# Saliency Detection: A Boolean Map Approach

Jianming Zhang Stan Sclaroff
Department of Computer Science, Boston University
{jmzhang,sclaroff}@bu.edu

## Abstract

*A novel Boolean Map based Saliency (BMS) model is proposed. An image is characterized by a set of binary images, which are generated by randomly thresholding the image's color channels. Based on a Gestalt principle of figure-ground segregation, BMS computes saliency maps by analyzing the topological structure of Boolean maps. BMS is simple to implement and efficient to run. Despite its simplicity, BMS consistently achieves state-of-the-art performance compared with ten leading methods on five eye tracking datasets. Furthermore, BMS is also shown to be advantageous in salient object detection.*

## 1. Introduction

In this paper, we focus on the bottom-up saliency detection problem. The main goal is to compute a saliency map that topographically represents the level of saliency for visual attention. Computing such saliency maps has recently raised a great amount of research interest (see [4] for a review) and has been shown to be beneficial in many applications, *e.g.* image segmentation [12], object recognition [32] and visual tracking [28].

Many previous works have exploited the contrast and the rarity properties of local image patches for saliency detection [19, 6, 3]. However, these properties have limited ability to model some global perceptual phenomena [23] known to be relevant to the deployment of visual attention. One such global perception mechanism is figure-ground segregation. As Gestalt psychological studies suggest, figures are more likely to be attended to than background elements [31, 29] and the figure-ground assignment can occur without focal attention [22]. Neuroscience findings also show that certain responses in monkey and human brains involved in shape perception are critically dependent on figure-ground assignment [2, 26], indicating that this process may start early in the visual system.

Fig. 1 shows an example that global cues for figure-ground segregation can help in saliency detection. A natural image along with eye tracking data is displayed in Fig. 1 (a), where the bird is naturally perceived as the foreground and the rest as the background. The eye fixations are concentrated on the bird, corresponding well to this figure-ground assignment. However, without the awareness of this global structure, rarity based models [6, 3] falsely assign high saliency values to the edge area between the trees and the sky in the background, because of the rarity of high contrast regions in this image (Fig. 1 (b)).

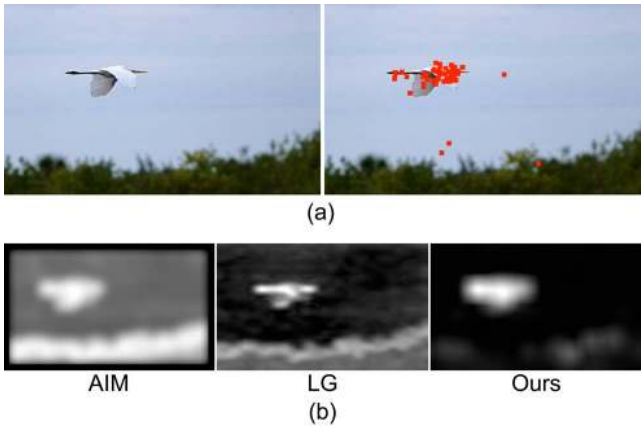


Figure 1: (a) Image from the MIT dataset [20] (left) and its eye tracking data (right). (b) Saliency maps estimated by (from left to right) AIM [6], LG [3] and our method. AIM and LG measure an image patch's saliency based on its rarity. Our method, based on global structural information, is less responsive to the elements in the background.

In this work, we present a novel Boolean Map based Saliency model (BMS), which leverages global topological cues that are known to help in perceptual figure-ground segregation. As Gestalt psychological studies suggest, several factors are likely to influence figure-ground segregation, *e.g.* size, surroundedness, convexity and symmetry [30]. In this paper, we explore the surroundedness cue for saliency detection. The essence of surroundedness is the enclosure topological relationship between the figure and the ground, which is well defined and invariant to various transformations. To measure the surroundedness, BMS characterizes an image by a set of Boolean maps. In BMS, an attention map is efficiently computed by binary image processing techniques to activate regions with closed outer

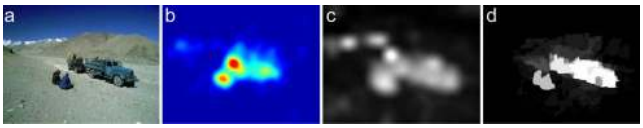


Figure 2: (a) is a sample image from the ImgSal eye tracking dataset [27]; (b) is the ground truth eye fixation heat map; (c) and (d) are the saliency maps generated by BMS for eye fixation prediction and salient object detection respectively.

contours on a given Boolean map. Then saliency is modeled as the expected attention level given the set of randomly sampled Boolean maps. The expected attention map, *i.e.* the mean attention map, is a full-resolution preliminary saliency map that can be further processed for a specific task such as eye fixation prediction or salient object detection [5]. Fig. 2 shows two types of saliency maps of BMS for eye fixation prediction and salient object detection.

We evaluate BMS against ten state-of-the-art saliency models on five benchmark eye tracking datasets. The compared models include some very recent ones that were shown to lead on some of the datasets tested in this paper. Despite its simplicity, BMS is the only method that consistently achieves state-of-the-art performance on all five benchmark datasets. We also show with both qualitative and quantitative results that the outputs of BMS are useful in salient object detection.

## 2. Related Works

A majority of the previous saliency models use center-surround filters or image statistics to identify salient patches that are complex (local complexity/contrast) or rare in their appearance (rarity/improbability). Center-surround difference is used in [19] to detect conspicuous regions on multi-scale feature maps, followed by a normalization and fusion of the resulting conspicuity maps. The negative logarithm of the probability, known as Shannon's self-information, is used to measure the improbability of a local patch as a bottom-up saliency cue in [6] and [39]. Moreover, [18] mensurates the improbability of a local region by "Bayesian surprise", a concept that aims to quantify how data modify the prior beliefs of the observer. Recently, [10] uses a hierarchically whitened feature space, where the square of the vector norms serves as a saliency metric to measure how far a pixel feature vector deviates from the center of the data. Besides the contrast and rarity priors for saliency, local symmetry has also been used by [25].

Unlike models based on properties like contrast, rarity and symmetry, another family of saliency models are based on spectral domain analysis [15, 14, 33, 27]. However, [27] shows that some previous spectral analysis based methods are in some sense equivalent to a local gradient operator plus Gaussian blurring on natural images, and thus cannot detect large salient regions very well. To overcome this limitation, a method based on spectral scale-space analysis is proposed by [27].

Some models employ machine learning to learn saliency. Kienzel *et al.* [21] learn a kernel support vector machine (SVM) for image patches based on eye tracking data. Judd *et al.* [20] train a SVM using a combination of low, middle and high level features, and the saliency classification is done in a pixel-by-pixel manner.

Unlike the previous approaches, the proposed BMS does not rely on center-surround filtering, statistical analysis of features, spectral transforms, off-line learning, or multi-scale processing. Instead, it makes use of topological structural information, which is scale-invariant and known to have a strong influence on visual attention [37, 8]. Most of the aforementioned models do not reflect this aspect.

Only a few attempts have been made to leverage the topological structure of a scene for saliency detection. In [13], Markov chain graphs are constructed based on low level feature maps and intermediate results, and their equilibrium distributions are used as the outputs of activation and normalization. In [36], a local patch's saliency is measured on a graphical model, by its shortest distance to the image borders. The edge weights of the graphical model are computed based on local dissimilarity and are specifically tailored to the needs of object segmentation. Compared with [13, 36], BMS utilizes the topological cues through Boolean maps, in a more explicit and much simpler way.

The salient region detection method of [16] also employs a feature channel thresholding step. However, thresholding is applied to each feature channel only once to extract regions of interest in a deterministic fashion, and subsequent processing critically depends on the original image. In contrast, BMS computes saliency entirely based on the set of randomly thresholded Boolean maps.

## 3. Boolean Map based Saliency

To derive a bottom-up saliency model, we borrow the Boolean Map concept that was put forward in the Boolean Map Theory of visual attention [17], where an observer's momentary conscious awareness of a scene can be represented by a Boolean Map. We assume that Boolean maps in BMS are generated from randomly selected feature channels, and the influence of a Boolean map $B$ on visual attention can be represented by an Attention Map $A(B)$, which highlights regions on $B$ that attract visual attention. Then the saliency is modeled by the mean attention map $\bar{A}$ over randomly generated Boolean maps:

$$\bar{A} = \int A(B)p(B|I)dB \qquad (1)$$

where $I$ is the input image. $\bar{A}$ can be further post-processed to form a final saliency map $S$ for some specific task.

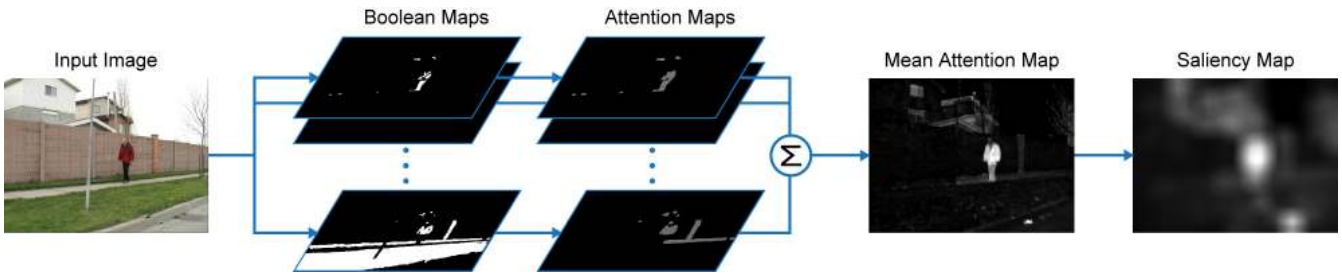


Figure 3: The Pipeline of BMS.

The pipeline of BMS is illustrated in Fig. 3. Given an image $I$, a set of Boolean maps $\mathbf{B} = \{B_1, B_2, \ldots, B_n\}$ is generated. Based on a Gestalt principle of figure-ground segregation, an attention map $A_i$ is computed for each Boolean map $B_i$. Then a mean attention map $\bar{A}$ is obtained through a linear combination of the resulting attention maps. Finally, some post-processing is applied on the mean attention map to output a saliency map $S$. Each step will be described in the following sections.

### 3.1. Generation of Boolean Maps

BMS generates a set of Boolean maps by randomly thresholding the input image's feature maps, according to the prior distributions over the feature channels and the threshold:

$$B_i = \texttt{THRESH}(\phi(I), \theta), \quad \phi \sim p_\phi, \theta \sim p_\theta. \tag{2}$$

The function $\texttt{THRESH}(., \theta)$ assigns 1 to a pixel if its value on the input map is greater than $\theta$, and 0 otherwise. $\phi(I)$ denotes a feature map of $I$, whose values are assumed to range between 0 to 255. $p_\phi$ and $p_\theta$ denote the prior distributions of $\phi$ and $\theta$ respectively. Feature channels can consist of multiple features like color, orientation, depth, motion, *etc*. In this work, we demonstrate in an implementation using only color for still images.

Given that $\phi(I)$ is a color channel of $I$, without loss of generality, the threshold $\theta$ is drawn from a uniform distribution over $[0, 255]$, because the the effect of any distribution of $\theta$ with accumulative distribution $F(\theta)$, is equivalent to applying a mapping function $255 \cdot F^{-1}(.)$ on each color channel. Therefore, given an image, the distribution of generated Boolean maps is solely determined by the choice of color space and the prior distribution for color channel selection.

Boolean maps should be generated in such a way that more salient regions have higher chances to be separated from the surrounding background. Given a uniform distribution of the threshold $\theta$, an ideal color space for BMS should be the one whose distance metric reflects the visual difference between colors. Therefore, we choose CIE *Lab* color space, which is known for its perceptual uniformity. The range of each channel in *Lab* space is translated and scaled to be $[0, 255]$ for the sake of consistency.

We assume that the three channels of *Lab* space play equally important roles in visual perception. To generate Boolean maps for an image, we simply enumerate the three channels and sample the threshold $\theta$ from 0 to 255 by a fixed step size $\delta$. An inverted copy of each Boolean map is also included in the output, in order to account for the inverted region selection. An opening operation with kernel $\omega_o$ is then applied to each Boolean map for noise removal.

### 3.2. Attention Map Computation

Given a Boolean map $B$, BMS computes an attention map $A(B)$ based on a Gestalt principle for figure-ground segregation: surrounded regions are more likely to be perceived as figures [30]. Surroundedness in a Boolean map is well defined as a property of a connected region (either of value 1 or 0) that has a closed outer contour. Under this definition, only regions connected to the image borders are not surrounded. To compute the attention map, BMS assigns 1 to the union of surrounded regions, and 0 to the rest of the map. This operation can be efficiently implemented by using Flood Fill algorithm to mask out all the pixels connected to the image borders.

The resultant attention maps need to be normalized before the linear combination step, so that attention maps with small concentrated active areas will receive more emphasis. Different normalization schemes have been used for similar purposes in previous works [19, 13]. For eye fixation prediction, BMS uses simple L2-normalization, *i.e.* dividing a vectorized map by its L2-norm, to emphasize attention maps with small active areas. Compared with L1-normalization, L2-normalization is less sensitive to attention maps with extremely small active areas, which will otherwise dominate the fusion process. To further penalize attention maps with small scattered active areas, we dilate the attention map with kernel width $\omega_{d1}$ before normalization.

All the attention maps are linearly combined into a full-resolution mean attention map $\bar{A}$. The mean attention maps can be further processed for a specific task. The whole algorithm of BMS is summarized in Algorithm 1.

| Datasets | No. Images | No. Viewers | Features |
|---|---|---|---|
| MIT [20] | 1003 | 15 | Daily life indoor and outdoor pictures; Portraits. |
| Toronto [6] | 120 | 20 | A large portion of images do not contain particular regions of interest. |
| Kootstra [24] | 100 | 31 | Five categories of images: 12 animals, 12 cars and streets, 16 buildings, 20 flowers and plants, and 40 natural scenes. |
| Cerf [7] | 181 | 8 | The objects of interest are faces and some other small objects like cell phone, toys, *etc*. |
| ImgSal [27] | 235 | 21 | Six categories: 50/80/60 with large/medium/small salient regions; 15 with clustering background; 15 with repeating distractors; 15 with both large and small salient regions. |

Table 1: Description of the Eye Tracking Datasets.

**Alg. 1** $S = \texttt{BMS}(I)$

$\mathbf{B} = \{\}$
**for** each color channel map $\{\phi_k(I) : k = 1, 2, 3\}$ in *Lab* space
  **for** $\theta = 0 : \delta : 255$
    $B = \texttt{THRESH}(\phi_k(I), \theta)$
    $\widetilde{B} = \texttt{INVERT}(B)$
    add $\texttt{OPENING}(B, \omega_o)$ and $\texttt{OPENING}(\widetilde{B}, \omega_o)$ to $\mathbf{B}$
**for** each $B_k \in \mathbf{B}$
  $A_k = \texttt{ZEROS}(B_k.\text{size}())$
  set $A_k(i, j) = 1$ if $B_k(i, j)$ belongs to a surrounded region
  $A_k = \texttt{DILATION}(A_k, \omega_{d1})$
  $A_k = \texttt{NORMALIZE}(A)$
$\bar{A} = \frac{1}{n} \sum_{k=1}^{n} A_k$
$S = \texttt{POST\_PROCESS}(\bar{A})$
**return** $S$

## 4. Eye Fixation Prediction

In this section, we evaluate the performance of BMS in eye fixation prediction.

**Implementation Details.** Each input image is first resize to 600 pixels in width, and the kernel's width of the opening operation $\omega_o$ is fixed at 5 pixels. The sampling step size $\delta$ is set to 8 and the dilation kernel width $\omega_{d1}$ is fixed at 7. We post-process $\bar{A}$ to produce the saliency map $S$ by Gaussian blurring with standard deviation (STD) $\sigma$. However, strong Gaussian blur will remove small peaks on the mean attention map, which is sometimes undesirable. To control for this factor, we use a dilation operation with kernel width $\omega_{d2}$ before Gaussian blur. We do not find this dilation operation improves the performance of other compared methods. By experiment, we have found setting $\sigma$ to 20 and $\omega_{d2}$ to 23 usually works well. We fix these parameters in the following experiments. The source code is available on our website[1].

### 4.1. Experimental Setup

We have quantitatively evaluated our algorithm in comparison with ten state-of-the-art saliency methods shown in Table 2. The code for these baseline methods is available on authors' websites[2], and we used the default configuration set by the authors. When evaluating Judd's model [20], we removed the features from the object detectors for a fair comparison, and this also slightly improves the shuffled-AUC scores of Judd's model.

**Datasets.** The methods are evaluated on five benchmark eye tracking data sets: MIT [20] (MIT data set), Toronto [6], Kootstra [24], Cerf [7] (FIFA data set) and ImgSal [27]. These datasets are available on the authors' websites. Some statistics and features of these datasets are summarized in Table 1.

**Evaluation Metric.** One of the most widely used metrics for saliency method evaluation is the ROC Area Under the Curve (AUC) metric. However, factors such as border cut and center-bias setting have been shown to have a dramatic influence over AUC [34, 39]. For example, in [39], it has been shown that a static Gaussian blob has an average ROC score of 0.80 on the Toronto dataset, exceeding many state-of-the-art methods, without using any bottom-up features in the images. To control for these factors, we adopt the shuffled-AUC proposed by [34, 39], which has become a standard evaluation method used in many recent works [33, 14, 3, 10]. Under the shuffled-AUC metric, a perfect prediction will give an AUC of 1.0, while any static saliency map will give a score of approximately 0.5. We use the implementation of shuffled-AUC by the authors of [33, 3]. For details of the implementation of shuffled-AUC, we refer the readers to [39].

### 4.2. Results

AUC scores are sensitive to the level of blurring applied on the saliency maps. As in [14, 3], we smooth the saliency maps of each method by varying the Gaussian blur standard deviation (STD), and show in Fig. 4 its influence on the average shuffled-AUC scores of each method on different datasets. The optimal scores of each model together with the corresponding Gaussian blur STD are reported in Table 2. We also report the results of BMS using RGB color space.

BMS achieves state-of-the-art performance, with the best average AUC scores, on all the five datasets (see Table 2). The choice of color space for BMS has a consistent effect on each dataset. By substituting RGB space for *Lab* space, the average score of BMS degrades by more than 0.01 on each dataset. This result agrees with the analysis in Section 3.1.

Evaluation on different datasets gives different ranks of methods, because these datasets vary in many aspects, such as the selection of visual stimuli, the composition of

[1] http://www.cs.bu.edu/groups/ivc/software/BMS/
[2] For Itti's model, we use the improved version by [13].

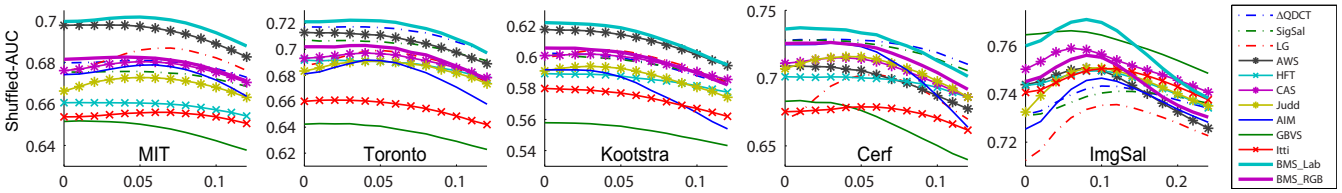


Figure 4: **Average Shuffled-AUC against the STD of Gaussian Blur**. X-axis represents the Gaussian blur standard deviation (STD) in image width and Y-axis represents the average shuffled-AUC score on one dataset.

| Dataset | BMS *Lab* | BMS RGB | ΔQDCT [33] | SigSal [14] | LG [3] | AWS [10] | HFT [27] | CAS [11] | Judd [20] | AIM [6] | GBVS [13] | Itti [19] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MIT [20] | .7017 | .6825 | .6808 | .6756 | .6868 | .6979 | .6606 | .6803 | .6726 | .6787 | .6518 | .6559 |
| opt. $\sigma$ | .05 | .04 | .04 | .04 | .07 | .01 | 0.1 | .05 | .05 | .06 | .01 | .06 |
| Toronto [6] | .7221 | .7029 | .7176 | .7068 | .6888 | .7130 | .6914 | .6970 | .6910 | .6913 | .6430 | .6610 |
| opt. $\sigma$ | .03 | .04 | .03 | .00 | .05 | .01 | .02 | .04 | .05 | .04 | .02 | .03 |
| Kootstra [24] | .6220 | .6057 | .6025 | .6013 | .6046 | .6174 | .5891 | .6021 | .5941 | .5922 | .5579 | .5798 |
| opt. $\sigma$ | .00 | .00 | .00 | .00 | .04 | .00 | .01 | .03 | .03 | .01 | .01 | .00 |
| Cerf [7] | .7365 | .7257 | .7286 | .7281 | .7026 | .7091 | .7011 | .7151 | .7159 | .7251 | .6830 | .6787 |
| opt. $\sigma$ | .01 | .01 | .03 | .01 | .06 | .01 | .03 | .04 | .04 | .03 | .01 | .05 |
| ImgSal [27] | .7712 | .7562 | .7434 | .7412 | .7357 | .7510 | .7498 | .7591 | .7510 | .7467 | .7665 | .7507 |
| opt. $\sigma$ | .08 | .08 | .10 | .12 | .12 | .08 | .10 | .06 | .10 | .10 | .06 | .10 |
| Avg. | .7107 | .6946 | .6946 | .6906 | .6837 | .6977 | .6784 | .6907 | .6849 | .6868 | .6604 | .6652 |

Table 2: **Average Shuffled-AUC with Optimal Blurring**. Optimal average shuffled-AUC of each method with the corresponding Gaussian blur STD is reported. The highest score on each dataset is shown in red color; the second and third highest are underlined. As in [39], we repeat the shuffling for 20 times and compute the standard deviation of each average shuffled-AUC, which ranges from 1E-4 to 5E-4.

participants and the experimental environment. Although several compared methods have similar performance as BMS on some of the datasets, *e.g.* AWS [10] on MIT and Kootstra, ΔQDCT [33] on Toronto and Cerf, GBVS [13] on ImgSal, none of them achieves top performance on all five datasets. We note that for all the compared methods, their average AUC scores are worse than those of BMS by more than 0.025 on at least one datset. All the methods perform drastically worse on the Kootstra dataset, whose inter-observer consistency has shown to be low [3]. On average, BMS, either using *Lab* or RGB, and the recent methods like AWS [10], ΔQDCT [33], CAS [11] and SigSal[14] give better performance than the others. GBVS [13] has significantly worse AUC scores on the MIT, Toronto, Kootstra and Cerf datasets. Interstingly, it ranks the 2nd on the ImgSal dataset.

In our experiments, we found that BMS tends to be less distracted by background clutter or highly textured background elements than most of the compared methods, and it is capable of highlighting the interior regions of salient objects of different scales without resorting to any multi-scale processing. Fig. 5 shows some examples. The input images are roughly arranged in ascending order of the size of their salient regions. In these examples, most of the compared methods are more influenced by the cluttered and highly textured areas in the background. Moreover, they tend to favor the boundaries rather than the interior regions of large salient objects, like the car and the STOP sign in the last two examples, even with the help of multi-scale processing [33, 3, 10, 11, 13, 19].

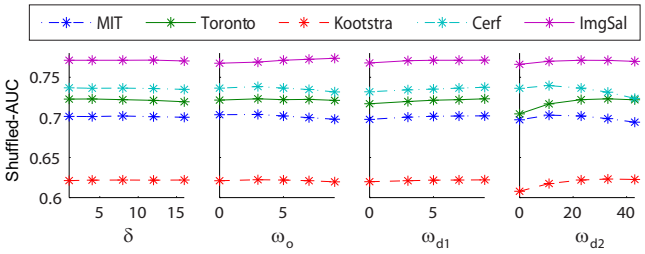


Figure 6: Parameter Analysis.

**Parameter Analysis**. Five parameters are involved in the implementation of BMS: sample step $\delta$, kernel widths of opening operation $\omega_o$, kernel widths of two dilation operations $\omega_{d1}$ and $\omega_{d2}$ , and the Gaussian blur STD $\sigma$. The influence of Gaussian blur has already been shown in Fig. 4. Fig. 6 displays the influences of the other four parameters on the average AUC scores on each dataset. Overall, BMS is not very sensitive to these parameters except the dilation kernel width $\omega_{d2}$ in the post-precessing step. The influence of $\omega_{d2}$ is dataset dependent. Having a slight dilation before the final smoothing improves the AUC scores on all the datasets, while setting $\omega_{d2}$ to greater than 20 only improves the average AUC scores on the Toronto and Kootstra dataset. The sample step size has a direct impact on the runtime, since the time complexity of BMS grows linearly with number of Boolean maps. On average, the AUC scores start to drop slightly when $\delta$ is greater than 12. Applying an opening operation over Boolean maps does not significantly change the average AUC scores on most of the datasets, but the score on the ImgSal dataset improves by more than 0.006 when $\omega_o = 9$ (the standard deviation

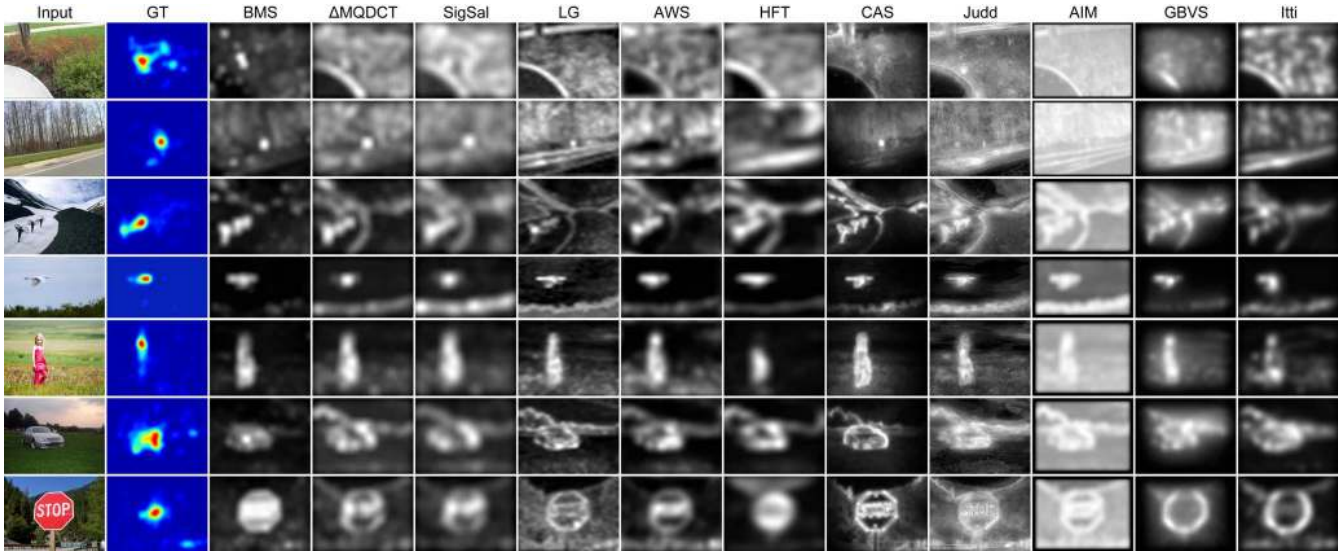


Figure 5: **Saliency maps of different methods**. The first two columns are the input images and their fixation heat maps from the Toronto (1st and 2nd rows), MIT (4th row) and ImgSal datasets (the rest). The fixation heat maps are computed by applying Gaussian blur on the raw eye fixation maps. The rest columns show the saliency maps from BMS and the compared methods. Images are roughly arranged in ascending order of the size of their salient regions.

of the average shuffled-AUC with regard to the shuffling is less than 1E-4 on this dataset.). Applying a dilation operation over the attention maps improves the AUC scores on average, but the improvement drops when $\omega_{d1}$ is greater than 7.

**Runtime Performance.** BMS is implemented in C++. On average it takes BMS 0.38s to process a $600 \times 400$ image using a 2.5GHz dual-core 32-bit Windows desktop computer with 2GB memory. All the compared models are implemented in Matlab or Matlab+C. The average time taken by the compared methods to process a $600 \times 400$ image on the same machine is listed as follows [3]: CAS [11] 78s, LG [3] 13s, AWS [10] 10s, Judd [20] 6.5s, AIM [6] 4.8s, GBVS [13] 1.1s, ΔQDCT [33] 0.49s, Itti [19] 0.43s, HFT [27] 0.27s and SigSal [14] 0.12s.

## 5. Salient Object Detection

In this section, we show that BMS is also useful in salient object detection. Salient object detection aims at segmenting salient objects from the background. Models for salient object detection have different emphasis compared with models for eye fixation prediction. Because eye fixations are sparsely distributed and possess some level of uncertainty, the corresponding saliency maps are usually highly blurred and very selective. However, salient object detection requires object level segmentation, which means the corresponding saliency map should be high-resolution with uniformly highlighted salient regions and clearly defined region boundaries.

We use the same sample step size $\delta$ as before, but set $\omega_o$ to 13 to have more small isolated areas removed from the Boolean maps. We also turn off the dilation operation in the attention maps computation (*i.e.* $\omega_{d1} = 1$) to enhance the accuracy of attention maps. Attention maps are not normalized before linear combination, which can be thought of as implicitly using a $L_\infty$-normalization. In this way, object regions of different sizes will be more evenly highlighted. Then we post-process the mean attention maps of BMS using an opening-by-reconstruction operation followed by a closing-by-reconstruction operation [35] with kernel radius 15, in order to smooth the saliency maps but keep the boundary details.

We quantitatively evaluate BMS on the ASD dataset [1], which comprises 1000 images and ground-truth segmentation masks. BMS is compared with six state-of-the-art salient object detection methods (HSal [38], GSSP, GSGD [36], RC, HC [9] and FT [1]), as well as some leading models for eye fixation prediction. Similar to previous works [1, 36], we binarize the saliency maps at a fixed threshold and compute the average precision and recall (PR) for each method. By varying the threshold of binarization, a PR curve can be obtained for each method.

Fig. 7 shows the PR curves of different methods on the ASD dataset. According to these results, BMS is comparable with HSal [38] and GSSP [36], and significantly better than other previous methods on the ASD dataset. Compared with HSal [38], BMS gives lower precision when the recall rate is below 92%, but achieves slightly better precision as the recall rate increases. Region based methods, GSSP [36]

[3] Note that some compared methods implicitly down-sampled input images before processing. The runtime reported here is based on their default settings.

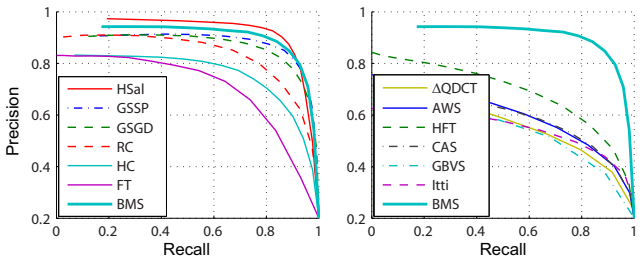


Figure 7: Precision-Recall curves on the ASD dataset [1].

and RC [9] , perform better than their grid/histogram based variants, GSGD [36] and HC [9]. Leading models for eye fixation prediction perform significantly worse than salient object detection methods. Some sample images and results are displayed in Fig. 8.

The ImgSal dataset [27] used in the previous section also has ground-truth salient regions labeled by 19 subjects. We show a couple of results on this dataset in Fig. 9. The labeled salient regions of this dataset are not very precise, and thus unsuitable for quantitative evaluation using the PR metric.

## 6. Conclusion and Future Work

In this work, a novel Boolean Map based Saliency model is proposed to leverage the surroundedness cue that helps in figure-ground segregation. The model borrows the concept of Boolean map from the Boolean Map Theory of visual attention [17], and characterizes an image by a set of Boolean maps. This representation leads to an efficient algorithm for saliency detection. BMS is the only model that consistently achieves state-of-the-art performance on five benchmark eye tracking datasets, and it is also shown to be useful in salient object detection.

We demonstrate the effectiveness of BMS using only color channels, while BMS should also be applicable to other feature channels, such as orientation, depth and motion. Testing on other features remains for future work. Another interesting direction for future work is to improve the attention map computation by incorporating more saliency cues like convexity, symmetry and familiarity. This may help to redeem the limitation that salient regions that touch the image borders cannot be well detected using the surroundedness cue alone.

**Acknowledgments**. This work was supported in part through grants from the US National Science Foundation #1029430, #0910908, and #0855065.

## References

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

[2] G. C. Baylis, J. Driver, et al. Shape-coding in it cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nature Neuroscience*, 4:937–942, 2001.

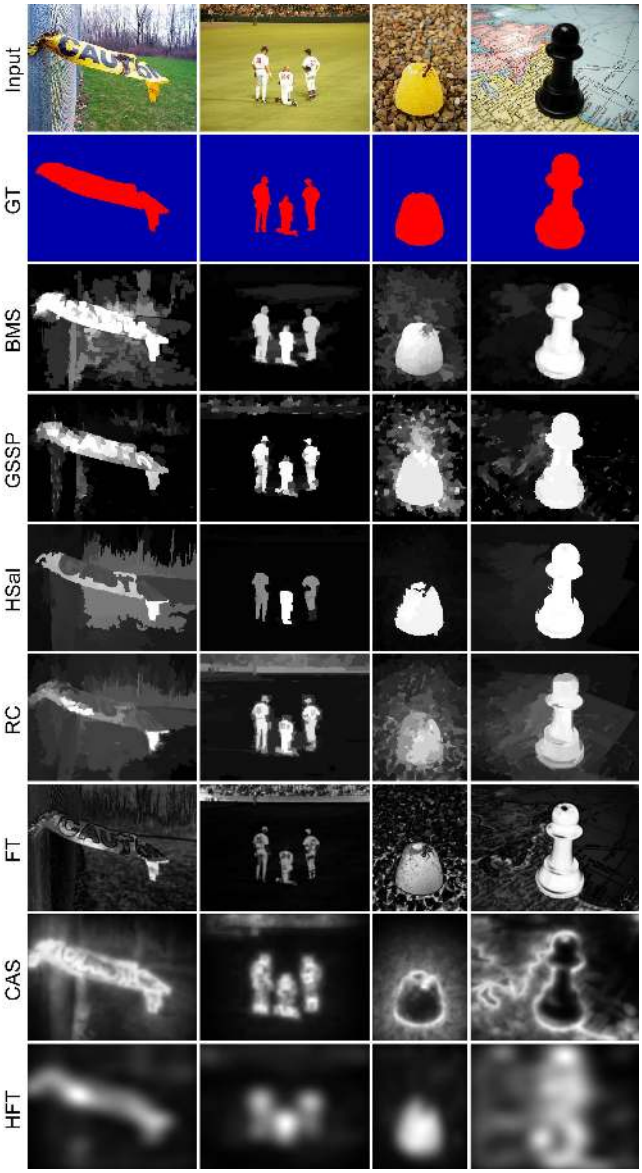


Figure 8: Saliency maps on the ASD dataset.

[3] A. Borji and L. Itti. Exploiting local and global patch rarities for saliency detection. In *CVPR*, 2012.

[4] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *PAMI*, 2012.

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

[6] N. Bruce and J. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3), 2009.

[7] M. Cerf, J. Harel, W. Einhäuser, and C. Koch. Predicting human gaze using low-level saliency combined with face detection. In *NIPS*, 2008.

[8] L. Chen. Topological structure in visual perception. *Science*, 1982.

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

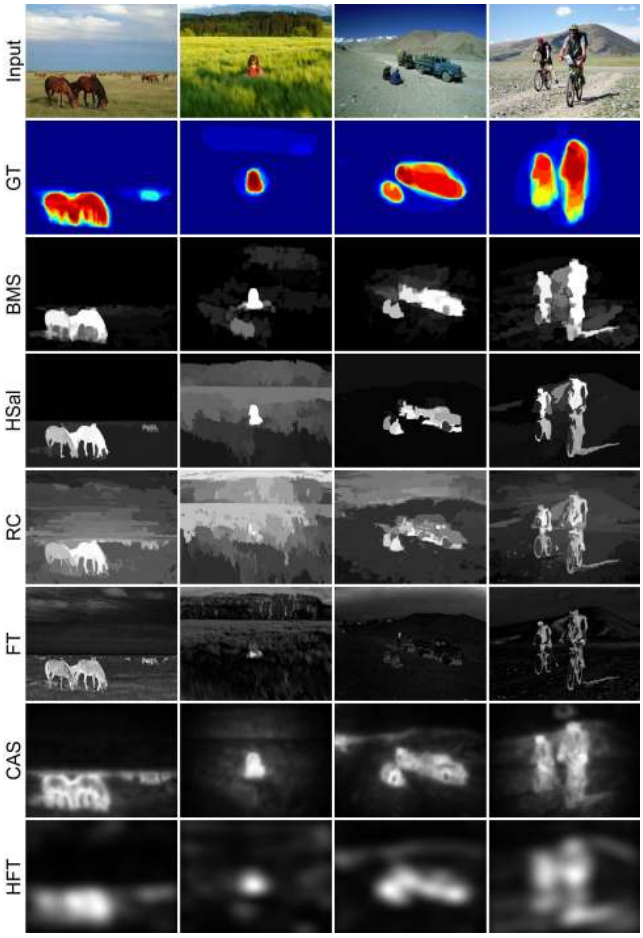


Figure 9: Saliency maps on the ImgSal dataset. We cannot show saliency maps of GSSP [36] because its code is not publicly available.


[10] A. Garcia-Diaz, X. Vidal, X. Pardo, and R. Dosil. Saliency from hierarchical adaptation through decorrelation and variance normalization. *IVC*, 2011.

[11] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *PAMI*, 34(10), 2012.

[12] J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. *Trans. Circuits and Systems for Video Technology*, 16(1), 2006.

[13] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, 2007.

[14] X. Hou, J. Harel, and C. Koch. Image signature: Highlighting sparse salient regions. *PAMI*, 34(1), 2012.

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

[16] Y. Hu, X. Xie, W.-Y. Ma, L.-T. Chia, and D. Rajan. Salient region detection using weighted feature maps based on the human visual attention model. In *Pacific Rim Conference on Advances in Multimedia Information Processing*, 2004.

[17] L. Huang and H. Pashler. A boolean map theory of visual attention. *Psychological review*, 114(3):599, 2007.

[18] L. Itti and P. Baldi. Bayesian surprise attracts human attention. In *NIPS*, 2006.

[19] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11):1254–1259, 1998.

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

[21] W. Kienzle, F. Wichmann, B. Schölkopf, and M. Franz. A nonparametric approach to bottom-up visual saliency. In *NIPS*, 2007.

[22] R. Kimchi and M. A. Peterson. Figure-ground segmentation can occur without attention. *Psychological Science*, 19(7):660–668, 2008.

[23] K. Koffka. Principles of Gestalt psychology. 1935.

[24] G. Kootstra, A. Nederveen, and B. De Boer. Paying attention to symmetry. In *BMCV*, 2008.

[25] G. Kootstra and L. Schomaker. Prediction of human eye fixations using symmetry. In *Proc. of the 31st Annual Conf. of the Cognitive Science Society (CogSci09)*, 2009.

[26] Z. Kourtzi and N. Kanwisher. Representation of perceived object shape by the human lateral occipital complex. *Science*, 293(5534):1506–1509, 2001.

[27] J. Li, M. D. Levine, X. An, X. Xu, and H. He. Visual saliency based on scale-space analysis in the frequency domain. *PAMI*, 35(4), 2013.

[28] V. Mahadevan and N. Vasconcelos. Saliency-based discriminant tracking. In *CVPR*, 2009.

[29] V. Mazza, M. Turatto, and C. Umilta. Foreground–background segmentation and attention: A change blindness study. *Psychological Research*, 69(3):201–210, 2005.

[30] S. E. Palmer. *Vision science: Photons to phenomenology.* The MIT press, 1999.

[31] E. Rubin. Figure and ground. *Readings in Perception*, pages 194–203, 1958.

[32] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? In *CVPR*, 2004.

[33] B. Schauerte and R. Stiefelhagen. Quaternion-based spectral saliency detection for eye fixation prediction. In *ECCV*, 2012.

[34] B. Tatler, R. Baddeley, I. Gilchrist, et al. Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45(5):643–659, 2005.

[35] L. Vincent. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *TIP*, 2(2), 1993.

[36] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *ECCV*, 2012.

[37] J. Wolfe and T. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6):495–501, 2004.

[38] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *CVPR*, 2013.

[39] L. Zhang, M. Tong, T. Marks, H. Shan, and G. Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7), 2008.