Saliency Detection via Cellular Automata

Yao Qin, Huchuan Lu, Yiqun Xu and He Wang
Dalian University of Technology

Abstract

In this paper, we introduce Cellular Automata–a dynamic evolution model to intuitively detect the salient object. First, we construct a background-based map using color and space contrast with the clustered boundary seeds. Then, a novel propagation mechanism dependent on Cellular Automata is proposed to exploit the intrinsic relevance of similar regions through interactions with neighbors. Impact factor matrix and coherence matrix are constructed to balance the influential power towards each cell's next state. The saliency values of all cells will be renovated simultaneously according to the proposed updating rule. It's surprising to find out that parallel evolution can improve all the existing methods to a similar level regardless of their original results. Finally, we present an integration algorithm in the Bayesian framework to take advantage of multiple saliency maps. Extensive experiments on six public datasets demonstrate that the proposed algorithm outperforms state-of-the-art methods.

1. Introduction

Recently, saliency detection aimed at finding out the most important part of an image has become increasingly popular in computer vision [1, 14]. As a pre-processing procedure, saliency detection can be used for many vision tasks, such as visual tracking [26], object retargeting [11, 39], image categorization [37] and image segmentation [35].

Generally, methods of saliency detection can be categorized as either top-down or bottom-up approaches. Top-down methods [3, 27, 31, 50] are task-driven and require supervised learning with manually labeled ground truth. To better distinguish salient objects from background, high-level information and supervised methods are incorporated to improve the accuracy of saliency map. In contrast, bottom-up methods [15, 18, 22, 40, 41, 48] usually exploit low-level cues such as features, colors and spatial distances to construct saliency maps. One of the most used principles, *contrast prior*, is to take the color contrast or geodesic distance against surroundings as a region's saliency[7, 8, 18, 21, 22, 33, 43]. In addition, several recent methods formulate their algorithms based on *boundary prior*, assuming that regions along the image boundary are more likely to be the background [19, 23, 44, 49]. Admittedly, it is highly possible for the image border to be the background, which have been proved in [5, 36]. However, it is not appropriate to sort all nodes on the boundary into one category as most previous methods. If the object appears on the image boundary, the chosen background seeds will be imprecise and directly lead to the inaccuracy of results.

In this paper, we propose effective methods to address the aforementioned problem. Firstly, we apply the K-means algorithm to classify the image border into different clusters. Due to the compactness of boundary clusters, we can generate different color distinction maps with complementary advantages and integrate them by taking spatial distance into consideration. Secondly, a novel propagation method based on Cellular Automata [42] is introduced to enforce saliency consistency among similar image patches. Through interactions with neighbors, boundary cells misclassified as background seeds will automatically modify their saliency values. Furthermore, we use Single-layer Cellular Automata to optimize the existing methods and achieve favorable results.

Many effective methods have been established to deal with saliency detection and each of them has their own superiorities [7, 17, 23, 33, 52]. In order to take advantage of different methods, we propose an integration model called Multi-layer Cellular Automata. With different saliency maps as original inputs, the integrated result outperforms all the previous state-of-the-arts.

In summary, the main contributions of our work include:
1). We propose an efficient algorithm to integrate global distance matrices and apply Cellular Automata to optimize the prior maps via exploiting local similarity.
2). Single-layer Cellular Automata can greatly improve all the state-of-the-art methods to a similar precision level and is insensitive to the previous maps.
3). Multi-layer Cellular Automata can integrate multiple saliency maps into a more favorable result under the Bayes framework.

2. Related works

Recently, more and more bottom-up methods prefer to construct the saliency map by choosing the image boundary as the background seeds. Considering the connectivity of regions in the background, Wei *et al*. [44] define each region's saliency value as the shortest-path distance towards the boundary. In [19], the contrast against image boundary is used as a new regional feature vector to characterize the background. In [49], Yang *et al*. compute the saliency of image regions according to their relevance to boundary patches via manifold ranking. In [52], a more robust boundary-based measure is proposed, which takes the spatial layout of image patches into consideration.

In addition, some effective algorithms have been proposed in the Bayesian framework. In [34], Rahtu *et al*. first apply Bayesian theory to optimize saliency maps and achieve better results. In [46, 47], Xie *et al*. use the low level visual cues derived from the convex hull to compute the observation likelihood. In [51], the salient object can naturally emerge under a Bayesian framework due to the self-information of visual features. Besides, Li *et al*. [23] attain saliency maps through dense and sparse reconstruction and propose a Bayesian algorithm to combine saliency maps. All of them demonstrate the effectivity of Bayesian theory in the optimization of saliency detection.

Cellular Automata, which was first put forward in [42], is a dynamic system with simple construction but complex self-organizing behaviour. The model consists of a lattice of cells with discrete states, which evolve in discrete time steps according to definite rules. Each cell's next state will be determined by its current state and the states of its nearest neighbors. Cellular Automata has been applied to simulate the evolution process of many complicated dynamic systems [4, 9, 10, 28, 29]. Considering that salient objects tend to be clustered, we apply Cellular Automata to exploit the intrinsic relationship of neighbors and reduce the difference in similar regions. Combined with Bayesian theory, Cellular Automata is introduced into this field as a propagation mechanism which can lay the roots for the optimization of saliency maps.

3. Proposed Algorithm

In this section, we first construct global color distinction and spatial distance matrix based on clustered boundary seeds and integrate them into a background-based map. Then, a novel propagation method based on Cellular Automata is proposed to intuitively exploit the intrinsic relevance of similar regions. We further discuss its great efficiency and robustness in optimizing other existing methods.

3.1. Global Distance Matrix Integration

To better capture intrinsic structural information and improve computational efficiency, an input image is segmented into N small superpixels by the simple linear iterative clustering (SLIC) algorithm [2]. The mean color features and coordinates of pixels are used to describe each superpixel. In order to achieve more optimal background seeds, we apply the K-means algorithm to divide the image boundary into K clusters based on their CIE LAB color features. We empirically set the number of boundary clusters $K = 3$ in this paper. The number of boundary superpixels belonging to cluster k is represented as p^k $(k = 1, 2, \cdots, K)$. Based on K different clusters, we can construct K different global color distinction (GCD) maps. The element $s_{k,i}$ in the GCD matrix $\boldsymbol{S} = [s_{k,i}]_{K\times N}$ represents the saliency of superpixel i in the k-th GCD map and is computed as:

$$s_{k,i} = \frac{1}{p^k}\sum_{j=1}^{p^k}\frac{1}{e^{-\frac{\|\boldsymbol{c}_i,\boldsymbol{c}_j\|}{2{\sigma_1}^2}}+\beta} \tag{1}$$

where $\|\ \boldsymbol{c}_i, \boldsymbol{c}_j\ \|$ is the Euclidean Distance between the superpixel i and j in CIE LAB color space. We set the balance weight $\sigma_1 = 0.2$ and $\beta = 10$. The result is insensitive to $\beta \in [7, 15]$.

We can see from Figure 1 that GCD maps constructed upon the boundary clusters are not satisfying, but each of them has certain superpixels with high precision. Due to the compactness of optimized boundary clusters, K GCD maps are complementary to each other. And a superpixel's saliency value is more accurate when it is computed based on the nearest background clusters. Therefore, we introduce global spacial distance (GSD) matrix $\boldsymbol{W} = [w_{k,i}]_{K\times N}$ to balance the importance of different GCD maps. $w_{k,i}$ represents the spacial distance between superpixel i and all background seeds in the k-th cluster. It is computed as:

$$w_{k,i} = \frac{1}{p^k}\sum_{j=1}^{p^k} e^{\frac{-\|\boldsymbol{r}_i,\boldsymbol{r}_j\|_2^2}{2{\sigma_2}^2}} \tag{2}$$

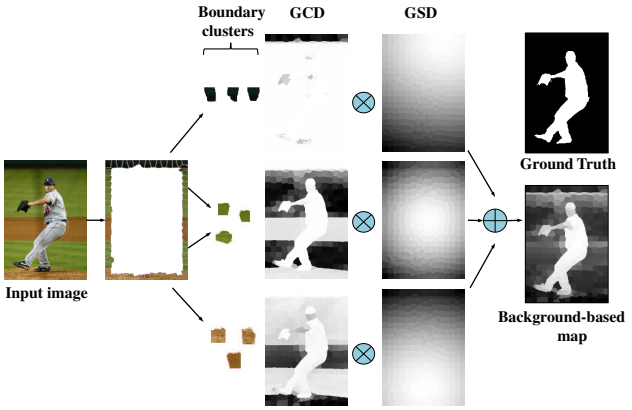


Figure 1. Integration of global color distinction (GCD) maps in Eqn 1 and global spatial distance (GSD) maps in Eqn 2 .

where $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are the coordinates of the superpixel i and j, σ_2 is a constant to control the strength of weight and robust in $[1.1, 1.5]$. In this work, we set $\sigma_2 = 1.3$. Then, the background-based map $\boldsymbol{S}^{bg} = [S_1^{bg}, \cdots, S_N^{bg}]^T$ is constructed by combining the geodesic information $w_{k,i}$ with the color information $s_{k,i}$:

$$S_i^{bg} = \sum_{k=1}^{K} w_{k,i} \times s_{k,i} \tag{3}$$

As Figure 1 shows, the geodesic constraint enforced on GCD maps greatly facilitates saliency accuracy by strengthening the contrast in local regions. Through effectively integrating the superiorities of different GCD maps, the background-based map is much more convincing and precise.

3.2. Parallel Evolution via Cellular Automata

In Single-layer Cellular Automata (SCA), each cell denotes a superpixel generated by the SLIC algorithm. We make three major modifications to the previous models [38, 42]. Firstly, the states of cells in most existing Cellular Automata models are discrete [30, 45]. However, in this paper, we use the saliency value of each superpixel as its state, which is continuous between 0 and 1. Secondly, we give a broader definition of neighborhood which is similar to the concept of z-layer neighborhood (here $z = 2$) in graph theory. A cell's newly defined neighbors include cells surrounding it as well as sharing common boundaries with its adjacent cells. Also we consider that superpixels on the image boundaries are all connected to each other because all of them serve as background seeds. Finally, unlike the well applied Cellular Automata models, the influences of all neighbors are not fixed but closely related to the similarity between any pair of cells in color space.

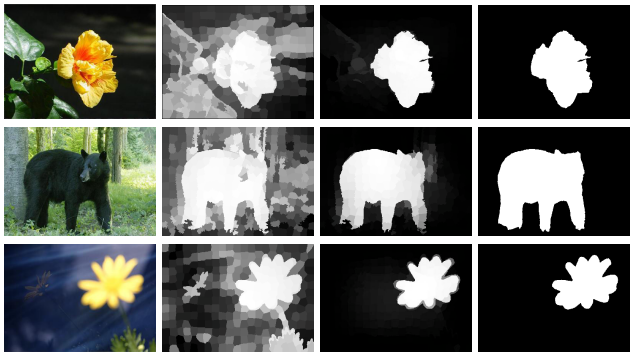
3.2.1 Impact Factor Matrix

It is intuitive to accept that neighbors with more similar color features have a greater influence on the cell's next state. The similarity of any pair of superpixels is measured by a defined distance in CIE LAB color space. We construct impact factor matrix $\boldsymbol{F} = [f_{ij}]_{N \times N}$ by defining the impact factor f_{ij} of superpixel i to j as:

$$f_{ij} = \begin{cases} \exp(\frac{-\|\boldsymbol{c_i}, \boldsymbol{c_j}\|}{{\sigma_3}^2}) & j \in NB(i) \\ 0 & i = j \text{ or otherwise} \end{cases} \tag{4}$$

where $\| \boldsymbol{c_i}, \boldsymbol{c_j} \|$ denotes the Euclidean Distance in CIE LAB color space between the superpixel i and j, σ_3 is a parameter to control strength of similarity. We set ${\sigma_3}^2 = 0.1$ as in [49]. $NB(i)$ is the set of neighbors of cell i. In order to normalize impact factor matrix, a degree matrix $\boldsymbol{D} = diag\{d_1, d_2, \cdots, d_N\}$ is generated, where $d_i = \sum_j f_{ij}$. Finally, a row-normalized impact factor matrix can be clearly calculated as follows:

$$\boldsymbol{F}^* = \boldsymbol{D^{-1}} \cdot \boldsymbol{F} \tag{5}$$



(a) (b) (c) (d)

Figure 2. Effects of Single-layer Cellular Automata. (a) Input images. (b) Background-based maps. (c) Background-based maps optimized via Single-layer Cellular Automata. (d) Ground truth.

3.2.2 Coherence Matrix

Considering that each cell's next state is determined by its current state as well as its neighbors', we need to balance the importance of the two decisive factors. For one thing, if a superpixel is quite different from all neighbors in color space, its next state will be primarily relied on itself. For the other, if a cell is similar to neighbors, it is more likely to be assimilated by the local environment. To this end, we build a coherence matrix $\boldsymbol{C} = diag\{c_1, c_2, \cdots, c_N\}$ to better promote the evolution of all cells. Each cell's coherence towards its current state is calculated as:

$$c_i = \frac{1}{max(f_{ij})} \tag{6}$$

In order to control $c_i \in [b, a + b]$, we construct the coherence matrix $\boldsymbol{C}^* = diag\{c_1^*, c_2^*, \cdots, c_N^*\}$ by the formulation as:

$$c_i^* = a \cdot \frac{c_i - min(c_j)}{max(c_j) - min(c_j)} + b \tag{7}$$

where $j = 1, 2, \cdots, N$. We set the constant a and b as 0.6 and 0.2. If a is fixed to 0.6, our results are insensitive to the interval when $b \in [0.1, 0.3]$. With coherence matrix $\boldsymbol{C}^*$, each cell can automatically evolve into a more accurate and steady state. And the salient object can be more easily detected under the influence of neighbors.

3.2.3 Synchronous Updating Rule

In Single-layer Cellular Automata, all cells update their states simultaneously according to the updating rule. Given an impact factor matrix and coherence matrix, the synchronous updating rule $f : S^{NB} \to S$ is defined as follows:

$$\boldsymbol{S}^{t+1} = \boldsymbol{C^*} \cdot \boldsymbol{S}^t + (\boldsymbol{I} - \boldsymbol{C^*}) \cdot \boldsymbol{F}^* \cdot \boldsymbol{S}^t \tag{8}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{C^*}$ and $\boldsymbol{F}^*$ are coherence matrix and impact factor matrix respectively. The initial $\boldsymbol{S}^t$ when $t = 0$ is $\boldsymbol{S}^{bg}$ in Eqn 3, and the ultimate saliency map after N_1 time steps (a time step is defined as one traversal iteration through all cells) is denoted as $\boldsymbol{S}^{N_1}$.

(a) (b) (c)

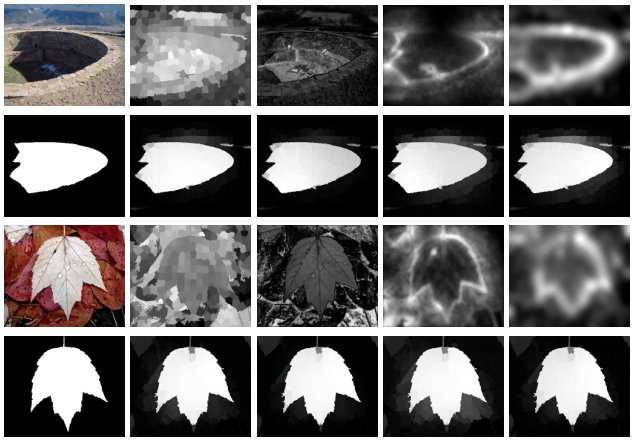
Figure 3. Saliency maps when objects touch the image boundaries. (a) Input images. (b) Saliency maps achieved by BSCA. (c) Ground truth.

We propose the updating rule based on the intrinsic characteristics of most images. Firstly, superpixels belonging to the foreground usually share similar color features. Via exploiting the intrinsic relationship in the neighborhood, Single-layer Cellular Automata can enhance saliency consistency among similar regions and form a steady local environment. Secondly, there is a great difference between the object and its surrounding background in color space. Influenced by similar neighbors, a clear boundary will naturally emerge between the object and the background. We denote the background-based map optimized via Single-layer Cellular Automata as BSCA. Figure 2 shows that Single-layer Cellular Automata can uniformly highlight the foreground and suppress the background.

In addition, Single-layer Cellular Automata can effectively deal with the problem mentioned in Section 1. When salient superpixels are selected as the background seeds by mistake, they will automatically increase their saliency values under the influence of local environment. Figure 3 shows that when the object touches the image boundary, the results achieved by our algorithm are still satisfying.

3.2.4 Optimization of State-of-the-Arts

Due to the connectivity and compactness of the object, the salient part of an image will naturally emerge after evolution. Moreover, we surprisingly find out that even if the background-based map is poorly constructed, the salient object can still be precisely detected via Singly-layer Cellular Automata, as exemplified in Figure 4 (b). Therefore, we use several classic methods as the prior maps and refresh them according to the synchronous updating rule. The saliency maps achieved by different methods are taken as S^t when $t = 0$ in Eqn 8. The optimized results via Single-layer Cellular Automata are shown in Figure 4. We can see that even though the original results are not satisfying, all of them are greatly improved to a similar accuracy level after evolution.



(a) (b) Ours (c) FT (d) CA (e) IT

Figure 4. Comparison of different methods and their optimized version after parallel evolutions. (a) The first and third row are input images. The second and fourth row are the ground truth. (b)-(e) The first and third row are the original results of different methods. From left to right: our background-based maps, saliency maps generated by FT [1], CA [13], IT [16]. The second and fourth row are their optimized results via Single-layer Cellular Automata.

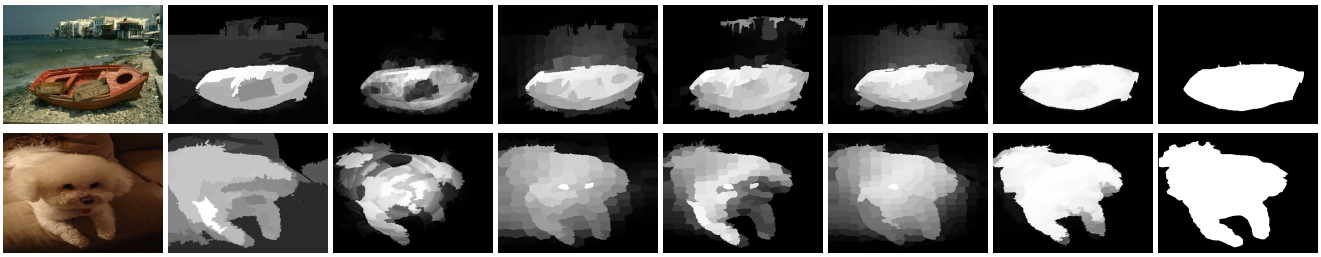
That means our method is independent to prior maps and can make an effective optimization towards state-of-the-art methods.

4. Multi-layer Cellular Automata

Many innovative methods by far have been put forward to deal with saliency detection. And different methods have their own advantages and disadvantages. In order to take advantage of the superiority of each method, we propose an effective method to incorporate M saliency maps generated by M state-of-the-art methods, each of which serves as a layer of Cellular Automata.

In Multi-layer Cellular Automata (MCA), each cell represents a pixel and the number of all pixels in an image is denoted as H. The saliency values consist of the set of cells' states. Different from the definition of neighborhood in Section 3.2, in Multi-layer Cellular Automata, pixels with the same coordinates in different maps are neighbors. That is, for any cell on a saliency map, it may have $M-1$ neighbors on other maps and we assume that all neighbors have the same influential power to determine the cell's next state. The saliency value of pixel i stands for its probability to be the foreground F, denoted as $P(i \in F) = S_i$, while $1 - S_i$ stands for its possibility to be the background B, denoted as $P(i \in B) = 1 - S_i$. We binarize each map with an adaptive threshold generated by OTSU [32]. The threshold is only related to the initial saliency map and remains the same all the time. The threshold of the m-th saliency map is denoted as γ_m.

If the pixel i is measured as foreground after segmentation, it will be denoted as $\eta_i = +1$ and similarly, $\eta_i = -1$



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5. Effects of the Bayesian integration via Multi-layer Cellular Automata. (a) Input image. (b)-(f) Saliency maps generated respectively by HS [48], DSR [23], MR [49], wCO [52] and our algorithm BSCA. (g) The integrated result S^{N_2}. (h) Ground truth.

represents that it is binarized as background. That a pixel is measured or binarized as foreground doesn't mean that it actually belongs to the foreground because the segmentation may not always be correct. If pixel i belongs to the foreground, the probability that one of its neighboring pixel j (with the same coordinates on another saliency map) is measured as foreground is $\lambda = P(\eta_j = +1|i \in F)$. Correspondingly, the probability $\mu = P(\eta_j = -1|i \in B)$ represents that the pixel j is measured as B under the condition that pixel i belongs to the background. It is reasonable to assume that λ is a constant and the same as μ. Then the posterior probability $P(i \in F|\eta_j = +1)$ can be calculated as follows:

$$P(i \in F|\eta_j = +1) \propto P(i \in F)P(\eta_j = +1|i \in F) = S_i \cdot \lambda \tag{9}$$

In order to get rid of the normalizing constant, we define the prior ratio $\Lambda(i \in F)$ as:

$$\Lambda(i \in F) = \frac{P(i \in F)}{P(i \in B)} = \frac{S_i}{1 - S_i} \tag{10}$$

and then the posterior ratio $\Lambda(i \in F|\eta_j = +1)$ turns into:

$$\Lambda(i \in F|\eta_j = +1) = \frac{P(i \in F|\eta_j = +1)}{P(i \in B|\eta_j = +1)} = \frac{S_i}{1 - S_i} \cdot \frac{\lambda}{1 - \mu} \tag{11}$$

Notice that the first term is the prior ratio and it is easier to deal with the logarithm of Λ because the changes in log-odds $l = ln(\Lambda)$ will be additive. So we have:

$$l(i \in F|\eta_j = +1) = l(i \in F) + ln(\frac{\lambda}{1 - \mu}) \tag{12}$$

In this paper, the prior and posterior ratio $\Lambda(i \in F)$ and $\Lambda(i \in F|\eta_j = +1)$ are also defined as:

$$\Lambda(i \in F) = \frac{S_i^t}{1 - S_i^t},\ \Lambda(i \in F|\eta_j = +1) = \frac{S_i^{t+1}}{1 - S_i^{t+1}} \tag{13}$$

where S_i^t means the saliency value of pixel i at time t. And we define the synchronous updating rule $f : S^{M-1} \to S$ as:

$$l(\boldsymbol{S}_m^{t+1}) = l(\boldsymbol{S}_m^t) + \sum_{\substack{k=1 \\ k \neq m}}^{M} sign(\boldsymbol{S}_k^t - \gamma_k \cdot \mathbf{1}) \cdot ln(\frac{\lambda}{1 - \lambda}) \tag{14}$$

where $\boldsymbol{S}_m^t = [S_{m1}^t, \cdots, S_{mH}^t]^T$ represents the saliency value of all cells on the m-th map at time t, and the matrix $\mathbf{1} = [1, 1, \cdots, 1]^T$ have H elements. Intuitively, if a pixel observes that its neighbors are binarized as foreground, it ought to increase its saliency value. Therefore, Eqn 14 requires $\lambda > 0.5$ and then $ln(\frac{\lambda}{1-\lambda}) > 0$. In this paper, we empirically set $ln(\frac{\lambda}{1-\lambda}) = 0.15$. After N_2 time steps, the final integrated saliency map $\boldsymbol{S}^{N_2}$ is calculated as:

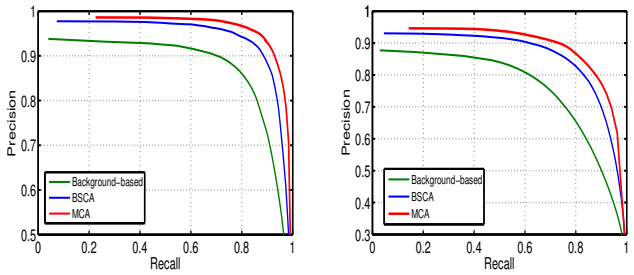
$$\boldsymbol{S}^{N_2} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{S}_m^{N_2} \tag{15}$$

In this paper, we use Multi-layer Cellular Automata to integrate saliency maps generated by HS [48], DSR [23], MR [49], wCO [52] and our algorithm BSCA. From Figure 5 we can clearly see that the detected object on the integrated map is uniformly highlighted and much more close to the ground truth.

5. Experimental Evaluation

We evaluate the proposed method on six standard datasets: ASD [1], MSRA-5000 [25], THUS [6], ECSSD [48], PASCAL-S [24] and DUT-OMRON [49]. ASD is the most widely used dataset and is relatively simple. MSRA-5000 contains more comprehensive images with complex background. THUS is the largest dataset which consists of 10000 images. ECSSD contains 1000 images with multiple objects of different sizes. Some of the images come from the challenging Berkeley-300 dataset. The PASCAL-S dataset ascends from the validation set of PASCAL VOC 2010 [12] segmentation challenge and contains 850 natural images surrounded by complex background. The last Dut-OMRON contains 5168 challenging images with pixelwise ground truth annotations.

We compare our algorithm with the most classic or the newest methods including IT98 [16], FT09 [1], CA10 [13], RC11 [8], XL13 [47], LR12 [36], HS13 [48], UFO13 [20], DSR13 [23], MR13 [49], wCO14 [52]. The results of different methods are provided by authors or achieved by running available codes or softwares. The code of our proposed algorithm can be found at our project site.



(a) (b)

Figure 6. Effects of our proposed algorithm. (a) is PR curves on ASD dataset. (b) is PR curves on MSRA-5000 dataset.

5.1. Parameters and Evaluation Metrics

Implementation Details We set the number of superpixels $N = 300$ in all experiments. In the Single-layer Cellular Automata, the number of time steps $N_1 = 20$. In Multi-layer Cellular Automata, the number of time steps $N_2 = 5$. N_1 and N_2 are determined respectively by the convergence time of Single-layer and Multi-layer Cellular Automata. The results will not change any more once the dynamic systems achieve stabilities.

Evaluation Metrics We evaluate all methods by standard precision-recall curves obtained by binarizing the saliency map with a threshold sliding from 0 to 255 and compare the binary maps with the ground truth. In many cases, high precision and recall are both required and therefore F-measure is put forward as the overall performance measurement:

$$F_\beta = \frac{(1+\beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall} \tag{16}$$

where we set $\beta^2 = 0.3$ as suggested in [1] to emphasize the precision. As complementary to PR curves, we also introduce the mean absolute error (MAE) which calculates the average difference between the saliency map and the ground truth GT in pixel level:

$$MAE = \frac{1}{H}\sum_{h=1}^{H} |S(h) - GT(h)| \tag{17}$$

This measure indicates how similar a saliency map is to the ground truth, and is of great importance for different applications, such as image segmentation and cropping [33].

5.2. Validation of the Proposed Algorithm

To demonstrate the effectiveness of our proposed algorithm, we test the results on the standard dataset ASD and MSRA-5000. PR curves in Figure 6 show that: 1) The background-based maps are already satisfying; 2) Single-layer Cellular Automata can greatly improve the precision of the background-based maps. 3) Results integrated by Multi-layer Cellular Automata are better. Similar results are also achieved on other datasets but not presented here to be succinct.

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

As is shown in Figure 7, our proposed method BSCA performs favorably against existing algorithms with higher precision and recall values on five different datasets. And it has a wider range of high F-measure compared to others. Furthermore, the fairly low MAEs displayed in Table 1 indicate the similarity between our saliency maps and the ground truth. Several saliency maps are shown in Figure 8 for visual comparison of our method with other results.

Improvement of the state-of-the-arts In Section 3.2.4, we conclude that results generated by different methods can be effectively optimized via Single Cellular Automata. PR curves in Figure 7 and MAEs in Table 1 compare various saliency methods and their optimized results on different datasets. Both PR curves and MAEs demonstrate that SCA can greatly improve any existing results to a similar performance level. Even though the original saliency maps are not satisfying, the optimized results are comparable to the state-of-the-arts.

Integration of the state-of-the-arts In Section 4, we propose a novel method to integrate several state-of-the-art methods with top performance. PR curves in Figure 7 strongly prove the effectiveness and robustness of Multi-layer Cellular Automata which outperforms all the existing methods on five datasets. And the F-measure curves of MCA in Figure 7 are fixed at high values which are insensitive to the selective thresholds. In addition, the mean absolute errors of Multi-layer Cellular Automata are always the lowest on different datasets as Table 1 displays. The fairly low mean absolute errors indicate that the integrated results are quite similar to the ground truth. We can observe from Figure 8 that saliency maps generated by MCA are almost the same as the ground truth.

5.4. Run Time

The algorithm BSCA takes on average 0.470s except the time for generating superpixels to process one image from ASD dataset via Matlab with a PC equipped with a i7-4790k 4.00 GHz CPU and 16GB RAM. The main algorithm SCA spends about 0.284s to process one image. Furthermore, Multi-layer Cellular Automata (MCA) only takes on average 0.043s to integrate different methods and can achieve a much better saliency map.

6. Conclusion

In this paper, we propose a novel bottom-up method to construct a background-based map, which takes both global color and spatial distance matrices into consideration. Based upon Cellular Automata, an intuitive updating mechanism is designed to exploit the intrinsic connectivity of salient objects through interactions with neighbors. This context-based propagation can improve any given state-of-the-art results to a similar level with higher ac-

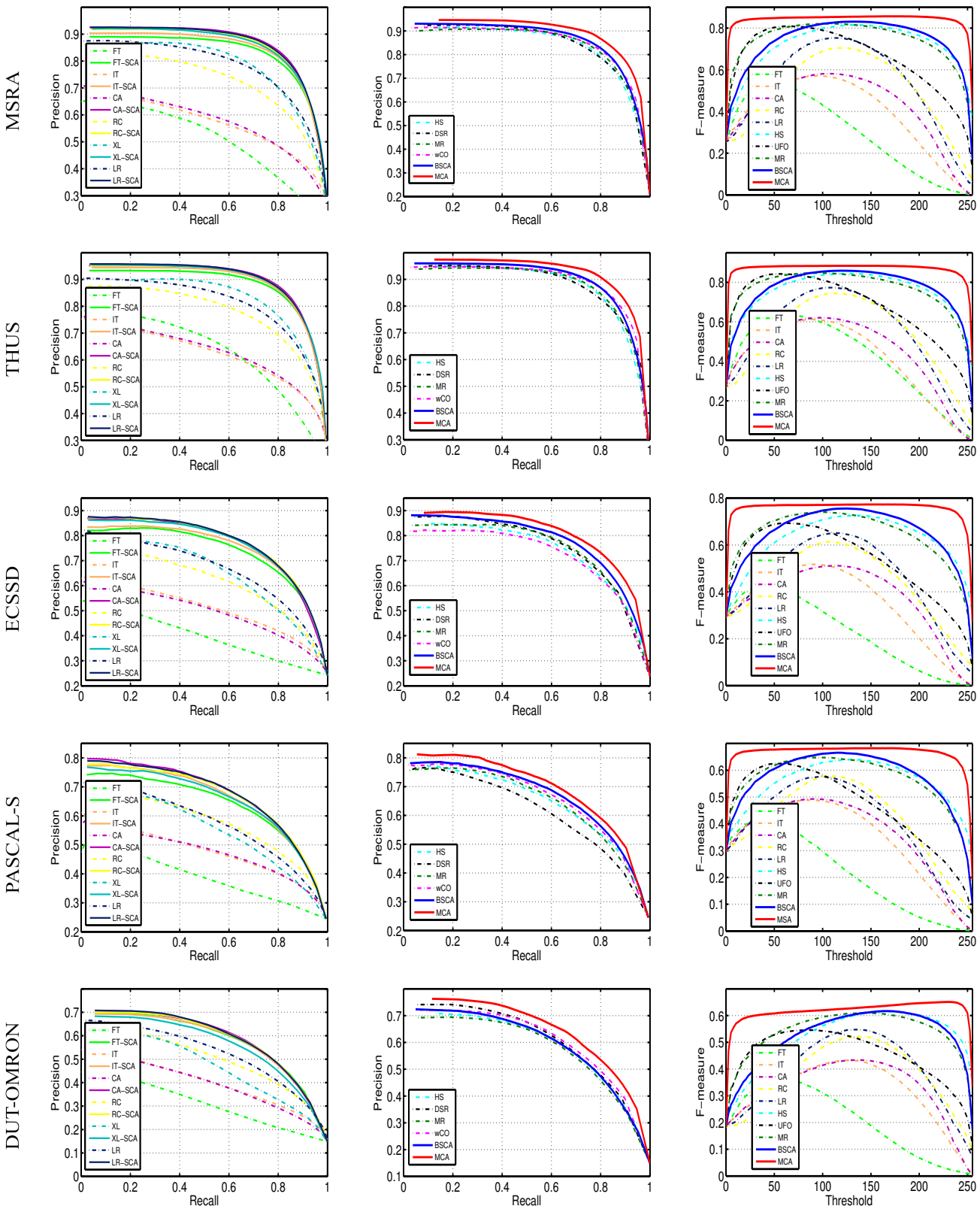
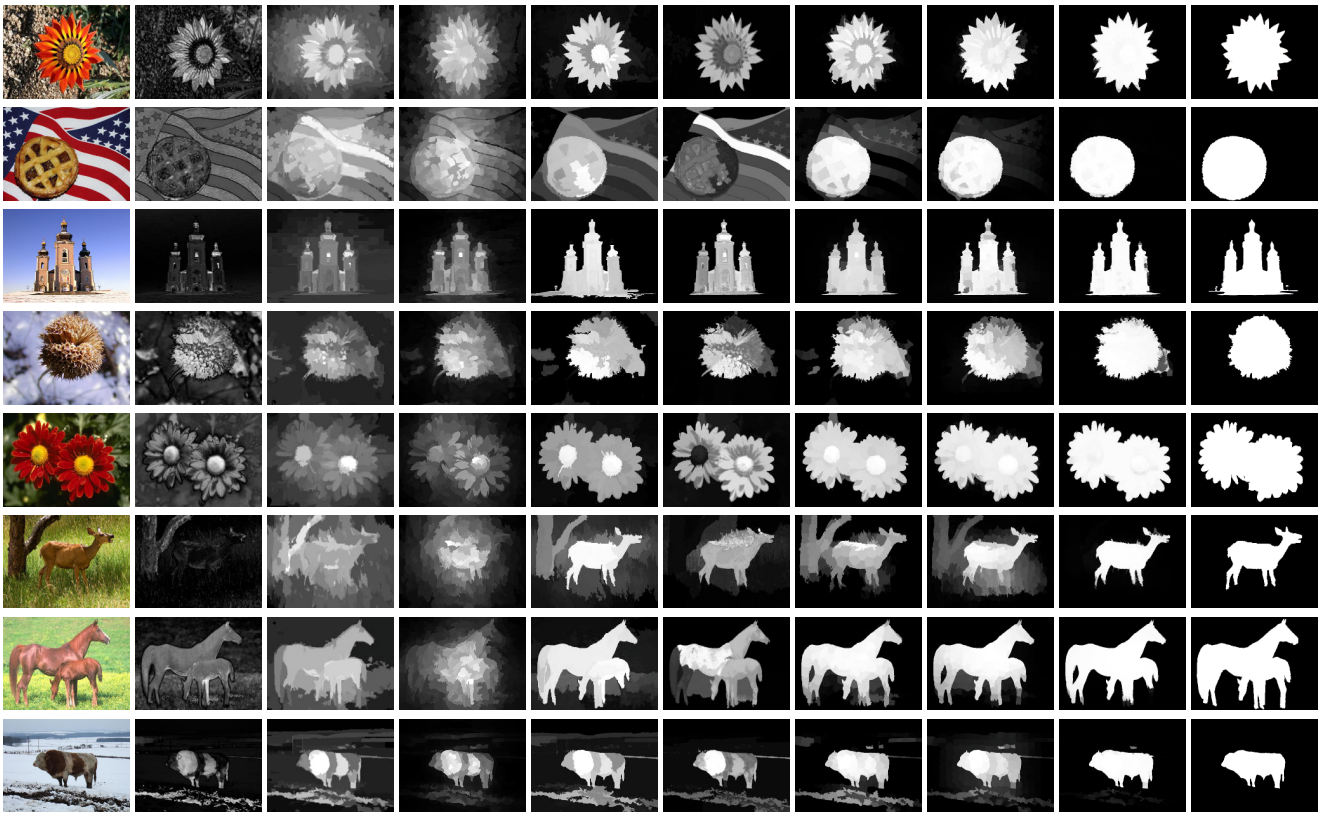


Figure 7. PR curves and F-measure curves of different methods and their optimized version via Single-layer Cellular Automata(-SCA). From top to bottom: MSRA, THUS, ECSSD, PASCAL-S and DUT-OMRON are tested.

	FT[1]	IT[16]	CA[13]	RC[8]	XL[47]	LR [36]	HS[48]	UFO [20]	BSCA	MCA
ASD	0.205	0.235	0.234	0.235	0.137	0.185	0.115	0.110	**0.086**	**0.039**
ASD*	0.083	0.098	0.093	0.084	0.082	0.083	0.073	0.073	-	-
MSRA	0.230	0.249	0.250	0.263	0.184	0.221	0.162	0.146	**0.131**	**0.078**
MSRA*	0.132	0.137	0.134	0.130	0.127	0.128	0.121	0.118	-	-
THUS	0.235	0.241	0.237	0.252	0.164	0.224	0.149	0.147	**0.125**	**0.076**
THUS*	0.128	0.125	0.129	0.125	0.121	0.126	0.117	0.116	-	-
ECSSD	0.272	0.291	0.310	0.302	0.259	0.276	0.228	0.205	**0.183**	**0.134**
ECSSD*	0.188	0.186	0.186	0.182	0.181	0.183	0.179	0.180	-	-
PASCAL-S	0.288	0.298	0.302	0.314	0.289	0.288	0.264	0.233	**0.225**	**0.180**
PASCAL-S*	0.227	0.225	0.225	0.220	0.226	0.223	0.220	0.215	-	-
DUT-OMRON	0.217	0.256	0.255	0.294	0.282	0.264	0.233	**0.180**	0.196	**0.138**
DUT-OMRON*	0.180	0.186	0.189	0.184	0.194	0.185	0.181	0.180	-	-

Table 1. The MAEs of different methods and their optimized versions via Single-layer Cellular Automata. The original results of different methods are displayed in the row of ASD, MSRA, THUS, ECSSD, PASCAL-S and DUT-OMRON. The best two results are shown in **red** and **blue** respectively. And their optimized results are displayed in the row of ASD* , MSRA* , THUS* , ECSSD*, PASCAL-S* and DUT-OMRON*.



(a) Input (b) FT (c) RC (d) LR (e) HS (f) UFO (g) wCO (h) BSCA (i) MCA (j) GT

Figure 8. Comparison of saliency maps on different datasets. BSCA: The background-based maps optimized by Single-layer Cellular Automata. MCA: The integrated saliency maps via Multi-layer Cellular Automata. GT: Ground Truth.

curacy. Furthermore, we propose an integration method named Multi-layer Cellular Automata within the Bayesian inference framework. It can take advantage of the superiorities of different state-of-the-art saliency maps and incorporate them into a more discriminative saliency map with higher precision and recall. Experimental results demonstrate that the superior performance of our algorithms compared to other existing methods.

Acknowledgements. The paper is supported by the Natural Science Foundation of China #61472060 and the Fundamental Research Funds for the Central Universities under Grant DUT14YQ101.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1597–1604. IEEE, 2009. 1, 4, 5, 6, 8

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels. Technical report, 2010. 2

[3] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 73–80. IEEE, 2010. 1

[4] M. Batty. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. The MIT press, 2007. 2

[5] A. Borji, D. N. Sihite, and L. Itti. Salient object detection: A benchmark. In *Computer Vision–ECCV 2012*, pages 414–429. Springer, 2012. 1

[6] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu. Salient object detection and segmentation. *Image*, 2(3):9, 2011. 5

[7] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook. Efficient salient region detection with soft image abstraction. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1529–1536. IEEE, 2013. 1

[8] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global contrast based salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 409–416. IEEE, 2011. 1, 5, 8

[9] B. Chopard and M. Droz. *Cellular automata modeling of physical systems*, volume 24. Cambridge University Press Cambridge, 1998. 2

[10] R. Cowburn and M. Welland. Room temperature magnetic quantum cellular automata. *Science*, 287(5457):1466–1468, 2000. 2

[11] Y. Ding, J. Xiao, and J. Yu. Importance filtering for image retargeting. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 89–96. IEEE, 2011. 1

[12] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 5

[13] S. Goferman and A. L Tal. context-aware saliency detection. *Computer*, 2010. 4, 5, 8

[14] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(10):1915–1926, 2012. 1

[15] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007. 1

[16] 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. 4, 5, 8

[17] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang. Saliency detection via absorbing markov chain. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1665–1672. IEEE, 2013. 1

[18] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li. Automatic salient object segmentation based on context and shape prior. In *BMVC*, volume 3, page 7, 2011. 1

[19] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2083–2090. IEEE, 2013. 1, 2

[20] P. Jiang, H. Ling, J. Yu, and J. Peng. Salient region detection by ufo: Uniqueness, focusness and objectness. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1976–1983. IEEE, 2013. 5, 8

[21] Z. Jiang and L. S. Davis. Submodular salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2043–2050. IEEE, 2013. 1

[22] D. A. Klein and S. Frintrop. Center-surround divergence of feature statistics for salient object detection. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2214–2219. IEEE, 2011. 1

[23] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang. Saliency detection via dense and sparse reconstruction. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 2976–2983. IEEE, 2013. 1, 2, 5

[24] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille. The secrets of salient object segmentation. CVPR, 2014. 5

[25] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(2):353–367, 2011. 5

[26] V. Mahadevan and N. Vasconcelos. Saliency-based discriminant tracking. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1007–1013. IEEE, 2009. 1

[27] L. Marchesotti, C. Cifarelli, and G. Csurka. A framework for visual saliency detection with applications to image thumbnailing. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2232–2239. IEEE, 2009. 1

[28] C. Maria de Almeida, M. Batty, A. M. Vieira Monteiro, G. Câmara, B. S. Soares-Filho, G. C. Cerqueira, and C. L. Pennachin. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems*, 27(5):481–509, 2003. 2

[29] A. C. Martins. Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 19(04):617–624, 2008. 2

[30] J. v. Neumann and A. W. Burks. Theory of self-reproducing automata. 1966. 3

[31] A. Y. Ng, M. I. Jordan, Y. Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002. 1

[32] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975. 4

[33] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 733–740. IEEE, 2012. 1, 6

[34] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä. Segmenting salient objects from images and videos. In *Computer Vision–ECCV 2010*, pages 366–379. Springer, 2010. 2

[35] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 309–314. ACM, 2004. 1

[36] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 853–860. IEEE, 2012. 1, 5, 8

[37] C. Siagian and L. Itti. Rapid biologically-inspired scene classification using features shared with visual attention. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(2):300–312, 2007. 1

[38] A. R. Smith III. Real-time language recognition by one-dimensional cellular automata. *Journal of Computer and System Sciences*, 6(3):233–253, 1972. 3

[39] J. Sun and H. Ling. Scale and object aware image retargeting for thumbnail browsing. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1511–1518. IEEE, 2011. 1

[40] J. Sun, H. Lu, and S. Li. Saliency detection based on integration of boundary and soft-segmentation. In *ICIP*, 2012. 1

[41] N. Tong, H. Lu, Y. Zhang, and X. Ruan. Salient object detection via global and local cues. *Pattern Recognition*, doi:10.1016/j.patcog.2014.12.005, 2014. 1

[42] J. Von Neumann. The general and logical theory of automata. *Cerebral mechanisms in behavior*, 1:41, 1951. 1, 2, 3

[43] L. Wang, J. Xue, N. Zheng, and G. Hua. Automatic salient object extraction with contextual cue. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 105–112. IEEE, 2011. 1

[44] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *Computer Vision–ECCV 2012*, pages 29–42. Springer, 2012. 1, 2

[45] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of modern physics*, 55(3):601, 1983. 3

[46] Y. Xie and H. Lu. Visual saliency detection based on bayesian model. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 645–648. IEEE, 2011. 2

[47] Y. Xie, H. Lu, and M.-H. Yang. Bayesian saliency via low and mid level cues. *Image Processing, IEEE Transactions on*, 22(5):1689–1698, 2013. 2, 5, 8

[48] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1155–1162. IEEE, 2013. 1, 5, 8

[49] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3166–3173. IEEE, 2013. 1, 2, 3, 5

[50] J. Yang and M.-H. Yang. Top-down visual saliency via joint crf and dictionary learning. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2296–2303. IEEE, 2012. 1

[51] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32, 2008. 2

[52] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency optimization from robust background detection. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2814–2821. IEEE, 2014. 1, 2, 5