The Secrets of Salient Object Segmentation
Supplementary Materials

Yin Li*
Georgia Tech
yli440@gatech.edu

Xiaodi Hou*
Caltech
xiaodi.hou@gmail.com

Christof Koch
Allen Institute for Brain Science
christofk@alleninstitute.org

James M. Rehg
Georgia Tech
rehg@cc.gatech.edu

Alan L. Yuille
UCLA
yuille@stat.ucla.edu

1. Implementation Details

1.1. Features for Salient Object Segmentation

Given a proposed object candidate segment and its fixation map, we learn a function that estimates the overlapping score (intersection over union) of the region with respect to its ground-truth salient object mask. Without loss of generality, we use the term *fixation energy* to refer to pixel values in the fixation map. For human fixations, the energy is a discrete value as the number of fixations at current location. For algorithm generated fixation maps, the energy is a continuous value ranging from 0 to 1 as the probability of a fixation at current location. As mentioned in Sec 4.2 of the paper, we extract two types of features for each segment, shape features (15 dimensions) and fixation features (18 dimensions).

The details of the 33 dimensional feature can be found in Table. 1. In particular, the *Fixation Energy Ratio* is the defined as the sum of fixation energy within the segment, divided by the sum of fixation energy of the whole image.

Shape Features	Dims
Area	1
Centroid	2
Convex Area	1
Euler Number	1
Perimeter	1
Major/Minor Axes Length	2
Eccentricity	1
Orientation	1
Equivalent Diameter	1
Solidity	1
Extent	1
Width/Height	2
Fixation Features	Dims
Min/Max Fixation Energy	2
Mean Fixation Energy	1
Weighted Fixation Centroid	2
Fixation Energy Ratio	1
Histogram of Fixations	12

Table 1. Shape and fixation features used in our model.

1.2. Training using Random Forest

We use random regression forest to learn the scoring function. A random regression forest is an ensemble of decision trees. For each branch node, a feature is selected from a random subset of all features and a decision boundary is set by minimizing the Minimum Square Error (MSE). The leaf nodes keep the mean value of all training samples that end up in the node. And the final result is a weighted average of all leaf nodes that a testing sample reaches. We choose random forest since our feature vector contains discrete values (Euler Number), which can be easily handled in a decision tree. For all our experiments, we train a random forest using 30 trees, where each node uses 6 feature dimensions. We use the first 40% of the images for training. The rest is used for testing.

2. Quantitative Results

The full benchmarking results of all algorithms on all 3 datasets are shown in Fig. 1, Fig. 2, and Fig. 3. In particular, for all CPMC related algorithms we choose the top $K = 10$ segments (same as what we have mentioned in the paper). The results are grouped into 4 categories:

Salient Object Algorithms refer to the 4 algorithms FT[1], GC[3], PCAS[9], and SF[10] that are originally proposed for salient object segmentation.

*These authors contribute equally.

CPMC + Fixation Algorithms refer to the model proposed in our paper. We choose the top 10 segments for each image, and assign scores based on the fixation map of 7 algorithms: `AIM`[2], `AWS`[4], `DVA`[7], `GBVS`[5], `ITTI`[8], `SIG`[6], and `SUN`[11].

Original Fixation Algorithms refer to the 7 fixation algorithms. To cancel the effect of center bias, we add a fixed center bias with $\sigma = 0.4$ of the image width, to each generated fixation map.

Baseline Models refers to 4 other models. `CPMC Ranking` refers to the original rankings of CPMC, with the same choice of $K = 10$. `CPMC+Human Fixations` refers to a variation of our model that replaces the algorithm fixation maps with human fixations -supposedly, human map should reflect the saliency of the scene more accurately than algorithms. `CPMC Best` refers to the salient object maps generated by greedily selecting the best CPMC segments with respect to the ground truth. This score estimates the upper limit of any algorithm that is based on CPMC segmentation. Finally `GT Seg + Human Fixations` refers to the method that uses ground-truth segmentations plus human fixations. This score validates the strong connection between the fixation task and the salient object segmentation task.

The actual F-measures of each algorithm on each dataset is reported in Tab. 2. Our results consistent outperform all state-of-the-art methods on all datasets. In Fig. 4, we present the F-Measure of our model at different choices of K on all 3 different datasets. Our method converges quickly with a small number of segments. We achieve superior results using only the first 10 segments.

3. Qualitative Results

In this section, we provide qualitative results of salient object segmentation generated by our methods as well as others. Fig. 5-7 demonstrate our results using 7 different fixation prediction algorithms in comparison to CPMC ranking function and major salient object segmentation methods, on FT, IS and PARSCAL-S, respectively. We also present our results using human fixations in PASCAL-S and IS. To illustrate the strength and weakness of our method, we select the results based on the F-measures of each image. For each dataset, the average F-measure of our method decreases from top row to the bottom row. Our method is able to capture the full region of an salient object. We notice that our method does not segment most of the small salient objects. This is largely due to the output from CPMC using sparse uniform seeds. An object is missing if it does not contain a seed in CPMC.

Salient Object	FT	IS	PASCAL-S
FT	0.7427	0.4736	0.4325
GC	0.8383	0.6261	0.6072
PCAS	0.8646	0.6558	0.6275
SF	0.8850	0.5555	0.5557
CPMC + Fixation	FT	IS	PASCAL-S
AIM	0.8920	0.6728	0.7204
AWS	0.8998	0.7241	0.7224
DVA	0.8700	0.6377	0.7112
GBVS	0.9097	0.7264	0.7457
ITTI	0.8950	0.6827	0.7288
SIG	0.8908	0.7255	0.7214
SUN	0.8635	0.6249	0.7058
Orig. Fixation	FT	IS	PASCAL-S
AIM	0.6858	0.4804	0.6267
AWS	0.7228	0.6033	0.5084
DVA	0.6534	0.4795	0.5426
GBVS	0.6899	0.5333	0.6383
ITTI	0.6544	0.4431	0.6228
SIG	0.6741	0.6110	0.5897
SUN	0.6692	0.3881	0.5482
Fix	N/A	0.6972	0.6781
Baseline Models	FT	IS	PASCAL-S
CPMC Ranking	0.4421	0.5287	0.6339
CPMC + Human	N/A	0.7863	0.7756
CPMC Best	0.9496	0.8416	0.8699
GT Seg. + Human	N/A	N/A	0.9201

Table 2. The F-measures of all algorithms on all 3 datasets.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604. IEEE, 2009. 1

[2] N. Bruce and J. Tsotsos. Saliency based on information maximization. In *NIPS*, pages 155–162, 2005. 2

[3] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global contrast based salient region detection. In *CVPR*, pages 409–416. IEEE, 2011. 1

[4] A. Garcia-Diaz, V. Leborán, X. R. Fdez-Vidal, and X. M. Pardo. On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. *Journal of Vision*, 12(6), 2012. 2

[5] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, pages 545–552, 2006. 2

[6] X. Hou, J. Harel, and C. Koch. Image signature: Highlighting sparse salient regions. *TPAMI, IEEE*, 34(1):194–201, 2012. 2

[7] X. Hou and L. Zhang. Dynamic visual attention: Searching for coding length increments. In *NIPS*, pages 681–688, 2008. 2

[8] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI, IEEE*, 20(11):1254–1259, 1998. 2

[9] R. Margolin, A. Tal, and L. Zelnik-Manor. What makes a patch distinct? In *CVPR*. IEEE, 2013. 1

[10] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, pages 733–740. IEEE, 2012. 1

[11] 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), 2008. 2

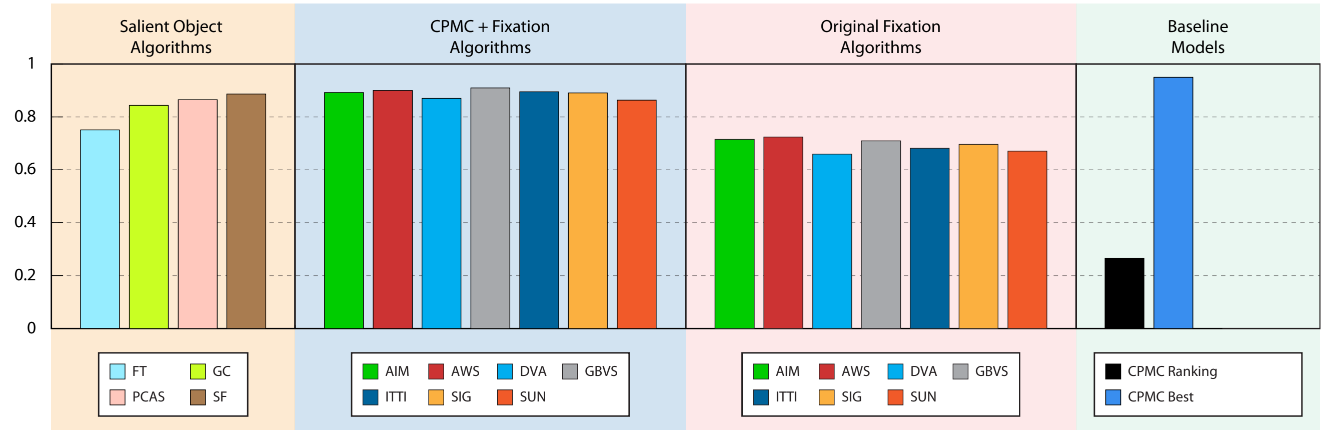

Figure 1. The F-measures of all algorithms on FT dataset. Due to the absence of human fixations on this dataset, as well as the full segmentation ground-truth, we are unable to evaluate `CPMC + Human Fixations` and `GT Seg + Human Fixations`.

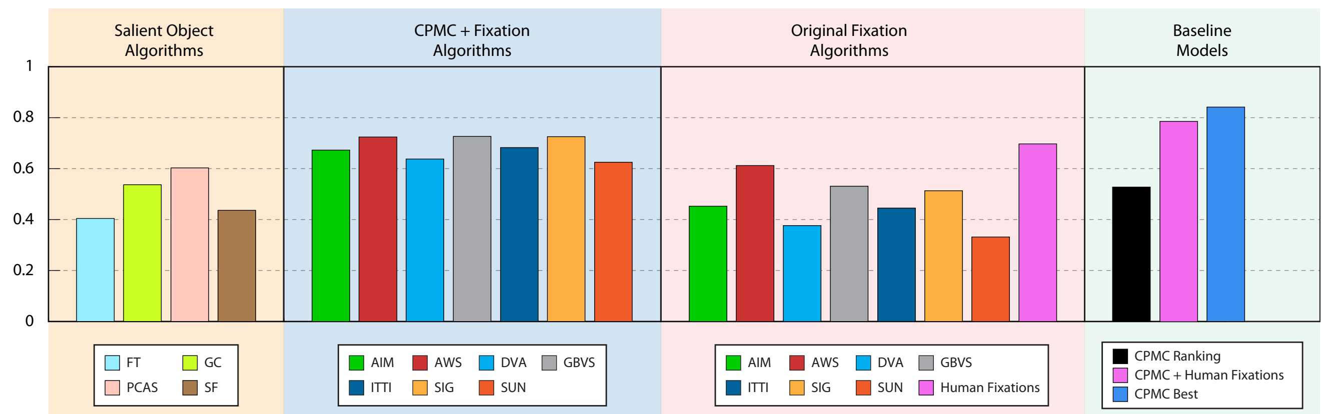

Figure 2. The F-measures of all algorithms on IS dataset. Due to the absence of full segmentation ground-truth of this dataset, we are unable to evaluate `GT Seg + Human Fixations`.

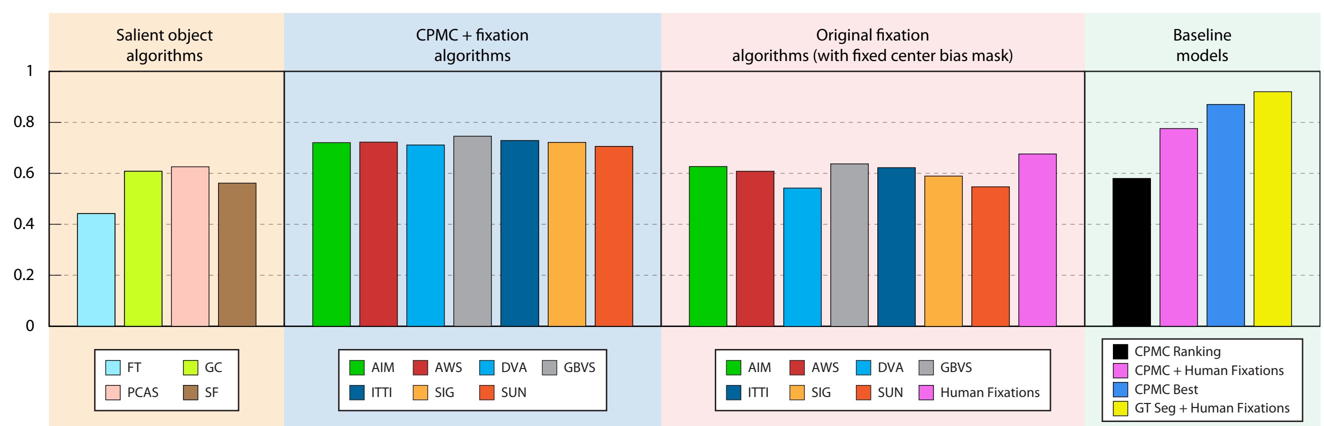

Figure 3. The F-measures of all algorithms on PASCAL-S dataset.

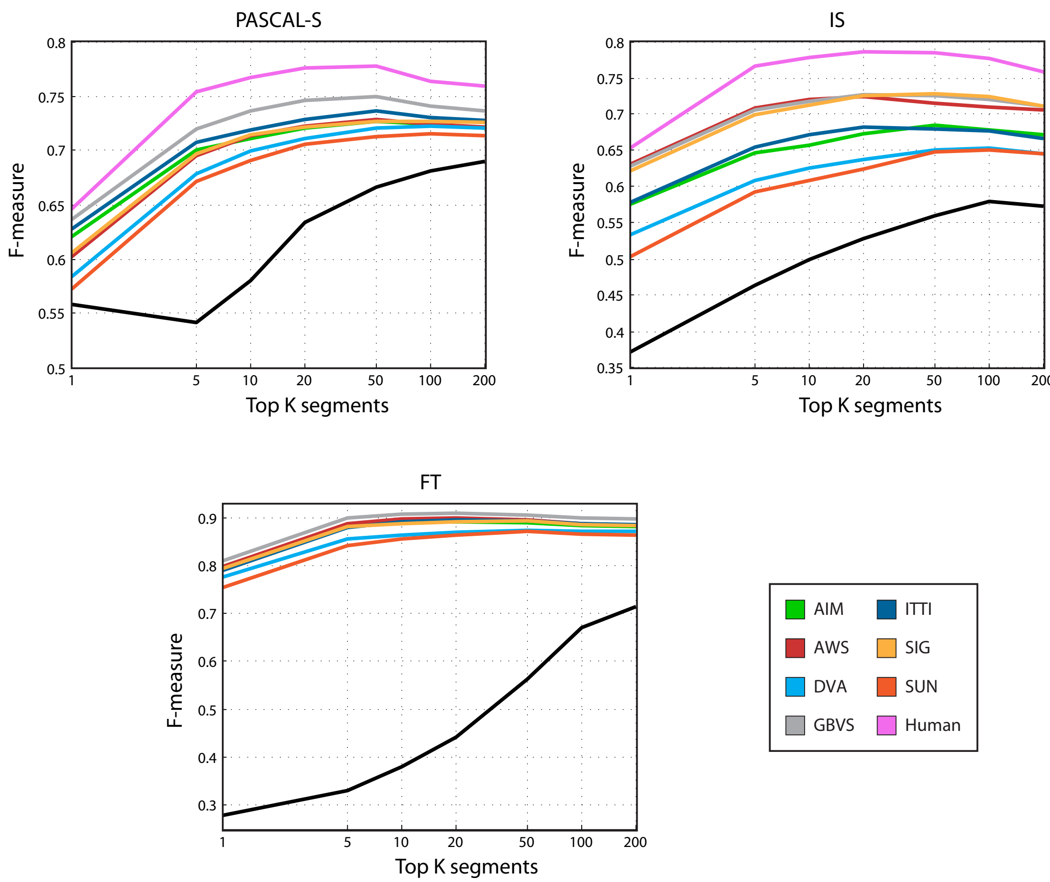

Figure 4. The F-Measure of our method at different choices of K using different fixations. Note that FT dataset does not come with human fixation data.

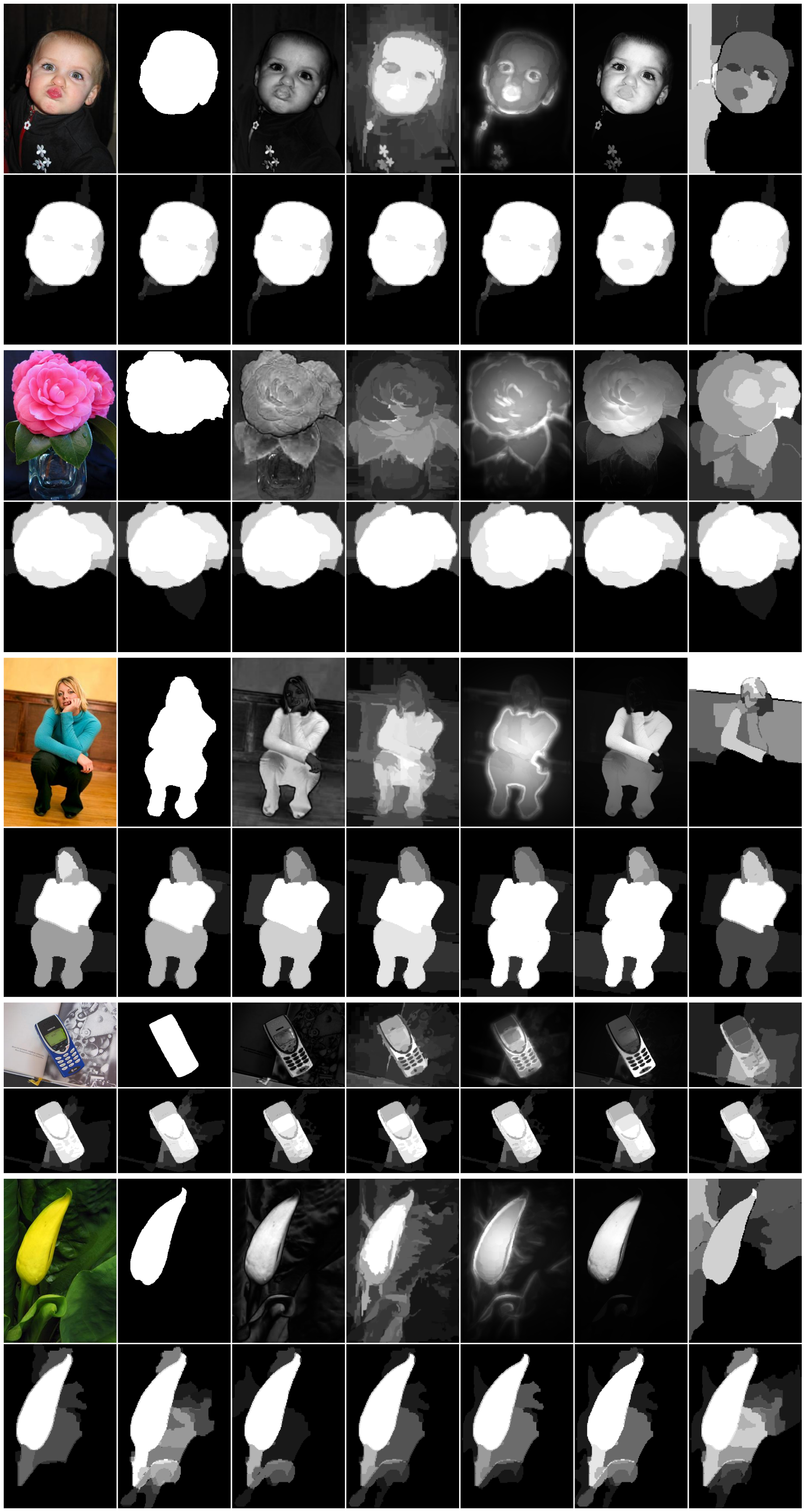

Figure 5. Visualization of salient object segmentation results on FT. Each two-row compares the results of one image. The first row includes results from existing methods (Left to Right): Original image, Ground-truth mask, FT, GC, PCAS, SF and CPMC ranking; The second row shows results of our method using different fixations (Left to Right): AIM, AWS, DVA, GBVS, ITTI, SIG and SUN. We are not able to report results using human fixations. The images are selected by sorting the F-measure of our results in a decreasing order.

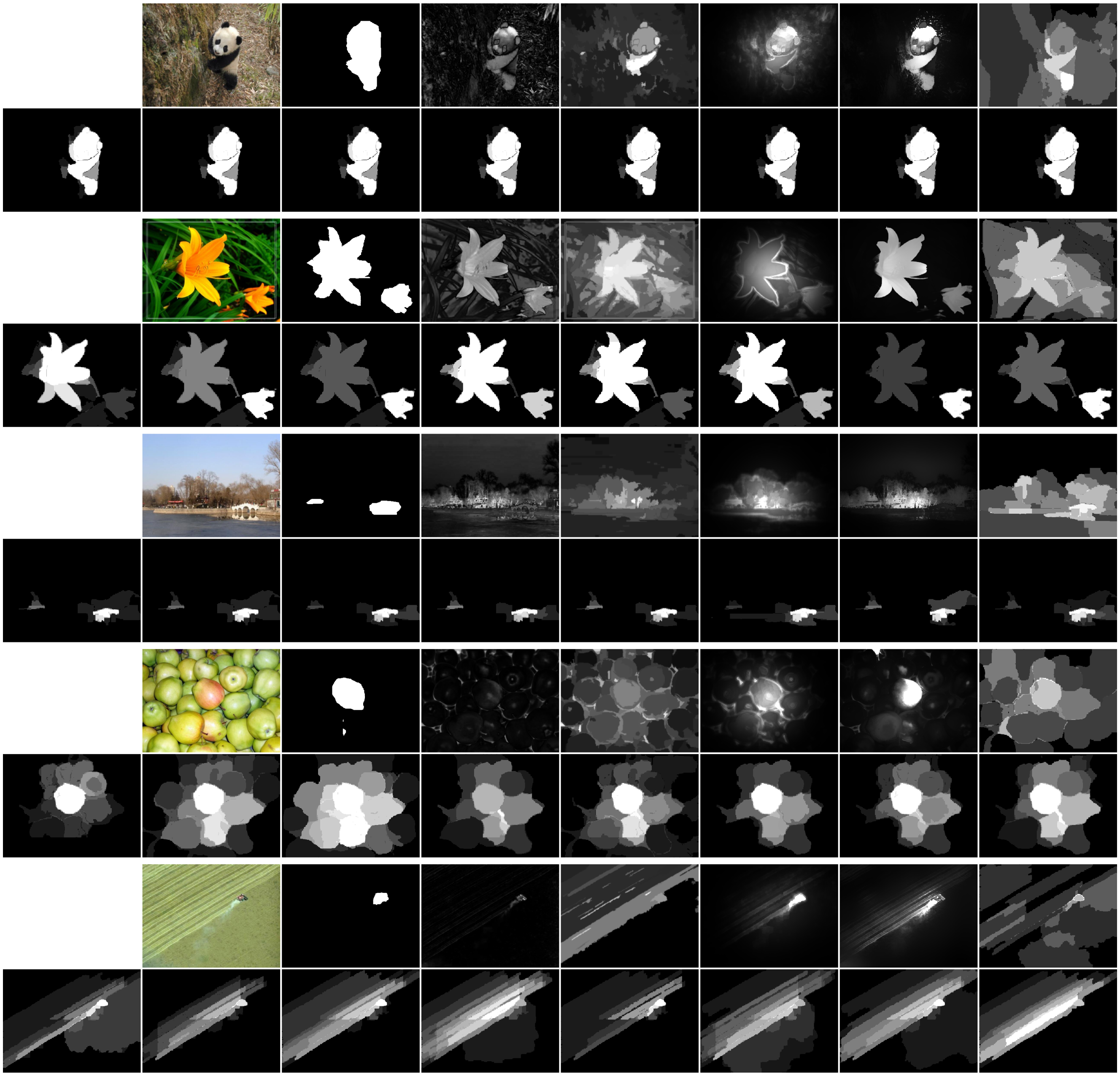

Figure 6. Visualization of salient object segmentation results on IS. Each two-row compares the results of one image. The first row includes results from existing methods (Left to Right): Original image, Ground-truth mask, FT, GC, PCAS, SF and CPMC ranking; The second row shows results of our method using different fixations (Left to Right): Human Fixation, AIM, AWS, DVA, GBVS, ITTI, SIG and SUN. The images are selected by sorting the F-measure of our results in a decreasing order. We notice that IS favors sparse saliency maps, since it contains a significant portion of small salient objects.

Figure 7. Visualization of salient object segmentation results on PARSCAL-S. Each two-row compares the results of one image. The first row includes results from existing methods (Left to Right): Original image, Ground-truth mask, FT, GC, PCAS, SF and CPMC ranking; The second row shows results of our method using different fixations (Left to Right): Human Fixation, AIM, AWS, DVA, GBVS, ITTI, SIG and SUN. The images are selected by sorting the F-measure of our results in a decreasing order.