IMPROVING OPEN-WORLD CLASS-AGNOSTIC OBJECT DETECTORS VIA FEATURE DISTILLATION WITH STUDENT-AWARE ADAPTATION

Supplementary Material

A. IMPLEMENTATION DETAILS

Our method. We implemented our method using the MMDetection framework [1]. In the adaptation phase, projectors and auxiliary modules were trained over 16 epochs using the SGD optimizer with an initial learning rate of 0.005, a weight decay of 0.0001, and a batch size of 8. The learning rate was reduced by a factor of 10 at 12 and 14 epochs, respectively. Following GOOD [2], we employed strong data augmentation, which involved AutoAugment [3] including random resizing, flip, and cropping. In the transfer phase, a student model was optimized using both detection loss and distillation loss with the aforementioned learning settings. Following VkD [4], we implemented distillation loss L_{kd} as follows:

$$L_{kd}(g^{T\rightarrow S}, g^{S}) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{d_i}\mathrm{SmoothL1}(g_i^{T\rightarrow S}, \phi_i(g_i^{S})), \quad (1)$$

where $g_i^{T\rightarrow S}$ and g_i^{S} are the i-th feature maps of the auxiliary neck and the student's neck, respectively; M is the number of feature maps; d_i is the number of elements in the i-th feature maps; ϕ_i denotes an orthogonal projection layer [4]; and $\mathrm{SmoothL1}$ represents the smooth L_1 loss.

Implementation of ATSS [5] and FCOS [6] based on OLN [7]. Following OLN [7], which proposes objects using IoU and centerness scores, we implemented OLN-based ATSS by replacing the classification head in the original ATSS with an IoU head, which predicts IoU scores of predefined bbox anchors. The OLN-based ATSS proposes objects using the outputs of the IoU head and the centerness head, the latter being included in the original ATSS. For FCOS, which does not involve predefined bbox anchors, we introduced IoU head instead of the classification head and trained it to predict IoU scores for the predicted bboxes. Similar to the OLN-based ATSS, we used the outputs of the IoU and centerness heads for proposing in the OLN-based FCOS.

B. ADDITIONAL EXPERIMENTAL RESULTS

Detection performance on learned known object categories. Table 1 presents the AR@100 scores for COCO category objects using two existing methods, both with and without our approach, across three detectors on our modified benchmark. As shown in Table 1, introducing our method enhanced the AR@100 scores for COCO category objects by at least 0.1 points. This enhancement indicates that our method achieves performance improvement on unknown objects without degrading detection performance on seen objects.

Table 1. Impact of integrating the proposed method into two existing methods on known object detection performance. AR represents Average Recall over multiple IoU thresholds (0.5:0.95) with the proposal number of 100. The superscript 's', 'm' and 'l' denote the evaluation for small, medium large size of objects, respectively. The results were evaluated on COCO category objects in the LVIS validation set under our modified benchmark.

(a) Faster R-CNN

Method	AR	$\mathrm{AR^s}$	$\mathrm{AR^m}$	$\mathrm{AR^l}$
OLN [7]	38.3	23.5	54.7	67.2
w/ Ours	**38.7 (+0.4)**	24.0	55.1	67.5
GOOD [2]	34.6	20.2	51.7	60.8
w/ Ours	**34.7 (+0.1)**	20.7	51.3	61.5

(b) ATSS

Method	AR	$\mathrm{AR^s}$	$\mathrm{AR^m}$	$\mathrm{AR^l}$
OLN [7]	34.8	17.7	55.0	66.2
w/ Ours	**35.7 (+0.9)**	18.3	56.0	68.2
GOOD [2]	32.3	15.2	52.5	63.7
w/ Ours	**32.9 (+0.6)**	15.9	52.6	65.0

(c) FCOS

Method	AR	$\mathrm{AR^s}$	$\mathrm{AR^m}$	$\mathrm{AR^l}$
OLN [7]	33.9	19.2	50.7	61.6
w/ Ours	**35.4 (+1.5)**	20.7	52.2	63.4
GOOD [2]	30.7	16.5	47.6	56.9
w/ Ours	**31.0 (+0.3)**	16.9	47.8	56.9

C. REFERENCES

[1] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, Chen C. L., and D. Lin, "MMDetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.

[2] H. Huang, A. Geiger, and D. Zhang, "GOOD: exploring geometric cues for detecting objects in an open world," in *International Conference on Learning Representations*, 2023.

[3] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 113–123.

[4] R. Miles, I. Elezi, and J. Deng, "VkD: Improving knowledge

distillation using orthogonal projections," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15720–15730.

[5] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9759–9768.

[6] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: A simple and strong anchor-free object detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 1922–1933, 2020.

[7] D. Kim, T.-Y. Lin, A. Angelova, I. S. Kweon, and W. Kuo, "Learning open-world object proposals without learning to classify," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5453–5460, 2022.