BASNet: Boundary-Aware Salient Object Detection – Supplementary Material

Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan and Martin Jagersand
University of Alberta, Canada
{xuebin,vincent.zhang,chuang8,cgao3,masood1,mj7}@ualberta.ca

This supplementary material provides complementary quantitative and qualitative analysis of the proposed Boundary-Aware Salient Object Detection Network (BASNet).

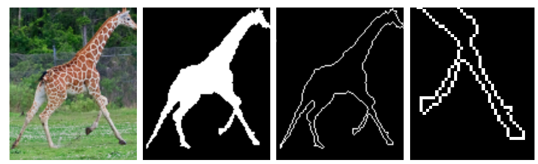
1. With/without CRF Post-Processing

Conditional Random Field (CRF) [5] is a frequently used post-processing procedure in image segmentation. Here, we show the comparison of our method and PiCANetR (which is the best-performing existing method) with/without CRF. Table 1 illustrates the quantitative comparison. The performance improvement achieved by adding CRF post-processing to our method is very limited (see 1st and 2nd rows in Table 1). The main reason is that the probability maps produced by our method usually have high confidence and clear boundaries, which are already very close to or even better than the CRF post-processed results. Although PiCANetR is slightly improved by CRF, its result is still not as good as ours in terms of $maxF_\beta$, $relaxF^b_\beta$ and MAE. Fig. 1 (b) and (c) show the qualitative comparison of our method without/with CRF post-processing. Their visual differences are negligible. Fig. 1 (d) and (e) illustrate the difference between PiCANetR and PiCANetRC (PiCANeR with CRF). While CRF gives higher confidence of the saliency map, it introduces salt-and-pepper noises around the boundaries. Additionally, CRF post-processing is usually slow and approximately doubles the time costs. Therefore, we do not suggest to use CRF post-processing for our method.

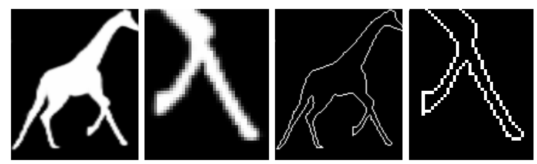
2. Complete Qualitative Comparison

In this section, we provide a comprehensive qualitative comparison of our method with other 15 methods (see Fig. 4) on challenging images with low contrast (1st and 2nd columns), fine structures (3rd and 4th columns), large object touching image boundaries (5th and 6th columns), complex object boundaries (7th and 8th columns), cluttered foreground and background (last two columns).

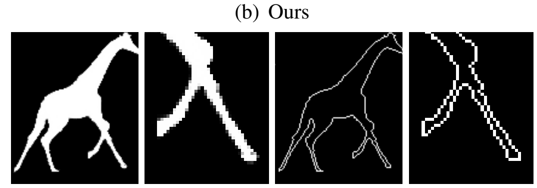
Fig. 5 shows more challenging cases, including large ob-



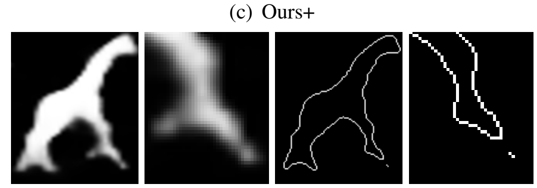
(a) im/GT



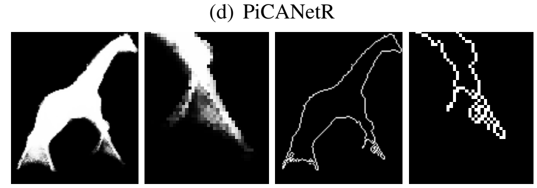
(b) Ours



(c) Ours+



(d) PiCANetR

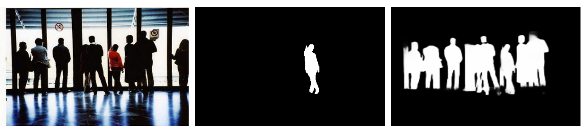


(e) PiCANetRC

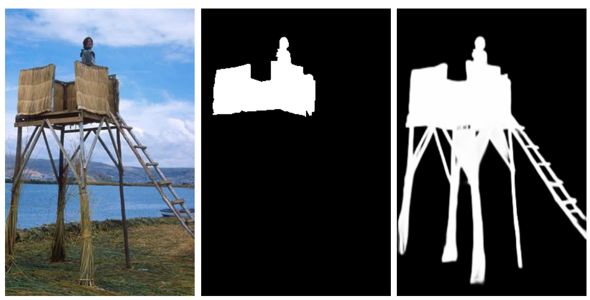
Figure 1. Illustration of CRF post-processing on our method and PiCANetR. (a) shows the input image, ground truth (GT), ground truth boundary map and zoom-in view of the boundary map. “Ours+” and “PiCANetRC” are with CRF post-processing. In (b), (c), (d) and (e), from left to right are predicted salient object, zoom-in view of salient object, boundary map of predicted salient object and zoom-in view of boundary map, respectively.

Table 1. Comparison of the results with/without CRF post-processing of our method and PiCANetR [10] on six datasets in terms of the maximum F-measure $maxF_\beta$ (larger is better), the relaxed boundary F-measure $relaxF_\beta^b$ (larger is better) and the MAE (smaller is better). Red, Green, and Blue indicate the best, second best and third best performance. "Ours+" and "PiCANetRC" are with CRF.

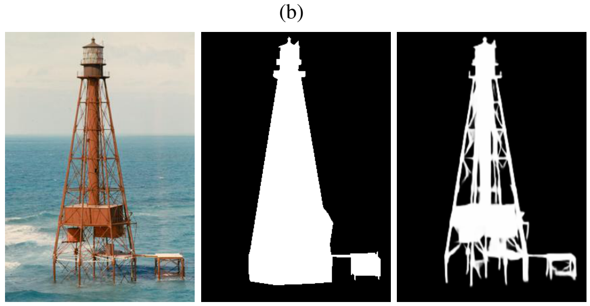
-	SOD [12]			ECSSD [16]			DUT-OMRON [17]			PASCAL-S [9]			HKU-IS [6]			DUTS-TE [13]		
Method	$maxF_\beta$	$relaxF_\beta^b$	MAE	$maxF_\beta$	$relaxF_\beta^b$	MAE	$maxF_\beta$	$relaxF_\beta^b$	MAE	$maxF_\beta$	$relaxF_\beta^b$	MAE	$maxF_\beta$	$relaxF_\beta^b$	MAE	$maxF_\beta$	$relaxF_\beta^b$	MAE
Ours+	0.852	0.592	0.113	0.944	0.830	0.034	0.807	0.695	0.055	0.852	0.663	0.074	0.931	0.812	0.029	0.861	0.759	0.046
Ours	0.851	0.603	0.114	0.942	0.826	0.037	0.805	0.694	0.056	0.854	0.660	0.076	0.928	0.807	0.032	0.860	0.758	0.047
PiCANetRC[10]	0.855	0.514	0.096	0.940	0.775	0.035	0.804	0.629	0.054	0.859	0.605	0.064	0.927	0.766	0.031	0.867	0.699	0.040
PiCANetR [10]	0.856	0.528	0.104	0.935	0.775	0.046	0.803	0.632	0.065	0.857	0.598	0.076	0.918	0.765	0.043	0.860	0.696	0.050



(a)



(b)



(c)

Figure 2. Defective cases. From left to right are input images, ground truth and our results.

ject with cluttered backgrounds (1st column), object with fine structures (2nd column), hollow object (3rd column), objects with thin structures (4th-10th columns) and multiple objects (11th-13th columns).

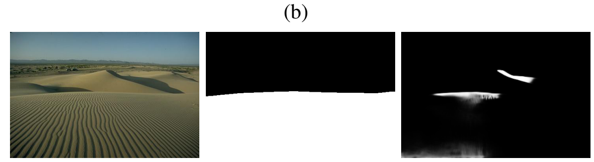
These two figures demonstrate that our method is able to handle various challenging cases and produce accurate salient objects with high quality boundaries.

3. Defective and Failure Cases

The term "defective" refers to the cases where our results are inconsistent with the ground truth. Fig. 2 illustrates some of the defective results of our BASNet compared with the ground truth. However, as can be seen in Fig. 2, these defective cases are not necessarily inferior results. It depends on the practical applications. For example, only coarse regional segmentation is required in camera autofocusing while more details/boundaries are preferred in applications like image matting and editing. Furthermore, the salient objects in certain images could be ambiguous. Hence, avoiding these kinds of "defective" cases could yield better performance measures but is often unnecessary in the real-world applications.

(a)

(b)



(c)

Figure 3. Failure cases. From left to right are input images, ground truth and our results.

Besides, we show several typical failure cases of our method in Fig. 3. Fig. 3(a) shows that our method is not able to correctly segment large objects with salient subregions. Fig. 3(b) illustrates that our method fails to handle large complex scene with too many objects. Fig. 3(c) shows that our method fails in images with no obvious foreground objects. It is worth noting that these failure cases also hard to most of the other state-of-the-art methods. Therefore, there is still a large room for the improvement of our BASNet.

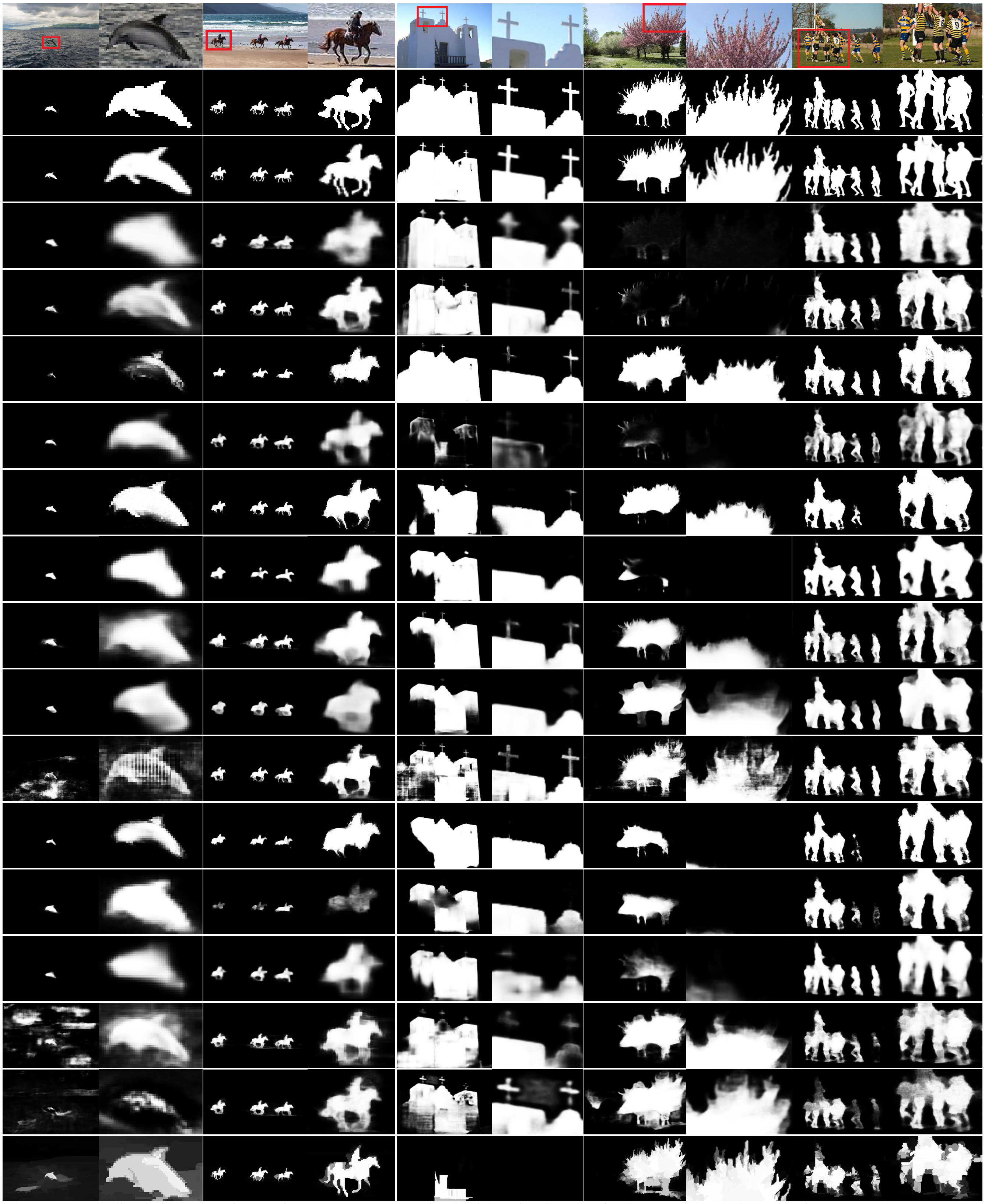


Figure 4. Qualitative comparison of the proposed method with 15 other methods. Each sample occupies two columns. The 2nd column of each sample is the zoom-in view. From top to bottom are **input image**, **ground truth**, **Ours**, **PiCANetR** [10], **BMPM** [18], **R^{3}Net** [2], **PAGRN** [22], **RADF** [4], **DGRL** [15], **RAS** [1], **C2S** [8], **LFR** [19], **DSS** [3], **NLDF** [11], **SRM** [14], **Amulet** [20], **UCF** [21], **MDF** [7], respectively.

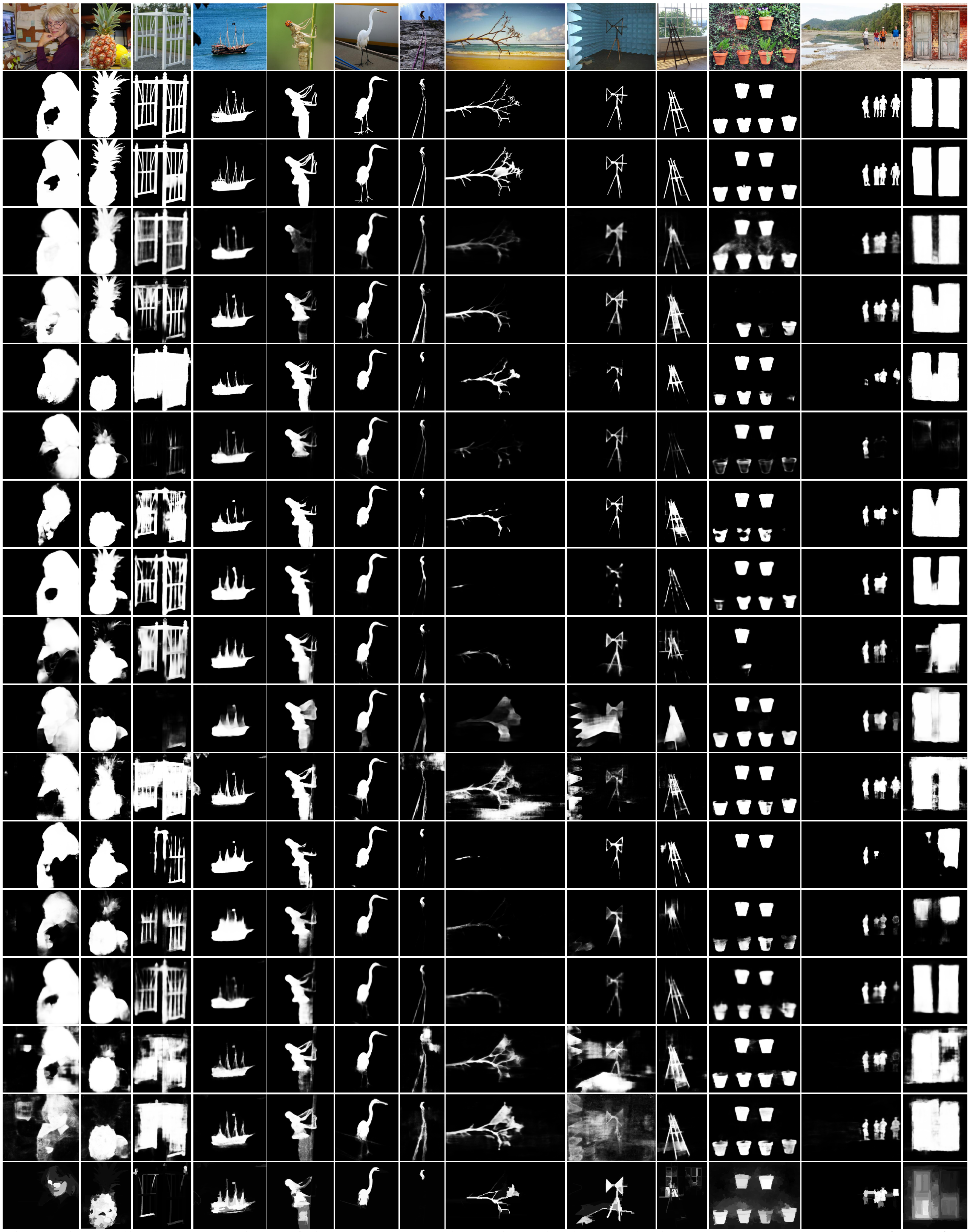


Figure 5. Qualitative comparison of the proposed method with 15 other methods. From top to bottom are **input image**, **ground truth**, **Ours**, **PiCANetR** [10], **BMPM** [18], **R^{3}Net** [2], **PAGRN** [22], **RADF** [4], **DGRL** [15], **RAS** [1], **C2S** [8], **LFR** [19], **DSS** [3], **NLDF** [11], **SRM** [14], **Amulet** [20], **UCF** [21], **MDF** [7], respectively.

References

[1] Shuhan Chen, Xiuli Tan, Ben Wang, and Xuelong Hu. Reverse attention for salient object detection. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IX*, pages 236–252, 2018.

[2] Zijun Deng, Xiaowei Hu, Lei Zhu, Xuemiao Xu, Jing Qin, Guoqiang Han, and Pheng-Ann Heng. R3net: Recurrent residual refinement network for saliency detection. IJCAI, 2018.

[3] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. Deeply supervised salient object detection with short connections. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5300–5309. IEEE, 2017.

[4] Xiaowei Hu, Lei Zhu, Jing Qin, Chi-Wing Fu, and Pheng-Ann Heng. Recurrently aggregating deep features for salient object detection. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 6943–6950, 2018.

[5] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011.

[6] Guanbin Li and Yizhou Yu. Visual saliency based on multiscale deep features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5455–5463, 2015.

[7] Guanbin Li and Yizhou Yu. Visual saliency detection based on multiscale deep cnn features. *IEEE Transactions on Image Processing*, 25(11):5012–5024, 2016.

[8] Xin Li, Fan Yang, Hong Cheng, Wei Liu, and Dinggang Shen. Contour knowledge transfer for salient object detection. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XV*, pages 370–385, 2018.

[9] Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille. The secrets of salient object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 280–287, 2014.

[10] Nian Liu, Junwei Han, and Ming-Hsuan Yang. Picanet: Learning pixel-wise contextual attention for saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3089–3098, 2018.

[11] Zhiming Luo, Akshaya Mishra, Andrew Achkar, Justin Eichel, Shaozi Li, and Pierre-Marc Jodoin. Non-local deep features for salient object detection. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 6593–6601. IEEE, 2017.

[12] Vida Movahedi and James H Elder. Design and perceptual validation of performance measures for salient object segmentation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 49–56. IEEE, 2010.

[13] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.(CVPR)*, pages 136–145, 2017.

[14] Tiantian Wang, Ali Borji, Lihe Zhang, Pingping Zhang, and Huchuan Lu. A stagewise refinement model for detecting salient objects in images. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 4039–4048, 2017.

[15] Tiantian Wang, Lihe Zhang, Shuo Wang, Huchuan Lu, Gang Yang, Xiang Ruan, and Ali Borji. Detect globally, refine locally: A novel approach to saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3127–3135, 2018.

[16] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1155–1162, 2013.

[17] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3166–3173, 2013.

[18] Lu Zhang, Ju Dai, Huchuan Lu, You He, and Gang Wang. A bi-directional message passing model for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1741–1750, 2018.

[19] Pingping Zhang, Wei Liu, Huchuan Lu, and Chunhua Shen. Salient object detection by lossless feature reflection. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 1149–1155, 2018.

[20] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Xiang Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 202–211, 2017.

[21] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Baocai Yin. Learning uncertain convolutional features for accurate saliency detection. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 212–221, 2017.

[22] Xiaoning Zhang, Tiantian Wang, Jinqing Qi, Huchuan Lu, and Gang Wang. Progressive attention guided recurrent network for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 714–722, 2018.