# Saliency Optimization from Robust Background Detection

Wangjiang Zhu*
Tsinghua University
wangjiang88119@gmail.com

Shuang Liang†
Tongji University
shuangliang@tongji.edu.cn

Yichen Wei, Jian Sun
Microsoft Research
{yichenw, jiansun}@microsoft.com

## Abstract

*Recent progresses in salient object detection have exploited the boundary prior, or background information, to assist other saliency cues such as contrast, achieving state-of-the-art results. However, their usage of boundary prior is very simple, fragile, and the integration with other cues is mostly heuristic. In this work, we present new methods to address these issues. First, we propose a robust background measure, called* boundary connectivity. *It characterizes the spatial layout of image regions with respect to image boundaries and is much more robust. It has an intuitive geometrical interpretation and presents unique benefits that are absent in previous saliency measures. Second, we propose a principled optimization framework to integrate multiple low level cues, including our background measure, to obtain clean and uniform saliency maps. Our formulation is intuitive, efficient and achieves state-of-the-art results on several benchmark datasets.*

## 1. Introduction

Recent years have witnessed rapidly increasing interest in salient object detection [2]. It is motivated by the importance of saliency detection in applications such as object aware image retargeting [5, 11], image cropping [13] and object segmentation [20]. Due to the absence of high level knowledge, all bottom up methods rely on assumptions about the properties of objects and backgrounds. The most widely utilized assumption is that appearance contrasts between objects and their surrounding regions are high. This is called *contrast prior* and is used in almost all saliency methods [25, 19, 22, 6, 16, 14, 17, 26, 28, 8, 9, 29, 3].

Besides contrast prior, several recent approaches [26, 8, 29] exploit *boundary prior* [26], *i.e.*, image boundary regions are mostly backgrounds, to enhance saliency computation. Such methods achieve state-of-the-art results, suggesting that boundary prior is effective. However, we observe two drawbacks. The first is they simply treat all image boundary as background. This is fragile and may fail even when the object only slightly touches the boundary. The second is their usage of boundary prior is mostly heuristic. It is unclear how it should be integrated with other cues for saliency computation.

This work presents new methods to address the above two problems. Our first contribution is a novel and reliable background measure, called *boundary connectivity*. Instead of assuming the image boundary is background [8, 29], or an image patch is background if it can easily be connected to the image boundary [26], the proposed measure states that *an image patch is background only when the region it belongs to is heavily connected to the image boundary*. This measure is more robust as it characterizes the spatial layout of image regions with respect to image boundaries. In fact, it has an intuitive geometrical interpretation and thus is stable with respect to image content variations. This property provides unique benefits that are absent in previously used saliency measures. For instance, boundary connectivity has similar distributions of values across images and are directly comparable. It can detect the background at a high precision with decent recall using a single threshold. It naturally handles purely background images without objects. Specifically, it can significantly enhance traditional contrast computation. We describe and discuss this in Section 3.

It is well known that the integration of multiple low level cues can produce better results. Yet, this is usually done in heuristic ways [25, 17, 2, 28], *e.g.*, weighted summation or multiplication. Our second contribution is a principled framework that regards saliency estimation as a global optimization problem. The cost function is defined to directly achieve the goal of salient object detection: object regions are constrained to take high saliency using foreground cues; background regions are constrained to take low saliency using the proposed background measure; a smoothness constraint ensures that the saliency map is uniform on flat regions. All constraints are in linear form and the optimal saliency map is solved by efficient least-square optimization. Our optimization framework combines low level cues in an intuitive, straightforward and efficient manner.

*This work was done while Wangjiang Zhu was an intern at Microsoft Research Asia.

†Corresponding author.

1063-6919/14 $31.00 © 2014 IEEE
DOI 10.1109/CVPR.2014.360



IEEE computer society

This makes it fundamentally different from complex CRF/MRF optimization methods that combine multiple saliency maps [15, 28], or those adapted from other optimization problems [23, 29, 9]. Section 4 describes our optimization method.

In Section 5, extensive comparisons on several benchmark datasets and superior experimental results verify the effectiveness of the proposed approach.

## 2. Related Work

Another research direction for visual saliency analysis [12, 7, 4, 27, 24, 21] aims to predict human visual attention areas. Such works are more inspired by biological visual models and are evaluated on sparse human eye fixation data instead of object/background labelings. We do not discuss such works due to these differences. In the following we briefly review previous works from the two viewpoints of interest in this paper: the usage of boundary prior and optimization methods for salient object detection.

Some early works use the so called *center prior* to bias the image center region with higher saliency. Usually, center prior is realized as a gaussian fall-off map. It is either directly combined with other cues as weights [25, 28, 3], or used as a feature in learning-based methods [24, 8]. This makes strict assumptions about the object size and location in the image. From an opposite perspective, recent works [26, 8, 29] introduce *boundary prior* and treat image boundary regions as background. In [8], the contrast against the image boundary is used as a feature in learning. In [29], saliency estimation is formulated as a ranking and retrieval problem and the boundary patches are used as background queries. In [26], an image patch's saliency is defined as the shortest-path distance to the image boundary, observing that background regions can easily be connected to the image boundary while foreground regions cannot. These approaches work better for off-center objects but are still fragile and can fail even when an object only slightly touches the boundary[1]. In contrast, the proposed new method takes more spatial layout characteristics of background regions into consideration and is therefore more robust.

Most methods implement and combine low level cues heuristically. Recently, a few approaches have adopted more principled global optimization. In [15], multiple saliency maps from different methods are aggregated into a better one. Similarly, in [28], saliency maps computed on multiple scales of image segmentation are combined. These methods adopt a complex CRF/MRF formulation and the process is usually slow. The work in [23] treats salient objects as sparse noises and solves a low rank matrix recovery problem instead. The work in [29] ranks the similarity of image patches via graph-based manifold ranking. The work in [9] models salient region selection as the facility location problem and maximizes the sub-modular objective function. These methods adapt viewpoints and optimization techniques from other problems for saliency estimation. Unlike all the aforementioned methods, our optimization directly integrates low level cues in an intuitive and effective manner.

[1] A simple "1D-saliency" method is proposed in [26] to alleviate this problem, but it is highly heuristic and not robust. See [26] for more details.

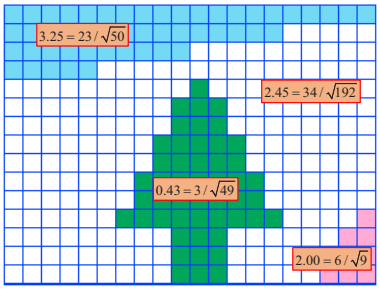



Figure 1. (Better viewed in color) An illustrative example of boundary connectivity. The synthetic image consists of four regions with their boundary connectivity values (Eq.(1)) overlaid. The boundary connectivity is large for background regions and small for object regions.

## 3. Boundary Connectivity: a Robust Background Measure

We first derive our new background measure from a conceptual perspective and then describe an effective computation method. We further discuss the unique benefits originating from its intuitive geometrical interpretation.

### 3.1. Conceptual Definition

We observe that object and background regions in natural images are quite different in their spatial layout, *i.e.*, *object regions are much less connected to image boundaries than background ones*. This is exemplified in Figure 1. The synthetic image consists of four regions. From human perception, the green region is clearly a salient object as it is large, compact and only slightly touches the image boundary. The blue and white regions are clearly backgrounds as they significantly touch the image boundary. Only a small amount of the pink region touches the image boundary, but as its size is also small it looks more like a partially cropped object, and therefore is not salient.

We propose a measure to quantify how heavily a region $R$ is connected to the image boundaries, called *boundary connectivity*. It is defined as

$$BndCon(R) = \frac{|\{p|p \in R, p \in Bnd\}|}{\sqrt{|\{p|p \in R\}|}} \qquad (1)$$

where $Bnd$ is the set of image boundary patches and $p$ is an image patch. It has an intuitive geometrical interpreta-

tion: *it is the ratio of a region's perimeter on the boundary to the region's overall perimeter, or square root of its area*. Note that we use the square root of the area to achieve scale-invariance: the measure remains stable across different image patch resolutions. As illustrated in Figure 1, the boundary connectivity is usually large for background regions and small for object regions.

### 3.2. Effective Computation

The definition in Eq.(1) is intuitive but difficult to compute because image segmentation itself is a challenging and unsolved problem. Using a hard segmentation not only involves the difficult problem of algorithm/parameter selection, but also introduces undesirable discontinuous artifacts along the region boundaries.

We point out that an accurate hard image segmentation is unnecessary. Instead, we propose a "soft" approach. The image is first abstracted as a set of nearly regular superpixels using the SLIC method [18]. Empirically, we find 200 superpixels are enough for a typical $300*400$ resolution image. Superpixel result examples are shown in Figure 5(a).

We then construct an undirected weighted graph by connecting all adjacent superpixels $(p, q)$ and assigning their weight $d_{app}(p, q)$ as the Euclidean distance between their average colors in the CIE-Lab color space. The geodesic distance between any two superpixels $d_{geo}(p, q)$ is defined as the accumulated edge weights along their shortest path on the graph

$$d_{geo}(p, q) = \min_{p_1=p, p_2, \ldots, p_n=q} \sum_{i=1}^{n-1} d_{app}(p_i, p_{i+1}) \quad (2)$$

For convenience we define $d_{geo}(p, p) = 0$. Then we define the "spanning area" of each superpixel $p$ as

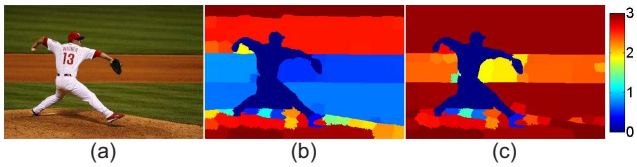
$$Area(p) = \sum_{i=1}^{N} exp(-\frac{d_{geo}^2(p, p_i)}{2\sigma_{clr}^2}) = \sum_{i=1}^{N} S(p, p_i), \quad (3)$$

where $N$ is the number of superpixels.

Eq.(3) computes a soft area of the region that $p$ belongs to. To see that, we note the operand $S(p, p_i)$ in the summation is in $(0, 1]$ and characterizes how much superpixel $p_i$ contributes to $p$'s area. When $p_i$ and $p$ are in a flat region, $d_{geo}(p, p_i) = 0$ and $S(p, p_i) = 1$, ensuring that $p_i$ adds a unit area to the area of $p$. When $p_i$ and $p$ are in different regions, there exists at least one strong edge ($d_{app}(*, *) \gg 3\sigma_{clr}$) on their shortest path and $S(p, p_i) \approx 0$, ensuring that $p_i$ does not contribute to $p$'s area. Experimentally, we find that the performance is stable when parameter $\sigma_{clr}$ is within $[5, 15]$. We set $\sigma_{clr} = 10$ in the experiments.

Similarly, we define the *length along the boundary* as

$$Len_{bnd}(p) = \sum_{i=1}^{N} S(p, p_i) \cdot \delta(p_i \in Bnd) \quad (4)$$




(a) (b) (c)

Figure 2. (Better viewed in color) Enhancement by connecting image boundaries: (a) input image; (b) boundary connectivity without linking boundary patches; (c) improved boundary connectivity by linking boundary patches.

where $\delta(\cdot)$ is 1 for superpixels on the image boundary and 0 otherwise.

Finally we compute the boundary connectivity in a similar spirit as in Eq.(1),

$$BndCon(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}} \quad (5)$$

We further add edges between any two boundary superpixels. It enlarges the boundary connectivity values of background regions and has little effect on the object regions. This is useful when a physically connected background region is separated due to occlusion of foreground objects, as illustrated in Figure 2.

To compute Eq.(5), the shortest paths between all superpixel pairs are efficiently calculated using Johnson's algorithm [10] as our graph is very sparse. For 200 superpixels, this takes less than 0.05 seconds.

### 3.3. Unique Benefits

The clear geometrical interpretation makes boundary connectivity robust to image appearance variations and stable across different images. To show this, we plot the distributions of this measure on four benchmarks on ground truth object and background regions, respectively, in Figure 3. This clearly shows that the distribution is stable across different benchmarks. The objects and backgrounds are clearly separated. Most background superpixels have values $> 1$ and most object superpixels have values close to 0.

This property provides unique benefits that are absent in previous works. As shown in Table 1, when using a single threshold of 2, the proposed measure can detect backgrounds with very high precision and decent recall on all datasets. By contrast, previous saliency measures are incapable of achieving such good uniformity, since they are usually more sensitive to image appearance variations and vary significantly across images. The absolute value of previous saliency measures is therefore much less meaningful.

Moreover, an interesting result is that our measure can naturally handle pure background images, while previous methods cannot, as exemplified in Figure 4.

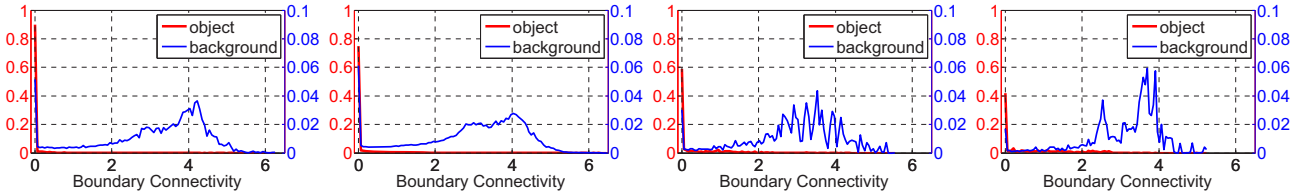



Figure 3. (Better viewed in color) The distribution of boundary connectivity of ground truth object and background regions on four benchmarks. From left to right: ASD [19], MSRA [25], SED1 [1] and SED2 [1]. Note that we use different y-axis scales for object and background for better visualization.

| Benchmark | Boundary Connectivity | | Geodesic Saliency | |
|---|---|---|---|---|
| | Precision | Recall | Precision | Recall |
| ASD [19] | 99.7% | 80.7% | 99.7% | 57.4% |
| MSRA [25] | 98.3% | 77.3% | 98.3% | 63.6% |
| SED1 [1] | 97.4% | 81.4% | 96.5% | 69.6% |
| SED2 [1] | 95.8% | 88.4% | 94.7% | 65.7% |

Table 1. Background precision/recall for superpixels with boundary connectivity $> 2$ on four benchmarks. For comparison, we treat geodesic saliency [26] as a background measure and show its recall at the same precision. Note, on SED1 and SED2, we cannot obtain the same high precision, so the max precision is given.

**Background Weighted Contrast** This highly reliable background measure provides useful information for saliency estimation. Specifically, we show that it can greatly enhance the traditional contrast computation.

Many works use the region contrast against its surroundings as a saliency cue, which is computed as the summation of its appearance distance to all other regions, weighted by their spatial distances [22, 16, 17, 28]. In this fashion, a superpixel's contrast in our notation can be written as

$$Ctr(p) = \sum_{i=1}^{N} d_{app}(p, p_i) w_{spa}(p, p_i) \tag{6}$$

where $w_{spa}(p, p_i) = exp(-\frac{d^2_{spa}(p,p_i)}{2\sigma^2_{spa}})$. $d_{spa}(p, p_i)$ is the distance between the centers of superpixel $p$ and $p_i$, and $\sigma_{spa} = 0.25$ as in [17].

We extend Eq. (6) by introducing a background probability $w_i^{bg}$ as a new weighting term. The probability $w_i^{bg}$ is mapped from the boundary connectivity value of superpixel $p_i$. It is close to 1 when boundary connectivity is large, and close to 0 when it is small. The definition is

$$w_i^{bg} = 1 - exp(-\frac{BndCon^2(p_i)}{2\sigma^2_{bndCon}}) \tag{7}$$

We empirically set $\sigma_{bndCon} = 1$. Our results are insensitive to this parameter when $\sigma_{bndCon} \in [0.5, 2.5]$.

The enhanced contrast, called *background weighted contrast*, is defined as

$$wCtr(p) = \sum_{i=1}^{N} d_{app}(p, p_i) w_{spa}(p, p_i) w_i^{bg} \tag{8}$$




Figure 4. (Better viewed in color) A pure background image case. (a) input image. (b) result of one of the state-of-the-art methods [29]. It is hard to tell whether the detected salient regions are really salient. (c) boundary connectivity, clearly suggesting that there is no object as all values $> 2$.

According to Eq.(8), the object regions receive high $w_i^{bg}$ from the background regions and their contrast is enhanced. On the contrary, the background regions receive small $w_i^{bg}$ from the object regions and the contrast is attenuated. This asymmetrical behavior effectively enlarges the contrast difference between the object and background regions. Such improvement is clearly observed in Figure 5. The original contrast map (Eq.(6) and Figure 5(b)) is messy due to complex backgrounds. With the background probability map as weights (Figure 5(c)), the enhanced contrast map clearly separates the object from the background (Figure 5(d)). We point out that, this is only possible with our highly reliable background detection.

The background probability in Eq.(7) and the enhanced contrast in Eq.(8) are complementary as they characterize the background and the object regions, respectively. Yet, both are still bumpy and noisy. In the next section, we present a principled framework to integrate these measures and generate the final clean saliency map, as in Figure 5(e).

## 4. Saliency Optimization

To combine multiple saliency cues or measures, previous works simply use weighted summation or multiplication. This is heuristic and hard for generalization. Also, although the ideal output of salient object detection is a clean binary object/background segmentation, such as the widely used ground truth in performance evaluation, most previous methods were not explicitly developed towards this goal.

In this work, we propose a principled framework that intuitively integrates low level cues and directly aims for this




Figure 5. The pipeline of our method. (a) input images with superpixel boundaries overlaid. (b) contrast maps using Eq.(6). Note that certain background regions have higher contrast than object regions. (c) background probability weight in Eq.(7); (d) background weighted contrast using Eq.(8). The object regions are more highlighted. (e) optimized saliency maps by minimizing Eq.(9). (f) ground truth.

goal. We model the salient object detection problem as the optimization of the saliency values of all image superpixels. The objective cost function is designed to assign the object region value 1 and the background region value 0, respectively. The optimal saliency map is then obtained by minimizing the cost function.

Let the saliency values of $N$ superpixels be $\{s_i\}_{i=1}^N$. Our cost function is thus defined as

$$\underbrace{\sum_{i=1}^{N} w_i^{bg} s_i^2}_{\text{background}} + \underbrace{\sum_{i=1}^{N} w_i^{fg} (s_i - 1)^2}_{\text{foreground}} + \underbrace{\sum_{i,j} w_{ij} (s_i - s_j)^2}_{\text{smoothness}} \quad (9)$$

The three terms define costs from different constraints. The *background* term encourages a superpixel $p_i$ with large background probability $w_i^{bg}$ (Eq. (7)) to take a small value $s_i$ (close to 0). As stated above, $w_i^{bg}$ is of high accuracy derived from our reliable and stable background detection.

Similarly, the *foreground* term encourages a superpixel $p_i$ with large foreground probability $w_i^{fg}$ to take a large value $s_i$ (close to 1). Note that for $w_i^{fg}$ we can essentially use any meaningful saliency measure or a combination of them. In Figure 8, we compare several state-of-the-art methods as well as the background weighted contrast in Eq.(8) as a simple baseline (all normalized to $[0, 1]$ for each image). Surprisingly we found out that although those measures have very different accuracies, after optimization they all improve significantly, and to a similar accuracy level. This is due to our proposed background measure and the optimization framework.

The last *smoothness* term encourages continuous saliency values. For every adjacent superpixel pair $(i, j)$, the weight $w_{ij}$ is defined as

$$w_{ij} = exp(-\frac{d_{app}^2(p_i, p_j)}{2\sigma_{clr}^2}) + \mu \quad (10)$$

It is large in flat regions and small at region boundaries. Note that $\sigma_{clr}$ is defined in Eq.(3). The parameter $\mu$ is a small constant (empirically set to 0.1) to regularize the optimization in cluttered image regions. It is useful to erase small noise in both background and foreground terms.

The three terms are all squared errors and the optimal saliency map is computed by least-square. The optimization takes 3 millisecond for 200 superpixels in our tests. This is much more efficient than previous CRF/MRF based optimization methods [25, 15, 28]. Figure 5 shows the optimized results.

## 5. Experiments

We use the standard benchmark datasets: ASD [19], MSRA [25], SED1 [1] and SED2 [1]. ASD [19] is widely used in almost all methods and is relatively simple. The other three datasets are more challenging. MSRA [25] contains many images with complex backgrounds and low contrast objects. SED1 and SED2 [1] contain objects of largely different sizes and locations. Note that we obtain the pixel-wise labeling of the MSRA dataset from [8].

For performance evaluation, we use standard precision-recall curves (PR curves). A curve is obtained by normalizing the saliency map to $[0, 255]$, generating binary masks with a threshold sliding from 0 to 255, and comparing the binary masks against the ground truth. The curves are then averaged on each dataset.

Although commonly used, PR curves are limited in that they only consider whether the object saliency is higher than the background saliency. Therefore, we also introduce the mean absolute error (MAE) into the evaluation. It is the average per-pixel difference between the binary ground truth and the saliency map, normalized to $[0, 1]$. It directly measures how close a saliency map is to the ground truth and is more meaningful for applications such as object segmentation or cropping. This measure is also used in recent meth-

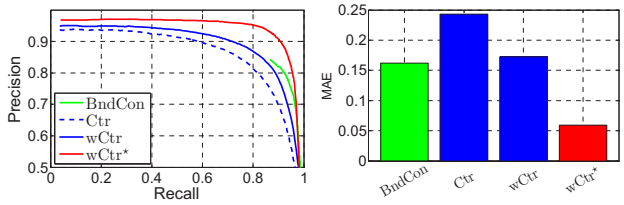



Figure 6. Comparison of PR curves (left) and MAE (right) on ASD [19] dataset. Note that we use $wCtr^*$ to denote the optimized version of $wCtr$ using Eq.( 9).

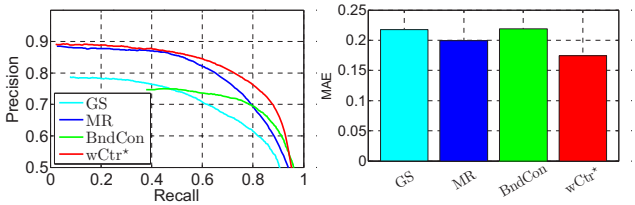



Figure 7. PR curves (left) and MAE (right) on MSRA-hard dataset.

| Method | SF [17] | GS [26] | HS [28] | MR [29] | SIA [3] | Ours |
|---|---|---|---|---|---|---|
| Time | 0.16 | 0.21 | 0.59 | 0.25 | 0.09 | 0.25 |
| Code | C++ | Matlab | C++ | Matlab | C++ | Matlab |

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

ods [17, 3] and found complementary to PR curves.

We compare with the most recent state-of-the-art methods, including saliency filter(SF) [17], geodesic saliency(GS-SP, short for GS) [26], soft image abstraction(SIA) [3], hierarchical saliency(HS) [28] and manifold ranking(MR) [29]. Among these, SF [17] and SIA [3] combine low level cues in straightforward ways; GS [26] and MR [29] use boundary prior; HS [28] and MR [29] use global optimization and are the best algorithms up to now. There are many other methods, and their results are mostly inferior to the aforementioned methods. The code for our algorithm and other algorithms we implement is all available online.

**Validation of the proposed approach** To verify the effectiveness of the proposed boundary connectivity measure and saliency optimization, we use the standard dataset ASD. Results in Figure 6 show that 1) boundary connectivity already achieves decent accuracy[2]; 2) background weighted contrast (Eq.(8)) is much better than the traditional one (Eq.(6)); 3) optimization significantly improves the previous two cues. Similar conclusions are also observed on other datasets but omitted here for brevity.

To show the robustness of boundary connectivity, we compare with two methods that also use boundary prior (GS [26] and MR [29]). We created a subset of 657 images from MSRA [25], called MSRA-hard, where objects touch the image boundaries. Results in Figure 7 show 1) boundary connectivity already exceeds GS [26]; 2) the optimized result is significantly better than MR [29].

**Integration and comparison with state-of-the-art** As mentioned in Section 4, our optimization framework can integrate any saliency measure as the foreground term. Figure 8 reports both PR curves and MAEs for various saliency methods on four datasets, with before and after optimization compared. Both PR curves and MAEs show that all methods are significantly improved to a similar performance level. The big improvements clearly verify that the proposed background measure and optimization is highly effective. Especially, we find that our weighted contrast (Eq.(8)) can lead to performances comparable to using other sophisticated saliency measures, such as [28, 29]. This is very meaningful when the simplicity and efficiency of the weighted contrast is considered.

Example results of previous methods (no optimization) and our optimization using background weighted contrast are shown in Figure 9.

**Running time** In Table 2, we compare average running time on ASD [19] with other state-of-the-art algorithms mentioned above. We implement GS [26] and MR [29] on our own, and use the authors' code for other algorithms. For GS [26], we use the same superpixel segmentation [18], resulting smaller time cost as reported in [26].

## 6. Conclusions

We present a novel background measure with intuitive and clear geometrical interpretation. Its robustness makes it especially useful for high accuracy background detection and saliency estimation. The proposed optimization framework effectively and efficiently combines other saliency cues with the proposed background cue, achieving the state-of-the-art results. It can be further generalized to incorporate more constraints, which we will consider for future works on this subject.

## Acknowledgement

This work is supported by The National Science Foundation of China (No.61305091), The Fundamental Research Funds for the Central Universities (No.2100219038), and Shanghai Pujiang Program (No.13PJ1408200).

## References

[1] S. Alpert, M. Galun, R. Basri, and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. In *CVPR*, 2007. 4, 5, 7

[2] A. Borji, D.N.Sihite, and L.Itti. Salient object detection: A benchmark. In *ECCV*, 2012. 1

[3] 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. 1, 2, 6

[2] We normalize and inverse the boundary connectivity map and use it as a saliency map.

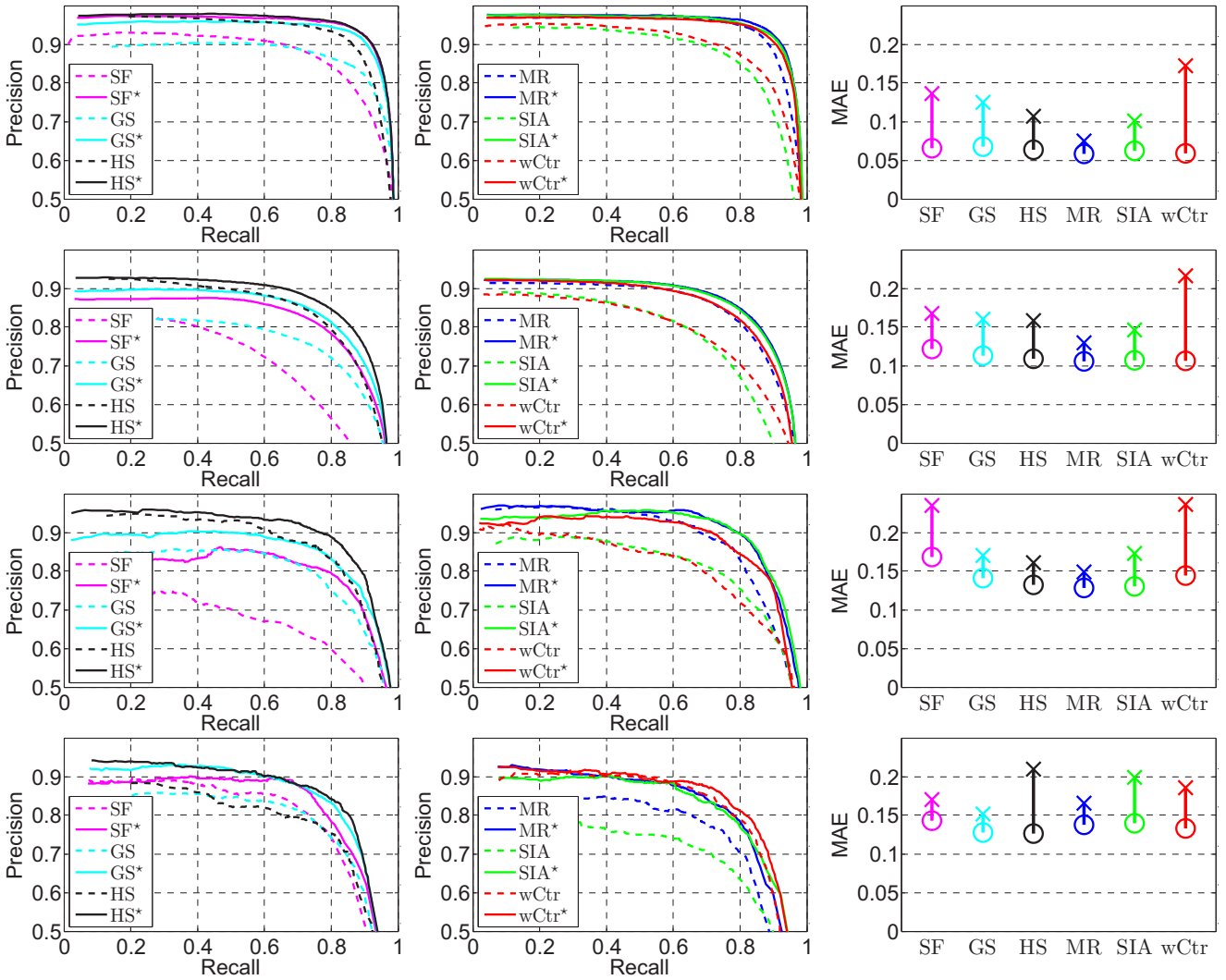



Figure 8. PR curves and MAEs of different methods and their optimized version(∗). From top to bottom: ASD [19], MSRA [25], SED1 [1], and SED2 [1] are tested. The first two columns compare PR curves and the last column directly shows MAE drops from state-of-the-art methods (x) to their corresponding optimized results (o).


[4] D.Gao, V.Mahadevan, and N.Vasconcelos. The discriminant center-surround hypothesis for bottom-up saliency. In *NIPS*, 2007. 2

[5] Y. Ding, J. Xiao, and J. Yu. Importance filtering for image retargeting. In *CVPR*, 2011. 1

[6] D.Klein and S.Frintrop. Center-surround divergence of feature statistics for salient object detection. In *ICCV*, 2011. 1

[7] J.Harel, C.Koch, and P.Perona. Graph-based visual saliency. In *NIPS*, 2006. 2

[8] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. In *CVPR*, 2013. 1, 2, 5

[9] Z. Jiang and L. S. Davis. Submodular salient region detection. In *CVPR*, 2013. 1, 2

[10] Johnson and D. B. Efficient algorithms for shortest paths in sparse networks. *J. ACM*, 24(1):1–13, 1977. 3

[11] J.Sun and H.Ling. Scale and object aware image retargeting for thumbnail browsing. In *ICCV*, 2011. 1

[12] L.Itti, C.Koch, and E.Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998. 2

[13] L.Marchesotti, C.Cifarelli, and G.Csurka. A framework for visual saliency detection with applications to image thumbnailing. In *ICCV*, 2009. 1

[14] L.Wang, J.Xue, N.Zheng, and G.Hua. Automatic salient object extraction with contextual cue. In *ICCV*, 2011. 1

[15] L. Mai, Y. Niu, and F. Liu. Saliency aggregation: A data-driven approach. In *CVPR*, 2013. 2, 5

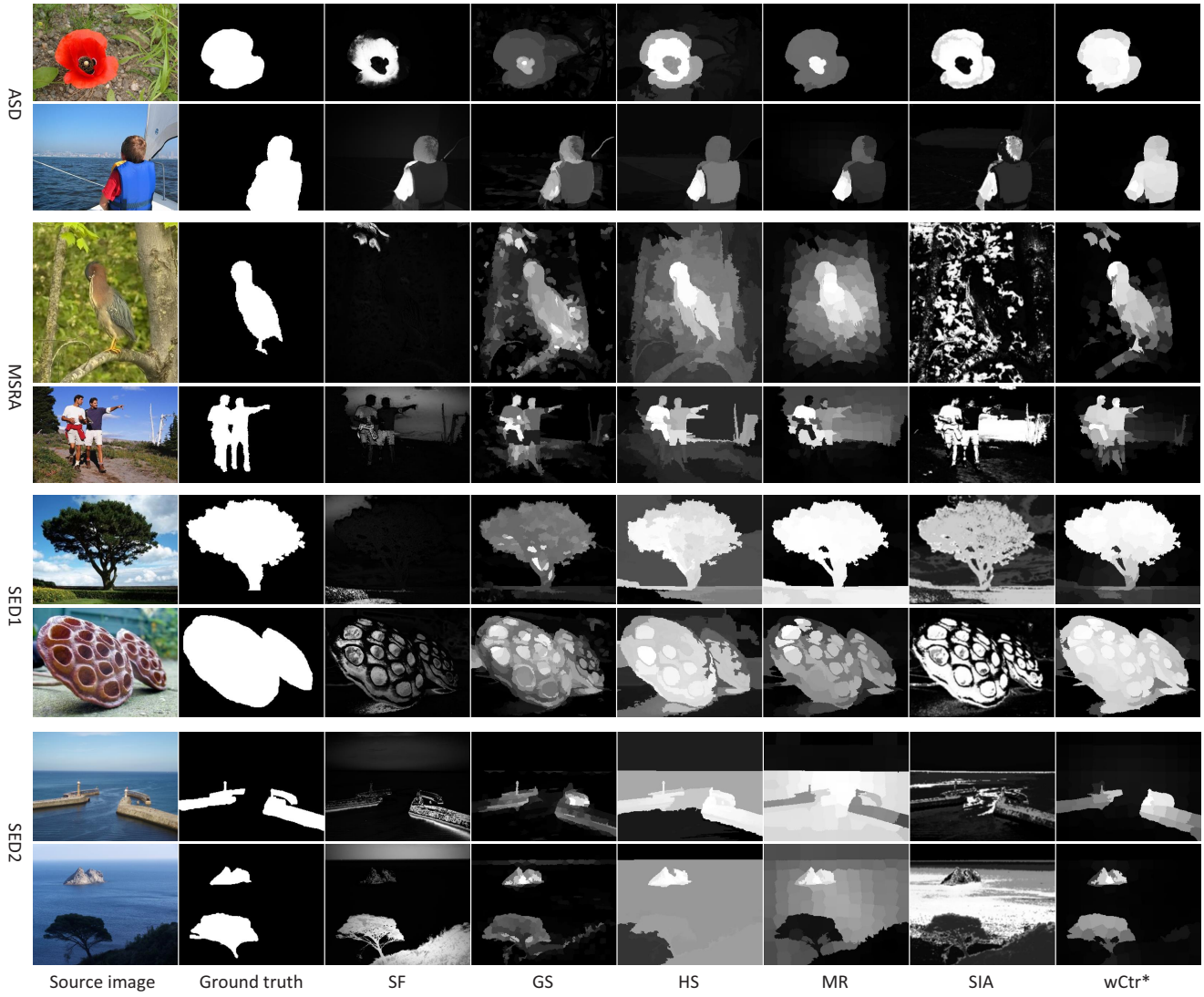



Figure 9. Example results of different methods on four datasets.


[16] M.Cheng, G.Zhang, N.Mitra, X.Huang, and S.Hu. Global contrast based salient region detection. In *CVPR*, 2011. 1, 4

[17] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012. 1, 4, 6

[18] R.Achanta, A.Shaji, K.Smith, A.Lucchi, P.Fua, and S.Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2281, 2012. 3, 6

[19] R.Achanta, S.Hemami, F.Estrada, and S.Susstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009. 1, 4, 5, 6, 7

[20] C. Rother, V. Kolmogorov, and A. Blake. "grabcut"cinteractive foreground extraction using iterated graph cuts. In *SIGGRAPH*, 2004. 1

[21] R.Valenti, N.Sebe, and T.Gevers. Image saliency by isocentric curvedness and color. In *ICCV*, 2009. 2

[22] S.Goferman, L.manor, and A.Tal. Context-aware saliency detection. In *CVPR*, 2010. 1, 4

[23] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, 2012. 2

[24] T.Judd, K.Ehinger, F.Durand, and A.Torralba. Learning to predict where humans look. In *ICCV*, 2009. 2

[25] T.Liu, J.Sun, N.Zheng, X.Tang, and H.Shum. Learning to detect a salient object. In *CVPR*, 2007. 1, 2, 4, 5, 6, 7

[26] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *ECCV*, 2012. 1, 2, 4, 6

[27] X.Hou and L.Zhang. Saliency detection: A spectral residual approach. In *CVPR*, 2007. 2

[28] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *CVPR*, 2013. 1, 2, 4, 5, 6

[29] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013. 1, 2, 4, 6