

Supplementary Material for “Exploring Bottom-up and Top-down Cues with Attentive Learning for Webly Supervised Object Detection”

Zhonghua Wu¹ Qingyi Tao^{1,2} Guosheng Lin*¹ Jianfei Cai^{1,3}

¹Nanyang Technological University ²NVIDIA AI Technology Center ³Monash University

{zhonghua001, qtao002}@e.ntu.edu.sg {gslin}@ntu.edu.sg {jianfei.cai}@monash.edu

1. Detail results on PASCAL VOC dataset

We show the comparison results in Table A, B and C. The bottom parts of Tables show the AP results of each category for our proposed webly supervised object detection (WebSOD) and its variants under the three different novel/base class splits.

Table A. Object detection performance (AP) for the 1st split of the novel and base categories on the PASCAL VOC 2007 dataset.

Methods	Novel						Base															
	bird	bus	cow	mbike	sofa	mean	aero	bike	boat	bottle	car	cat	chair	table	dog	horse	person	plant	sheep	train	tv	mean
ZAOD [3]	17.8	42.9	20.3	43.8	37.3	32.4	40.6	30.1	15.9	6.4	40.5	31.5	11.4	27.4	15.7	24.1	8.9	12.2	17.7	32.1	31.0	23.0
WSDDN [1]	31.5	64.5	35.7	55.6	40.7	45.6	39.4	50.1	16.3	12.6	42.8	42.6	10.1	24.9	38.2	34.4	9.4	14.7	30.2	54.7	46.9	31.6
ZLDN [5]	50.1	62.7	57.8	68.2	56.1	58.9	55.4	68.5	16.8	20.8	66.8	56.5	2.1	47.5	40.1	69.7	21.6	27.2	53.4	52.5	58.2	43.8
WSOD ² [4]	61.5	73.4	71.9	71.4	60.9	67.8	68.2	70.7	42.3	28.0	69.3	52.3	32.7	42.8	57.9	73.8	25.5	29.2	61.6	56.5	70.7	52.1
Few Shot [2]	30.0	62.7	43.2	60.6	39.6	47.2	65.3	73.5	54.7	39.5	75.7	81.1	35.3	62.5	72.8	78.8	68.6	41.5	59.2	76.2	69.2	63.6
Base WebSOD	50.1	61.9	78.1	56.2	45.0	58.3	66.6	73.7	59.7	60.1	78.0	85.0	44.8	63.0	79.5	75.6	75.4	41.1	73.1	76.0	62.8	67.6
WebSOD + ACL	52.5	63.3	79.5	58.3	48.7	60.5	67.4	76.3	61.0	60.7	80.0	82.9	46.1	61.9	81.0	78.7	75.7	45.5	74.2	77.6	65.9	69.0
WebSOD + ACL + FT	56.6	60.9	73.9	54.8	44.8	58.2	78.9	83.7	67.6	67.0	83.9	86.0	54.2	70.9	81.8	84.1	77.6	48.7	78.7	81.0	74.9	74.6
WebSOD + ACL + RFR	56.8	66.2	80.3	57.3	48.2	61.8	69.9	76.2	61.0	60.6	82.9	84.6	48.7	64.4	82.5	80.8	76.5	46.0	76.6	80.8	68.8	70.7
Fully Supervised	79.1	86.1	85.8	84.8	79.8	83.1	79.4	85.7	72.0	68.4	87.7	88.4	63.0	71.0	87.8	86.9	82.3	52.1	82.1	87.1	76.9	78.1

Table B. Object detection performance (AP) for the 2nd split of the novel and base categories on the PASCAL VOC 2007 dataset.

Methods	Novel						Base															
	bike	bird	boat	horse	sofa	mean	bike	bird	boat	bus	car	cat	chair	table	dog	mbike	person	plant	sheep	train	tv	mean
ZAOD [3]	40.6	6.4	20.3	24.1	37.3	25.7	30.1	17.8	15.9	42.9	40.5	31.5	11.4	27.4	15.7	43.8	8.9	12.2	17.7	32.1	31.0	25.3
WSDDN [1]	39.4	12.6	35.7	34.4	40.7	32.6	50.1	31.5	16.3	64.5	42.8	42.6	10.1	24.9	38.2	55.6	9.4	14.7	30.2	54.7	46.9	35.5
ZLDN [5]	55.4	20.8	57.8	69.7	56.1	51.9	68.5	50.1	16.8	62.7	66.8	56.5	2.1	47.5	40.1	68.2	21.6	27.2	53.4	52.5	58.2	46.1
WSOD ² [4]	68.2	28.0	71.9	73.8	60.9	60.6	70.7	61.5	42.3	73.4	69.3	52.3	32.7	42.8	57.9	71.4	25.5	29.2	61.6	56.5	70.7	54.5
Few Shot [2]	43.2	13.9	41.5	58.1	39.2	39.2	74.1	63.8	52.0	75.5	77.6	81.8	35.6	57.9	68.2	77.6	68.0	37.9	62.4	76.9	71.3	65.4
Base WebSOD	50.0	9.3	80.5	75.9	45.3	52.2	72.2	74.3	58.8	83.0	83.2	84.2	41.9	63.1	82.2	73.5	75.4	41.8	75.9	78.5	63.3	70.1
WebSOD + ACL	52.8	10.3	80.8	74.3	46.9	53.0	74.5	74.4	62.1	81.9	85.0	87.0	45.2	64.2	81.4	74.0	75.5	42.8	76.9	78.4	64.6	71.2
WebSOD + ACL + RFR	53.5	10.6	81.2	74.4	50.2	54.0	76.4	75.3	61.6	81.7	86.0	87.5	48.2	64.9	82.9	76.5	76.1	43.6	78.0	80.8	67.5	72.5
Fully Supervised	79.4	68.4	85.8	86.9	79.8	80.6	85.7	79.1	72.0	86.1	87.7	88.4	63.0	71.0	87.8	84.8	82.3	52.1	82.1	87.1	76.9	79.1

Table C. Object detection performance (AP) for the 3rd split of the novel and base categories on the PASCAL VOC 2007 dataset.

Methods	Novel						Base															
	boat	cat	mbike	sheep	sofa	mean	aero	bike	bird	bottle	bus	car	chair	cow	table	dog	horse	person	plant	train	tv	mean
ZAOD [3]	15.9	31.5	43.8	17.7	37.3	29.2	40.6	30.1	17.8	6.4	34.4	24.2	5.7	20.3	22.3	24.9	29.1	7.8	9.4	22.6	26.0	21.4
WSDDN [1]	16.3	42.6	55.6	30.2	40.7	37.1	39.4	50.1	31.5	12.6	64.5	42.8	10.1	35.7	24.9	38.2	34.4	9.4	14.7	54.7	46.9	34.0
ZLDN [5]	16.8	56.5	68.2	53.4	56.1	50.2	55.4	68.5	50.1	20.8	62.7	66.8	2.1	57.8	47.5	40.1	69.7	21.6	27.2	52.5	58.2	46.7
WSOD ² [4]	42.3	52.3	71.4	61.6	60.9	57.7	68.2	70.7	61.5	28.0	73.4	69.3	32.7	71.9	42.8	57.9	73.8	25.5	29.2	56.5	70.7	55.5
Few Shot [2]	20.1	51.8	55.6	42.4	36.6	41.3	68.4	71.4	66.6	37.0	75.0	76.2	35.7	52.6	60.6	66.7	79.7	68.9	40.7	76.5	68.6	63.0
Base WebSOD	32.5	80.7	56.3	67.8	52.5	58.0	64.6	65.0	71.8	57.5	82.4	83.0	42.5	79.5	48.9	76.8	77.0	73.7	42.0	77.1	65.6	67.2
WebSOD + ACL	30.9	77.7	58.9	72.9	53.8	58.9	66.2	68.6	71.0	54.1	78.0	83.3	38.1	82.3	53.3	79.2	77.6	73.6	43.4	73.7	67.4	67.3
WebSOD + ACL + RFR	35.6	78.0	59.7	75.1	51.7	60.0	70.9	70.5	74.1	54.0	82.7	85.4	43.3	83.4	60.1	81.3	79.9	75.5	45.0	81.2	69.7	70.5
Fully Supervised	72.0	88.4	84.8	82.1	79.8	81.4	79.4	85.7	79.1	68.4	86.1	87.7	63.0	85.8	71.0	87.8	86.9	82.3	52.1	87.1	76.9	79.3

References

[1] Hakan Bilen and Andrea Vedaldi. Weakly supervised deep detection networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2846–2854, 2016.

*Corresponding author: G. Lin (e-mail: gslin@ntu.edu.sg)

- [2] Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-shot object detection via feature reweighting. *arXiv preprint arXiv:1812.01866*, 2018.
- [3] Qingyi Tao, Hao Yang, and Jianfei Cai. Zero-annotation object detection with web knowledge transfer. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 369–384, 2018.
- [4] Zhaoyang Zeng, Bei Liu, Jianlong Fu, Hongyang Chao, and Lei Zhang. Wsod2: Learning bottom-up and top-down objectness distillation for weakly-supervised object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8292–8300, 2019.
- [5] Xiaopeng Zhang, Jiashi Feng, Hongkai Xiong, and Qi Tian. Zigzag learning for weakly supervised object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4262–4270, 2018.