Supplementary Material

MARS: Model-agnostic Biased Object Removal without Additional Supervision for Weakly-Supervised Semantic Segmentation

Sanghyun Jo[1] In-Jae Yu[2] Kyungsu Kim[3*]
[1]OGQ, Seoul, Korea [2]Samsung Electronics, Suwon, Korea
[3]Department of Data Convergence and Future Medicine, Sungkyunkwan University, Seoul, Korea
{shjo.april, ijyu.phd, kskim.doc}@gmail.com

A. Additional Analysis

A.1. Examples of All Biased Objects

In Fig. 1, we introduce two observations: (1) The severe FP of some classes causes the performance gap between existing WSSS methods [22, 8] and FSSS, (2) 35% of all classes (*i.e.*, problematic classes) activate target objects (*e.g.*, boat, train, bird, and aeroplane) with biased objects (*e.g.*, sea, railroad, rock, and vapour trail). Following Fig. 1(c), we present additional examples of biased objects for all problematic classes in Fig. A. We hope that our detailed analysis of the biased problem in WSSS encourages other researchers to develop more robust and future WSSS approaches related to the biased problem.

A.2. Effect of Selecting Debiased Centroids

In Sec. 3.2, we present selecting target objects among separated objects of all images after disentangling target and biased objects using the USS-based clustering in Sec. 3.2. To evaluate the accuracy of debiased centroids, we measure how many selected centroids are target centroids among separated centroids of all images for each class in Fig. 14. Following Fig. 8 in Sec. 4.3, we employ the T-SNE [21] and the same criterion to classify target and biased centroids using pixel-wise annotations. In our experiments, the minimum accuracy for all classes on the PASCAL VOC 2012 $train$ dataset is 85%. These results mean that the proposed selection using background information from other images successfully chooses target centroids in the group of target and biased centroids.

A.3. Additional Category-wise Improvements

In line with Fig. 9, we evaluate per-class improvements of four WSSS methods [1, 22, 12, 8] with our method. All WSSS methods with ours show consistent improvements for top-3 classes (*i.e.*, bicycle, train, and boat) in our FP analysis in Fig. 1(b). Also, the performance of non-problematic classes (*e.g.*, person, dog, and cat) are improved by removing minor inconsistent objects (*e.g.*, legs of the horse) when complementing debiased labels in Sec. 3.4. However, a few categories (*e.g.*, chair, dining table, and potted plant) show inconsistent improvements due to the poor quality of initial WSSS labels. As a result, our method improves less when the WSSS method performs erroneously, albeit our method improves performance for most categories.

A.4. Qualitative Analysis with Existing Approaches

In addition to the quantitative comparison (see Tab. 5), Fig. F illustrates a qualitative comparison of our method, ADELE [16], and W-OoD [13] using two WSSS methods [22, 12]. ADELE [16] enlarges biased pixels since it enforces consistency of all classes without considering biased objects (the fourth column). Meanwhile, W-OoD [13] removes biased objects (*e.g.*, railroad) by utilizing extra images collected from human annotators, but it increases FN for most classes (*e.g.*, train and aeroplane) due to implicitly training biased objects with collected images (the seventh column). Unlike these studies, to find biased pixels in WSSS labels, we first match biased objects with background information from other images by utilizing the USS features. Our MARS then complements biased pixels with the model's predictions to prevent increasing FN of non-biased pixels (*e.g.*, legs of animals) in the fifth and eighth columns. Therefore, our method achieves the fully-automatic biased removal by explicitly eliminating biased objects in pseudo labels.

A.5. USS Drawback for Multi-class Scenario.

Sec. 3.3 aims at identifying biased pixels by leveraging the debiased centroids. Notably, USS fails to distinguish

*Correspondence to

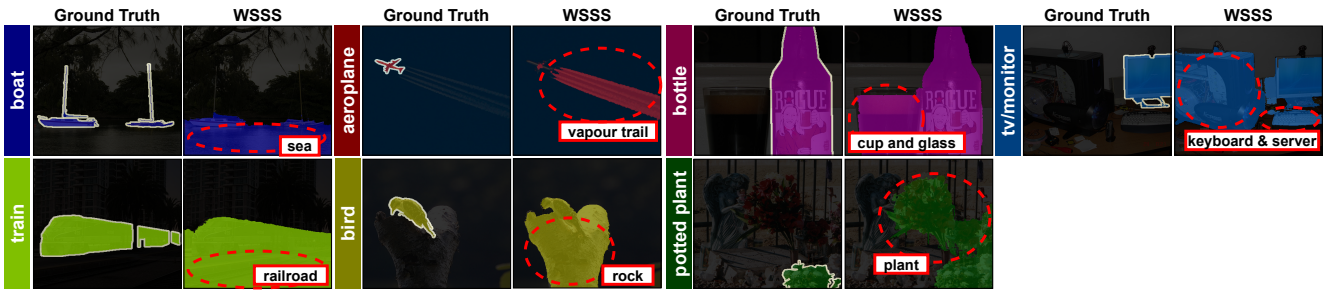

Figure A. Examples of all biased objects on the VOC dataset. Red dotted circles indicate the false activation of biased objects.

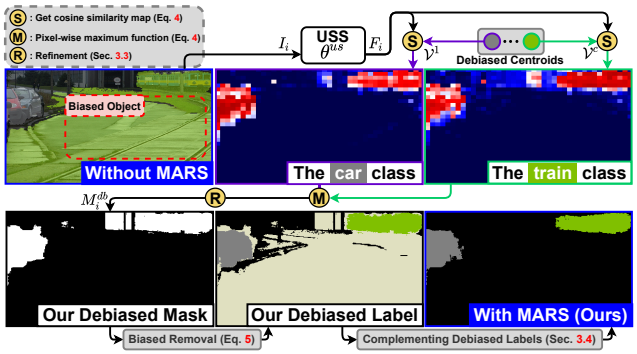

Figure B. The shortcoming of USS in a multi-class scenario. USS removes biased objects (*e.g.*, railroad) but fails to separate between similar classes (*e.g.*, car and train). To preserve multi-class information, we unify them and complement debiased labels.

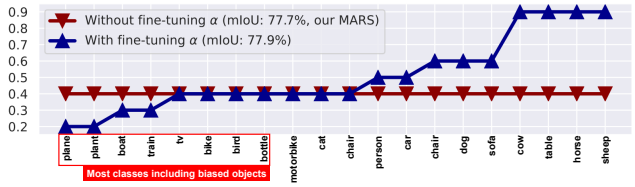

Figure C. Analysis of fine-tuning α on the PASCAL VOC 2012 validation set.

similar classes (*e.g.*, car and train) within the same super-category (*e.g.*, vehicle) in Fig. B. Consequently, we apply the pixel-wise maximum function to produce a binary mask in Eq. 4.

A.6. Related to Class-aware α and Background.

Since biased objects (*e.g.*, sea) appear in images across all foreground classes (*e.g.*, car and boat classes), we use a single shared background class. In Fig. 10(a), our method shows robustness against variations in K_{bg}. We also set K_{fg} to 2 to separate between target and biased objects. Fig. C shows 1) tuning α results in a marginal improvement, and 2) the optimal α is contingent upon the presence or absence of biased objects. Nevertheless, in Fig. 10(b), we use a single α value for automatically eliminating biased objects.

B. Additional Results

B.1. Quantitative Results

We present per-class segmentation results for two popular benchmarks in Tabs A, B, and C. Our method significantly improves performance of train (+29.1%) and boat (+9.1%) classes, which suffer from the biased problem in Fig. 1, versus the previous state-of-the-art method (*i.e.*, RS+EPM [8]). Also, we first demonstrate performance improvements for most classes including biased objects on the MS COCO 2014 dataset. When analyzing performance of our method on the MS COCO 2014 dataset, we find some classes (*e.g.*, surfboard, tennis racket, and train) that contain biased objects (*e.g.*, sea, tennis court, and railroad), causing performance degradation in existing WSSS methods [7, 8]. By contrast, without additional human supervision, our method achieves significant improvements for most classes including surfboard (+44.3%), tennis racket (+43%), and train (+24.6%) versus the latest WSSS method [8].

B.2. Qualitative Results

The qualitative segmentation results produced by the latest method [8] and our MARS are displayed in Fig. G. Our MARS performs well in various objects or multiple instances and can achieve satisfactory segmentation performance in challenging scenes. Specifically, our method removes biased objects for problematic classes (*e.g.*, railroad in train, lake in boat, tennis court in tennis racket, and sea in surfboard), covers more object regions for large-scale objects (*e.g.*, horse, car, and dining table), and captures the accurate boundaries of small-scale objects (*e.g.*, bird) by complementing debiased labels with online predictions and considering the model's uncertainty. Our method shows superior performance in the qualitative and quantitative comparison with the previous state-of-the-art method (*i.e.*, RS+EPM [8]), demonstrating the effectiveness of our MARS for the real-world dataset with multiple labels and complex relationships.

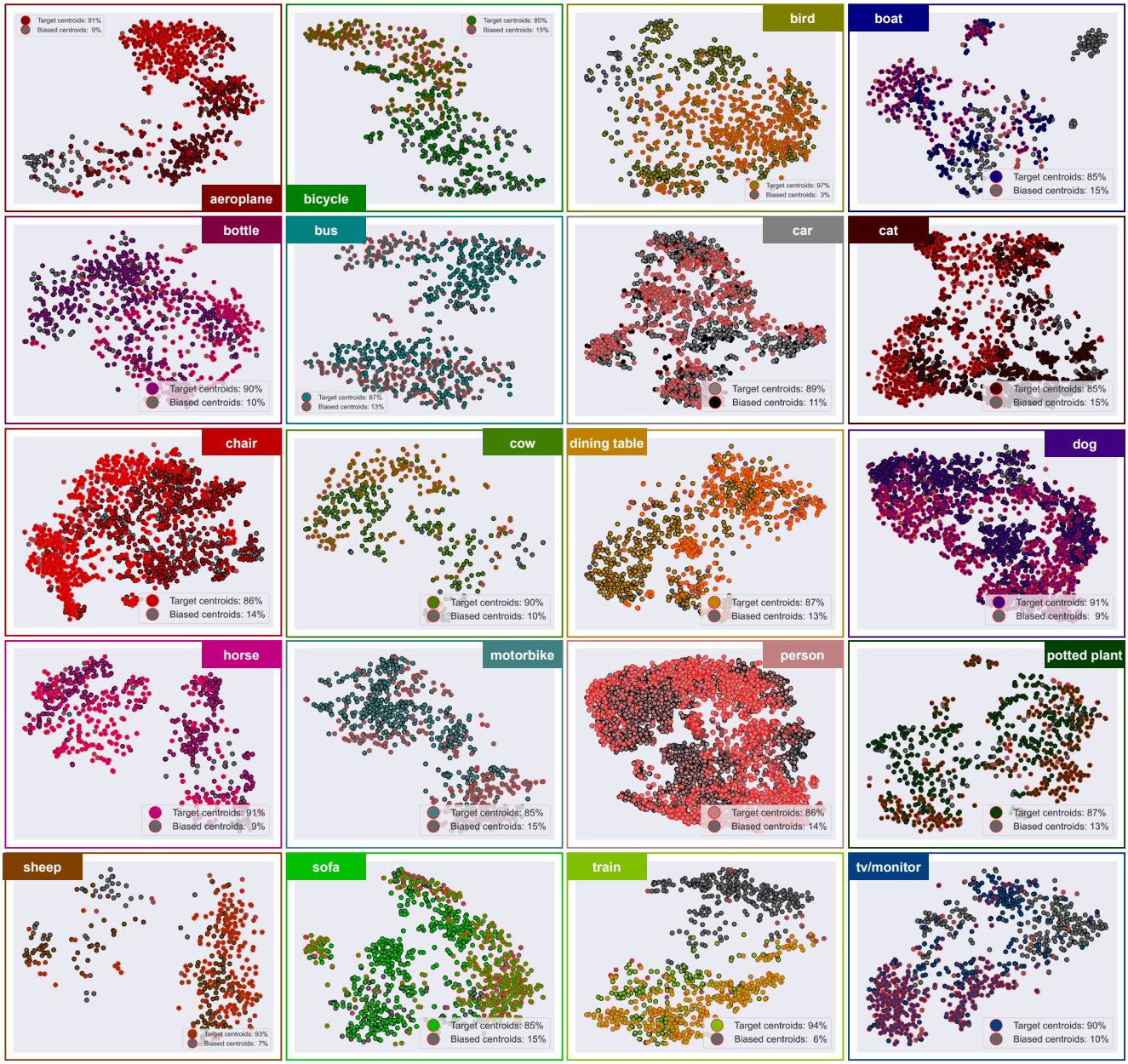

Figure D. Visualization of selecting debiased centroids for all classes on the PASCAL VOC 2012 *train* set. Red circles are selected centroids by our method. The average ratio of target centroids is more than 85%, showing the effectiveness of the proposed selection.

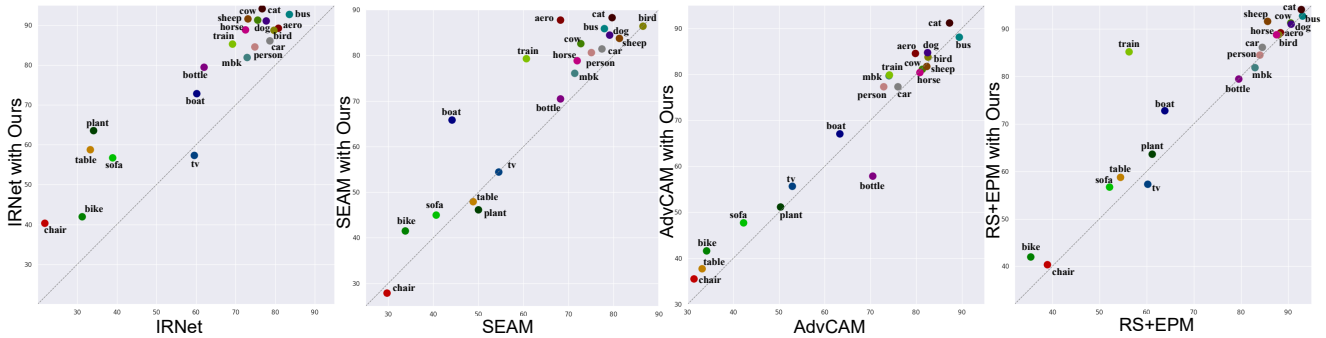

Figure E. Category-wise comparison with IRNet [1], SEAM [22], AdvCAM [12], RS+EPM [8], and ours in terms of the IoU (%) on PASCAL VOC 2012 *train* set.

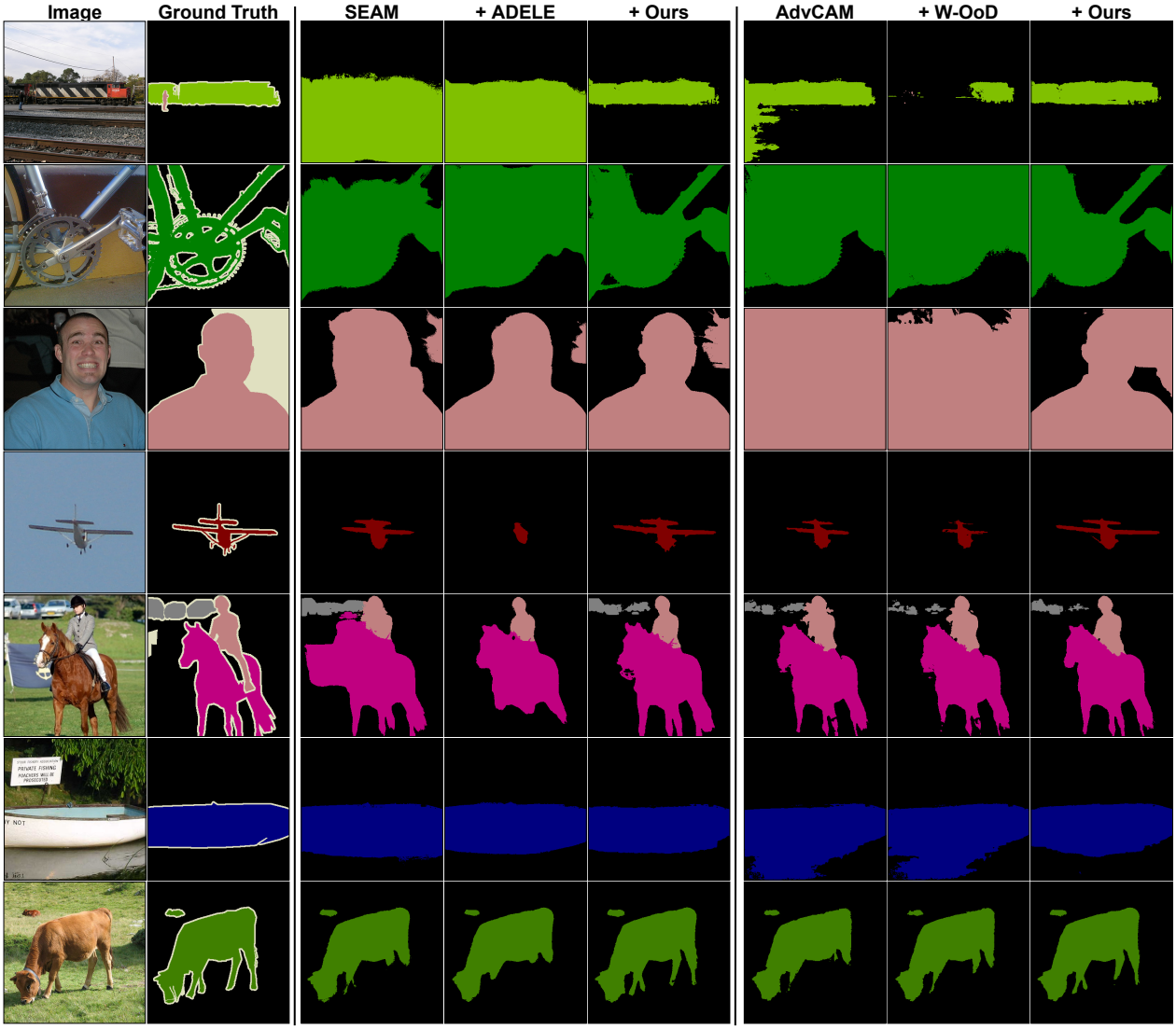

Figure F. Examples of final segmentation results on PASCAL VOC 2012 *val* set for SEAM [22], ADELE [16], AdvCAM [12], W-OoD [13], and Ours.

Table A. Class-specific performance comparisons with WSSS methods in terms of IoUs (%) on the PASCAL VOC 2012 *val* set.

Method	bkg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbk	person	plant	sheep	sofa	train	tv	mIoU
EM ICCV'15 [17]	67.2	29.2	17.6	28.6	22.2	29.6	47.0	44.0	44.2	14.6	35.1	24.9	41.0	34.8	41.6	32.1	24.8	37.4	24.0	38.1	31.6	33.8
MIL-LSE CVPR'15 [18]	79.6	50.2	21.6	40.9	34.9	40.5	45.9	51.5	60.6	12.6	51.2	11.6	56.8	52.9	44.8	42.7	31.2	55.4	21.5	38.8	36.9	42.0
SEC ECCV'16 [9]	82.4	62.9	26.4	61.6	27.6	38.1	66.6	62.7	75.2	22.1	53.5	28.3	65.8	57.8	62.3	52.5	32.5	62.6	32.1	45.4	45.3	50.7
TransferNet CVPR'16 [5]	85.3	68.5	26.4	69.8	36.7	49.1	68.4	55.8	77.3	6.2	75.2	14.3	69.8	71.5	61.1	31.9	25.5	74.6	33.8	49.6	43.7	52.1
CRF-RNN CVPR'17 [19]	85.8	65.2	29.4	63.8	31.2	37.2	69.6	64.3	76.2	21.4	56.3	29.8	68.2	60.6	66.2	55.8	30.8	66.1	34.9	48.8	47.1	52.8
WebCrawl CVPR'17 [6]	87.0	69.3	32.2	70.2	31.2	58.4	73.6	68.5	76.5	26.8	63.8	29.1	73.5	69.5	66.5	70.4	46.8	72.1	27.3	57.4	50.2	58.1
CIAN AAAI'20 [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
SSDD ICCV'19 [20]	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
PSA CVPR'18 [2]	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
FickleNet CVPR'19 [11]	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
RRM AAAI'20 [24]	87.9	75.9	31.7	78.3	54.6	62.2	80.5	73.7	71.2	30.5	67.4	40.9	71.8	66.2	70.3	72.6	49.0	70.7	38.4	62.7	58.4	62.6
SSSS CVPR'20 [3]	88.7	70.4	35.1	75.7	51.9	65.8	71.9	64.2	81.1	30.8	73.3	28.1	81.6	69.1	62.6	74.8	48.6	71.0	40.1	68.5	64.3	62.7
SEAM CVPR'20 [22]	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
AdvCAM CVPR'21 [12]	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
CPN ICCV'21 [25]	89.9	75.0	32.9	87.8	60.9	69.4	87.7	79.4	88.9	28.0	80.9	34.8	83.4	79.6	74.6	66.9	56.4	82.6	44.9	73.1	45.7	67.8
RIB NeurIPS'21 [10]	90.3	76.2	33.7	82.5	64.9	73.1	88.4	78.6	88.7	32.3	80.1	37.5	83.6	79.7	75.8	71.8	47.5	84.3	44.6	65.9	54.9	68.3
AMN CVPR'22 [14]	90.6	79.0	33.5	83.5	60.5	74.9	90.0	81.3	86.6	30.6	80.9	53.8	80.2	79.6	74.6	75.5	54.7	83.5	46.1	63.1	57.5	69.5
ADELE CVPR'22 [16]	91.1	77.6	33.0	88.9	67.1	71.7	88.8	82.5	89.0	26.6	83.8	44.6	84.4	77.8	74.8	78.5	43.8	84.8	44.6	56.1	65.3	69.3
W-OoD CVPR'22 [13]	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
RCA CVPR'22 [26]	91.8	88.4	39.1	85.1	69.0	75.7	86.6	82.3	89.1	28.1	81.9	37.9	85.9	79.4	82.1	78.6	47.7	84.4	34.9	75.4	58.6	70.6
SANCE CVPR'22 [15]	91.4	78.4	33.0	87.6	61.9	**79.6**	90.6	82.0	92.4	33.3	76.9	**59.7**	86.4	78.0	76.9	77.7	61.1	79.4	47.5	62.1	53.3	70.9
MCTformer CVPR'22 [23]	91.9	78.3	39.5	**89.9**	55.9	76.7	81.8	79.0	90.7	32.6	87.1	57.2	87.0	84.6	77.4	79.2	55.1	89.2	47.2	70.4	58.8	71.9
RS+EPM Arxiv'22 [8]	92.2	88.4	35.4	87.9	63.8	79.5	**93.0**	84.5	92.7	39.0	90.5	54.5	90.6	87.5	**83.0**	84.0	61.1	85.6	52.1	56.2	60.2	74.4
MARS (Ours)	**94.1**	**89.3**	**42.0**	88.8	**72.9**	79.5	92.7	**86.2**	**94.2**	**40.3**	**91.4**	58.8	**91.1**	**88.9**	81.9	**84.6**	**63.6**	**91.7**	**56.7**	**85.3**	57.3	**77.7**

Table B. Class-specific performance comparisons with WSSS methods in terms of IoUs (%) on the PASCAL VOC 2012 *test* set.

Method	bkg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbk	person	plant	sheep	sofa	train	tv	mIoU
EM ICCV'15 [17]	76.3	37.1	21.9	41.6	26.1	38.5	50.8	44.9	48.9	16.7	40.8	29.4	47.1	45.8	54.8	28.2	30.0	44.0	29.2	34.3	46.0	39.6
MIL-LSE CVPR'15 [18]	78.7	48.0	21.2	31.1	28.4	35.1	51.4	55.5	52.8	7.8	56.2	19.9	53.8	50.3	40.0	38.6	27.8	51.8	24.7	33.3	46.3	40.6
SEC ECCV'16 [9]	83.5	56.4	28.5	64.1	23.6	46.5	70.6	58.5	71.3	23.2	54.0	28.0	68.1	62.1	70.0	55.0	38.4	58.0	39.9	38.4	48.3	51.7
TransferNet CVPR'16 [5]	85.7	70.1	27.8	73.7	37.3	44.8	71.4	53.8	73.0	6.7	62.9	12.4	68.4	73.7	65.9	27.9	23.5	72.3	38.9	45.9	39.2	51.2
CRF-RNN CVPR'17 [19]	85.7	58.8	30.5	67.6	24.7	44.7	74.8	61.8	73.7	22.9	57.4	27.5	71.3	64.8	72.4	57.3	37.3	60.4	42.8	42.2	50.6	53.7
WebCrawl CVPR'17 [6]	87.2	63.9	32.8	72.4	26.7	64.0	72.1	70.5	77.8	23.9	63.6	32.1	77.2	75.3	76.2	71.5	45.0	68.8	35.5	46.2	49.3	58.7
PSA CVPR'18 [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 CVPR'19 [11]	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 ICCV'19 [20]	89.5	71.8	31.4	79.3	47.3	64.2	79.9	74.6	84.9	30.8	73.5	58.2	82.7	73.4	76.4	69.9	37.4	80.5	54.5	65.7	50.3	65.5
RRM AAAI'20 [24]	87.8	77.5	30.8	71.7	36.0	64.2	75.3	70.4	81.7	29.3	70.4	52.0	78.6	73.8	74.4	72.1	54.2	75.2	50.6	42.0	52.5	62.9
SSSS CVPR'20 [3]	88.7	70.4	35.1	75.7	51.9	65.8	71.9	64.2	81.1	30.8	73.3	28.1	81.6	69.1	62.6	74.8	48.6	71.0	40.1	68.5	**64.3**	62.7
SEAM CVPR'20 [22]	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
AdvCAM CVPR'21 [12]	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
CPN ICCV'21 [25]	90.4	79.8	32.9	85.7	52.8	66.3	87.2	81.3	87.6	28.2	79.7	50.1	82.9	80.4	78.8	70.6	51.1	83.4	55.4	68.5	44.6	68.5
RIB NeurIPS'21 [10]	90.4	80.5	32.8	84.9	59.4	69.3	87.2	83.5	88.3	31.1	80.4	44.0	84.4	82.3	80.9	70.7	43.5	84.9	55.9	59.0	47.3	68.6
AMN CVPR'22 [14]	90.7	82.8	32.4	84.8	59.4	70.0	86.7	83.0	86.9	30.1	79.2	56.6	83.0	81.9	78.3	72.7	52.9	81.4	59.8	53.1	56.4	69.6
W-OoD CVPR'22 [13]	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
RCA CVPR'22 [26]	92.1	86.6	40.0	90.1	60.4	68.2	89.8	82.3	87.0	27.2	86.4	32.0	85.3	88.1	83.2	78.0	59.2	86.7	45.0	71.3	52.5	71.0
SANCE CVPR'22 [15]	91.6	82.6	33.6	89.1	60.6	**76.0**	91.8	83.0	90.9	33.5	80.2	64.7	87.1	82.3	81.7	78.3	58.5	82.9	**60.9**	53.9	53.5	72.2
MCTformer CVPR'22 [23]	92.3	84.4	37.2	82.8	60.0	72.8	78.0	79.0	89.4	31.7	84.5	59.1	85.3	83.8	79.2	81.0	53.9	85.3	60.5	65.7	57.7	71.6
RS+EPM Arxiv'22 [8]	91.9	89.7	37.3	88.0	62.5	72.1	93.5	85.6	90.2	36.3	**88.3**	62.5	86.3	**89.1**	82.9	81.2	59.7	89.2	56.2	44.5	59.4	73.6
MARS (Ours)	**93.7**	**93.3**	**40.3**	**90.8**	**70.8**	71.7	**94.0**	**86.3**	**93.9**	**40.4**	87.6	**67.6**	**90.0**	87.3	**83.9**	**83.1**	**64.2**	**89.5**	59.6	79.0	55.1	**77.2**

Table C. Class-specific performance comparisons with WSSS methods in terms of IoUs (%) on the MS COCO 2014 *val* set.

Class	SEC [9]	DSRG [7]	RS+EPM [8]	MARS (Ours)	Class	SEC [9]	DSRG [7]	RS+EPM [8]	MARS (Ours)
background	74.3	80.6	83.6	**83.7**	wine glass	22.3	24.0	39.8	**45.5**
person	43.6	-	**74.9**	56.8	cup	17.9	20.4	38.9	**42.0**
bicycle	24.2	30.4	55.0	**59.2**	fork	1.8	0.0	**4.9**	1.7
car	15.9	22.1	50.1	**52.0**	knife	1.4	5.0	**9.0**	6.4
motorcycle	52.1	54.2	72.9	**75.2**	spoon	0.6	0.5	**1.1**	0.9
airplane	36.6	45.2	76.5	**79.6**	bowl	12.5	**18.8**	11.3	14.1
bus	37.7	38.7	72.5	**76.8**	banana	43.6	46.4	67.0	**67.7**
train	30.1	33.2	47.4	**72.0**	apple	23.6	24.3	**49.2**	47.9
truck	24.1	25.9	46.5	**54.1**	sandwich	22.8	24.5	33.7	**34.9**
boat	17.3	20.6	44.1	**52.1**	orange	44.3	41.2	62.3	**62.5**
traffic light	16.7	16.1	**60.8**	53.8	broccoli	36.8	35.7	**50.4**	45.9
fire hydrant	55.9	60.4	80.3	**80.9**	carrot	6.7	15.3	**35.0**	31.7
stop sign	48.4	51.0	**84.1**	76.8	hot dog	31.2	24.9	48.3	**51.5**
parking meter	25.2	26.3	**77.8**	74.8	pizza	50.9	56.2	**68.6**	68.0
bench	16.4	22.3	41.2	**47.2**	donut	32.8	34.2	62.3	**64.9**
bird	34.7	41.5	62.6	**72.3**	cake	12.0	6.9	48.3	**53.3**
cat	57.2	62.2	79.2	**80.9**	chair	7.8	9.7	28.9	**30.3**
dog	45.2	55.6	73.3	**76.3**	couch	5.6	17.7	44.9	**49.1**
horse	34.4	42.3	76.1	**78.2**	potted plant	6.2	14.3	16.9	**20.6**
sheep	40.3	47.1	80.0	**83.5**	bed	23.4	32.4	53.6	**55.9**
cow	41.4	49.3	79.3	**83.2**	dining table	0.0	3.8	**24.6**	17.4
elephant	62.9	67.1	85.6	**87.7**	toilet	38.5	43.6	71.1	**76.5**
bear	59.1	62.6	82.9	**87.5**	tv	19.2	25.3	49.9	**54.9**
zebra	59.8	63.2	87.0	**87.9**	laptop	20.1	21.1	56.6	**64.5**
giraffe	48.8	54.3	82.2	**83.4**	mouse	3.5	0.9	**17.4**	12.9
backpack	0.3	0.2	9.4	**11.9**	remote	17.5	20.6	54.8	**55.3**
umbrella	26.0	35.3	73.4	**77.1**	keyboard	12.5	12.3	48.8	**51.8**
handbag	0.5	0.7	4.6	**8.4**	cell phone	32.1	33.0	60.8	**64.6**
tie	6.5	7.0	17.2	**18.4**	microwave	8.2	11.2	43.6	**56.9**
suitcase	16.7	23.4	53.9	**57.2**	oven	13.7	12.4	38.0	**43.5**
frisbee	12.3	13.0	**57.7**	57.5	toaster	0.0	0.0	0.0	0.0
skis	1.6	1.5	8.2	**10.8**	sink	10.8	17.8	36.9	**40.7**
snowboard	5.3	16.3	24.7	**27.7**	refrigerator	4.0	15.5	51.8	**63.4**
sports ball	7.9	9.8	**41.6**	40.4	book	0.4	12.3	27.3	**29.2**
kite	9.1	17.4	62.6	**63.8**	clock	17.8	20.7	**23.3**	19.8
baseball bat	1.0	**4.8**	1.5	1.6	vase	18.4	23.9	26.0	**31.0**
baseball glove	0.6	**1.2**	0.4	0.3	scissors	16.5	17.3	**47.1**	47.0
skateboard	7.1	14.4	34.8	**34.9**	teddy bear	47.0	46.3	68.8	**69.5**
surfboard	7.7	13.5	17.0	**61.3**	hair drier	0.0	0.0	0.0	0.0
tennis racket	9.1	6.8	9.0	**52.0**	toothbrush	2.8	2.0	19.7	**32.2**
bottle	13.2	22.3	**38.1**	36.6	**mIoU**	22.4	26.0	46.4	**49.4**

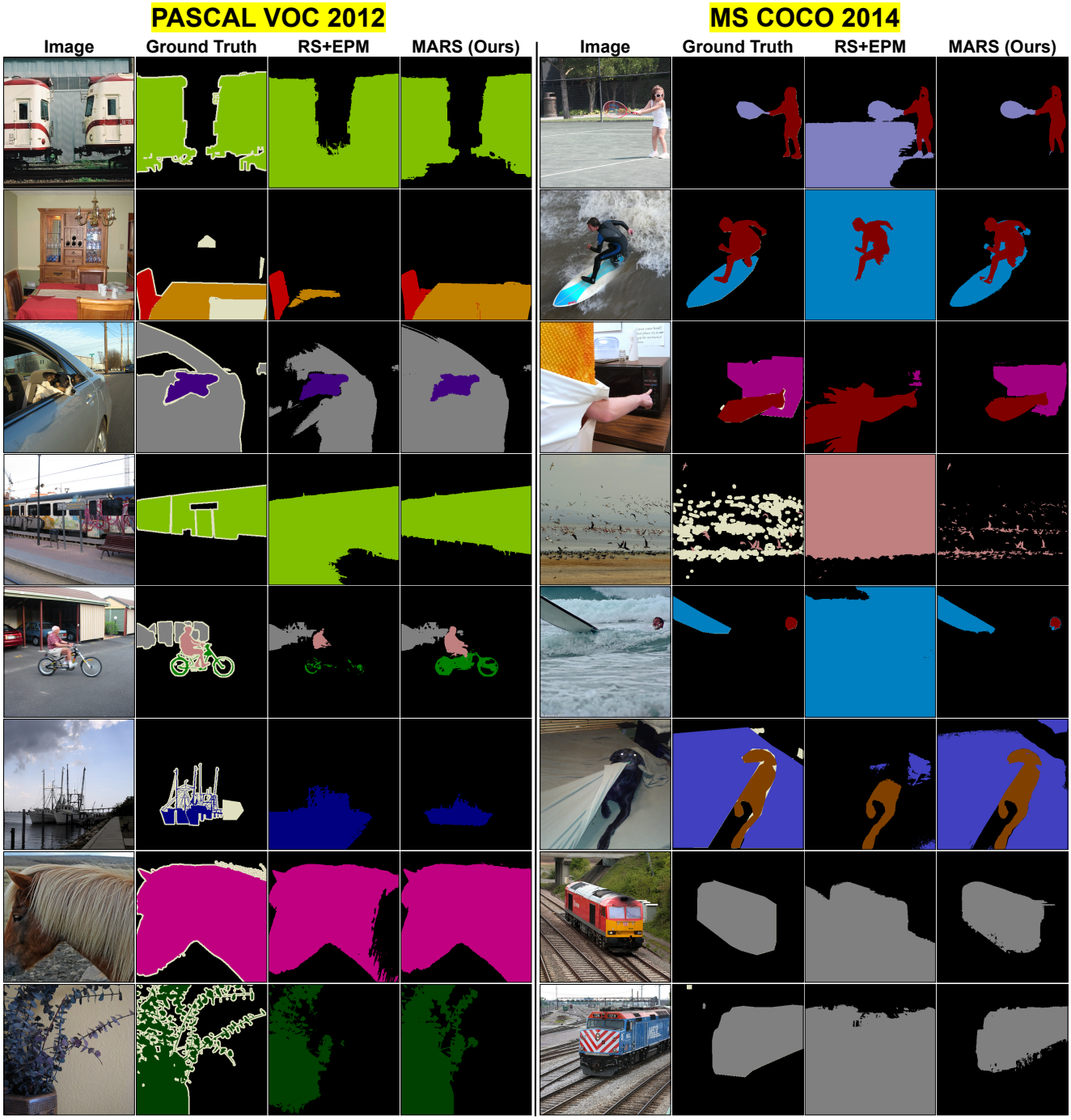

Figure G. Qualitative segmentation results of the latest method (*i.e.*, RS+EPM [8]) and the proposed MARS on PASCAL VOC 2012 and MS COCO 2014 validation sets.

References

[1] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. Weakly supervised learning of instance segmentation with inter-pixel relations. In *IEEE CVPR*, pages 2209–2218, 2019. 1, 3

[2] Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In *IEEE CVPR*, pages 4981–4990, 2018. 5

[3] Nikita Araslanov and Stefan Roth. Single-stage semantic segmentation from image labels. In *IEEE CVPR*, pages 4253–4262, 2020. 5

[4] Junsong Fan, Zhaoxiang Zhang, Tieniu Tan, Chunfeng Song, and Jun Xiao. Cian: Cross-image affinity net for weakly supervised semantic segmentation. In *AAAI*, volume 34, pages 10762–10769, 2020. 5

[5] Seunghoon Hong, Junhyuk Oh, Honglak Lee, and Bohyung Han. Learning transferrable knowledge for semantic segmentation with deep convolutional neural network. In *IEEE CVPR*, pages 3204–3212, 2016. 5

[6] Seunghoon Hong, Donghun Yeo, Suha Kwak, Honglak Lee, and Bohyung Han. Weakly supervised semantic segmentation using web-crawled videos. In *IEEE CVPR*, pages 7322–7330, 2017. 5

[7] Zilong Huang, Xinggang Wang, Jiasi Wang, Wenyu Liu, and Jingdong Wang. Weakly-supervised semantic segmentation network with deep seeded region growing. In *IEEE CVPR*, pages 7014–7023, 2018. 2, 6

[8] Sanghyun Jo, In-Jae Yu, and Kyungsu Kim. Recurseed and edgepredictmix: Single-stage learning is sufficient for weakly-supervised semantic segmentation. *arXiv preprint arXiv:2204.06754*, 2022. 1, 2, 3, 5, 6, 7

[9] Alexander Kolesnikov and Christoph H Lampert. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In *ECCV*, pages 695–711. Springer, 2016. 5, 6

[10] Jungbeom Lee, Jooyoung Choi, Jisoo Mok, and Sungroh Yoon. Reducing information bottleneck for weakly supervised semantic segmentation. *NeurIPS*, 34, 2021. 5

[11] Jungbeom Lee, Eunji Kim, Sungmin Lee, Jangho Lee, and Sungroh Yoon. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In *IEEE CVPR*, pages 5267–5276, 2019. 5

[12] Jungbeom Lee, Eunji Kim, and Sungroh Yoon. Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation. In *IEEE CVPR*, pages 4071–4080, 2021. 1, 3, 4, 5

[13] Jungbeom Lee, Seong Joon Oh, Sangdoo Yun, Junsuk Choe, Eunji Kim, and Sungroh Yoon. Weakly supervised semantic segmentation using out-of-distribution data. In *IEEE CVPR*, pages 16897–16906, 2022. 1, 4, 5

[14] Minhyun Lee, Dongseob Kim, and Hyunjung Shim. Threshold matters in wsss: Manipulating the activation for the robust and accurate segmentation model against thresholds. In *IEEE CVPR*, pages 4330–4339, 2022. 5

[15] Jing Li, Junsong Fan, and Zhaoxiang Zhang. Towards noiseless object contours for weakly supervised semantic segmentation. In *IEEE CVPR*, pages 16856–16865, 2022. 5

[16] Sheng Liu, Kangning Liu, Weicheng Zhu, Yiqiu Shen, and Carlos Fernandez-Granda. Adaptive early-learning correction for segmentation from noisy annotations. In *IEEE CVPR*, pages 2606–2616, 2022. 1, 4, 5

[17] George Papandreou, Liang-Chieh Chen, Kevin P Murphy, and Alan L Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *IEEE ICCV*, pages 1742–1750, 2015. 5

[18] Pedro O Pinheiro and Ronan Collobert. From image-level to pixel-level labeling with convolutional networks. In *IEEE CVPR*, pages 1713–1721, 2015. 5

[19] Anirban Roy and Sinisa Todorovic. Combining bottom-up, top-down, and smoothness cues for weakly supervised image segmentation. In *IEEE CVPR*, pages 3529–3538, 2017. 5

[20] Wataru Shimoda and Keiji Yanai. Self-supervised difference detection for weakly-supervised semantic segmentation. In *IEEE ICCV*, pages 5208–5217, 2019. 5

[21] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 1

[22] Yude Wang, Jie Zhang, Meina Kan, Shiguang Shan, and Xilin Chen. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In *IEEE CVPR*, pages 12275–12284, 2020. 1, 3, 4, 5

[23] Lian Xu, Wanli Ouyang, Mohammed Bennamoun, Farid Boussaid, and Dan Xu. Multi-class token transformer for weakly supervised semantic segmentation. In *IEEE CVPR*, pages 4310–4319, 2022. 5

[24] Bingfeng Zhang, Jimin Xiao, Yunchao Wei, Mingjie Sun, and Kaizhu Huang. Reliability does matter: An end-to-end weakly supervised semantic segmentation approach. In *AAAI*, volume 34, pages 12765–12772, 2020. 5

[25] Fei Zhang, Chaochen Gu, Chenyue Zhang, and Yuchao Dai. Complementary patch for weakly supervised semantic segmentation. In *IEEE ICCV*, pages 7242–7251, 2021. 5

[26] Tianfei Zhou, Meijie Zhang, Fang Zhao, and Jianwu Li. Regional semantic contrast and aggregation for weakly supervised semantic segmentation. In *IEEE CVPR*, pages 4299–4309, 2022. 5