ETHSeg: An Amodel Instance Segmentation Network and a Real-world Dataset for X-Ray Waste Inspection
– *Supplementary Material* –

Lingteng Qiu[1,2,3] Zhangyang Xiong[1,2] Xuhao Wang[2,3] Kenkun Liu[2] Yihan Li[2]
Guanying Chen[1,2] Xiaoguang Han[1,2*] Shuguang Cui[1,2,3]
[1]The Future Network of Intelligence Institute, CUHK-Shenzhen
[2]School of Science and Engineering, CUHK-Shenzhen [3]Shenzhen Research Institute of Big Data

1. More Quantitative Results on the Proposed Dataset

In this section, we present instance segmentation results for each category in Table 1. It is obvious that the results for all categories predicted by our method are consistently better than the baseline method, BCNet [6]. Our method also performs favorably against the state-of-the-art methods. These results clearly demonstrate the effectiveness of our easy-to-hard disassembling strategy and global structure guidance.

2. More Qualitative Results on the Proposed Dataset

Fig. 1 shows more qualitative comparisons between our ETHSeg and the state-of-the-art methods. We can see that our global structure guidance is helpful to predict contours that are closer to ground-truth edges, and our easy-to-hard disassembling strategy could not only alleviate the predicted areas belonging to false positive (FP) but also improve the accuracy of true positive (TP).

3. Visualization Results for the Easy-to-Hard Disassembling Strategy

To further verify the effectiveness of the proposed easy-to-hard disassembling strategy, we show qualitative results predicted by models trained with and without the easy-to-hard disassembling strategy. As shown in Fig. 2 and Fig. 3, the model trained with easy-to-hard disassembling strategy achieves more accurate results, demonstrating the effectiveness of our easy-to-hard disassembling strategy. Note that for the segmentation results predicted by top-down methods, we mainly compare the completeness of the segmentation area in each proposal since segmentation part is class-agnostic.

*corresponding author, hanxiaoguang@cuhk.edu.cn

References

[1] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee. Yolact: Real-time instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9157–9166, 2019.

[2] Z. Cai and N. Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018.

[3] H. Chen, K. Sun, Z. Tian, C. Shen, Y. Huang, and Y. Yan. Blendmask: Top-down meets bottom-up for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8573–8581, 2020.

[4] Y. Fang, S. Yang, X. Wang, Y. Li, C. Fang, Y. Shan, B. Feng, and W. Liu. Instances as queries. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6910–6919, 2021.

[5] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[6] L. Ke, Y.-W. Tai, and C.-K. Tang. Deep occlusion-aware instance segmentation with overlapping bilayers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4019–4028, 2021.

[7] Z. Tian, C. Shen, and H. Chen. Conditional convolutions for instance segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 282–298. Springer, 2020.

[8] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen. Solov2: Dynamic and fast instance segmentation. *arXiv preprint arXiv:2003.10152*, 2020.

Table 1. Instance segmentation results for each category on the proposed WIXray dataset.

Methods	**Backbone**	PlasticBottle	MealBox	Can	Carton	Glassbottle	Stick	FoodWaste	Tableware	HeatingPad	WasteDesiccant	WasteBattery	WasteBulb
Mask RCNN [5]	ResNet-101-FPN	43.61	39.83	68.32	25.31	69.19	1.03	37.03	6.49	53.10	42.91	52.70	74.79
Cascade Mask R-CNN [2]	ResNet-101-FPN	48.90	40.62	70.46	26.00	70.97	1.28	37.15	6.59	50.20	43.39	53.72	74.81
QueryInst [4]	ResNet-101-FPN	47.51	**42.18**	70.60	28.43	71.66	1.69	38.00	10.27	**51.78**	49.22	56.29	78.06
SOLOv2 [8]	ResNet-101-FPN	51.51	41.92	**71.54**	25.11	71.46	**3.05**	38.10	7.68	49.35	46.56	49.03	77.13
YOLACT [1]	ResNet-101-FPN	37.28	34.47	63.98	22.35	60.31	0.04	34.21	7.08	37.08	42.17	37.57	73.41
BlendMask [3]	ResNet-101-FPN	46.82	37.61	71.85	22.13	69.23	1.28	36.34	10.69	46.88	46.87	57.48	76.14
CondInst [7]	ResNet-101-FPN	47.66	37.29	69.68	22.05	68.68	2.13	37.04	10.47	45.07	45.68	58.89	**80.60**
BCNet [6]	ResNet-101-FPN	49.20	38.74	70.71	24.38	70.62	1.02	38.01	10.00	49.76	54.78	59.07	75.04
ETHSeg (ours)	ResNet-101-FPN	**50.36**	40.94	71.34	**28.83**	**71.66**	1.55	**40.15**	**12.21**	51.40	**56.43**	**59.73**	77.58

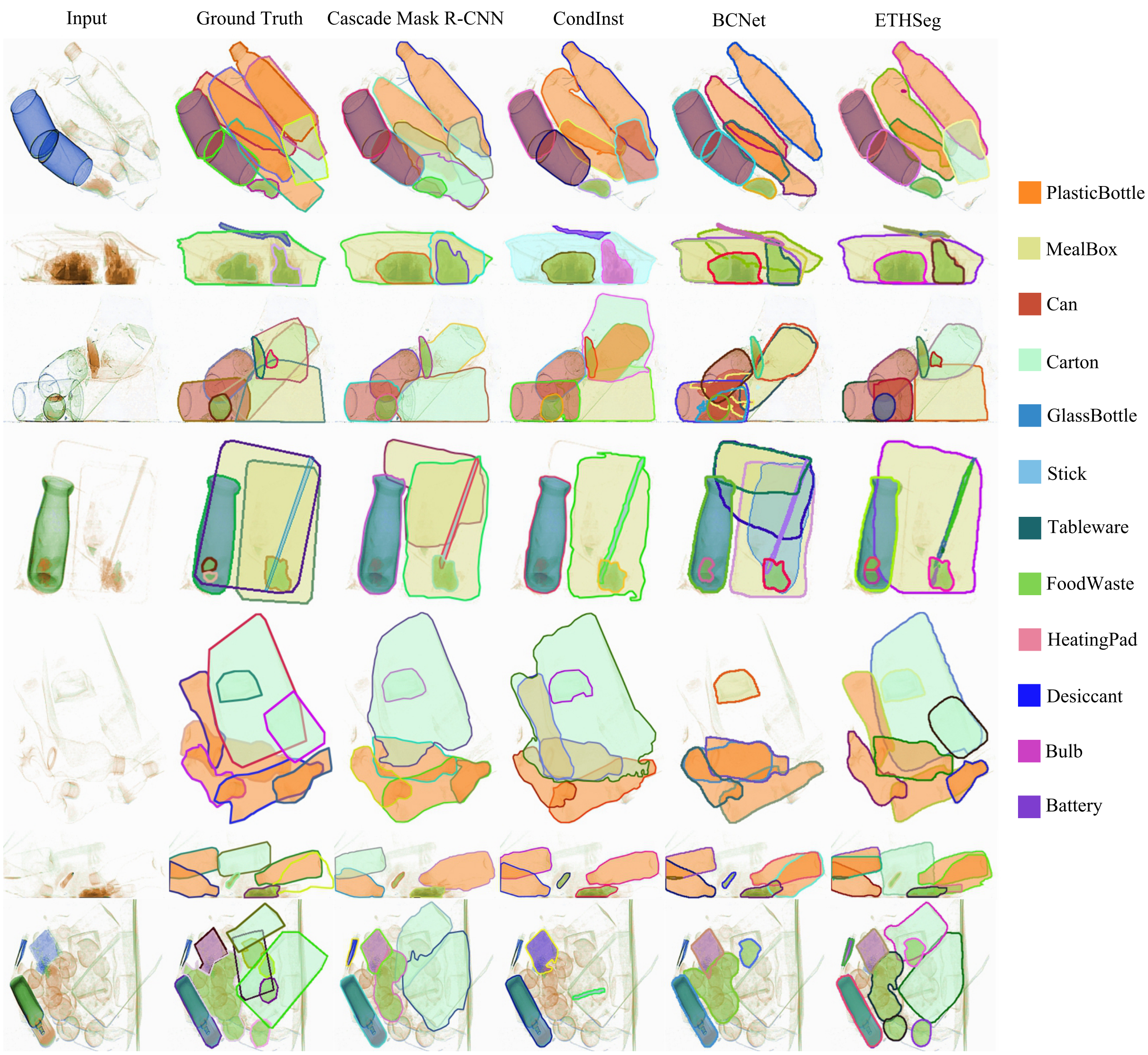

Figure 1. Qualitative comparison of instance segmentation on the proposed WIXray dataset. The mask color indicates the waste category and the boundary line is merely for identifying the instance contour.

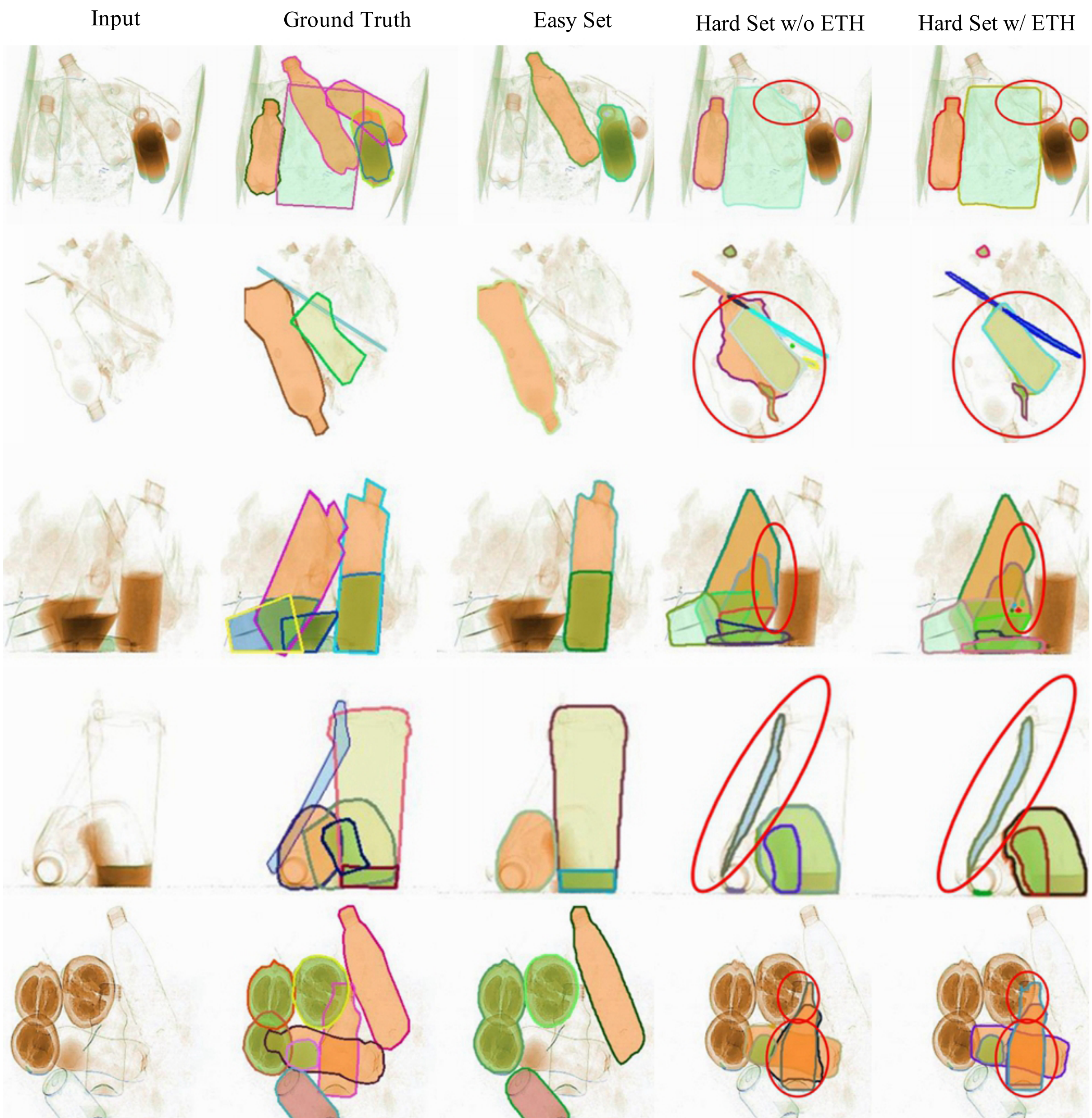

Figure 2. Visualization results for the easy-to-hard disassembling strategy (a). From left to right, each column represents input image, ground truth label, predicted easy set, predicted hard set w/o easy-to-hard disassembling, and predicted hard set w/ easy-to-hard disassembling strategy, respectively. The red cycles highlight the main different areas.

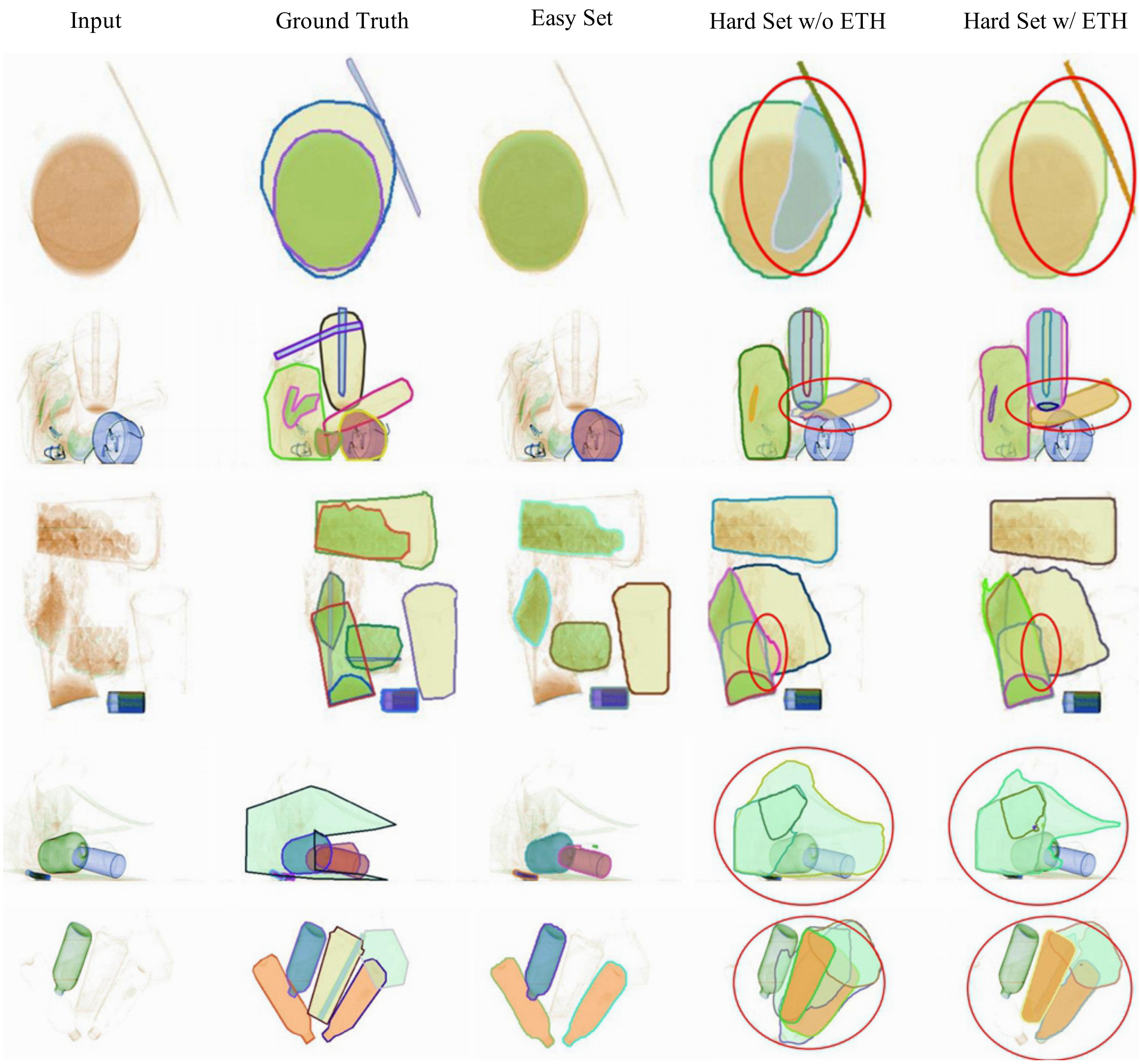

Figure 3. Visualization results for the easy-to-hard disassembling strategy (b). From left to right, each column represents input image, ground truth label, predicted easy set, predicted hard set w/o easy-to-hard disassembling, and predicted hard set w/ easy-to-hard disassembling strategy, respectively. The red cycles highlight the main different areas.