Supplementary Material for Few-Shot Object Detection by Knowledge Distillation Using Bag-of-Visual-Words Representations

Wenjie Pei[2,†], Shuang Wu[2,†], Dianwen Mei[2], Fanglin Chen[2], Jiandong Tian[3], and Guangming Lu[1,2,*]

[1] Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies
[2] Harbin Institute of Technology, Shenzhen, China
[3] Shenyang Institute of Automation, Chinese Academy of Sciences
{wenjiecoder, wushuang9811}@outlook.com, {178mdw, linwers}@gmail.com, luguangm@hit.edu.cn, tianjd@sia.cn

1 Results over Multiple Runs

To eliminate the effect of sample variance introduced by the random selection of few-shot training samples, we fine-tune our model over 10 random selections of few-shot training samples independently for each experimental settings (including different novel splits and shot numbers), and obtain the average results on PASCAL VOC dataset. As shown in Table 1, our method improves the performance of TFA++ [4] under all settings.

Table 1. Comparison with existing few-shot object detection methods using nAP50 as evaluation metric on three PASCAL VOC Novel Split sets. Results are averaged over 10 random runs. † indicates that model is evaluated using the released code.

Method / Shots	Novel Split 1					Novel Split 2					Novel Split 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
FRCN+ft [9]	9.9	15.6	21.6	28.0	35.6	9.4	13.8	17.4	21.9	29.8	8.1	13.9	19.0	23.9	31.0
TFA w/fc [5]	22.9	34.5	40.4	46.7	52.0	16.9	26.4	30.5	34.6	39.7	15.7	27.2	34.7	40.8	44.6
TFA w/cos [5]	25.3	36.4	42.1	47.9	52.8	18.3	27.5	30.9	34.1	39.5	17.9	27.2	34.3	40.8	45.6
FsDetView [7]	24.2	35.3	42.2	49.1	57.4	21.6	24.6	31.9	37.0	45.7	21.2	30.0	37.2	43.8	49.6
TIP [3]	27.7	36.5	43.3	50.2	59.6	22.7	30.1	33.8	40.9	46.9	21.7	30.6	38.1	44.5	50.9
DCNet [2]	33.9	37.4	43.7	51.1	59.6	23.2	24.8	30.6	36.7	46.6	**32.3**	**34.9**	39.7	42.6	50.7
FSCE [4]	32.9	44.0	46.8	52.9	59.7	**23.7**	**30.6**	38.4	43.0	48.5	22.6	33.4	39.5	47.3	54.0
TFA++† [4]	33.1	41.6	46.3	53.5	57.8	21.3	28.9	37.6	41.6	47.2	21.5	32.9	38.9	48.1	53.8
Ours (KD-TFA++)	**35.4**	**46.2**	**48.1**	**56.5**	**60.7**	22.8	30.2	**39.2**	**44.0**	**48.9**	25.2	33.9	**41.3**	**50.7**	**55.9**

2 Results on Base Classes

Table 2 shows the performance for base and novel classes on Novel Split 1 of PASCAL VOC dataset. Although AP for base classes (bAP50) is not our pri-

[†] Equal contribution.
[*] Corresponding author.

mary concern, our method makes competitive results. It can be observed that our method improves not only the performance of novel classes, but also the performance of base classes. These results demonstrate that our method can maintain the performance on previous knowledge without forgetting.

Table 2. Few-shot object detection results for base and novel classes on Novel Split 1 of PASCAL VOC dataset. † indicates that model is evaluated using the released code.

Shots	Method	bAP50	nAP50
3	FRCN+ft-full [9]	63.6	32.8
	Meta R-CNN [9]	64.8	35.0
	Baseline-FPN [6]	66.2	41.1
	MPSR [6]	67.8	51.4
	TFA w/cos [5]	**79.1**	44.7
	FSCE [4]	74.1	51.4
	TFA++† [4]	75.4	47.3
	Ours (KD-TFA++)	76.4	**52.5**
5	Baseline-FPN [6]	67.9	49.6
	MPSR [6]	68.4	55.2
	TFA w/cos [5]	77.0	55.6
	FSCE [4]	76.6	61.9
	TFA++† [4]	77.7	57.2
	Ours (KD-TFA++)	**79.0**	**62.1**
10	FRCN+ft-full [9]	61.3	45.6
	Meta R-CNN [9]	67.9	51.5
	Baseline-FPN [6]	70.0	56.9
	MPSR [6]	71.8	61.8
	TFA w/cos [5]	78.4	56.0
	TFA++† [4]	77.5	60.8
	Ours (KD-TFA++)	**78.6**	**64.2**

3 Comparison with More Baseline Methods

In Table 3 we integrate our method into two more baselines: TFA w/ fc [5] and Retentive R-CNN [1]. It can be observed that our method consistently boosts the performance, which shows the effectiveness of our method.

Table 3. Performance of integrating our method into more baselines in terms of nAP50 on PASCAL VOC Novel split 1.

Methods / Shots	nAP50				
	1	2	3	5	10
TFA w/ fc	36.8	29.1	43.6	55.7	57.0
Ours (KD-TFA w/ fc)	**41.6**	**40.5**	**48.3**	**56.2**	**59.9**
Retentive R-CNN	42.4	45.8	45.9	53.7	56.1
Ours (KD-Retentive R-CNN)	**48.7**	**48.4**	**51.7**	**58.7**	**60.3**

4 More Ablation Studies

Effect of the number of visual words. Table 4 shows the effect of the number of visual words. It can be observed that the performance first improves rapidly with the increase of visual words and then starts to degrade after 256. This is mainly resulted from the limited size of data corpus for learning the visual words.

Table 4. Effect of the number of visual words.

Number	nAP50		
	3	5	10
64	50.0	57.2	61.3
128	51.6	61.0	62.3
256	**52.5**	**62.1**	**64.2**
512	51.6	59.2	62.3

Knowledge distillation vs initialization of the object detector vs multi-task learning. We further explore other methods to learn a generalizable detector. As shown in Row 2 of Table 5, using the backbone pre-trained by PPC [8] to initialize the detector yields little improvement over the baseline. Row 3 shows that the performance degrades when using PPC for multi-task learning, presumably because PPC aims to distinguish between pixels, which is not entirely consistent with the objective of object detection.

Table 5. Performance of different ways of using PPC on VOC Novel Split 1.

Methods / Shots	nAP50		
	3	5	10
Baseline	47.2	57.2	60.8
Initialization	46.4	57.3	61.2
Multi-task Learning	45.9	55.1	60.3
Knowledge Distillation (Ours)	**52.5**	**62.1**	**64.2**

5 More Qualitative Detection Results

We provide more qualitative detection results under 10-shot setting of PASCAL VOC Novel Split1. As shown in Figure 1, our method reduces the appearance of each type of errors such as missing detections and misclassifying novel objects.

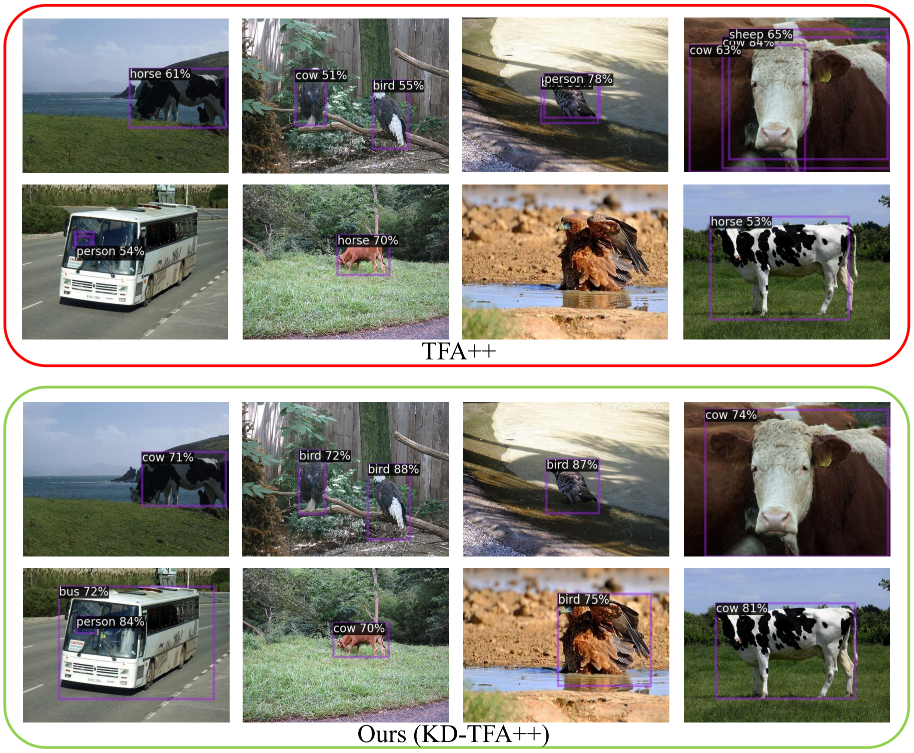

Fig. 1. Detection results of TFA++ [4] and our method under PASCAL VOC Novel Split1 10-shot setting.

References

1. Fan, Z., Ma, Y., Li, Z., Sun, J.: Generalized few-shot object detection without forgetting. In: CVPR (2021)
2. Hu, H., Bai, S., Li, A., Cui, J., Wang, L.: Dense relation distillation with context-aware aggregation for few-shot object detection. In: CVPR (2021)
3. Li, A., Li, Z.: Transformation invariant few-shot object detection. In: CVPR (2021)
4. Sun, B., Li, B., Cai, S., Yuan, Y., Zhang, C.: Fsce: Few-shot object detection via contrastive proposal encoding. In: CVPR (2021)
5. Wang, X., Huang, T., Gonzalez, J., Darrell, T., Yu, F.: Frustratingly simple few-shot object detection. In: ICML (2020)
6. Wu, J., Liu, S., Huang, D., Wang, Y.: Multi-scale positive sample refinement for few-shot object detection. In: ECCV (2020)
7. Xiao, Y., Marlet, R.: Few-shot object detection and viewpoint estimation for objects in the wild. In: ECCV (2020)
8. Xie, Z., Lin, Y., Zhang, Z., Cao, Y., Lin, S., Hu, H.: Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning. In: CVPR (2021)
9. Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., Lin, L.: Meta r-cnn: Towards general solver for instance-level low-shot learning. In: ICCV (2019)