# Image Co-Saliency Detection via Locally Adaptive Saliency Map Fusion

{Chung-Chi Tsai[1,2], Xiaoning Qian[1] and Yen-Yu Lin[2]} Texas A&M University[1] & Academia Sinica[2]

## Objective

The performance of co-saliency detection substantially relies on the explored visual cues. However, the optimal cues typically vary from region to region. Thus, we develop an approach considering

1. *intra-image saliency*,
2. *inter-image correspondence*,
3. *spatial consistency*,

and accomplish saliency detection with **locally adaptive saliency map fusion** via solving an energy optimization problem over a graph.

## Introduction

Different visual cues help provide complementary saliency information. To effectively integrate multiple cues, modern approaches carry out co-saliency detection by integrating multiple saliency maps via

1. *fixed-weight map-wise summation* [1],
2. *fixed-weight map-wise multiplication* [2],
3. *self-adaptive map-wise summation (SACS)* [3].

The above approaches neglect the fact that the goodness of saliency map is often region-dependent. Thus, our approach carry out **self-adaptive region-wise summation**, which yields better saliency maps with less false alarms and misses.

Imgs | GT | Intra-img cues [4] | Inter-img cues [1] | SACS [3] | Ours

## Proposed Approach

We formulate this into a constrained QP problem through the following steps:

1. Decompose images $I_1$ and $I_2$ into $N$ total superpixels.
2. Extraction features on superpixels and connect them into a joint graph.
3. Obtain saliency proposals by $M$ existing saliency detection methods.
4. Derive the optimal fusion weights $Y = [\mathbf{y}_1\ \mathbf{y}_2\ \dots\ \mathbf{y}_N] \in \mathbb{R}^{M \times N}$ via

$$\min_{Y} \quad \lambda_1 \sum_{v_i \in \mathcal{V}} U(\mathbf{y}_i) + \lambda_2 \sum_{v_i \in \mathcal{V}} V(\mathbf{y}_i) + \lambda_3 \sum_{e_{ij} \in \mathcal{E}} B(\mathbf{y}_i, \mathbf{y}_j) + \|Y\|_2^2 \qquad (1)$$

$$\text{s.t.} \quad \|\mathbf{y}_i\|_1 = 1, \mathbf{y}_i \geq \mathbf{0}, \ \text{for } 1 \leq i \leq N,$$

where $\mathbf{0}$ is a vector whose elements are zero, and $\lambda_1$, $\lambda_2$ and $\lambda_3$ are three positive constants. There are four terms introduced in Eq. (1).

(a) $U(\mathbf{y}_i)$ : This term leverages saliency map consensus to estimate the goodness of each saliency proposal on superpixel $i$.

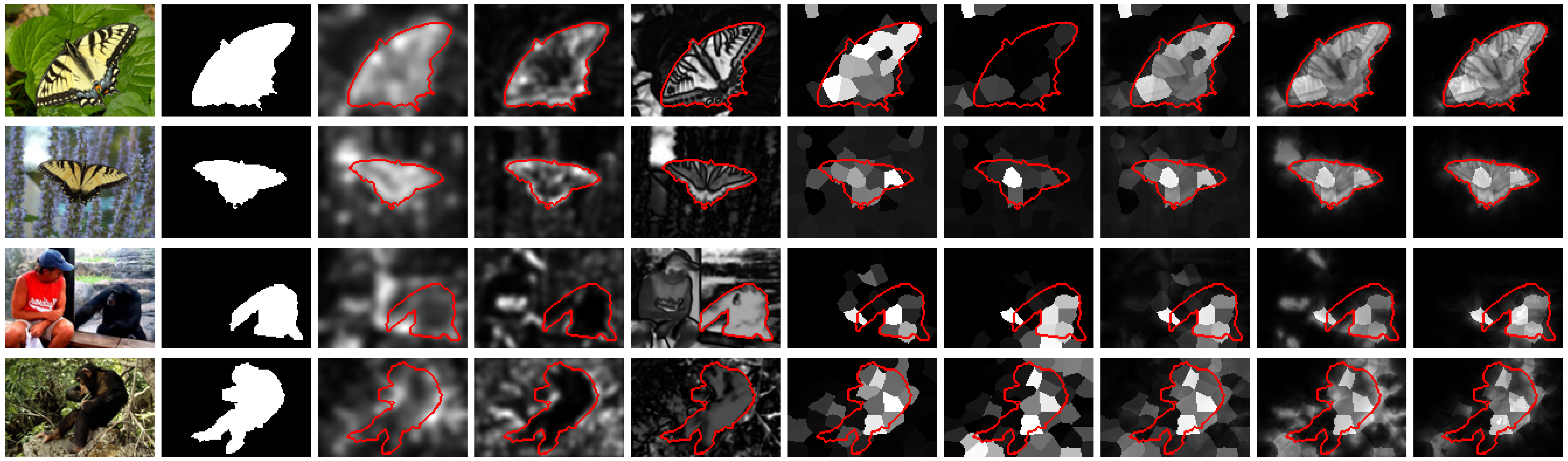
(b) $V(\mathbf{y}_i)$ : This term favors a saliency proposal on superpixel $i$ if the regional mean saliency is proportional to the inter-image correspondence.

(c) $B(\mathbf{y}_i, \mathbf{y}_j)$ : This term smooths the derived fusion weights for every connected superpixels.

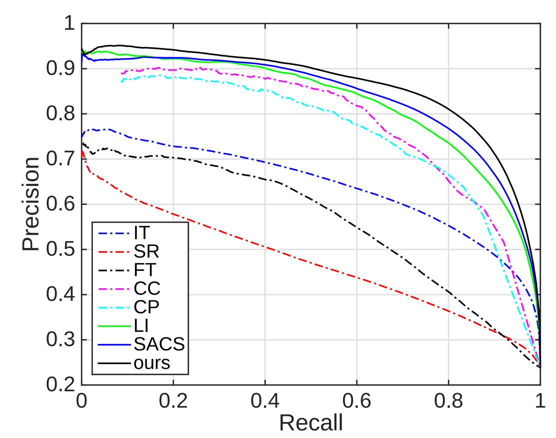
(d) $\|Y\|_2^2$ : This term avoids overconfidence on any saliency proposal.

## Experimental Result

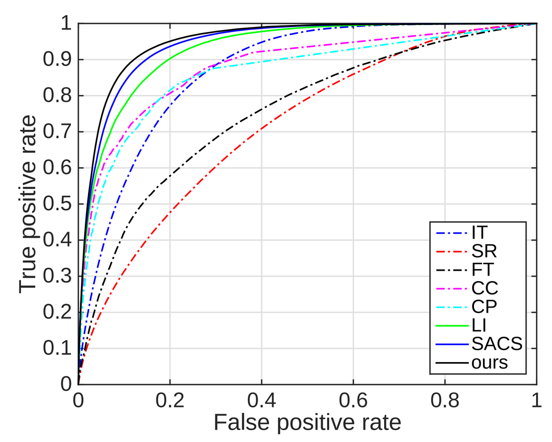
We evaluate on the Image Pair dataset [1] consisting of 105 image pairs by three different performance metrics (a) PR (b) ROC and (C) MAE.



Images | GT | IT [5] | SR [7] | FT [6] | CC [1] | CP [1] | LI [1] | SACS [3] | Ours




(a) PR curve




(b) ROC curve

| Method | IT [5] | SR [7] | FT [6] | CC [1] | CP [1] | LI [1] | SACS [3] | Ours |
|---|---|---|---|---|---|---|---|---|
| PR AUC | 0.640 | 0.471 | 0.559 | 0.702 | 0.681 | 0.824 | 0.836 | **0.861** |
| ROC AUC | 0.872 | 0.718 | 0.756 | 0.881 | 0.865 | 0.930 | 0.944 | **0.952** |
| MAE | 0.259 | 0.269 | 0.253 | 0.163 | 0.173 | 0.173 | 0.172 | **0.163** |

**Figure 1:** The performance of various approaches in 1) AUC of PR, 2) AUC of ROC, and 3) MAE. The higher the better in the first two measures. The lower the better in MAE.

## References


[1] Hongliang Li and King Ngi Ngan. A co-saliency model of image pairs. *IEEE Trans. on Image Processing*, 2011.

[2] Huazhu Fu, Xiaochun Cao, and Zhuowen Tu. Cluster-based co-saliency detection. *IEEE Trans. on Image Processing*, 2013.

[3] Xiaochun Cao, Zhiqiang Tao, Bao Zhang, Huazhu Fu, and Wei Feng. Self-adaptively weighted co-saliency detection via rank constraint. *IEEE Trans. on Image Processing*, 2014.

[4] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun. Saliency optimization from robust background detection. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2014.

[5] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998.

[6] Ravi Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2009.

[7] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2007.


## Contact Information


Name: Chung-Chi "Charles" Tsai
Email: chungchi@tamu.edu


## Acknowledgement

This work was partially supported by Award #1547557 from the National Science Foundation, Grants MOST 104-2628-E-001-001-MY2 and MOST 105-2221-E-001-030-MY2 from the Ministry of Science and Technology.