Appendix for Anti-Adversarially Manipulated Attributions for Weakly and Semi-Supervised Semantic Segmentation

Jungbeom Lee[1] Eunji Kim[1] Sungroh Yoon[1,2]
[1] Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea
[2] ASRI, INMC, ISRC, and Institute of Engineering Research, Seoul National University
{jbeom.lee93, kce407, sryoon}@snu.ac.kr

S1. Implementation Details

Details for Adversarial Climbing: Many recent studies [3, 11, 12] rely on the procedure of PSA [2] and IRN [1] for generating a CAM: a single image is flipped and resized with four different scales of $\{0.5, 1.0, 1.5, 2.0\}$, and the CAMs are extracted from those eight images. Those CAMs are aggregated into a single map by pixel-wise sum pooling.

We manipulate those eight images independently for adversarial climbing, resulting in eight localization maps ($\mathcal{A}$). We obtain a final map by aggregating those eight maps into a single map by pixel-wise sum pooling.

Details for Semantic Segmentation: We used the PyTorch implementation of DeepLab-v2-ResNet101[1] to train our segmentation network. We used multi-scale testing during inference time following [1, 2, 5, 6, 11]. Specifically, an input image is resized with four different scales of $\{0.5, 0.75, 1.0, 1.25\}$. These images are fed into the segmentation network independently, and the outputs are aggregated into a single map by pixel-wise max pooling, resulting in the final segmentation map. The experiments were performed on NVIDIA Tesla V100 GPUs.

S2. Additional Analysis

Threshold analysis: As mentioned in Section 4.2 of the main paper, we report the best initial seed performance by applying a range of thresholds to separate the foreground and background in the map $\mathcal{A}$. We present the effectiveness of this threshold by evaluating the initial seed, separated by a range of thresholds, in terms of mIoU. Figure S1(a) shows the mIoU of the initial seed obtained from the 'CAM', 'AdvCAM without regularization', and 'AdvCAM with regularization'. We select $t = 8$ for 'AdvCAM without regularization' and $t = 27$ for 'AdvCAM with regularization', which are the best values of t for each setting according to Figure 5(a) in the main paper.

Effects of suppressing other classes: Section 3.4 in the main paper has proposed two regularization terms: 1) suppressing other classes and 2) inhibiting excessive concentration. The effectiveness of the latter was dealt with in-depth in the main paper (please see Section 5). We will now focus on the effectiveness of suppressing other classes. To isolate the effect of this regularization procedure, we exclude the masking technique in all experiments here.

Figure S1(b) shows the mIoU of the initial seed for each adversarial iteration with and without the regularization of suppressing other classes. We can see that using this regularization technique provides better adversarial manipulation.

Comparison of per-class mIoU scores: Table S1 shows the per-class mIoU of our method and recently produced methods.

Additional mask examples on semantic segmentation. Figure S2 shows more examples of the semantic masks from FickleNet [5], IRN [1], CCT [9], and our method.

Additional examples of localization maps by adversarial climbing. Figure S3 shows additional examples of successive attribution maps obtained from images manipulated by iterative adversarial climbing.

[1] https://github.com/kazuto1011/deeplab-pytorch

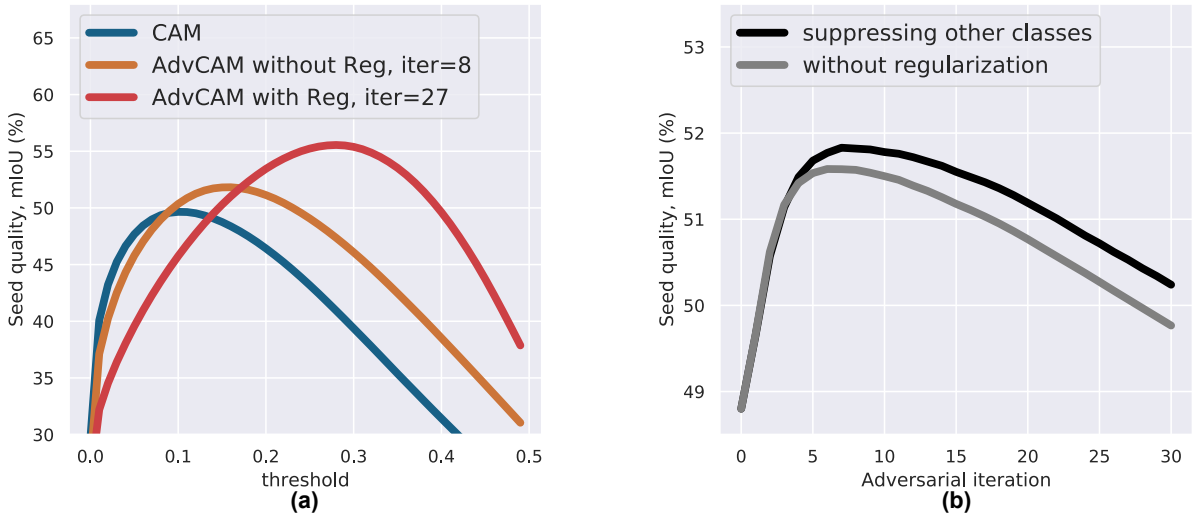

Figure S1: Examples of initial CAMs and successive localization maps obtained from images manipulated by iterative adversarial climbing.

	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:																						
GAIN [8]	87.6	76.7	33.9	74.5	58.5	61.7	75.9	72.9	78.6	18.8	70.8	14.1	68.7	69.6	69.5	71.3	41.5	66.5	16.4	70.2	48.7	59.4
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 [4]	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 [11]	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 [5]	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 [10]	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
Lee *et al.* [6]	90.8	82.2	35.1	82.4	72.2	71.4	82.7	75.0	86.9	18.3	74.2	29.6	81.1	79.2	74.7	76.4	44.2	78.6	35.4	72.8	63.0	66.5
BBAM [7]	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 (Ours, weak)	90.0	79.8	34.1	82.6	63.3	70.5	89.4	76.0	87.3	31.4	81.3	33.1	82.5	80.8	74.0	72.9	50.3	82.3	42.2	74.1	52.9	68.1
AdvCAM (Ours, semi)	94.4	91.7	65.6	89.1	72.4	72.8	93.4	86.0	90.4	37.5	90.6	58.6	84.5	88.9	83.3	84.9	62.0	81.6	49.5	85.9	71.8	77.8
Results on PASCAL VOC 2012 test images:																						
GAIN [8]	88.2	79.3	33.7	67.9	50.5	62.5	76.0	72.2	77.6	20.3	65.8	19.5	72.6	73.0	75.2	71.4	42.4	72.8	21.4	61.5	48.6	59.6
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 [5]	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	65.0
SSDD [10]	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
Lee *et al.* [6]	91.2	84.2	37.9	81.6	53.8	70.6	79.2	75.6	82.3	29.3	76.2	35.6	81.4	80.5	79.9	76.8	44.7	83.0	36.1	74.1	60.3	67.4
BBAM [7]	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 (Ours, weak)	90.1	81.2	33.6	80.4	52.4	66.6	87.1	80.5	87.2	28.9	80.1	38.5	84.0	83.0	79.5	71.9	47.5	80.8	59.1	65.4	49.7	68.0
AdvCAM (Ours, semi)	94.3	93.6	65.7	90.3	54.2	74.4	91.7	85.6	91.7	28.2	88.1	67.4	86.2	88.5	89.4	82.6	62.2	87.2	47.6	80.5	65.3	76.9

Table S1: Comparison of per-class mIoU scores.

Figure S2: Examples of predicted semantic masks for PASCAL VOC *val* images in weakly and semi-superivsed manner.

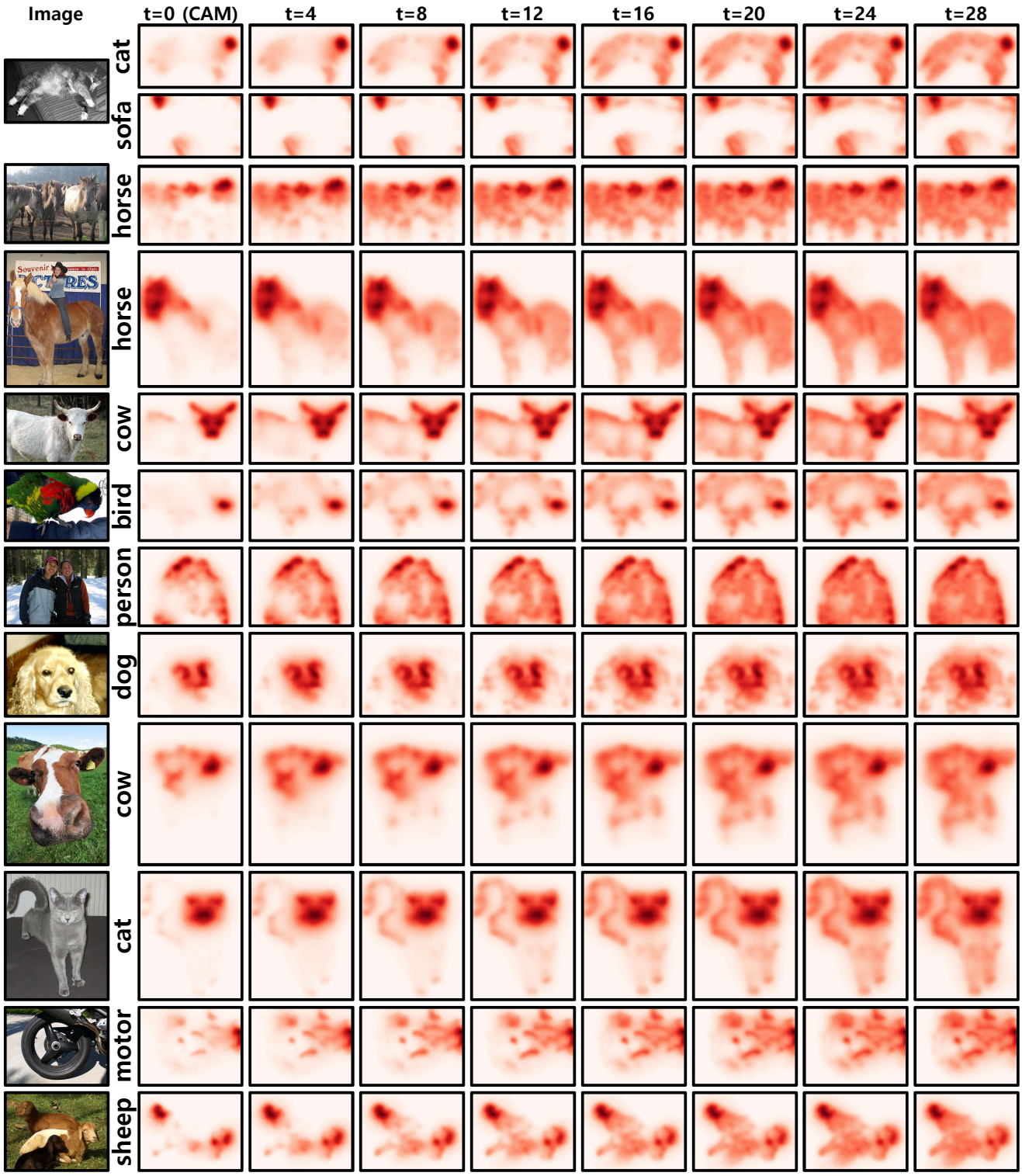

Figure S3: Examples of initial CAMs and successive localization maps obtained from images manipulated by iterative adversarial climbing.

References

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

[2] Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In *CVPR*, 2018. 1, 2

[3] Yu-Ting Chang, Qiaosong Wang, Wei-Chih Hung, Robinson Piramuthu, Yi-Hsuan Tsai, and Ming-Hsuan Yang. Weakly-supervised semantic segmentation via sub-category exploration. In *CVPR*, 2020. 1

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

[5] 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. 1, 2

[6] Jungbeom Lee, Eunji Kim, Sungmin Lee, Jangho Lee, and Sungroh Yoon. Frame-to-frame aggregation of active regions in web videos for weakly supervised semantic segmentation. In *ICCV*, 2019. 1, 2

[7] Jungbeom Lee, Jihun Yi, Chaehun Shin, and Sungroh Yoon. Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation. *arXiv preprint arXiv:2103.08907*, 2021. 2

[8] Kunpeng Li, Ziyan Wu, Kuan-Chuan Peng, Jan Ernst, and Yun Fu. Guided attention inference network. *IEEE transactions on pattern analysis and machine intelligence*, 42(12):2996–3010, 2019. 2

[9] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *CVPR*, 2020. 1

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

[11] Yude Wang, Jie Zhang, Meina Kan, Shiguang Shan, and Xilin Chen. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In *CVPR*, 2020. 1, 2

[12] Dong Zhang, Hanwang Zhang, Jinhui Tang, Xiansheng Hua, and Qianru Sun. Causal intervention for weakly-supervised semantic segmentation. 2020. 1