References

Ahn, J.; and Kwak, S. 2018. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4981–4990.

Chang, Y.-T.; Wang, Q.; Hung, W.-C.; Piramuthu, R.; Tsai, Y.-H.; and Yang, M.-H. 2020. Weakly-supervised semantic segmentation via sub-category exploration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8991–9000.

Lee, J.; Kim, E.; Lee, S.; Lee, J.; and Yoon, S. 2019. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5267–5276.

Wang, X.; You, S.; Li, X.; and Ma, H. 2018. Weakly-supervised semantic segmentation by iteratively mining common object features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1354–1362.

Zeng, Y.; Zhuge, Y.; Lu, H.; and Zhang, L. 2019. Joint learning of saliency detection and weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 7223–7233.

Method	bkg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv	mIoU
AffinityNet (Ahn and Kwak 2018)	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
Sub-Categories (Chang et al. 2020) (w/o CRF)	88.1	49.6	30.0	79.8	51.9	77.6	53.2	80.3	76.3	69.6	77.6	53.2	80.3	76.3	69.6	69.7	40.7	75.7	42.6	66.1	58.2	64.8
PuzzleCAM (w/o CRF)	88.5	75.6	43.7	88.0	58.8	85.5	43.6	90.9	80.0	79.6	85.5	43.6	90.9	80.0	79.6	477	68.5	90.7	52.7	61.6	43.2	69.5
BoundaryCAM (w/o CRF)	89.0	78.3	44.4	88.1	66.0	83.1	44.4	90.8	82.0	79.5	83.1	44.4	90.8	82.0	79.5	46.4	66.7	92.5	53.2	68.5	42.3	70.0
MCOF (Wang et al. 2018)	87.0	78.4	29.4	68.0	44.0	70.7	28.2	73.2	71.5	67.2	70.7	28.2	73.2	71.5	67.2	53.0	47.7	74.5	32.4	71.0	45.8	60.3
Zeng et al. (Zeng et al. 2019)	90.0	77.4	37.5	80.7	61.6	79.4	23.3	78.0	75.3	71.4	79.4	23.3	78.0	75.3	71.4	68.1	35.2	78.2	32.5	75.5	48.0	63.3
FickleNet (Lee et al. 2019)	89.5	76.6	32.6	74.6	51.5	73.4	47.4	78.2	74.0	68.8	73.4	47.4	78.2	74.0	68.8	73.2	47.8	79.9	37.0	57.3	64.6	64.9
Sub-Categories (Chang et al. 2020) (w/ CRF)	88.8	51.6	30.3	82.9	53.0	79.9	53.8	82.3	78.5	70.4	79.9	53.8	82.3	78.5	70.4	71.2	40.2	78.3	42.9	66.8	58.8	66.1
PuzzleCAM (w/ CRF)	88.6	76.4	44.1	88.7	59.2	86.4	43.9	91.4	80.4	79.7	86.4	43.9	91.4	80.4	79.7	47.8	68.6	91.5	52.9	61.5	43.2	69.7
BoundaryCAM (w/ CRF)	89.1	78.9	44.9	88.5	66.7	84.0	44.6	91.2	82.3	79.5	84.0	44.6	91.2	82.3	79.5	46.5	66.9	93.2	53.3	68.3	42.4	70.2

Table 1: Semantic segmentation performance on all classes of the VOC2012 training dataset for the final pseudo-labels. Bottom group contains results with CRF refinement, while the top group is without CRF.