Appendix for "Weakly Supervised Semantic Segmentation using Out-of-Distribution Data"

Jungbeom Lee[1]   Seong Joon Oh[2,3]   Sangdoo Yun[2]   Junsuk Choe[4]   Eunji Kim[1]   Sungroh Yoon[1,5,*]

[1]Department of Electrical and Computer Engineering, Seoul National University

[2]NAVER AI Lab   [3]University of Tübingen

[4]Department of Computer Science and Engineering, Sogang University

[5]Interdisciplinary Program in AI, AIIS, ASRI, INMC, and ISRC, Seoul National University

A1. Additional Analysis

Examples of dining table: We provide some examples of localization maps for "dining table" in Figure A1, mentioned in Section 4.3.3 in the main paper.

Clustered OoD samples: We provide examples of OoD samples clustered by the K-means clustering algorithm in Figure A2.

More examples: Figure A3 presents examples of localization maps obtained by IRN [1] and our method, for the PASCAL VOC dataset. Figure A4 shows examples of segmentation maps predicted by IRN [1], AdvCAM [5], and our method.

Hyper-parameter analysis: We analyze the sensitivity of the mIoU of the initial seed to τ and λ, hyper-parameters involved in the W-OoD training. Since τ and λ are dependent on each other, they must be searched jointly. As the value of τ increases, $\mathcal{K}$ increases, so the value of $\mathcal{L}_d$ increases. Therefore, λ must decrease accordingly.

Segmentation results in mIoU with smaller $|\mathcal{D}_{\mathbf{ood}}|$: In the main paper, we provide the quality of the initial seed by using smaller $|\mathcal{D}_{\text{ood}}|$. We here provide the final segmentation results with smaller $|\mathcal{D}_{\text{ood}}|$ in Table A1.

Comparison of per-class mIoU scores: Table A3 shows the per-class mIoU of the final segmentation obtained by our method and recently produced methods.

Table A1. Segmentation results with smaller $|\mathcal{D}_{\text{ood}}|$ on Pascal VOC 2012 *val* set.

$|\mathcal{D}_{\text{ood}}|$	0	20	500	5190
mIoU	67.5	68.1	69.0	69.8

Table A2. Effect of the values of two hyper-parameters τ and λ.

τ	λ			
10	0.01	0.015	0.02	0.025
	52.5	52.6	53.1	52.2
20	0.003	0.005	0.007	0.01
	52.2	52.7	53.3	52.3
30	0.001	0.002	0.003	0.004
	50.8	52.1	52.5	52.6
40	0.0003	0.0005	0.0007	0.001
	50.8	50.8	51.0	50.7

References

[1] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. Weakly supervised learning of instance segmentation with inter-pixel relations. In *CVPR*, 2019. 1, 4

[2] Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In *CVPR*, 2018. 2

[3] Junsong Fan, Zhaoxiang Zhang, and Tieniu Tan. Cian: Cross-image affinity net for weakly supervised semantic segmentation. *AAAI*, 2020. 2

[4] Jungbeom Lee, Eunji Kim, Sungmin Lee, Jangho Lee, and Sungroh Yoon. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In *CVPR*, 2019. 2

[5] Jungbeom Lee, Eunji Kim, and Sungroh Yoon. Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation. In *CVPR*, 2021. 1, 2, 3, 4

[6] Jungbeom Lee, Jihun Yi, Chaehun Shin, and Sungroh Yoon. Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation. In *CVPR*, 2021. 2

[7] Wataru Shimoda and Keiji Yanai. Self-supervised difference detection for weakly-supervised semantic segmentation. In *ICCV*, 2019. 2

*Correspondence to: Sungroh Yoon (sryoon@snu.ac.kr).

Table A3. Comparison of per-class mIoU scores for the Pascal VOC 2012 dataset.

|	bkg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv	mIOU
Results on PASCAL VOC 2012 validation images:																						
PSA [2]	88.2	68.2	30.6	81.1	49.6	61.0	77.8	66.1	75.1	29.0	66.0	40.2	80.4	62.0	70.4	73.7	42.5	70.7	42.6	68.1	51.6	61.7
CIAN [3]	88.2	79.5	32.6	75.7	56.8	72.1	85.3	72.9	81.7	27.6	73.3	39.8	76.4	77.0	74.9	66.8	46.6	81.0	29.1	60.4	53.3	64.3
SEAM [8]	88.8	68.5	33.3	85.7	40.4	67.3	78.9	76.3	81.9	29.1	75.5	48.1	79.9	73.8	71.4	75.2	48.9	79.8	40.9	58.2	53.0	64.5
FickleNet [4]	89.5	76.6	32.6	74.6	51.5	71.1	83.4	74.4	83.6	24.1	73.4	47.4	78.2	74.0	68.8	73.2	47.8	79.9	37.0	57.3	64.6	64.9
SSDD [7]	89.0	62.5	28.9	83.7	52.9	59.5	77.6	73.7	87.0	34.0	83.7	47.6	84.1	77.0	73.9	69.6	29.8	84.0	43.2	68.0	53.4	64.9
BBAM [6]	92.7	80.6	33.8	83.7	64.9	75.5	91.3	80.4	88.3	37.0	83.3	62.5	84.6	80.8	74.7	80.0	61.6	84.5	48.6	85.8	71.8	73.7
AdvCAM [5]	89.5	76.9	33.5	80.3	63.7	68.6	89.7	77.9	87.6	31.6	77.2	36.2	82.6	78.7	73.5	69.8	51.9	81.9	43.8	70.9	52.6	67.5
W-OoD (ResNet-101)	91.2	80.1	34.0	82.5	68.5	72.9	90.3	80.8	89.3	32.3	78.9	31.1	83.6	79.2	75.4	74.4	58.0	81.9	45.2	81.3	54.8	69.8
W-OoD (WideResNet-38)	91.0	80.1	34.1	88.1	64.8	68.3	87.4	84.4	89.8	30.1	87.8	34.7	87.5	85.9	79.8	75.0	56.4	84.5	47.8	80.4	46.4	70.7
Results on PASCAL VOC 2012 test images:																						
PSA [2]	89.1	70.6	31.6	77.2	42.2	68.9	79.1	66.5	74.9	29.6	68.7	56.1	82.1	64.8	78.6	73.5	50.8	70.7	47.7	63.9	51.1	63.7
FickleNet [4]	90.3	77.0	35.2	76.0	54.2	64.3	76.6	76.1	80.2	25.7	68.6	50.2	74.6	71.8	78.3	69.5	53.8	76.5	41.8	70.0	54.2	65.3
SSDD [7]	89.0	62.5	28.9	83.7	52.9	59.5	77.6	73.7	87.0	34.0	83.7	47.6	84.1	77.0	73.9	69.6	29.8	84.0	43.2	68.0	53.4	64.9
BBAM [6]	92.8	83.5	33.4	88.9	61.8	72.8	90.3	83.5	87.6	34.7	82.9	66.1	83.9	81.1	78.3	77.4	55.2	86.7	58.5	81.5	66.4	73.7
AdvCAM [5]	89.3	79.3	32.5	80.2	56.3	62.8	87.2	80.8	87.0	28.9	78.3	41.3	82.1	80.6	77.7	68.5	51.2	80.8	55.3	60.8	48.1	67.1[1]
W-OoD (ResNet-101)	91.4	85.3	32.8	79.8	59.0	68.4	88.1	82.2	88.3	27.4	76.7	38.7	84.3	81.1	80.3	72.8	57.8	82.4	59.5	79.5	52.6	69.9[2]
W-OoD (WideResNet-38)	90.9	83.1	35.6	89.0	61.5	63.0	86.2	80.8	89.9	29.6	79.6	40.1	82.1	81.0	82.6	74.0	60.1	85.3	58.0	71.9	47.0	70.1[3]
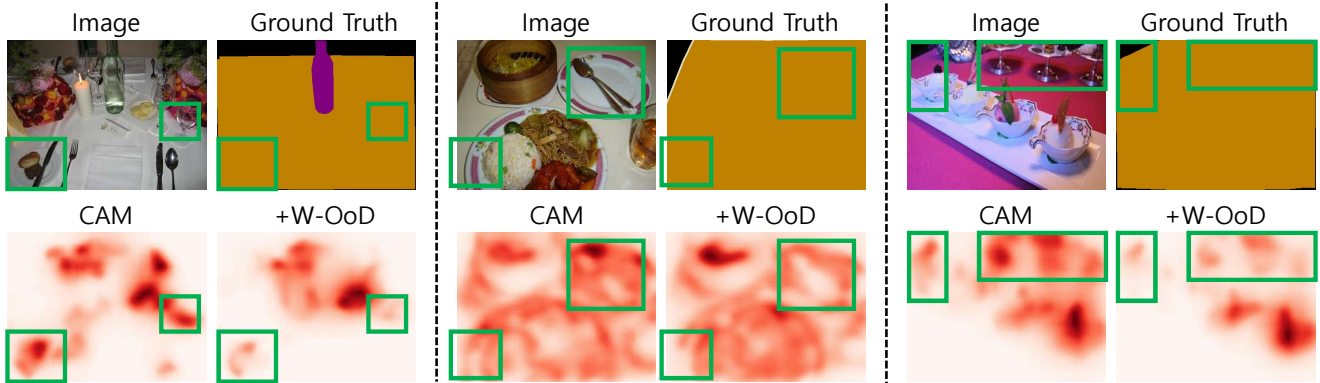
Figure A1. Example where the objects on the dining table are not identified as foreground by our method.

[8] Yude Wang, Jie Zhang, Meina Kan, Shiguang Shan, and Xilin Chen. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In *CVPR*, 2020. 2
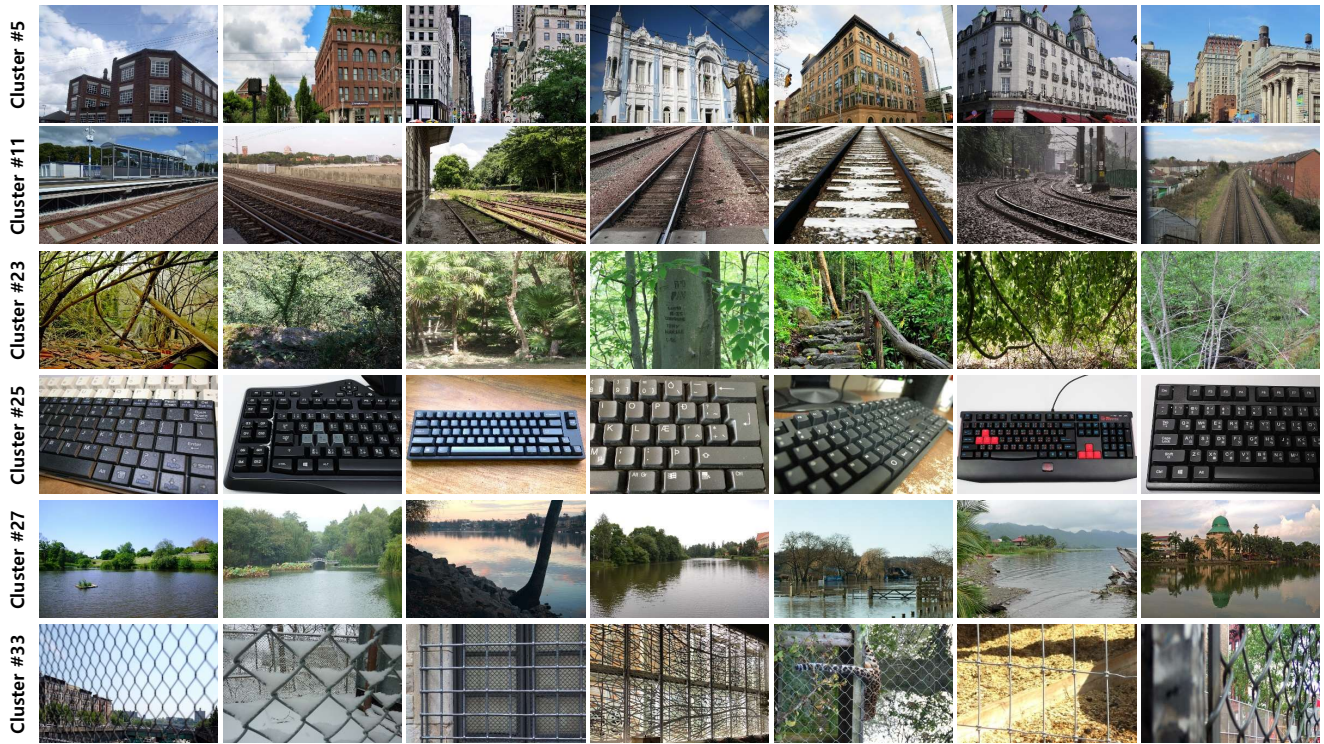
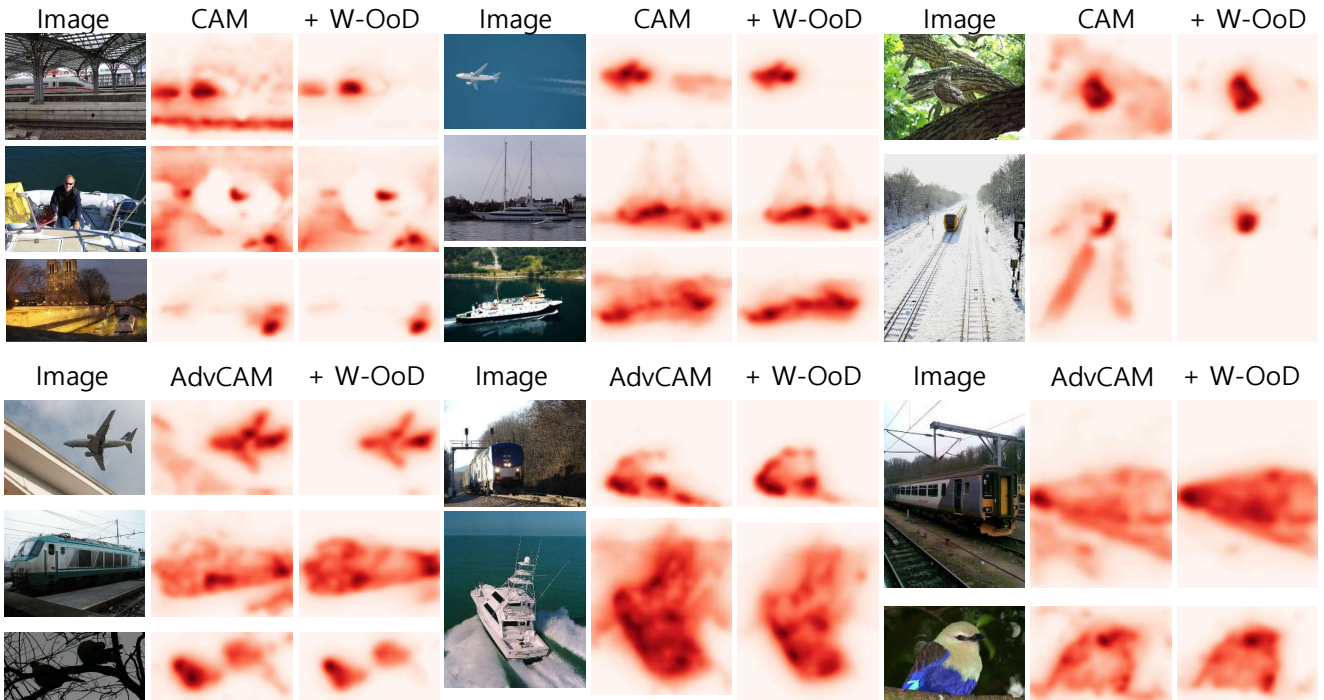Figure A2. Examples of OoD samples for each cluster, obtained by the K-means clustering algorithm..

Figure A3. Examples of localization maps obtained from CAM and CAM+W-OoD (upper), and AdvCAM [5] and AdvCAM+W-OoD (lower).
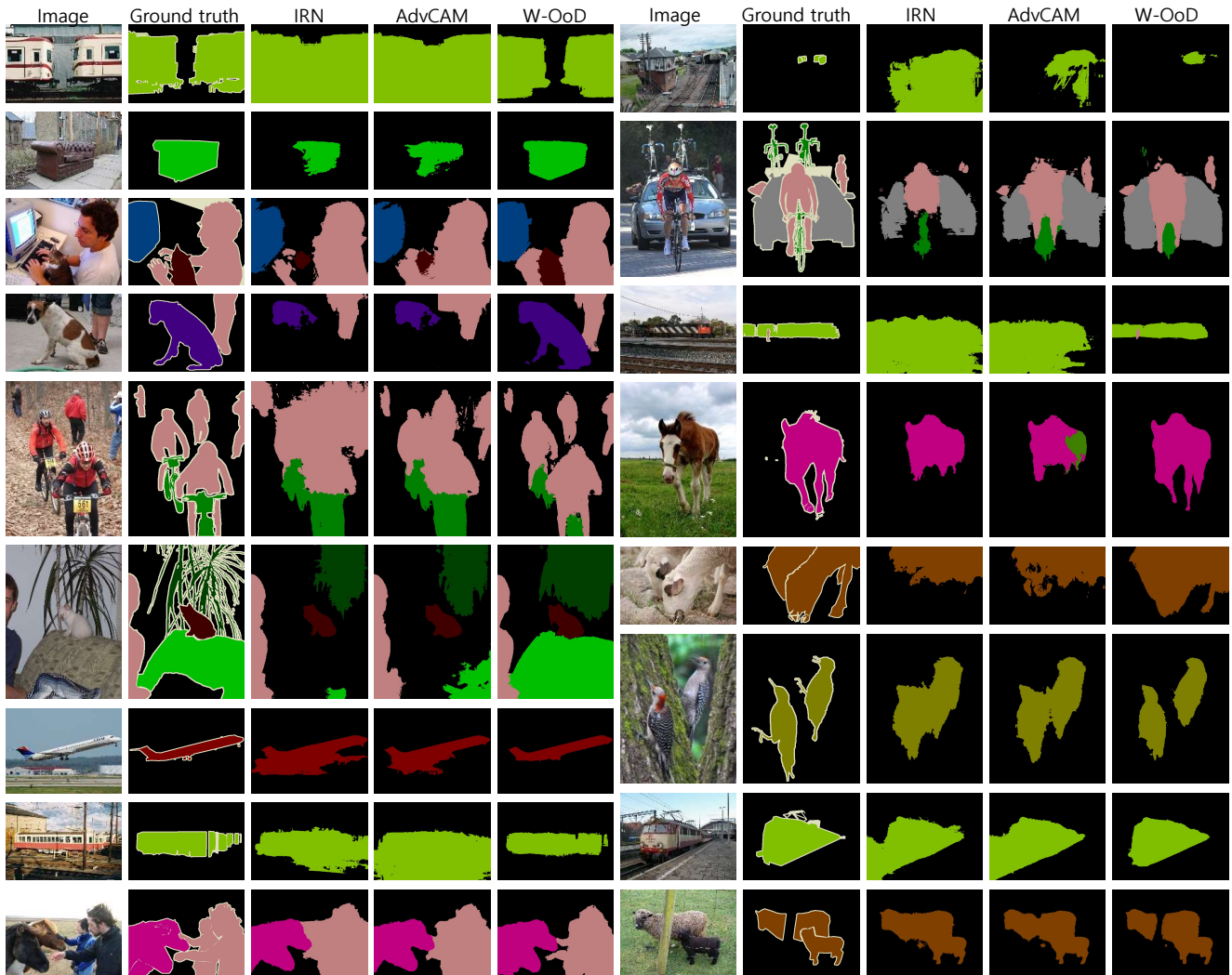
Figure A4. Examples of segmentation masks obtained by IRN [1], AdvCAM [5], and our method.