Saliency Propagation from Simple to Difficult

Chen Gong[1,2], Dacheng Tao[2], Wei Liu[3], S.J. Maybank[4], Meng Fang[2], Keren Fu[1], Jie Yang[1]
[1]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. [2]The Centre for Quantum Computation & Intelligent Systems, University of Technology, Sydney. [3]IBM T. J. Watson Research Center. [4]Birkbeck College, London.

Saliency detection aims to identify the most attractive object in an image. Recently, propagation methods [1, 2, 3] have gained much popularity in bottom-up saliency detection and achieved state-of-the-art performance. All these methods generate similar propagation sequences which are completely governed by the spatial relationship of image regions. In other words, unlabelled superpixels will be assigned saliency values as long as they are connected to the labelled superpixels. However, once encountering the inhomogeneous or incoherent adjacent superpixels, the conventional propagation sequences are misleading and likely to yield inaccurate results.

In this paper, we attempt to manipulate the propagation sequence for optimizing the propagation quality. Intuitively, we postpone the propagations to difficult regions and meanwhile advance the propagations to less ambiguous simple regions. A novel propagation algorithm employing the **T**eaching-to-**L**earn and **L**earning-to-**T**each (abbreviated as "TLLT") strategies (see Fig. 1) is proposed so that the unlabelled superpixels are logically propagated from simple to difficult. In the teaching-to-learn step, a teacher is designed to arrange the regions from simple to difficult and then assign the simplest regions to the learner. In the learning-to-teach step, the learner delivers its learning feedback to the teacher to assist the teacher to choose the proper subsequent simple regions.
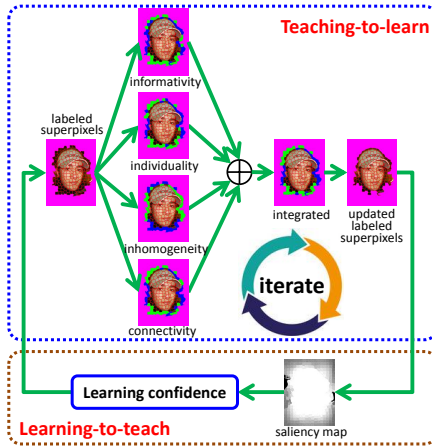

Figure 1: An illustration of our TLLT paradigm. In the teaching-to-learn step, based on a set of labeled superpixels (magenta) in an image, the teacher discriminates the adjacent unlabeled superpixels as difficult (blue superpixels) or simple (green superpixels) by fusing their informativity, individuality, inhomogeneity, and connectivity. Then simple superpixels are learned by the learner, and the labeled set is updated correspondingly. In the learning-to-teach step, the learner provides a learning feedback to the teacher to help decide the next curriculum.

Specifically, in the teaching-to-learn step, the teacher evaluates the propagation difficulty DS_i of an unlabeled superpixel s_i by considering its informativity INF_i, individuality IND_i, inhomogeneity IHM_i, and connectivity CON_i, namely:

$$DS_i = INF_i + IND_i + IHM_i + CON_i, \quad (1)$$

where

$$INF_i = H(s_i|\mathcal{L}) = \frac{1}{2}\ln(2\pi e \sigma_{i|\mathcal{L}}^2), \quad (2)$$

$$IND_i = IND(s_i, \mathcal{N}(s_i)) = \frac{1}{|\mathcal{N}(s_i)|}\sum_{j\in\mathcal{N}(s_i)}\left\|s_i^{color} - s_j^{color}\right\|, \quad (3)$$

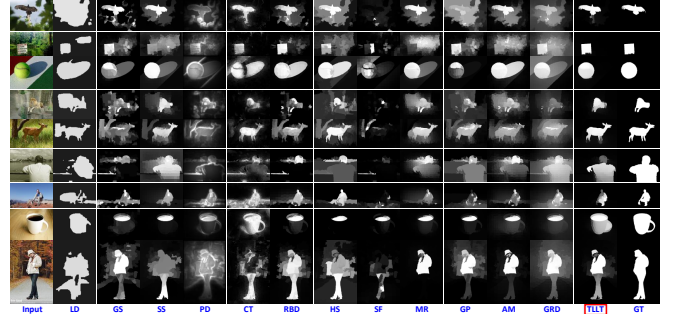$$IHM_i = \left(\frac{2}{b^2-b}\sum_{i=1}^{b}\sum_{j=i+1}^{b}Q_{ij}\right)^{-1}, \quad (4)$$

Figure 2: Visual comparisons of saliency maps generated by all the methods on some challenging images. The ground truth (GT) is presented in the last column.

$$CON_i = \frac{1}{l}\sum_{j\in\mathcal{L}}geo(s_i, s_j). \quad (5)$$

In (2), the informativity of a superpixel s_i is modelled by the conditional entropy $H(s_i|\mathcal{L})$, where $\mathcal{L}$ denotes the labelled set and $\sigma_{i|\mathcal{L}}^2$ denotes the conditional covariance of s_i's associated random variable f_i given $\mathcal{L}$. In (3), $\mathcal{N}(s_i)$ represents the neighboring superpixels of s_i and $|\mathcal{N}(s_i)|$ computes the amount of neighbors. In (4), the matrix Q encodes the pairwise correlations of b pixels contained by s_i in the LAB color feature space. In (5), $geo(s_i, s_j)$ represents the geodesic distance between s_i and s_j, and l is the number of labelled superpixels.

After the difficulty scores of all candidate superpixels are computed, the next step is to pick up the simplest $q^{(t)}$ superpixels as the t-th curriculum. This is completed by the learning-to-teach step. In detail, $q^{(t)}$ should be adjusted by considering the effect of previous learning. Therefore, by denoting $f_i^{(t-1)}$ as the s_i's saliency value after the $(t-1)$-th propagation, we define a confidence score to blindly evaluate the $(t-1)$-th learning performance, which is formulated as

$$ConfidenceScore = 1 - \frac{2}{q^{(t-1)}}\sum_{i=1}^{q^{(t-1)}}\min(f_i^{(t-1)}, 1-f_i^{(t-1)}), \quad (6)$$

and $q^{(t)}$ is finally computed by

$$q^{(t)} = \left\lceil\left|\mathcal{C}^{(t)}\right| \times ConfidenceScore\right\rceil, \quad (7)$$

where $\left|\mathcal{C}^{(t)}\right|$ is the number of candidate superpixels for the t-th propagation.

After the curriculum $\mathcal{T}^{(t)} = \left\{s_1, s_2, \cdots, s_{q^{(t)}}\right\}$ is specified, the learner will propagate the saliency values from $\mathcal{L}^{(t)}$ to $\mathcal{T}^{(t)}$ via the following expression:

$$f^{(t+1)} = M^{(t)}D^{-1}Wf^{(t)}, \quad (8)$$

where W is the adjacency matrix of the similarity graph between superpixels, and D is the diagonal degree matrix with $D_{ii} = \sum_j W_{ij}$. $M^{(t)}$ is a diagonal matrix with $M_{ii}^{(t)} = 1$ if $s_i \in \mathcal{L}^{(t)}\cup\mathcal{T}^{(t)}$, and $M_{ii}^{(t)} = 0$ otherwise. $f^{(t)} = \begin{pmatrix} f_1^t & \cdots & f_N^t \end{pmatrix}^T$ is an N-dimensional vector (N is the total amount of superpixels in the image) including the saliency values of all the superpixels. When the t-th iteration is completed, the labeled and unlabeled sets are updated as $\mathcal{L}^{(t+1)} = L^{(t)}\cup\mathcal{T}^{(t)}$ and $\mathcal{U}^{(t+1)} = \mathcal{U}^{(t)}\setminus\mathcal{T}^{(t)}$, respectively.

we compare the proposed TLLT with twelve representative methods on two popular saliency datasets. The results presented in Fig. 2 reveal that compared with existing methods TLLT renders more confident saliency maps with higher background suppression, yielding a better popping out of objects of interest.

[1] V. Gopalakrishnan, Y. Hu, and D. Rajan. Random walks on graphs to model saliency in images. In *CVPR*, 2009.

[2] Z. Ren, Y. Hu, L. Chia, and D. Rajan. Improved saliency detection based on superpixel clustering and saliency propagation. In *ACM MM*, 2010.

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