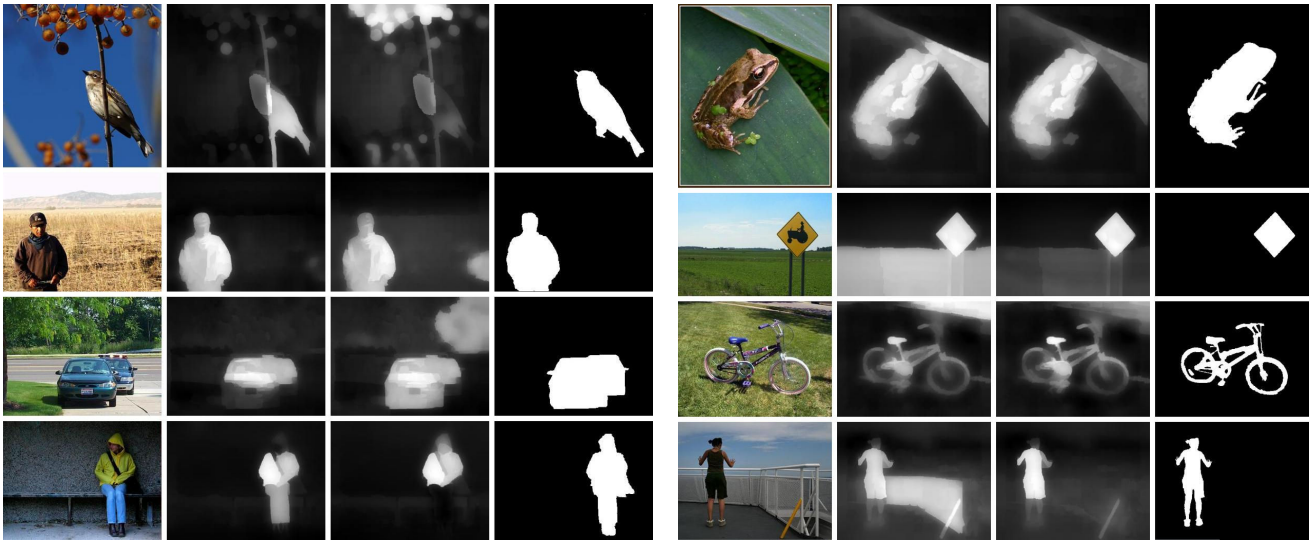
Salient Object Detection via Bootstrap Learning
Supplementary Materials

Na Tong[1], Huchuan Lu[1], Xiang Ruan[2] and Ming-Hsuan Yang[3]
[1]Dalian University of Technology [2]OMRON Corporation [3]University of California at Merced

We present some examples where the detection results of the RGB features are better than those of the CIELab features, or vice versa, as shown in Figure 1. The empirical results demonstrate that we can obtain better saliency maps by utilizing both RGB and CIELab features although they may be correlated.



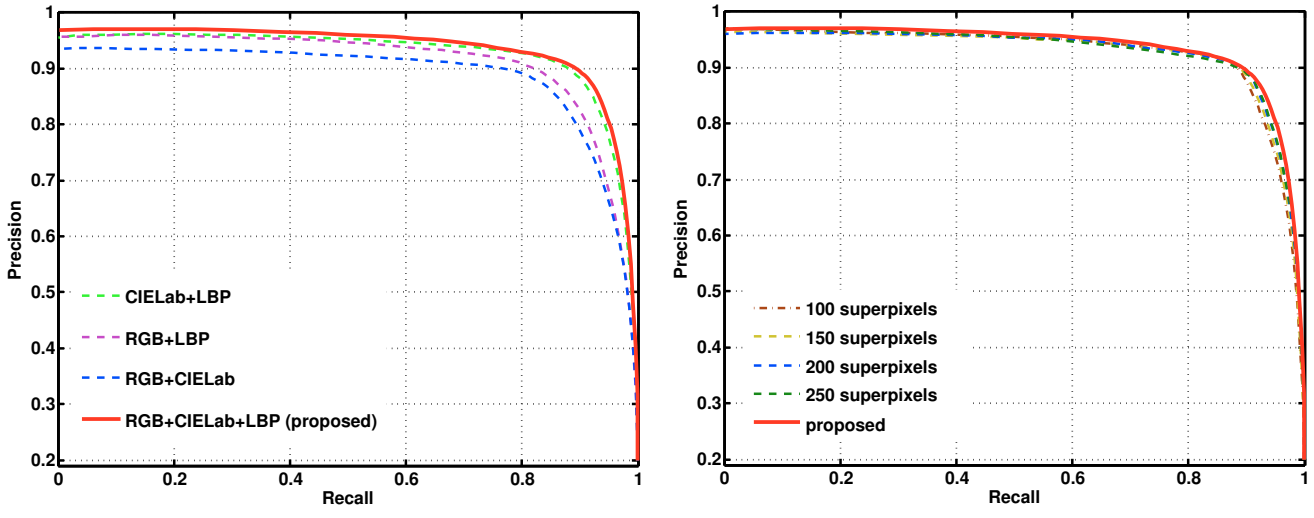
(a) (b)

Figure 1. Left to right in (a) and (b): input, saliency maps generated by utilizing the RGB and CIELab features, respectively (using linear SVM for both maps) and the ground truth. (a) shows the detection results using the RGB feature are better than those using the CIELab feature and vice versa in (b).

In Figure 2(a), we provide the Precision and Recall (P-R) curves of the proposed bootstrap learning algorithm when each of the three features (RGB, CIELab and LBP) is removed, respectively. These results show that each feature contributes to detect salient objects. In addition, we show the P-R curve of the proposed method in comparison with those of single-scale methods in Figure 2(b), which demonstrates the effects of the multiscale integration in the proposed method. These results show that the efficiency of the proposed method could be largely improved by adopting only single scale at the expense of some decrease in accuracy.

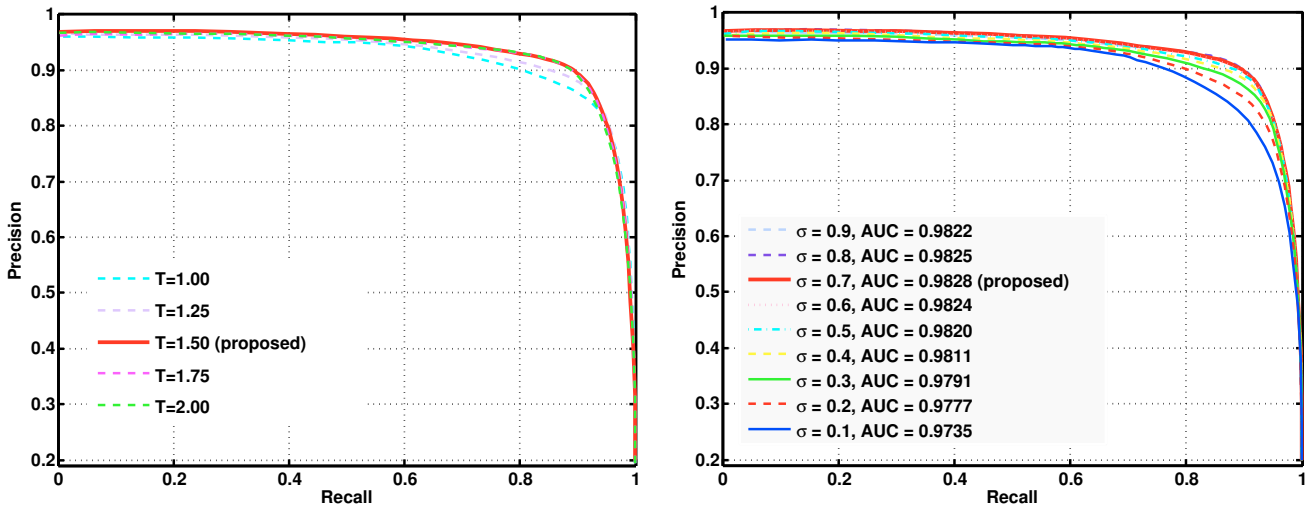
We clarify the details about the threshold setting for selecting the training set. As stated in Section 3.2 of the paper, we compute the average saliency value for each superpixel and set two thresholds to generate the training set containing both positive and negative samples. The superpixels with saliency values larger than the high threshold are labeled as the positive samples with $+1$ while those with saliency values smaller than the low threshold as the negative samples labeled with -1. The low threshold is set to 0.05 and it's fixed for all images since the training set is not sensitive to the low threshold. The high threshold is set empirically according to the results of Figure 3(a), which demonstrates that the proposed method with the threshold of 1.5 times the average value over the whole weak saliency map performs best.

In Figure 3(b), we further show the comparative results of the proposed methods with different values set for the parameter



(a) (b)

Figure 2. (a) shows P-R curves of the proposed algorithms when each of the three features (RGB, CIELab and LBP) is removed, respectively. (b) shows the P-R curves of the proposed method in comparison with those of single-scale methods, i.e., the input image is over-segmented into $100, 150, 200, 250$ superpixels, respectively.



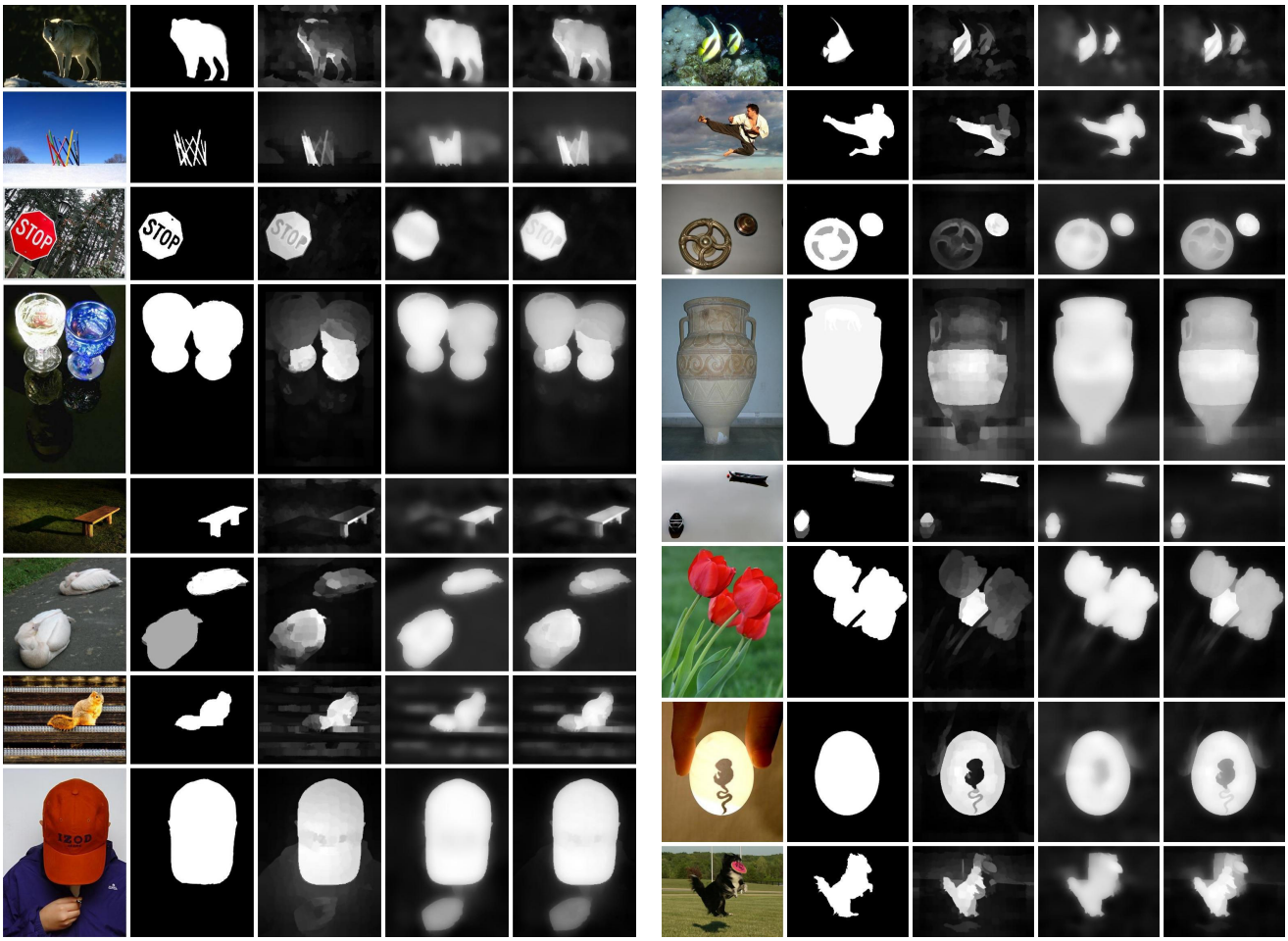
(a) (b)

Figure 3. (a) shows P-R curves of the proposed algorithms when the high threshold is set to different values (T times of the average value over the whole weak saliency map). (b) shows the P-R curves and the AUC values of the proposed methods when the parameter σ in (11) is set to different values.

σ in (11), which illustrates that both the weak and strong saliency models contribute to the final results while the strong model plays a more important role.

Figure 4 shows more saliency maps generated by the proposed method. These weak and strong saliency maps complement each other in which weak maps are able to detect fine details and strong maps are able to detect global shapes.

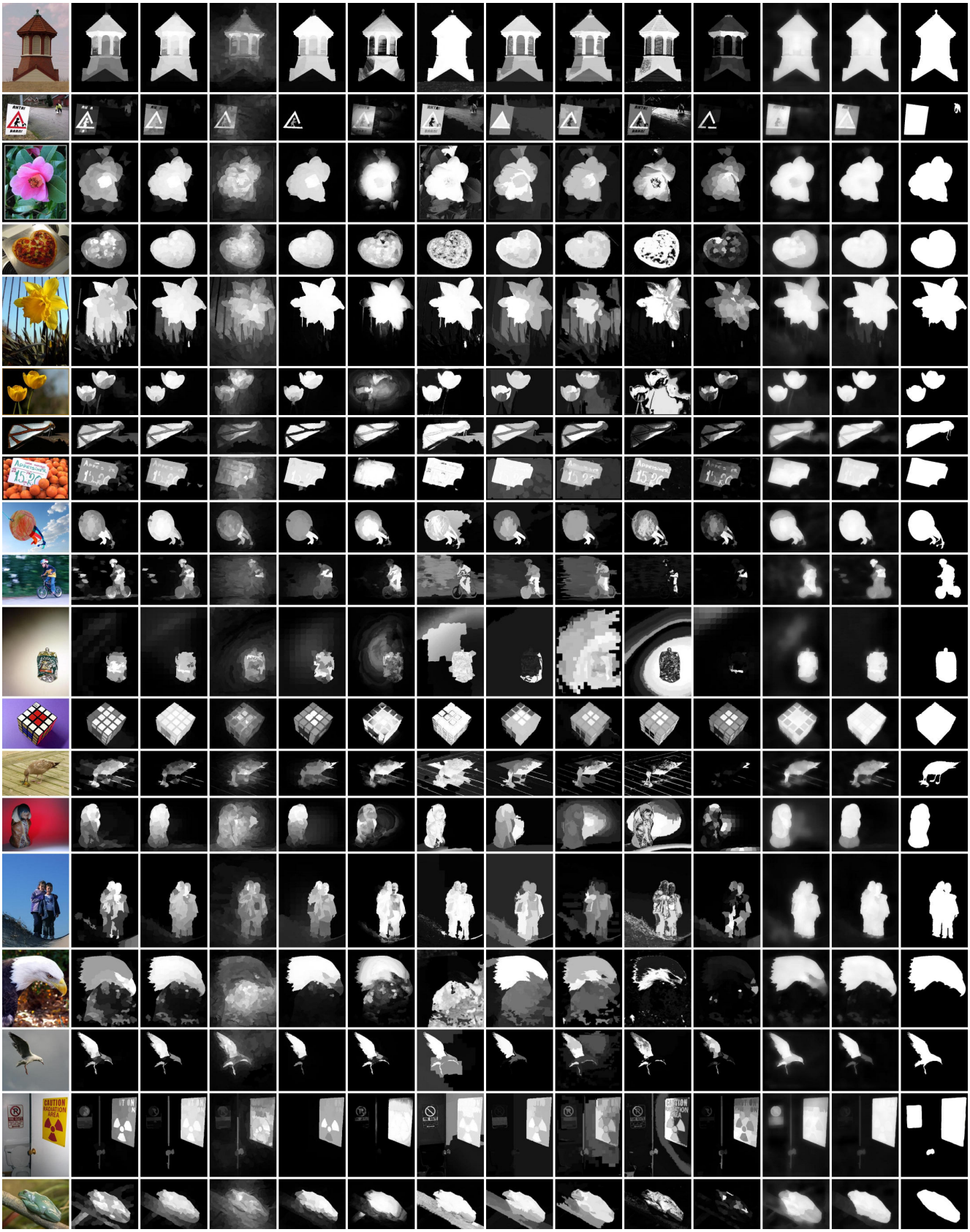
We present more examples of saliency maps generated by the proposed approach and several state-of-the-art methods (the number of these compared methods depends on whether the authors provide codes or provide results of their methods on the corresponding datasets) on five datasets in Figure 5-9, where "wCO-b" indicates the wCO [15] model bootstrapped by the proposed leaning approach and "GT" means the ground truth. Note that the ground truth results represented in grayscale in Figure 6-8 are of more reliable since they are the average annotations from multiple subjects. Compared with the state-of-the-arts, our methods ("Ours" and "wCO-b") are able to detect both the salient objects and backgrounds, and are robust



(a) (b)

Figure 4. Saliency maps generated by the proposed bootstrap learning method. Brighter pixels indicate higher saliency values. Left to right in (a) and (b): input, ground truth, weak saliency map, strong saliency map, and final saliency map.

to objects at different scales as well. The proposed bootstrap learning approach performs well when dealing with complex images, i.e., images with multiple objects or complex backgrounds that could be easily mistakenly detected as salient objects.

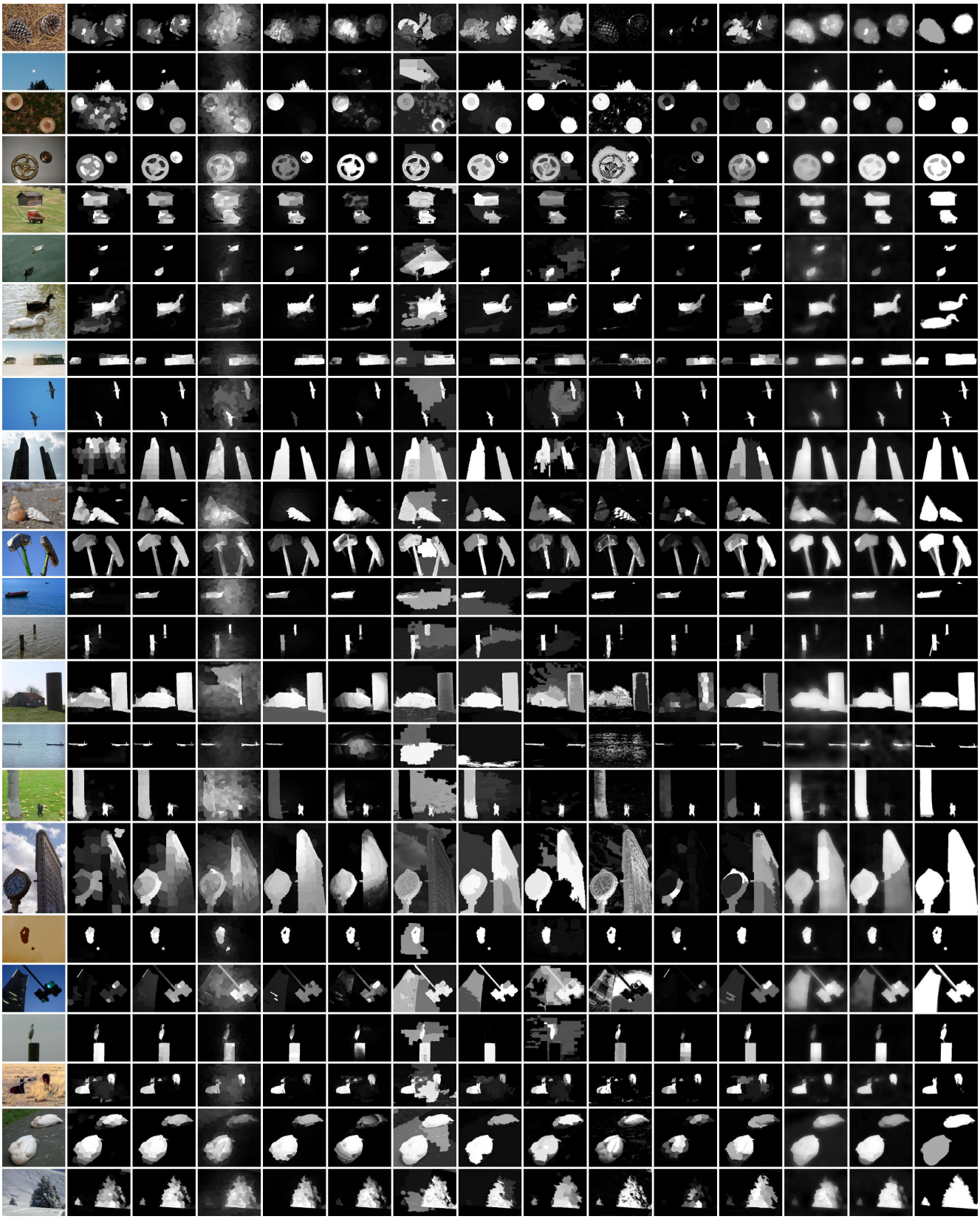


input GS_SP [11] wCO [15] LRMR [10] GMR [14] DSR [6] XL13 [12] HS [13] RC-J [3] GC [4] SF [9] Ours wCO-b GT

Figure 5. Comparison of our saliency maps with ten state-of-the-art methods on the ASD [1] dataset.

input GS_SP [11] wCO [15] LRMR [10] GMR [14] DSR [6] XL13 [12] HS [13] RC-J [3] GC [4] SF [9] Ours wCO-b GT

Figure 6. Comparison of our saliency maps with ten state-of-the-art methods on the SOD [8] dataset.



input GS_SP [11] wCO [15] LRMR [10] GMR [14] DSR [6] XL13 [12] HS [13] RC-J [3] GC [4] SF [9] DRFI [5] Ours wCO-b GT

Figure 7. Comparison of our saliency maps with eleven state-of-the-art methods on the SED2 [2] dataset.

input GS_SP [11] wCO [15] LRMR [10] GMR [14] DSR [6] XL13 [12] HS [13] RC-J [3] GC [4] SF [9] Ours wCO-b GT

Figure 8. Comparison of our saliency maps with ten state-of-the-art methods on the Pascal-S [7] dataset.

input GS_SP [11] wCO [15] LRMR [10] GMR [14] DSR [6] XL13 [12] HS [13] RC-J [3] GC [4] SF [9] Ours wCO-b GT

Figure 9. Comparison of our saliency maps with ten state-of-the-art methods on the THUS [3] dataset.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009. 4

[2] S. Alpert, M. Galun, R. Basri, and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. In *CVPR*, 2007. 6

[3] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu. Global contrast based salient region detection. *PAMI*, 37(3):569–582, 2015. 4, 5, 6, 7, 8

[4] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook. Efficient salient region detection with soft image abstraction. In *ICCV*, 2013. 4, 5, 6, 7, 8

[5] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. In *CVPR*, 2013. 6

[6] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang. Saliency detection via dense and sparse reconstruction. In *ICCV*, 2013. 4, 5, 6, 7, 8

[7] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille. The secrets of salient object segmentation. In *CVPR*, 2014. 7

[8] V. Movahedi and J. H. Elder. Design and perceptual validation of performance measures for salient object segmentation. In *POCV*, 2010. 5

[9] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012. 4, 5, 6, 7, 8

[10] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, 2012. 4, 5, 6, 7, 8

[11] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *ECCV*, 2012. 4, 5, 6, 7, 8

[12] Y. Xie, H. Lu, and M.-H. Yang. Bayesian saliency via low and mid level cues. *TIP*, 22(5):1689–1698, 2013. 4, 5, 6, 7, 8

[13] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *CVPR*, 2013. 4, 5, 6, 7, 8

[14] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013. 4, 5, 6, 7, 8

[15] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency optimization from robust background detection. In *CVPR*, 2014. 2, 4, 5, 6, 7, 8