Saliency Detection via Cellular Automata

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

Cellular Automata is a dynamic system with simple construction but complex self-organizing behaviour [5]. 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. In this paper, we introduce Cellular Automata as a propagation mechanism to intuitively detect the salient object. In addition, considering that some effective algorithms have been proposed in the Bayesian framework [4], we also combine Cellular Automata with Bayesian theory to integrate multiple saliency maps.

Firstly, we apply the K-means algorithm to classify the image border into K different clusters. Based on different boundary clusters, we can construct global color distinction maps and spacial distance maps as:

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

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

where p^k ($k = 1,2,\cdots,K$) is the number of boundary superpixels belonging to cluster k, $\| \boldsymbol{c}_i,\boldsymbol{c}_j \|$ is the Euclidean Distance between the superpixel i and j in CIE LAB color space, $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are the coordinates of the superpixel i and j. Then we can construct the background-based map $\boldsymbol{S}^{bg} = [S_1^{bg},\cdots,S_N^{bg}]^T$ by combining the geodesic information $w_{k,i}$ with the color information $s_{k,i}$:

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

In Single-layer Cellular Automata (SCA), each cell denotes a superpixel generated by the SLIC algorithm. We use the saliency value of each superpixel as its state. A cell's newly defined neighbors include cells surrounding it as well as sharing common boundaries with its adjacent cells. 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:

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

where $\| \boldsymbol{c_i},\boldsymbol{c_j} \|$ denotes the Euclidean Distance in CIE LAB color space between the superpixel i and j, $NB(i)$ is the set of neighbors of cell i. We do a row-normalization to $\boldsymbol{F}$ to achieve $\boldsymbol{F}^*$.

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 $\boldsymbol{C}^* = diag\{c_1^*,c_2^*,\cdots,c_N^*\}$ by the formulation as:

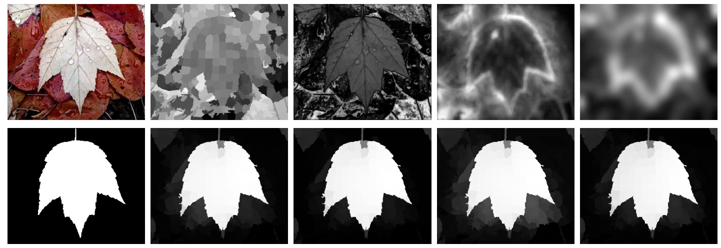
$$c_i^* = a\cdot\frac{c_i - min(c_j)}{max(c_j)-min(c_j)}+b \quad (5)$$

where $c_i = \frac{1}{max(f_{ij})}$ and $j = 1,2,\cdots,N$.

In Single-layer Cellular Automata, all cells update their states simultaneously according to the updating rule. In this paper, we define the synchronous updating rule $f : S^{NB} \to S$ as follows:

$$\boldsymbol{S}^{t+1} = \boldsymbol{C}^*\cdot\boldsymbol{S}^t + (\boldsymbol{I}-\boldsymbol{C}^*)\cdot\boldsymbol{F}^*\cdot\boldsymbol{S}^t \quad (6)$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{C}^*$ and $\boldsymbol{F}^*$ are coherence matrix and impact factor matrix respectively. We use our background-based map and several classic methods as the prior maps $S^{t=0}$ and refresh them according to the synchronous updating rule. The optimized results via SCA are shown in Figure 1. 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. That means our method is independent to prior maps.



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

Figure 1: Comparison of different methods and their optimized version after parallel evolutions. (a) The first row are input images. The second row are the ground truth. (b)-(e) The first row are the original results of different methods. From left to right: our background-based maps, saliency maps generated by FT [1], CA [2], IT [3]. The second row are their optimized results via Single-layer Cellular Automata.

In order to take advantage of the superiority of different saliency detection methods, we propose an effective algorithm named Multi-layer Cellular Automata (MCA) to incorporate M saliency maps generated by M state-of-the-art methods. In MCA, each cell represents a pixel and pixels with the same coordinates in different maps are neighbors. We define the synchronous updating rule $f : S^{M-1} \to S$ under the Bayesian framework as:

$$l(\boldsymbol{S}_m^{t+1}) = l(\boldsymbol{S}_m^t) + \sum_{\substack{k=1\\k\neq m}}^{M} sign(\boldsymbol{S}_k^t - \gamma_k\cdot\boldsymbol{1})\cdot ln(\frac{\lambda}{1-\lambda}) \quad (7)$$

where $\boldsymbol{S}_m^t$ represents the saliency value of all cells on the m-th map at time t and $l(\boldsymbol{S}_m^t) = ln(\frac{\boldsymbol{S}_m^t}{1-\boldsymbol{S}_m^t})$, γ_m denotes the adaptive threshold of the m-th saliency map. If the pixel i belongs to the foreground, the probability that one of its neighboring pixel j is binarized as foreground is denoted as λ. Intuitively, if a pixel observes that its neighbors are binarized as foreground, it ought to increase its saliency value. Therefore, Eqn 7 requires $\lambda > 0.5$ and then $ln(\frac{\lambda}{1-\lambda}) > 0$.

The iteration numbers of Single-layer Cellular Automata and Multi-layer Cellular Automata are determined respectively by the convergence time of the dynamic systems. The saliency map will not change any more once the system achieves the stability. Extensive experiments on six public datasets demonstrate that the proposed algorithm outperforms state-of-the-art methods.

[1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1597–1604. IEEE, 2009.

[2] S Goferman and A L Tal. context-aware saliency detection. *Computer*, 2010.

[3] Laurent Itti, Christof Koch, and Ernst 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] Xiaohui Li, Huchuan Lu, Lihe Zhang, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via dense and sparse reconstruction. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 2976–2983. IEEE, 2013.

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

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.