Segmentation Guided Local Proposal Fusion for Co-saliency Detection

Chung-Chi Tsai[1,2] Xiaoning Qian[1] Yen-Yu Lin[2]

[1] Texas A&M University [2]Academia Sinica

July 11, 2017

Introduction

Co-saliency is a weakly supervised extension of saliency detection by referencing inter-image cues in a set of images. Our paper addresses two issues hindering existing fusion-based image co-saliency detection

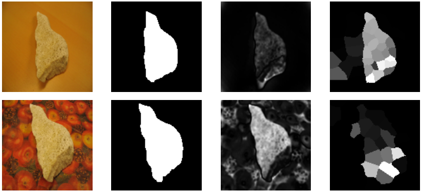
i. It has been shown that (co-)saliency fusion can generate stronger prediction. However, the optimal saliency proposal is region dependent [1], and the fusion process leads to blurred results.

ii. It has been shown that segmentation revealed "objectness" help recover sharp boundaries of salient objects. However, segmentation may suffer from significant intra-object variations.

In fact, "object segmentation" and "region-wise proposal fusion" can complement each other with our proposed unified optimization approach.

[1] Tsai *et al.*, "Image Co-saliency Detection via Locally Adaptive Saliency Map Fusion," in ICASSP 2017.

Introduction



Images | GT | Intra-img cue [2] | Inter-img cue [3]

i. Intra-image cue is from the color difference to the mean color, thus
   - The 1st stone is missing.
   - False alarm shows on the background of the 2nd stone.

ii. Inter-image cue is from the regional color similarity across images, thus
   - The 1st input shows false alarm due to similar background color.
   - The brighter side of the 2nd stone is missing.

iii.

iv.

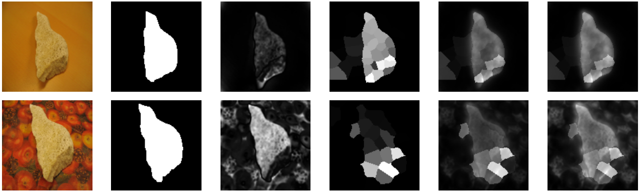
v.

---

[2] Achanta *et al.*, "Frequency-tuned salient region detection," in CVPR 2009.

[3] H. Li and K. N. Ngan, "A co-saliency model of image pairs," TIP 2011.

Introduction



Images | GT | Intra-img cue [4] | Inter-img cue [5] | Map-wise fusion [6] | Region-wise fusion

i. Intra-image cue is from the color difference to the mean color, thus
   - The 1st stone is missing.
   - False alarm shows on the background of the 2nd stone.

ii. Inter-image cue is from the regional color similarity across images, thus
   - The 1st input shows false alarm due to similar background color.
   - The brighter side of the 2nd stone is missing.

iii. Map-wise fusion gives better prediction via both the intra- and inter- cues.

iv. Our proposed region-wise fusion recovers the whole object region.

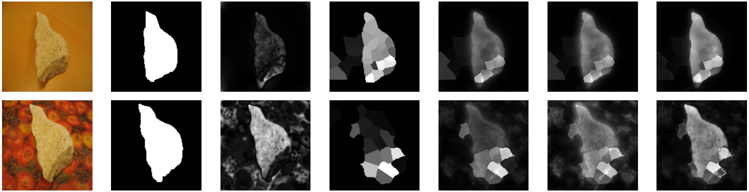
v.

[4] Achanta *et al.*, "Frequency-tuned salient region detection," in CVPR 2009.

[5] H. Li and K. N. Ngan, "A co-saliency model of image pairs," TIP 2011.

[6] Cao *et al.*, "Self-adaptively weighted co-saliency detection via rank constraint," TIP 2014.

Introduction



i. Intra-image cue is from the color difference to the mean color, thus
   - The 1st stone is missing.
   - False alarm shows on the background of the 2nd stone.
ii. Inter-image cue is from the regional color similarity across images, thus
   - The 1st input shows false alarm due to similar background color.
   - The brighter side of the 2nd stone is missing.
iii. Map-wise fusion gives better prediction via both the intra- and inter- cues.
iv. Our proposed region-wise fusion recovers the whole object region.
v. Segmentation guided fusion gives less false positive and sharper results.

[7] Achanta *et al.*, "Frequency-tuned salient region detection," in CVPR 2009.
[8] H. Li and K. N. Ngan, "A co-saliency model of image pairs," TIP 2011.
[9] Cao *et al.*, "Self-adaptively weighted co-saliency detection via rank constraint," TIP 2014.

Model Flowchart - Image preprocessing

Image preprocessing composed of two steps,

- Collect a set of (co-)saliency proposals (upper block).
- Superpixel extraction and graph construction (lower block).

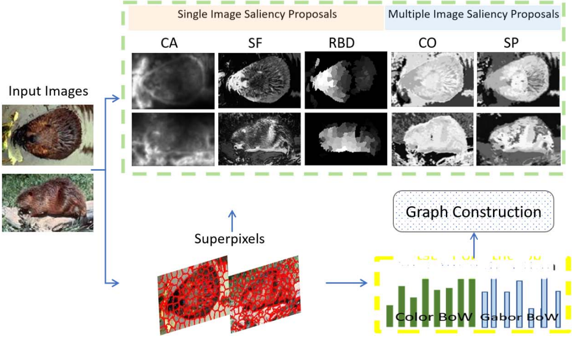


Figure 1: Model Flowchart

Model Flowchart - Co-saliency fusion

Conduct the locally adaptive saliency map fusion.

- Different parts of the object are more uniformly highlighted after the fusion.
- The objects' prior can be used for the image co-segmentation.

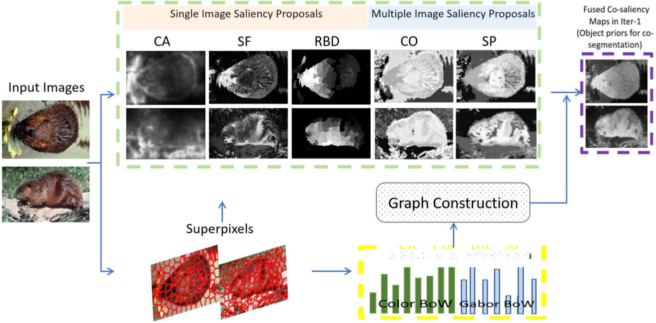


Figure 2: Model Flowchart

Model Flowchart - Co-segmentation

Conduct the image co-segmentation.

- The objectness evidence from co-segmentation provides effective guidance for the co-saliency fusion in the next iteration.

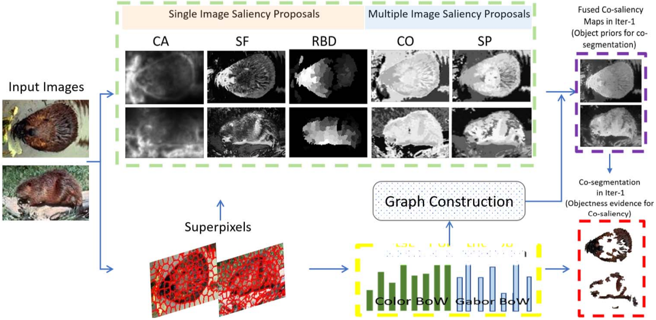


Figure 3: Model Flowchart

Model Flowchart - Alternative optimization

Through alternatively optimizing the co-saliency and co-segmentation process,

- Objectness priors are iteratively refined and fed back to guide the fusion.
- Better saliency maps gives better figure-background model for co-segmentation.

In the end, both tasks converge to a good point, thus no post-processing is required!

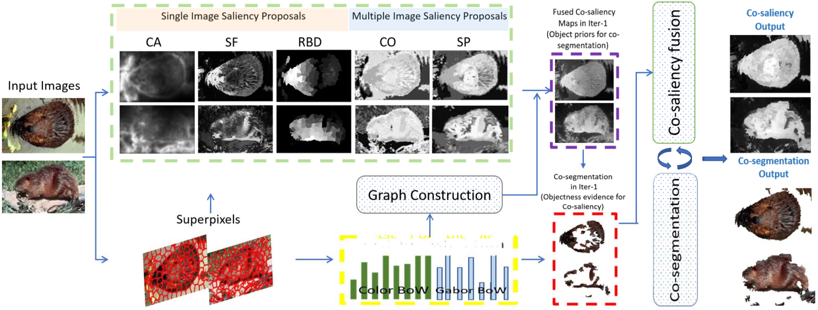


Figure 4: Model Flowchart

Progressive Improvement

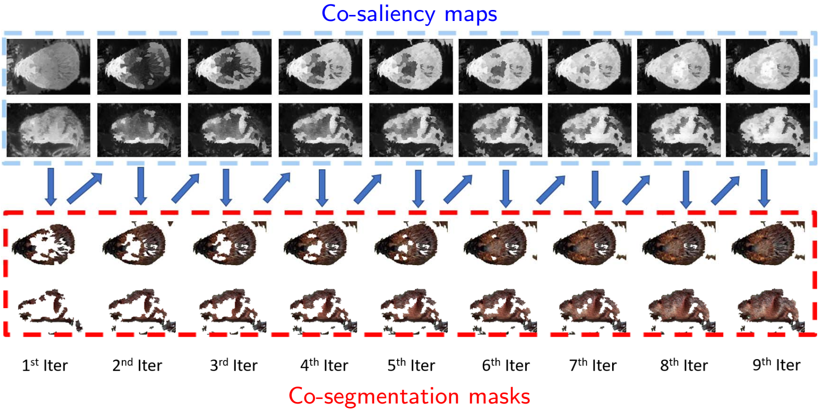


Figure 5: Co-saliency Detection and Co-segmentation results at different iteration

Progressive Improvement

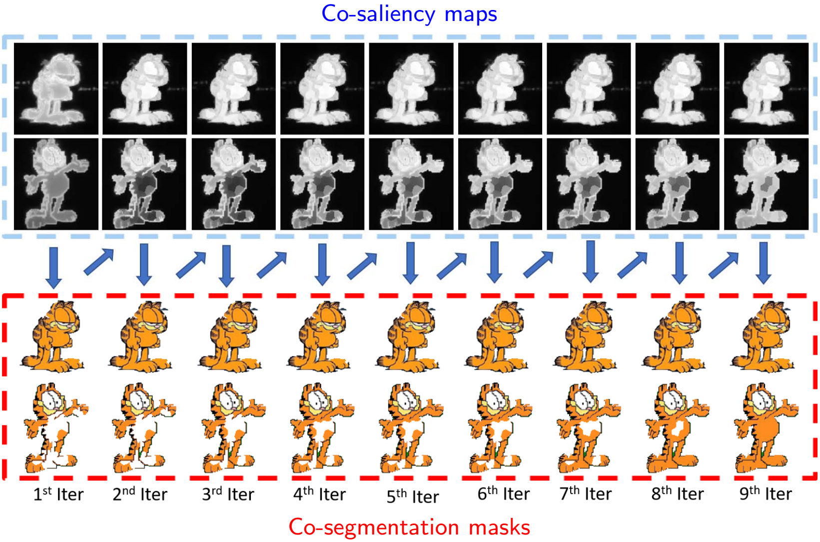


Figure 6: Co-saliency Detection and Co-segmentation results at different iteration

Proposed Approach - Optimization function

By alternatively minimizing the following energy function, we seek the optimal

i. weights $Y = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_N] \in \mathbb{R}^{M \times N}$ for superpixel-wise map fusion

ii. figure-background configuration $Z = [z_1\ z_2\ \ldots\ z_N] \in \mathbb{R}^N$ of co-segmentation.

$$\begin{aligned} J(Y, Z) = {} & \alpha_1 \sum_{v_i \in \mathcal{V}} U_1(\mathbf{y}_i) + \alpha_2 \sum_{v_i \in \mathcal{V}} U_2(z_i) + \alpha_3 \sum_{v_i \in \mathcal{V}} U_3(\mathbf{y}_i, z_i) \\ & + \beta_1 \sum_{e_{ij} \in \mathcal{E}} B_1(\mathbf{y}_i, \mathbf{y}_j) + \beta_2 \sum_{e_{ij} \in \mathcal{E}} B_2(z_i, z_j) + ||Y\|_2^2 \qquad (1) \\ \text{s.t.} \quad & \| \mathbf{y}_i \|_1 = 1,\ \mathbf{y_i} \geq \bar{\mathbf{0}}, z_i \in \{0, 1\},\ \text{for } 1 \leq i \leq N. \end{aligned}$$

$\bar{\mathbf{0}}$ is a zero vector, and α_1, α_2, α_3, β_1 and β_2 are five positive constants. Binary variable $z_i = 1$ if superpixel i belongs to the foreground, and 0 otherwise.

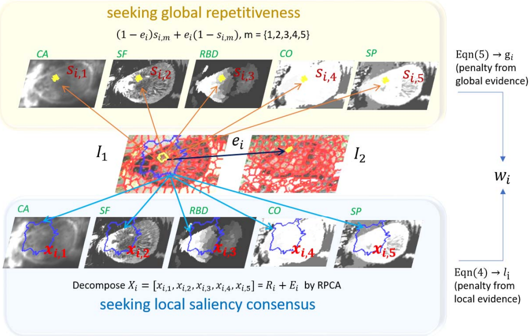
- $U_1(\mathbf{y}_i)$ and $B_1(\mathbf{y}_i, \mathbf{y}_j)$: unary and pairwise terms for co-saliency detection.
- $U_2(z_i)$ and $B_2(z_i, z_j)$: unary and pairwise terms for co-segmentation.
- $U_3(\mathbf{y}_i, z_i)$: This coupling term encourages coherence between the co-saliency map and the figure-ground segmentation.
- $||Y\|_2^2$: regularization term.

Proposed Approach - Co-saliency unary term $U_1(\mathbf{y}_i)$

– We abide to the fundamental co-saliency formula to design our unary term

$$\textit{Co-saliency} = \textit{Saliency} \times \textit{Repetitiveness}.$$

– Fusion weight $\mathbf{w_i}$ for each superpixel v_i on different saliency maps is computed from (1) local saliency consensus $\mathbf{l_i}$ (2) global repetitiveness cue $\mathbf{g_i}$.



– Considering all superpixels, the unary term becomes

$$\sum_{v_i \in \mathcal{V}} U_1(\mathbf{y}_i) = \sum_{i=1}^{N} \mathbf{w_i}^\top \mathbf{y_i} = \mathbf{tr}(\mathbf{W}^\top \mathbf{Y}), \tag{2}$$

where $\mathbf{w}_i = [w_{i,1} \ \ldots \ w_{i,M}]^\top$ and $W = [\mathbf{w_1} \ \ldots \ \mathbf{w_N}]$.

Proposed Approach - Co-segmentation unary term $U_2(z_i)$

– This term estimates the likelihood of superpixel v_i belonging to the common foreground in co-segmentation.

– Let superpixel v_i be represented by mean RGB color, *Gaussian mixture model* (GMM) is used to fit to the superpixels that are currently labeled as the foreground (F) and the background (B) superpixels of $I_k, k \in \{1, 2\}$.

$$\sum_{v_i \in \mathcal{V}} U_2(z_i) = \sum_{i=1}^{N} [p(v_i \in F|\mathbf{c_i})(1 - z_i) + p(v_i \in B|\mathbf{c_i}) z_i]. \quad (3)$$

– GMM $\theta_{\mathbf{f}}$ and $\theta_{\mathbf{b,k}}$ help predict the probability of superpixel i belonging to the foreground or background. Assuming $p(v_i \in F) = p(v_i \in B) = \frac{1}{2}$,

$$p(v_i \in F|\mathbf{c_i}) = \frac{p(\mathbf{c_i} \in F|\theta_{\mathbf{f}})p(v_i \in F)}{p(\mathbf{c_i}|\theta_{\mathbf{f}})p(v_i \in F) + \sum_{k=1}^{2} p(\mathbf{c_i}|\theta_{\mathbf{b,k}})\delta(v_i \in I_k)p(v_i \in B)}$$

, where $p(\cdot|\theta_{\mathbf{f}})$ and $p(\cdot|\theta_{\mathbf{b,k}})$ are the Gaussian probability distributions. And, $p(v_i \in B|\mathbf{c_i})$ is similarly set.

Proposed Approach - Binary term $B_1(\mathbf{y_i}, \mathbf{y_j})$ and $B_2(z_i, z_j)$

- $B_1(\mathbf{y_i}, \mathbf{y_j})$: This term encourages smooth weights Y between the connected superpixels in graph $\mathcal{G}$. It is defined as

$$\sum_{e_{ij} \in \mathcal{E}} B(\mathbf{y}_i, \mathbf{y}_j) = \sum_{e_{ij} \in \mathcal{E}} A(i,j) \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 = tr(YLY^\top), \tag{4}$$

- $B_2(z_i, z_j)$: This term enforces the spatial smoothness of co-segmentation results. It is defined as

$$\sum_{e_{ij} \in \mathcal{E}} B_2(z_i, z_j) = \sum_{e_{ij} \in \mathcal{E}} A(i,j) \|z_i - z_j\|_2^2 = tr(ZLZ^\top). \tag{5}$$

- L is the graph Laplacian of $\mathcal{G}$ with affinity matrix A.

Proposed Approach - Coupling term $U_3(\mathbf{y}_i, z_i)$

– This term encourages the coherence between the co-saliency maps and the co-segmentation result.

– Let s_i be mean saliency value of the fused map on superpixel v_i, and is represented as

$$s_i = \sum_{m=1}^{M} y_{i,m} s_{i,m} = \mathbf{y_i}^\top \mathbf{s_i}, \tag{6}$$

– This term penalizes the cases where one of s_i and z_i is large while the other is small, is defined as

$$\sum_{v_i \in \mathcal{V}} U_3(\mathbf{y_i}, z_i) = \sum_{i=1}^{N} s_i(1 - z_i) + (1 - s_i) z_i. \tag{7}$$

Proposed Approach - Optimization process (Co-saliency Detection)

Our proposed optimization function

$$J(Y,Z) = \alpha_1 \sum_{v_i \in \mathcal{V}} U_1(\mathbf{y}_i) + \alpha_2 \sum_{v_i \in \mathcal{V}} U_2(z_i) + \alpha_3 \sum_{v_i \in \mathcal{V}} U_3(\mathbf{y}_i, z_i)$$
$$+ \beta_1 \sum_{e_{ij} \in \mathcal{E}} B_1(\mathbf{y}_i, \mathbf{y}_j) + \beta_2 \sum_{e_{ij} \in \mathcal{E}} B_2(z_i, z_j) + ||Y||_2^2 \quad (8)$$
$$\text{s.t.} \quad \| \mathbf{y}_i \|_1 = 1,\ \mathbf{y}_i \geq \bar{\mathbf{0}}, z_i \in \{0,1\},\ \text{for } 1 \leq i \leq N.$$

By fixing Z, the optimization problem in (10) becomes

$$J(Y) = \alpha_1 \sum_{v_i \in \mathcal{V}} U_1(\mathbf{y}_i) + \beta_1 \sum_{e_{ij} \in \mathcal{E}} B_1(\mathbf{y}_i, \mathbf{y}_j)$$
$$+ \alpha_3 \sum_{v_i \in \mathcal{V}} U_3(\mathbf{y}_i, z_i) + ||Y||_2^2 \quad (9)$$
$$\text{s.t.} \quad \|\mathbf{y}_i\|_1 = 1, \mathbf{y}_i \geq \bar{\mathbf{0}},\ \text{for } 1 \leq i \leq N.$$

The above constrained optimization problem is a *quadratic programming* problem. We solve it by using the CVX [10].

[10] Grant and Boyd, "CVX users guide for CVX version 1.22," 2012.

Proposed Approach - Optimization process (Co-segmentation)

Our proposed optimization function

$$\begin{aligned} J(Y, Z) = {} & \alpha_1 \sum_{v_i \in \mathcal{V}} U_1(\mathbf{y}_i) + \alpha_2 \sum_{v_i \in \mathcal{V}} U_2(z_i) + \alpha_3 \sum_{v_i \in \mathcal{V}} U_3(\mathbf{y}_i, z_i) \\ & + \beta_1 \sum_{e_{ij} \in \mathcal{E}} B_1(\mathbf{y}_i, \mathbf{y}_j) + \beta_2 \sum_{e_{ij} \in \mathcal{E}} B_2(z_i, z_j) + \|Y\|_2^2 \qquad (10) \\ & \text{s.t.} \quad \| \mathbf{y}_i \|_1 = 1,\ \mathbf{y_i} \geq \bar{\mathbf{0}}, z_i \in \{0, 1\},\ \text{for } 1 \leq i \leq N. \end{aligned}$$

By fixing Y, the optimization task in (10) becomes

$$\begin{aligned} J(Z) = {} & \alpha_2 \sum_{v_i \in \mathcal{V}} U_2(z_i) + \beta_2 \sum_{e_{ij} \in \mathcal{E}} B_2(z_i, z_j) \\ & + \alpha_3 \sum_{v_i \in \mathcal{V}} U_3(\mathbf{y}_i, z_i) \qquad (11) \\ & \text{s.t.} \quad z_i \in \{0, 1\},\ \text{for } 1 \leq i \leq N. \end{aligned}$$

The energy function in (11) is graph representable and regular. Thus it can be efficiently minimized via graph cuts.

Proposed Approach - Implementation details

i. For initialization, we solve the weights Y for saliency map fusion with the coupling term U_3 removed.

ii. Then, the fused co-saliency maps are binarized with an adaptive threshold into foregrounds and backgrounds to initialize GMMs $\theta_{\mathbf{f}}$, $\theta_{\mathbf{b},1}$ and $\theta_{\mathbf{b},2}$ and enable the optimization of co-segmentation at the first iteration.

iii. In the alternating optimization process, the value of the objective function decreases and converges to a local optimum when solving (9) and (11) iteratively.

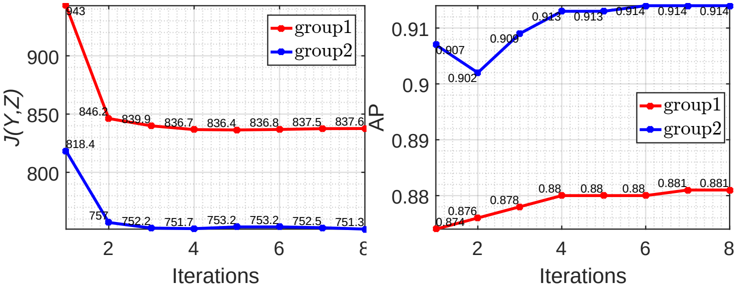
Experimental Settings

- Benchmark dataset: Image Pair dataset which is composed of 105 image pairs.
- We select two groups of saliency map proposals to test of proposed fusion method
  I. We followed Li's co-saliency detection [11] method by collecting
  Group 1: $\underbrace{\text{IT98/SR07/FT09}}_{\text{SISM}}/\underbrace{\text{CC11/CP11}}_{\text{MISM}}$
  II. We selected some state-of-the-art methods by collecting
  Group 2: $\underbrace{\text{CA10/SF12/RBD14}}_{\text{SISM}}/\underbrace{\text{CO13/SP13}}_{\text{MISM}}$
- Comparison candidates: We compare ours with other fusion methods.
  - CSM [Li,TIP2011] [11] $\rightarrow$ *fixed-weight map-wise addition*
  - SACS [Cao,TIP2014] [12] $\rightarrow$ *adaptive-weight map-wise addition*
  - Ours $\rightarrow$ *adaptive-weight region-wise addition*
- Performance metrics: "Average Precision(AP)" and "Area under the ROC curve(AUC)"

[11] H. Li and K. N. Ngan, "A co-saliency model of image pairs," TIP 2011.
[12] Cao *et al.*, "Self-adaptively weighted co-saliency detection via rank constraint," TIP 2014.

Experimental Results - Energy and AP curves

The energy curves and the AP curves converge rapidly after few iterations.

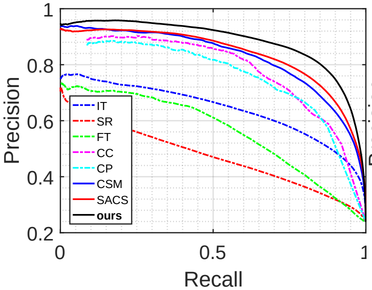


(a) (b)

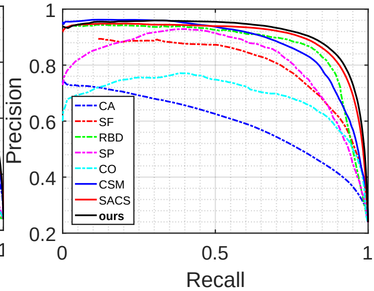
Figure 7: (a) The energy curves of the proposed optimization function (b) The AP curves, versus iterations, in two different saliency proposal groups.

Experimental Result - PR curves

Our fusion method outperforms the state-of-the-art fusion method and the adopted saliency maps.



(a)



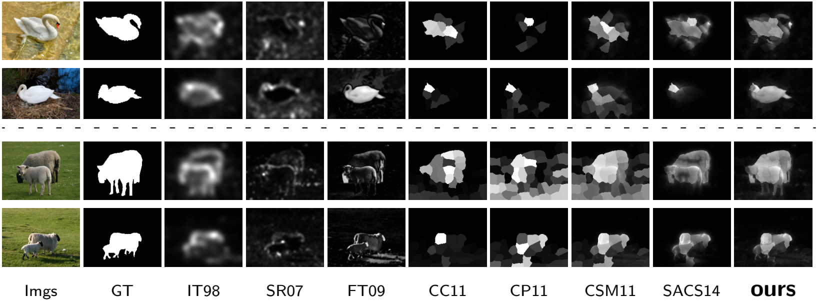
(b)

Figure 8: The PR curves of the evaluated approaches with the saliency proposals in (a) group 1 and (b) group 2.

Experimental Result - Visual result 1

Our performance gain over the best competing fusion method (SACS) is significant! (4.5% in AP and 1.3% in AUC)

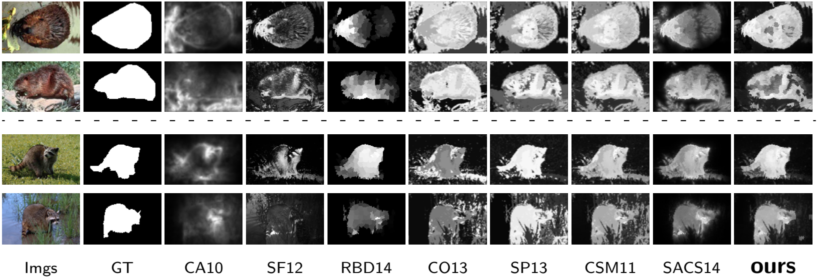
method	IT	SR	FT	CC	CP	CSM	SACS	**ours**
AP	0.640	0.471	0.559	0.702	0.681	0.824	0.836	**0.881**
AUC	0.872	0.718	0.756	0.881	0.865	0.930	0.944	**0.958**



Experimental Result - Visual result 2

Even though group 2 saliency proposals are less complementary, our proposed approach still outperforms the best competing fusion method (SACS)! (1.4% in AP and 0.4% in AUC)

method	CA	SF	RBD	SP	CO	CSM	SACS	**ours**
AP	0.595	0.701	0.847	0.813	0.692	0.879	0.900	**0.914**
AUC	0.843	0.922	0.936	0.915	0.886	0.948	0.970	**0.974**



Conclusion

i. We presented an unsupervised learning framework that carries out adaptive weight region-wise saliency proposal fusion.

ii. The joint optimization formulation gives a higher quality co-saliency map.

iii. Unlike existing models relying on additional post-processing to smooth the fused maps, our framework alreadymerged the advantages of such post-processing into our unified optimization process.

iv. In future, we plan to apply our algorithm to vision applications where saliency maps of high quality are appreciated, such as object recognition or scene understanding.

Thank you for listening.