency assignment for the non-salient pixels, i.e., the pixels erroneously marked as salient. It calculates the mean difference between the saliency map and the ground truth in pixel level.

C. Comparison with State-of-the-Art Methods

To validate the effectiveness of the proposed method, we test our method on three widely used public datasets and compare it with sixteen state-of-the art methods. The datasets include MSRA10K [48], PASCAL-S [37] and DUT-OMRON [13]. MSRA10K contains 10,000 images, and is an extension of two widely used datasets i.e., ASD [49] and MSRA5K [50] [8]. PASCAL-S includes 850 images, which is designed to avoid the dataset bias. DUT-OMRON contains 5,168 images. The three datasets, especially PASCAL-S and DUT-OMRON are the most challenging ones for saliency detection. The compared methods include PCAS [11], GC [10], UFO [8], RC [7], BMS [35], GS [12], MR [13], SO [29], MB+ [15], RRWR [14], TLLT [30], BSCA [33], MAP [28], LPS [32], GP [47] and MST [17]. Among them, PCAS [11], GC [10], UFO

Fig. 10: Visual comparisons of previous approaches to our method and ground truth (GT). Our method generates saliency maps closest to the ground truth (better view in colour).

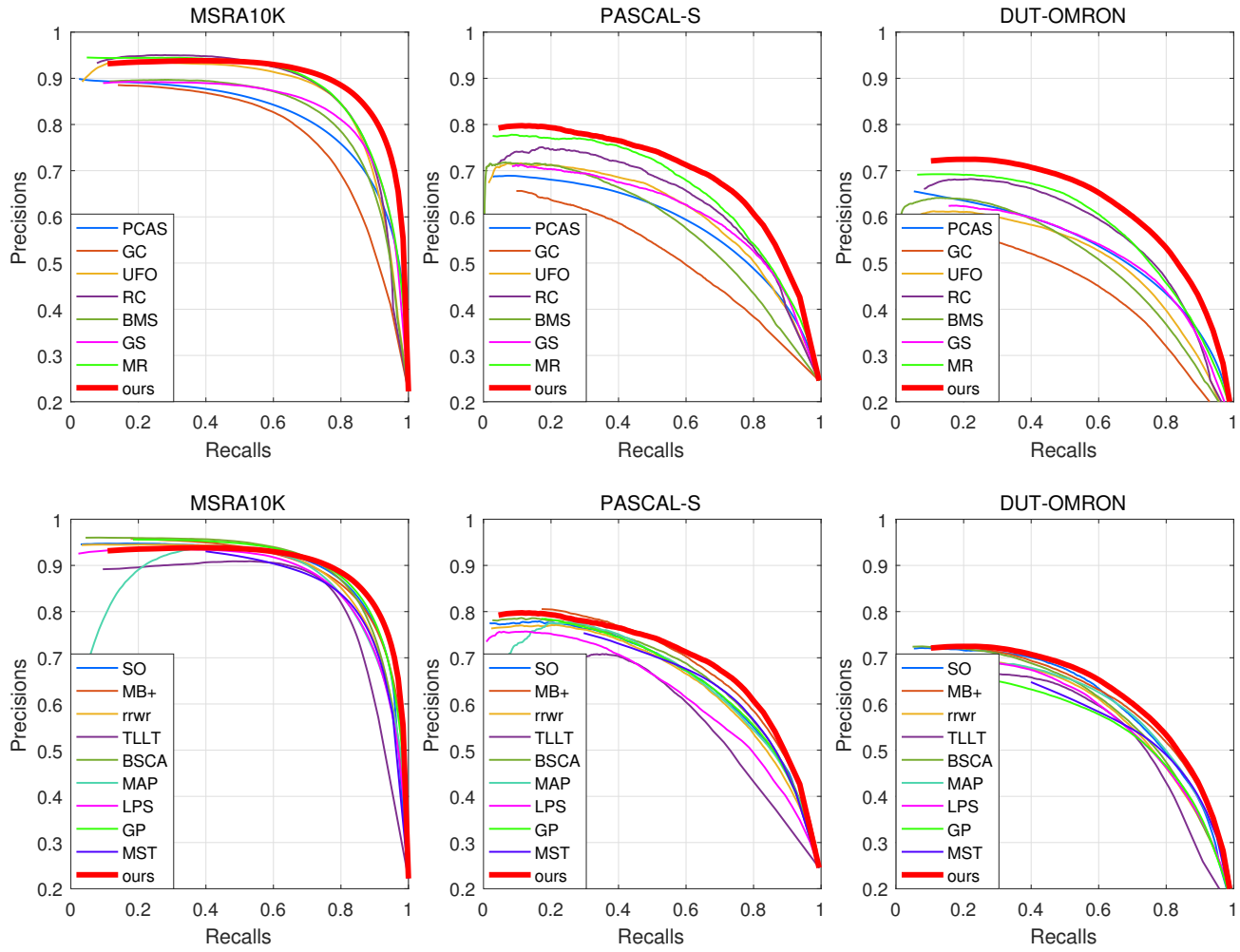


Fig. 11: Precision-recall curves of the compared methods.

[8] and RC [7] are biologically inspired methods. BMS [35] is psychology inspired while the rest are background prior based methods. Generally, saliency methods using deep learning networks such as [25] and [26] achieve better performances than the traditional unsupervised saliency methods. But they require large training data. Therefore, we only compare our method with traditional unsupervised saliency methods.

First, visual comparisons between the previous approaches and our method are performed. In Fig. 10, the detection results for nine examples are shown, in which the left three columns are from MSRA10K [48], the middle three columns are from PASCAL-S [37] and the last three columns are from DUT-OMRON [13]. For these nine examples which are challenging for previous methods, our method can generate saliency maps with uniformly salient objects and clear background.

Then quantitative comparisons between the previous approaches and our method are performed. Fig. 11 shows the PR curves of the compared methods on three datasets. On PASCAL-S [37] and DUT-OMRON [13], our method achieves the best performance. On MSRA10K [48], our method substantially outperforms the biologically inspired methods and the psychology inspired methods, and achieves comparable performance to background prior based methods.

Fig. 12 shows the ROC curves on the three datasets. It shows that our method consistently outperforms the compared methods on the three datasets. Table. I reports the F-measures, AUC and MAE of each method on the three datasets, where our method consistently achieves the best performance.

D. Validation of Contour Closure for Saliency Detection

To validate the effectiveness of the proposed saliency cues, we compare the closure completeness map S_{C}, the closure reliability map S_{R}, the refined closure completeness map $\hat{S}_{\mathrm{C}}$, the refined closure reliability map $\hat{S}_{\mathrm{R}}$ and the final fused saliency map S to five methods, i.e., RC [7], UFO [8], BMS [35], SO [29] and MB+ [15]. Among them, the first two methods are the state-of-the-art biologically inspired methods, and the third method is psychology inspired while the rest are the state-of-the-art background prior based methods. Table. II shows the F-measures, AUC and MAE of the compared methods. From the table we can see that (1) the refinement stage improves the performances substantially; (2) fusion slightly improves the performance; (3) compared to contrast based methods and psychology inspired method, both the closure completeness

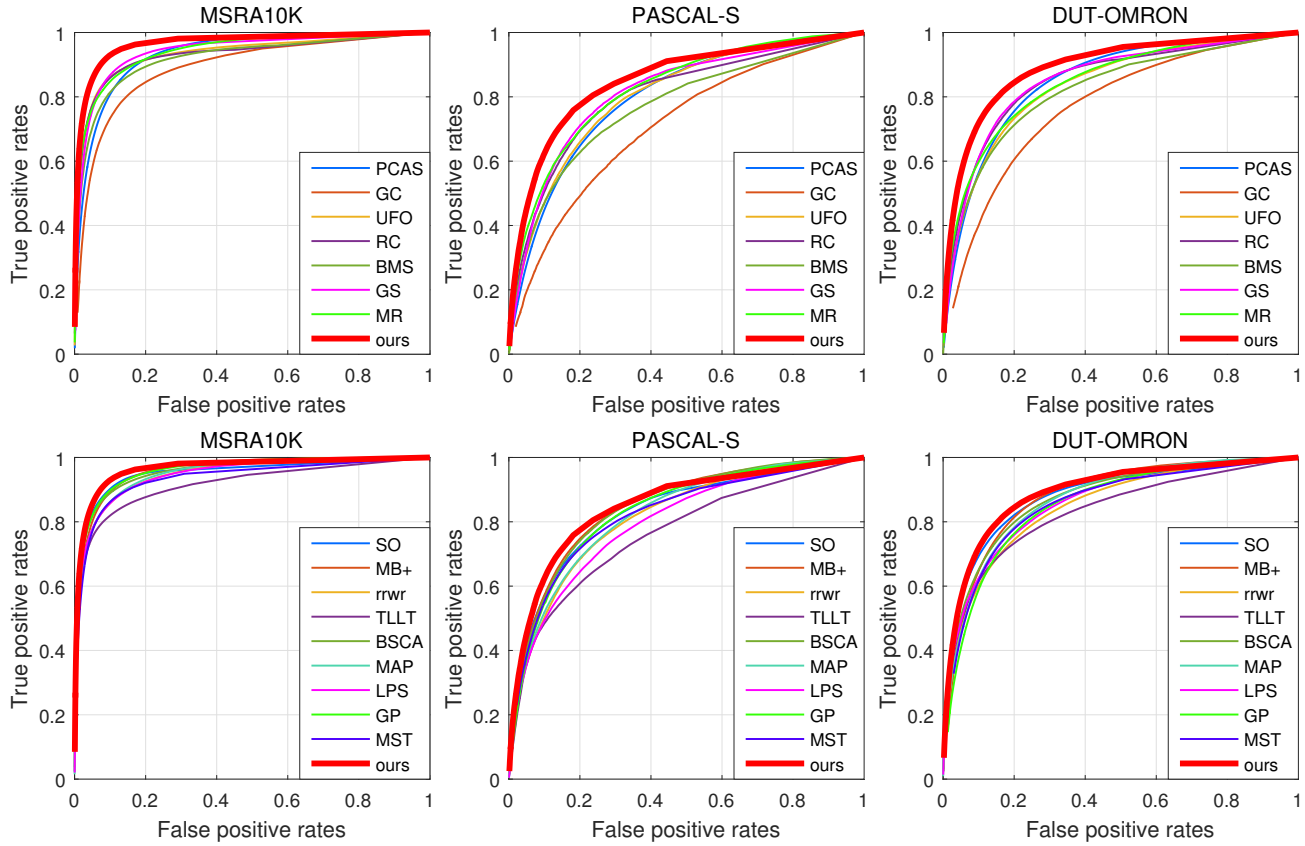


Fig. 12: ROC curves of the compared methods (better view in color).

Table I: F-measures, AUC and MAE of the compared methods on three datasets. Results marked in red are best ones while those in blue are second best ones (better view in color).

	MSRA10K			**PASCAL-S**			**DUT-OMRON**		
MODELS	F_β	AUC	MAE	F_β	AUC	MAE	F_β	AUC	MAE
PCAS [11]	0.745	0.934	0.185	0.526	0.799	0.248	0.462	0.852	0.207
GC [10]	0.724	0.889	0.150	0.447	0.708	0.268	0.417	0.768	0.219
UFO [8]	0.813	0.934	0.147	0.550	0.804	0.232	0.494	0.836	0.170
RC [7]	0.815	0.930	0.137	0.559	0.805	0.227	0.498	0.851	0.188
BMS [35]	0.783	0.918	0.151	0.528	0.770	0.234	0.451	0.821	0.160
GS [12]	0.784	0.941	0.139	0.559	0.815	0.224	0.465	0.853	0.173
MR [13]	0.832	0.944	0.127	0.602	0.825	0.223	0.526	0.846	0.188
SO [29]	0.837	0.953	0.108	0.596	0.831	0.201	0.527	**0.881**	**0.144**
MB+ [15]	0.829	0.959	**0.107**	**0.616**	**0.845**	0.198	0.521	0.879	0.168
RRWR [14]	0.836	0.948	0.124	0.587	0.816	0.229	0.529	0.849	0.185
TLLT [30]	0.801	0.916	0.114	0.535	0.764	0.210	**0.545**	0.832	**0.144**
BSCA [33]	0.833	0.957	0.125	0.595	0.844	0.224	0.509	0.874	0.191
MAP [28]	0.820	0.951	0.127	0.586	0.823	0.225	0.522	0.870	0.177
LPS [32]	0.812	0.947	0.124	0.544	0.798	0.220	0.538	0.861	0.145
GP [47]	**0.841**	**0.961**	0.123	0.590	0.833	0.232	0.502	0.852	0.209
MST [17]	0.814	0.934	**0.097**	0.606	0.817	**0.194**	0.518	0.850	0.161
Ours	**0.846**	**0.964**	**0.097**	**0.631**	**0.846**	**0.185**	**0.571**	**0.889**	**0.115**

map S_c and closure reliability map S_R outperforms other state-of-the-art biologically inspired methods and psychology inspired method on PASCAL-S and DUT-OMRON; (4) the refined closure completeness map $\hat{S}_c$ achieves the state-of-the-art level performances in terms of F-measures and AUC on three datasets, and better MAE than the compared state-of-the-art methods; (5) the refined closure reliability map $\hat{S}_R$ have obvious superiority compared to all the state of the arts in terms of F-measures, AUC and MAE on the three datasets; (6) after fusion, the proposed method achieves the best performances.

Though border information has been used in boundary connectivity [29], the closure completeness is different from boundary connectivity [29] in two aspects. First, the assumptions are different. We assume that regions bounded by completely closed contours are salient regions while regions bounded by opened contours are background. Thus, regions in the inner of the image tend to be salient regions while regions on the image boundaries tend to be background due to the gap along the image border. In [29], Zhu et al. assume

Table II: F-measures, AUC and MAE of the compared methods. Two top contrast based methods (RC [7] and UFO [8]), one top psychology inspired method (BMS [35]) and two top background prior based methods (GP [47] and MB+ [15]) are chosen to the comparison. The best one of those methods are marked using underlines. The better performances of the proposed methods compared to the state-of-the-arts are marked using $\uparrow$.

		S_{C}	S_{R}	$\hat{S}_{\mathrm{C}}$	$\hat{S}_{\mathrm{R}}$	S	RC [7]	UFO [8]	BMS [35]	GP [47]	MB+ [15]
MSRA10K	F_β	0.791	0.805	0.837	0.843 $\uparrow$	0.846 $\uparrow$	0.815	0.813	0.783	0.841	0.829
	AUC	0.929	0.947	0.959	0.967 $\uparrow$	0.964 $\uparrow$	0.930	0.934	0.918	0.961	0.959
	MAE	0.129	0.149	0.095 $\uparrow$	0.091 $\uparrow$	0.097 $\uparrow$	0.137	0.147	0.151	0.123	0.107
PASCAL-S	F_β	0.585	0.590	0.629 $\uparrow$	0.622 $\uparrow$	0.631 $\uparrow$	0.559	0.550	0.528	0.590	0.616
	AUC	0.821	0.843	0.838	0.848 $\uparrow$	0.846 $\uparrow$	0.805	0.804	0.770	0.833	0.845
	MAE	0.208	0.218	0.185 $\uparrow$	0.190 $\uparrow$	0.185 $\uparrow$	0.227	0.232	0.234	0.232	0.198
DUT-OMRON	F_β	0.555 $\uparrow$	0.562 $\uparrow$	0.540 $\uparrow$	0.537 $\uparrow$	0.571 $\uparrow$	0.498	0.494	0.451	0.502	0.521
	AUC	0.868	0.889 $\uparrow$	0.880 $\uparrow$	0.887 $\uparrow$	0.889 $\uparrow$	0.851	0.836	0.821	0.852	0.879
	MAE	0.124 $\uparrow$	0.144 $\uparrow$	0.131 $\uparrow$	0.147 $\uparrow$	0.115 $\uparrow$	0.188	0.170	0.160	0.209	0.168

Table III: Comparisons between (1) two saliency cues, i.e., the boundary connectivity (BndCon) [29] and the closure completeness S_C, (2) the refined BndCon using the proposed refinement and the refined S_C using the same refinement, i.e., $\hat{S}_{\mathrm{C}}$, (3) the whole model SO in [29] based on the BndCon and $\hat{S}_{\mathrm{C}}$.

	MSRA10K			**PASCAL-S**			**DUT-OMRON**		
Saliency cues	F_β	AUC	MAE	F_β	AUC	MAE	F_β	AUC	MAE
BndCon [29]	0.743	0.927	0.176	0.556	0.801	0.301	0.445	0.828	0.321
S_C	0.791	0.929	0.129	0.585	0.821	0.208	0.555	0.868	0.124
Refined BndCon	0.789	0.954	0.147	0.566	0.836	0.247	0.472	0.866	0.233
$\hat{S}_{\mathrm{C}}$	0.837	0.959	0.095	0.629	0.838	0.185	0.540	0.880	0.131
SO [29]	0.837	0.953	0.108	0.596	0.831	0.201	0.527	0.881	0.144
$\hat{S}_{\mathrm{C}}$	0.837	0.959	0.095	0.629	0.838	0.185	0.540	0.880	0.131

that most photographers seldom place the salient object on image borders. Thereby, regions on the image boundaries are background and regions that are easy to connect to image boundaries are also background. Second, the computation schemes are different. We estimate the closure completeness by the expectation of times that regions are bounded by closed contours over the whole hierarchical segmentation space. In [29], the boundary connectivity is measured by the ratio of the cost that a superpixel takes to reach the image boundary to the square root of the spanning area.

To validate that the closure completeness is more robust than boundary connectivity proposed by [29], the performance comparisons between them in three aspects on three datasets are reported in Table III. First, the two saliency cues without refinement are compared. On three datasets, the closure completeness achieves better F-measures, AUCs and MAEs consistently than boundary connectivity [29]. Especially, the F-measures on MSRA10K and DUT-OMRON are improved by 4.8% and 10% respectively. Second, the two saliency cues refined using the proposed refinement are compared. On three datasets, the refined closure completeness, i.e., $\hat{S}_{\mathrm{C}}$ achieves better performances than the refined boundary connectivity, especially on F-measure and MAE. Third, the whole model SO in [29] based on the boundary connectivity and $\hat{S}_{\mathrm{C}}$ are compared. On datasets PASCAL-S and DUT-OMRON, $\hat{S}_{\mathrm{C}}$ still outperforms SO [29] while the performances of them are comparable on dataset MSRA10K.

To evaluate the running time of the two cues, we perform all the experiments on a DELL R720 server with 2 x Intel Xeon E5-2650 v2 (2.6 GHz), 256 GB memory and 16 computing cores. Both of the two saliency cues are implemented in Matlab. The time they take is of the same order of magnitude. More specifically, for an image about size of 400×300, it takes 0.99s averagely to compute the closure completeness while it takes 0.61s to compute the boundary connectivity [29]. The main reason is that the hierarchical segmentation is performed prior to the closure completeness computation. Though the running time of the closure completeness is more than that of the boundary connectivity [29], the performance improvement that the closure completeness achieves is substantial.

E. Validation of Refinement for Saliency Detection

In Section III-C, a refinement procedure is introduced to further refine the coarse saliency maps. To validate the effectiveness of the refinement, we refine four state-of-the-art biologically inspired methods (RC [7], UFO [8], PCAS [11] and GC [10]), one psychology inspired method (BMS [35]) and three background prior based methods (SO [29], GP [47] and MB+ [15]). The F-measures, AUCs and MAEs before refinement and after refinement are reported in Fig. 13, Fig. 14 and Fig. 15 respectively. For biologically inspired methods and the psychology inspired methods, their performances on three datasets are substantially improved after the refinement. On the contrary, for background prior based methods, their performance changes after using the refinement procedure are not obvious. The reasons are: (1) the region connectivity and background prior that the refinement uses are complementary to the saliency cues that the biologically inspired methods and the psychology inspired methods use. Thus, the refinement improves their performances substantially; (2) the region connectivity and background prior that the refinement uses have been exploited for saliency in the background prior based

methods. So the refinement by reusing such two priors does not make obvious changes for performances of the background prior based methods.

F. Extension to Existing Background Prior based Methods

As mentioned in Sec. III-C, the proposed framework can be easily used to extend the previous background prior based methods and further improve their performances. To validate its extendability, by combining the novel saliency cues, we extend the most popular background prior based methods: BSCA [33], DSR [27], MR [13], SO [29], MAP [28], LPS [32], and GS [12]. The visual comparisons before combinations and after combinations are shown in Fig. 16. Obviously, the saliency maps generated by the improved versions of these methods are much closer to the ground truth. The F-measures and MAEs of various saliency methods and their combination versions on three datasets are reported in Fig. 17 and 18 respectively, which demonstrate that the performances of the extension versions are improved substantially to a similar performance level. After integrating the two improved saliency maps by adding, performances on DUT-OMRON are improved further.

V. Conclusion and Future Work

This paper aims to mimic the human visual attention mechanism which treats the salient object as a whole, rather than decomposes the salient object into small superpixels or patches. It proposes a framework for the holistic salient object detection. The proposed framework thoroughly explores two novel saliency cues about closure, i.e., the closure completeness and the closure reliability to measure the saliency. The contour closure is one of the global criteria that human visual system uses for perceptual organisation [51]. The two closure based cues are also strongly supported by the psychology finding that contour closure contributes to object based attention. The proposed framework can generate saliency maps with uniformly highlighted salient objects and clear background. Its effectiveness is demonstrated by experiments on three widely used datasets. Additionally, the proposed framework can be easily used to extend the background prior based methods and improve their performances to a similar level, which is demonstrated by experiments.

As our framework is closely relative to the hierarchical segmentation, unifying the task of salient object detection and the hierarchical segmentation in a framework such that they can be beneficial to each other is our future work. Additionally, deep learning based contour closure and a learning based fusion strategy to make better use of the two maps are suggested in the future.

Acknowledgment

Q. Liu is partially supported by the scholarship from China Scholarship Council (CSC). B. Zou and Z. Chen are partially supported by the National Natural Science Foundation of China under Grant No.61573380 and No. 61672542. X. Hong, J. Chen and G. Zhao are partially supported by Academy of Finland, Tekes Fidipro Program and Infotech Oulu. Xiaopeng Hong is also partially supported by the National Natural Science Foundation of China under Grant No. 61572205.

References

[1] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, "Saliency driven total variation segmentation," in *ICCV*. IEEE, 2009, pp. 817–824.
[2] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *TIP*, vol. 19, no. 1, pp. 185–198, 2010.
[3] S. Frintrop, G. M. Garcia, and A. B. Cremers, "A cognitive approach for object discovery," in *ICPR*. IEEE, 2014, pp. 2329–2334.
[4] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *TPAMI*, no. 11, pp. 1254–1259, 1998.
[5] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *CVPR*. IEEE, 2013, pp. 1155–1162.
[6] S. Frintrop, T. Werner, and G. M. García, "Traditional saliency reloaded: A good old model in new shape," in *CVPR*, 2015, pp. 82–90.
[7] M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. Hu, "Global contrast based salient region detection," *TPAMI*, vol. 37, no. 3, pp. 569–582, 2015.
[8] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *ICCV*, 2013, pp. 1976–1983.
[9] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*. IEEE, 2012, pp. 733–740.
[10] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *ICCV*, 2013, pp. 1529–1536.
[11] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *CVPR*, 2013, pp. 1139–1146.
[12] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *ECCV*. Springer, 2012, pp. 29–42.
[13] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *CVPR*, 2013, pp. 3166–3173.
[14] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. Dagan Feng, "Robust saliency detection via regularized random walks ranking," in *CVPR*, 2015, pp. 2710–2717.
[15] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, "Minimum barrier salient object detection at 80 fps," in *ICCV*, 2015, pp. 1404–1412.
[16] Q. Wang, W. Zheng, and R. Piramuthu, "Grab: Visual saliency via novel graph model and background priors," in *CVPR*, 2016, pp. 535–543.
[17] W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien, "Real-time salient object detection with a minimum spanning tree," in *CVPR*, 2016, pp. 2334–2342.
[18] I. Kovacs and B. Julesz, "A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation," *Proceedings of the National Academy of Sciences*, vol. 90, no. 16, pp. 7495–7497, 1993.
[19] K. Koffka, *Principles of Gestalt psychology*. Routledge, 2013, vol. 44.
[20] A. C. Marino and B. J. Scholl, "The role of closure in defining the objects of object-based attention," *Perception & Psychophysics*, vol. 67, no. 7, pp. 1140–1149, 2005.
[21] I. Kovacs and B. Julesz, "Perceptual sensitivity maps within globally defined visual shapes," *Nature*, vol. 370, no. 6491, pp. 644–646, 1994.
[22] J. H. Elder and S. W. Zucker, "Evidence for boundary-specific grouping," *Vision Research*, vol. 38, no. 1, pp. 143–152, 1998.
[23] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," in *NIPS*, 2008, pp. 497–504.
[24] D. Klein, S. Frintrop *et al.*, "Center-surround divergence of feature statistics for salient object detection," in *ICCV*. IEEE, 2011, pp. 2214–2219.
[25] N. Liu and J. Han, "Dhsnet: Deep hierarchical saliency network for salient object detection," in *CVPR*, 2016, pp. 678–686.
[26] G. Li and Y. Yu, "Deep contrast learning for salient object detection," pp. 478–487, 2016.
[27] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *CVPR*, 2013, pp. 2976–2983.
[28] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on markov absorption probabilities," *TIP*, vol. 24, no. 5, pp. 1639–1649, 2015.
[29] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *CVPR*, 2014, pp. 2814–2821.
[30] C. Gong, D. Tao, W. Liu, S. J. Maybank, M. Fang, K. Fu, and J. Yang, "Saliency propagation from simple to difficult," in *CVPR*, 2015, pp. 2531–2539.

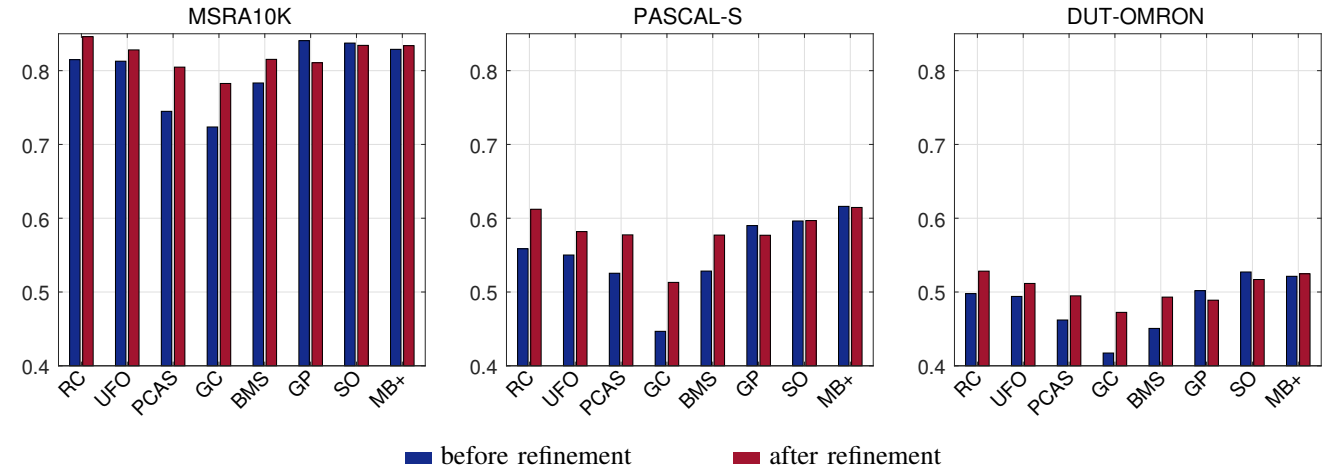


Fig. 13: F-measures of previous methods before refinement and after refinement on three datasets. RC [7], UFO [8], PCAS [11] and GC [10] are state-of-the-art biologically inspired methods. BMS [35] is psychology inspired. SO [29], GP [47] and MB+ [15] are background prior based methods. (better view in colour).

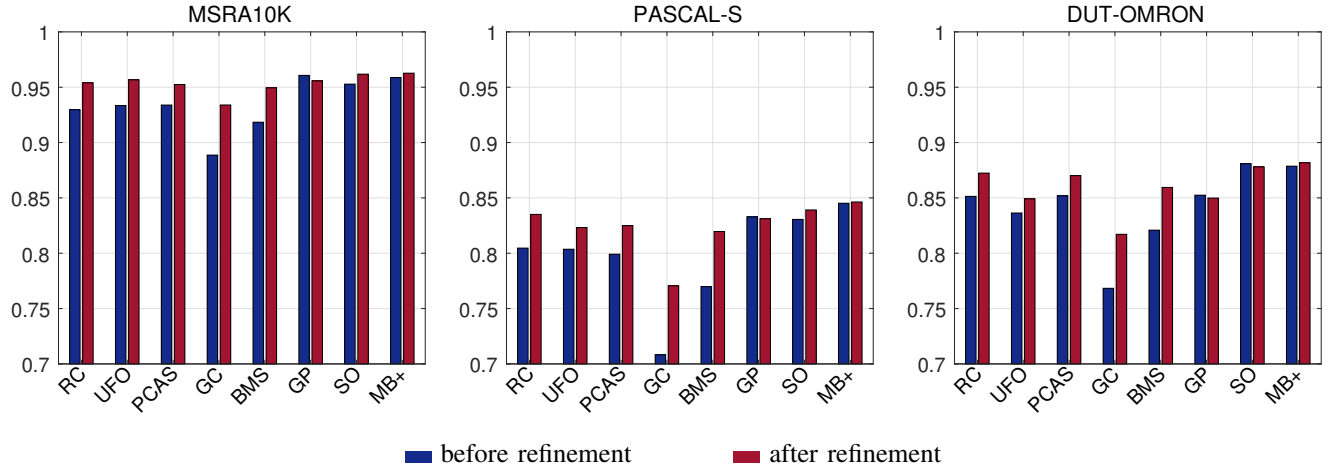


Fig. 14: AUCs of previous methods before refinement and after refinement on three datasets. (better view in colour).

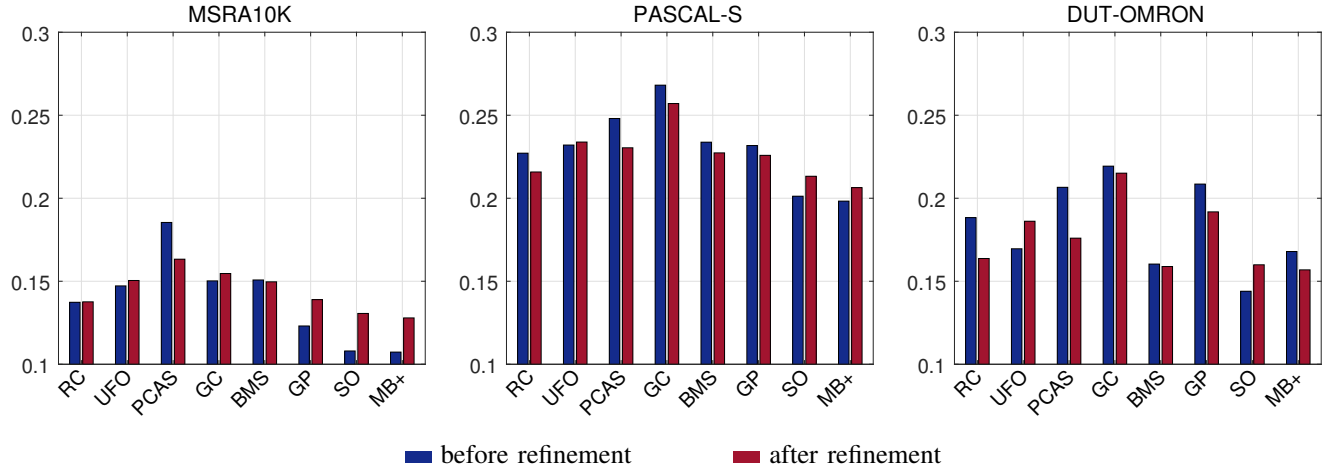


Fig. 15: MAEs of previous methods before refinement and after refinement on three datasets. (better view in colour).

[31] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *TCSVT*, vol. 25, no. 8, pp. 1309–1321, 2015.

[32] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation: Salient object detection in the wild," *TIP*, vol. 24, no. 10, pp. 3176–3186, 2015.

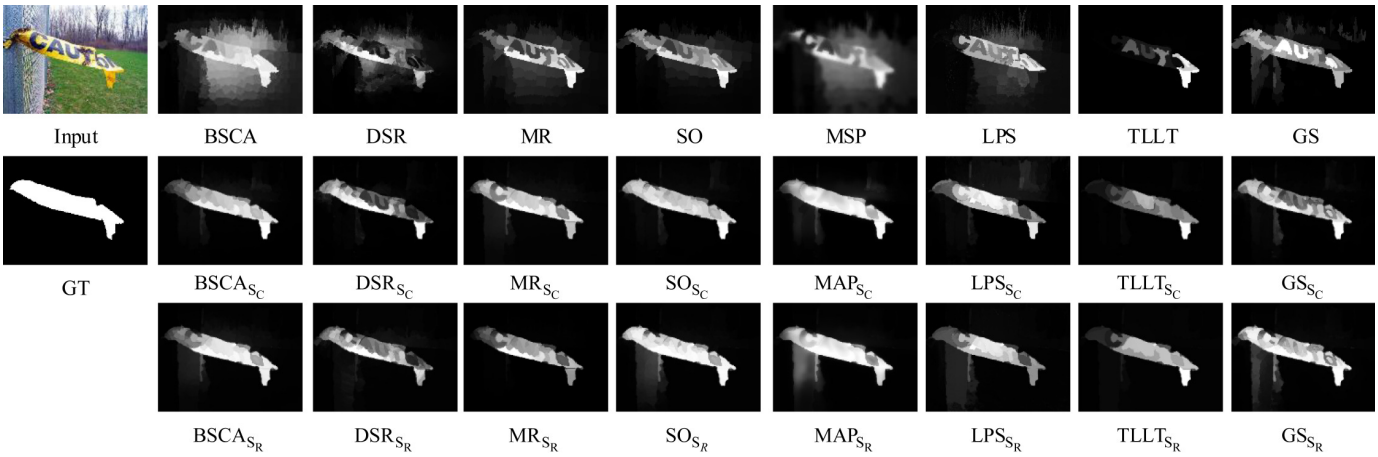


Fig. 16: Visual comparisons before combination and after combination. The first row: saliency maps generated by BSCA [33], DSR [27], MR [13], SO [29], MAP [28], LPS [32], TLLT [30], GS [12]. The last two rows: saliency maps after combining with S_C and S_R respectively. The figure shows that the existing background prior based methods after combining with our saliency cues are much closer to the ground truth.

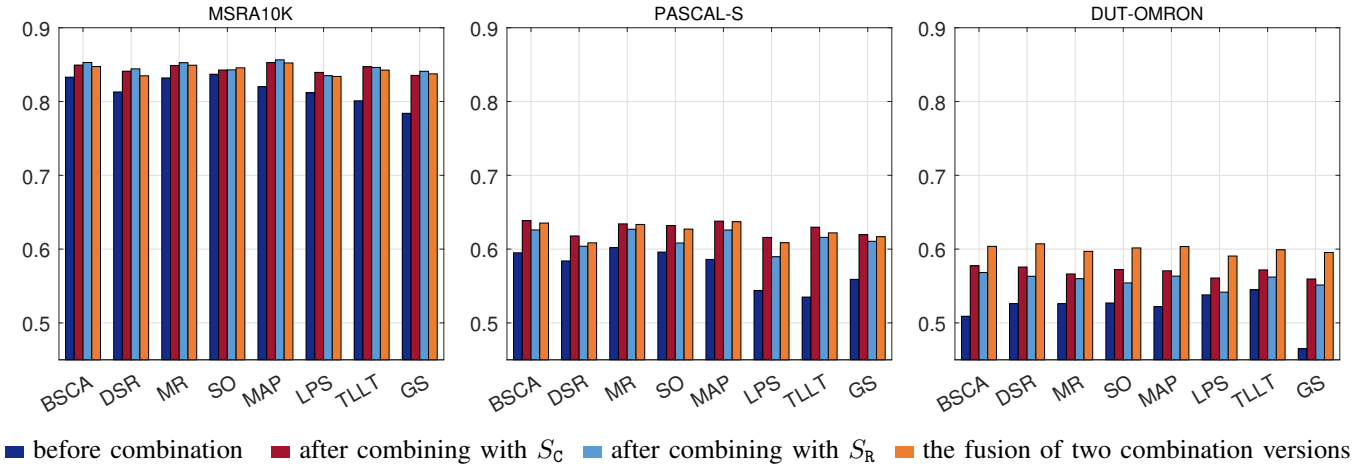


Fig. 17: F measures of background prior based methods before combination and after combination on three datasets (better view in colour).

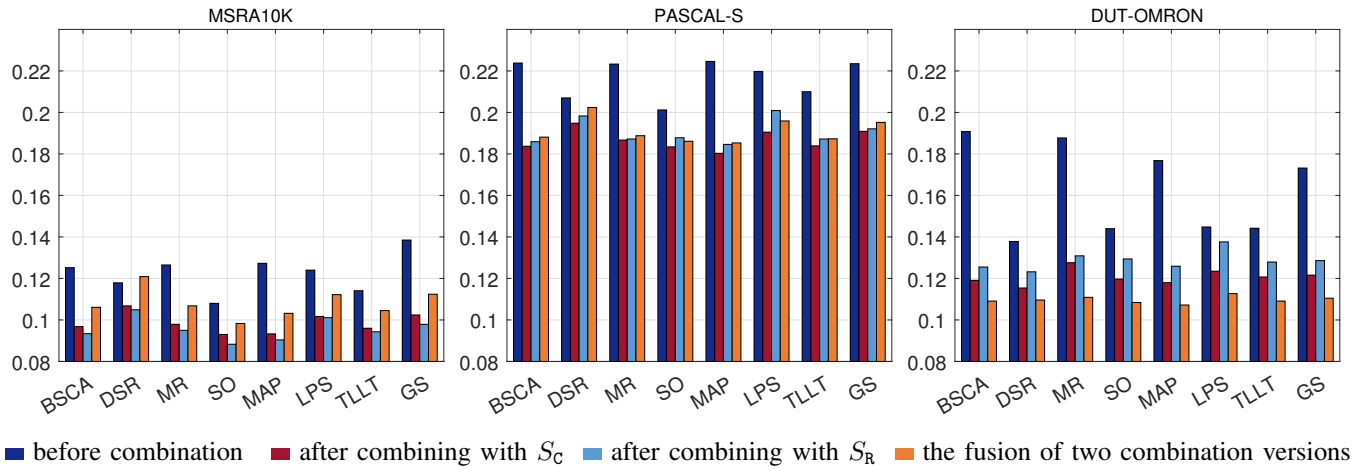


Fig. 18: MAEs of background prior based methods before combination and after combination on three datasets (better view in colour).

[33] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via cellular automata," in *CVPR*, 2015, pp. 110–119.
[34] Y. Lu, W. Zhang, H. Lu, and X. Xue, "Salient object detection using concavity context," in *ICCV*. IEEE, 2011, pp. 233–240.
[35] J. Zhang and S. Sclaroff, "Saliency detection: A boolean map approach," in *ICCV*, 2013, pp. 153–160.
[36] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *CVPR*. IEEE, 2013, pp. 2083–2090.
[37] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille, "The secrets of salient object segmentation," in *CVPR*, 2014, pp. 280–287.
[38] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *TPAMI*, vol. 34, no. 11, pp. 2189–2202, 2012.
[39] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr, "Bing: Binarized normed gradients for objectness estimation at 300fps," in *CVPR*, 2014, pp. 3286–3293.
[40] Y. Ming, H. Li, and X. He, "Connected contours: A new contour completion model that respects the closure effect," in *CVPR*, 2012, pp. 829–836.
[41] J. S. Stahl and S. Wang, "Edge grouping combining boundary and region information," *TIP*, vol. 16, no. 10, pp. 2590–2606, 2007.
[42] A. Levinshtein, C. Sminchisescu, and S. Dickinson, "Optimal contour closure by superpixel grouping," in *ECCV*. Springer, 2010, pp. 480–493.
[43] P. Arbelaez, "Boundary extraction in natural images using ultrametric contour maps," in *CVPR Workshop, 2006.* IEEE, 2006, pp. 182–182.
[44] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *TPAMI*, vol. 18, no. 12, pp. 1163–1173, 1996.
[45] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *TPAMI*, vol. 33, no. 5, pp. 898–916, 2011.
[46] P. Dollár and C. L. Zitnick, "Structured forests for fast edge detection," in *ICCV*. IEEE, 2013, pp. 1841–1848.
[47] P. Jiang, N. Vasconcelos, and J. Peng, "Generic promotion of diffusion-based salient object detection," in *ICCV*, 2015, pp. 217–225.
[48] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *TIP*, vol. 24, no. 12, pp. 5706–5722, 2015.
[49] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *CVPR*. IEEE, 2009, pp. 1597–1604.
[50] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *TPAMI*, vol. 33, no. 2, pp. 353–367, 2011.
[51] J. H. Elder and R. M. Goldberg, "Ecological statistics of gestalt laws for the perceptual organization of contours," *Journal of Vision*, vol. 2, no. 4, pp. 5–5, 2002.

Qing Liu received the bachelor degree in computer science and technology from the Central South University, Changsha, China, in 2011. She is currently a PhD candidate in Central South University. Her research interests include salient object detection and medical image analysis.

Dr. Xiaopeng Hong received his BEng and Ph.D. degree in computer application from Harbin Institute of Technology, Harbin, P. R. China, in 2004 and 2010 respectively. He has been a senior researcher in the Center for Machine Vision and Signal Analysis (CMVS), Faculty of Information Technology and Electrical Engineering, the University of Oulu, Finland since 2011. He has authored or co-authored more than 30 peer-reviewed articles in international journals and conferences, and has served as a reviewer for tens of journals and conferences. His current research interests include affective computing, saliency estimation, deep learning, pose and gaze estimation, texture classification, visual surveillance, and visual speech recognition.

Dr. Beiji Zou received the B.S., M.S., and Ph.D degrees from Zhejiang University in 1982, Qinghua University in 1984 and Hunan University in 2001 respectively. He is currently a Professor and served as the dean at the School of Information Science and Engineering at Central South University. His research interests include computer graphics and image processing.

Dr. Jie Chen received the M.S. and Ph.D. degrees from Harbin Institute of Technology, China, in 2002 and 2007, respectively. Since September 2007, he has been a senior researcher in the Machine Vision Group at the University of Oulu, Finland. In 2012 and 2015, he visited the Computer Vision Laboratory at the University of Maryland and School of Electrical and Computer Engineering at the Duke University respectively. Dr. Chen was a cochair of International Workshops at ACCV2014, CVPR2016, and ICCV2017. He was a guest editor of the special issue RoLoD: Robust local descriptors for computer vision for the journal of Neurocomputing. His research interests include pattern recognition, computer vision, machine learning, dynamic texture, deep learning, medical image analysis and watermarking. He is an associate editor of The Visual Computer. He has authored more than 50 papers in journals and conferences, and he is a member of the IEEE.

Zailiang Chen received the Ph.D. degree in computer science from Central South University in 2012. He is currently an Associate Professor with Central South University and the Associate Director of the Center for Ophthalmic Imaging Research of Central South University. In 2014, he was a visiting scholar at Stevens Institute of Technology, New Jersey, USA. He has authored or co-authored more than 30 papers in journals and conferences. His recent research interests include computer vision, medical image analysis and large-scale medical image processing.

Guoying Zhao (SM12) received the Ph.D. degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005. She is currently an Associate Professor with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland, where she has been a senior researcher since 2005. In 2011, she was selected to the highly competitive Academy Research Fellow position. She has authored or co-authored more than 160 papers in journals and conferences. Her papers have currently over 5700 citations in Google Scholar (h-index 33).She has served as area chairs for FG 2017 and WACV 2017 and is associate editor for Pattern Recognition and Image and Vision Computing Journals. She has lectured tutorials at ICPR 2006, ICCV 2009, and SCIA 2013, authored/edited three books and six special issues in journals. Dr. Zhao was a Co-Chair of 12 International Workshops at ECCV, ICCV, CVPR and ACCV, and two special sessions at FG13 and FG15. Her current research interests include image and video descriptors, facial-expression and micro-expression recognition, gait analysis, dynamic-texture recognition, human motion analysis, and person identification. Her research has been reported by Finnish TV programs, newspapers and MIT Technology Review.