Supplementary for
DAP: Detection-Aware Pre-training with Weak Supervision

A. Hyper-parameters and more visualization

During the pseudo box generation, we adopt the multi-threshold (0.2, 0.3, 0.4, 0.5) CAM thresholding strategy and choose the post-processing NMS IoU threshold as 0.8. Here we study how these hyper-parameter values impact the pseudo boxes and the downstream detection performance under the IN-1M DAP Faster RCNN ResNet-50 setting.

CAM threshold τ for finding salient regions. Threshold τ separates out the salient regions on the class activation maps as in Eq. (2) of the main paper. A higher τ may focus on the most activated regions, producing smaller bounding box instances on a single image. On the other hand, a lower τ may keep more parts of the objects, producing larger boxes and more complete objects. This is shown more clearly in Table A1 and Figure A1. As τ increases, the average number of mined objects per image increases, and the average size of boxes shrinks. The Multi (short for multiple-threshold) strategy introduced in the main paper is able to generate more pseudo boxes.

As for the downstream task accuracy, none of the single-threshold COCO results could match that of the multi-threshold strategy (≥ 0.25 AP drop), while for VOC07+12, the results of single threshold 0.4 and 0.5 are slightly better than Multi. COCO contains objects of diverse scales, while VOC contains mostly large objects only. This suggests that the multi-threshold strategy to find multiple pseudo boxes of different sizes is necessary to boost performance on COCO.

Table A1. Effect of the CAM Threshold τ on pseudo box generation and downstream task performance (NMS IoU=0.8) under the IN-1M DAP Faster RCNN ResNet-50 setting. The upper three rows show the average number of pseudo boxes per image, the average pseudo box width and height on ImageNet-1M. The bottom three rows show the downstream detection accuracy. "Multi" refers to the multiple-threshold strategy reported in the paper, which merges the box results from the 4 different thresholds. Other columns represent using a single threshold. We notice that a larger τ yields smaller boxes and "Multi" leads to the highest COCO AP.

CAM Threshold τ	0.2	0.3	0.4	0.5	Multi
Avg boxes / image	1.019	1.046	1.096	1.161	3.211
Avg box width	339.9	296.5	250.4	199.6	253.4
Avg box height	280.3	242.1	202.0	160.7	208.0
COCO $AP_{.5:.95}$	36.89	36.82	36.90	36.97	**37.25**
VOC07 $AP_{.5}$	79.47	79.57	79.13	79.06	**79.93**
VOC07+12 $AP_{.5}$	84.35	84.41	**84.76**	84.61	84.49

Table A2. Effect of NMS IoU threshold on pseudo box generation and downstream task performance under the IN-1M DAP Faster RCNN ResNet-50 setting. We vary the IoU threshold from 0.5 to 1.0. The IoU 0.8 column is the one reported in the main paper. IoU 1.0 refers to not doing the NMS post-processing.

NMS IoU	0.5	0.6	0.7	0.8	0.9	1.0
Avg boxes / image	2.047	2.308	2.661	3.211	3.966	4.321
Avg box width	258.9	256.2	254.2	253.4	260.9	269.6
Avg box height	215.6	212.3	209.7	208.0	212.3	219.3
COCO $AP_{.5:.95}$	37.12	36.98	37.13	**37.25**	**37.26**	37.02
VOC07 $AP_{.5}$	79.50	79.41	79.49	**79.93**	79.39	79.51
VOC07+12 $AP_{.5}$	84.09	**84.54**	84.30	84.49	84.32	84.36

NMS IoU threshold for pseudo box post-processing. The pseudo boxes generated from different connected components with multiple CAM thresholds are merged with non-maximum suppression (NMS) post-processing. The NMS has an IoU threshold hyper-parameter. We vary the IoU threshold from 0.5 to 1.0 (no NMS) to study its effect on the pseudo box statistics and the downstream task performance, as shown in Table A2. The boxes from different thresholds are visualized in Figure A2. We use the multi-threshold CAM strategy. A higher IoU threshold keeps more pseudo boxes, while a lower threshold eliminates more boxes. Different IoUs do not have a significant impact on the size of the pseudo boxes.

In terms of the downstream task accuracy, an IoU threshold of 0.8 or 0.9 achieves the best AP on COCO. On VOC07, IoU 0.8 achieves the best result, and on VOC07+12, IoU ranging from 0.6 to 1.0 gets similar $AP_{.5}$. Since IoU 0.8 delivers overall good performance while keeping only 3.2 pseudo boxes per image, we choose this value in the main experiments.

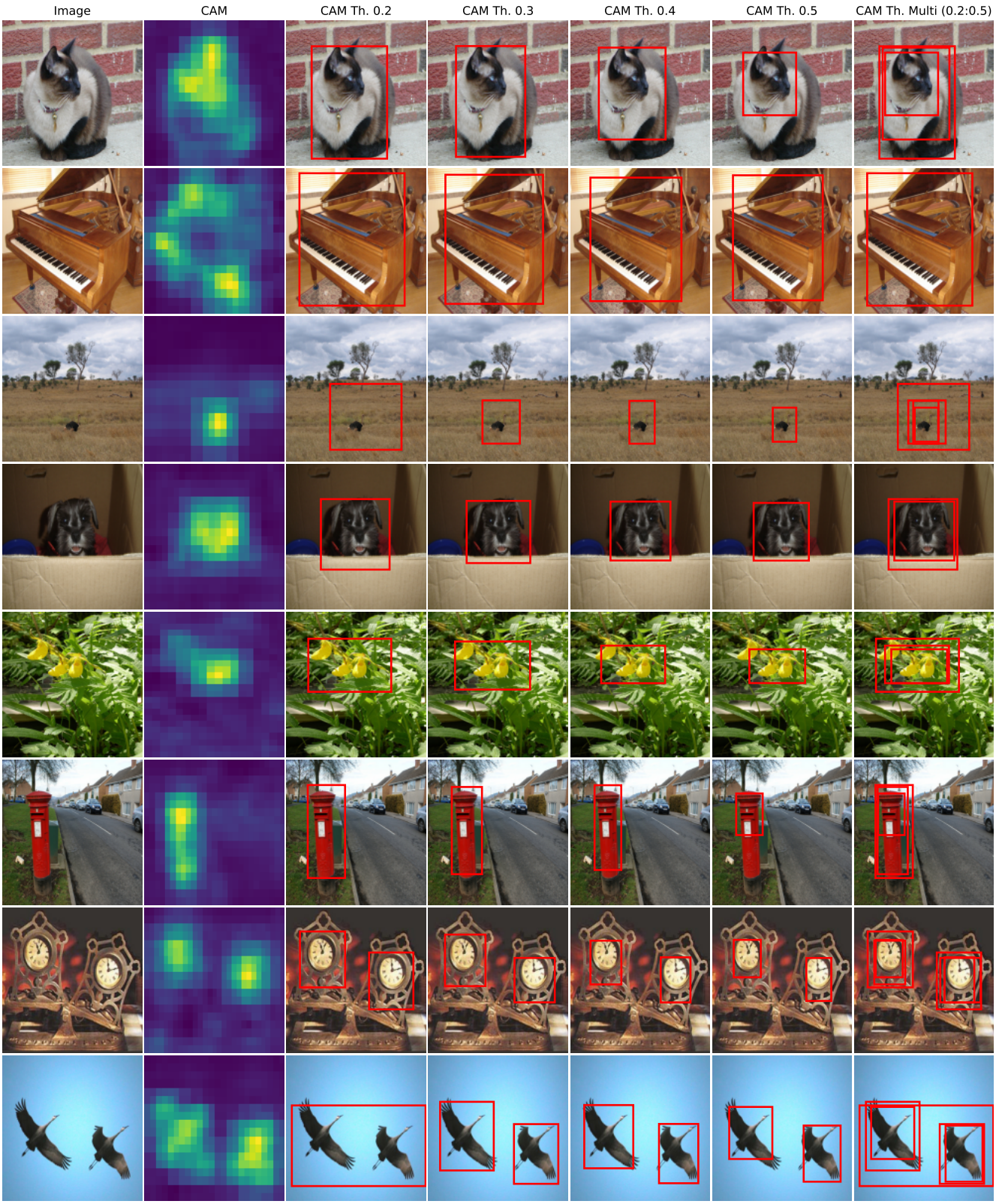

Figure A1. Visualization of pseudo boxes generated on ImageNet-1M with different CAM thresholds. In Row 1 and 6, lower threshold leads to more accurate boxes, while in Row 3, higher threshold can produce a tighter box. In the last row, higher threshold is required to discern the two birds. The final strategy in the main paper is "Multi", which combines boxes from different thresholds to improve recall.

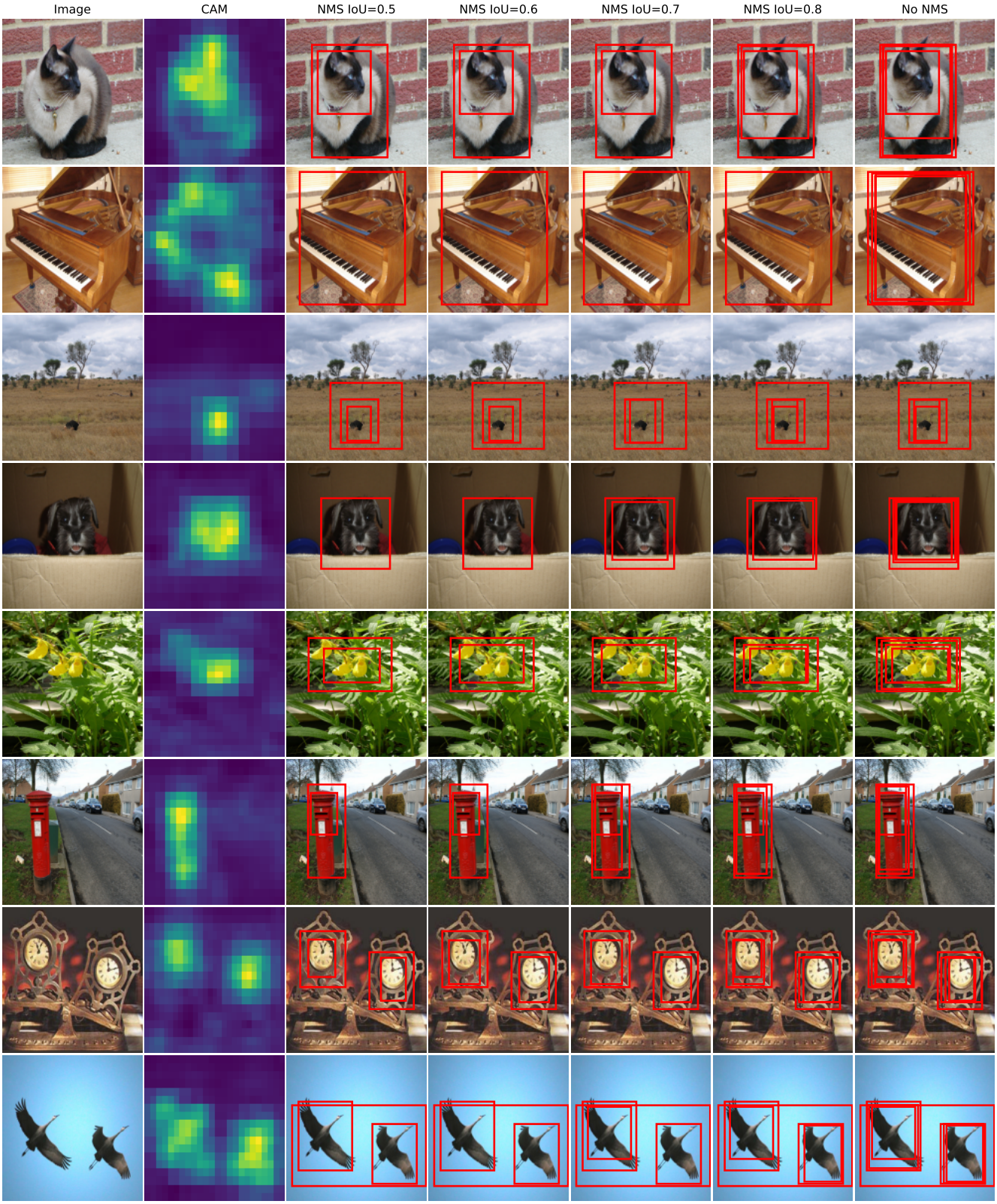

Figure A2. Visualization of pseudo boxes generated on ImageNet-1M with different NMS post-processing IoU thresholds. A smaller IoU threshold leads to fewer boxes. In the main paper, we set IoU as 0.8 to keep more boxes since it yields good overall downstream detection performance on COCO and VOC.