FSCE: Few-Shot Object Detection via Contrastive Proposal Encoding (Supplementary materials)

Bo Sun[1], Banghuai Li[2,*], Shengcai Cai[2], Ye Yuan[2], and Chi Zhang[2]
[1]University of Southern California
[2]MEGVII Technology
bos@usc.edu, {libanghuai,caishengcai,yuanye,zhangchi}@megvii.com

1. Average results over random runs

Few-shot object detection performance is inherently unstable and heavily depends on the randomly sampled training shots. Hence, [1] suggests evaluating few-shot detection performance over a series of random runs to obtain statistically reliable comparisons. In this supplementary material, we provide full benchmark results over n random runs for PASCAL VOC and COCO. We report averaged AP, AP50 and AP75 for novel classes (nAP, nAP50, nAP75) from all three splits from PASCAL VOC. For COCO, we also report the averaged AP for small ($nAPs$), medium ($nAPm$) and large ($nAPl$) novel objects. Following the practices in TFA [1], we calculate and report the 95% confidence interval (CI) for each metric we reported. The 95% CI is given by

$$CI_{0.95} = Z_{0.95} \cdot \frac{\sigma}{\sqrt{n}} \tag{1}$$

where $Z_{0.95} = 1.96$ is the Z-score for 95% CI, σ is the standard deviation, and n is the number of random runs. In our experiments, we perform n=10 random runs for both PASCAL VOC and COCO datasets.

2. Results for PASCAL VOC and COCO

We present the complete few-shot object detection benchmark results of our proposed FSCE over random runs. The main baseline we are comparing with is the baseline two-stage fine-tuning approach (TFA [1]). As shown in Table 1 for PASCAL VOC and Table 2 for COCO, FSCE significantly outperforms baseline TFA and other methods in almost all shots from all data splits. With up to +9% nAP50 on PASCAL VOC, and +3.2% nAP on COCO. Results averaged over repeated runs with randomly selected training shots, which are statistically stable and reliable, demonstrate the *state-of-the-art* few-shot object detection performance of our proposed method.

Code is available at https://github.com/MegviiDetection/FSCE.

*Corresponding author: libanghuai@megvii.com

References

[1] Xin Wang, Thomas E. Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. Frustratingly simple few-shot object detection. In *International Conference on Machine Learning (ICML)*, July 2020. 1, 2

[2] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell. Few-shot object detection via feature reweighting. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8419–8428, 2019. 2

[3] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Meta-Learning to Detect Rare Objects. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9924–9933, Seoul, Korea (South), October 2019. IEEE. 2

[4] X. Yan, Z. Chen, A. Xu, X. Wang, X. Liang, and L. Lin. Meta r-cnn: Towards general solver for instance-level low-shot learning. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9576–9585, 2019. 2

# Shots	Method	Novel Split 1			Novel Split 2			Novel Split 3		
		nAP	nAP50	nAP75	nAP	nAP50	nAP75	nAP	nAP50	nAP75
1	FRCN-baseline	6.0±0.7	9.9±1.2	6.3±0.8	5.0±0.6	9.4±1.2	4.5±0.7	4.5±0.7	8.1±1.3	4.2±0.7
	FSRW [2]	8.0±1.0	14.2±1.7	7.9±1.1	6.3±0.9	12.3±1.9	5.5±0.7	6.7±1.0	12.5±1.6	6.4±1.0
	TFA w/ fc [1]	12.2±1.6	22.9±2.5	11.6±1.9	8.1±1.2	16.9±2.3	6.6±1.1	7.8±1.1	15.7±2.1	6.5±1.0
	TFA w/ cos [1]	14.2±1.4	25.3±2.2	14.2±1.8	9.0±1.2	18.3±2.4	7.8±1.2	9.6±1.1	17.9±2.0	9.1±1.2
	FSCE (Ours)	**17.7±2.1**	**32.9±3.2**	**16.8±2.5**	**10.2±2.0**	**21.5±3.5**	**8.0±2.2**	**11.1±2.0**	**22.1±3.9**	**9.4±2.1**
2	FRCN-baseline	9.9±0.9	15.6±1.4	10.3±1.0	7.7±0.8	13.8±1.4	7.4±0.8	8.0±0.8	13.9±1.4	7.9±0.9
	FSRW [2]	13.2±1.0	23.6±1.7	12.7±1.1	9.9±0.7	19.6±1.3	8.8±0.6	11.3±0.7	21.3±1.0	10.6±0.8
	TFA w/ fc [1]	18.9±1.5	34.5±2.6	18.4±1.9	13.1±1.0	26.4±1.9	11.3±1.1	14.2±1.2	27.2±2.0	12.6±1.3
	TFA w/ cos [1]	21.7±1.0	36.4±1.6	22.8±1.3	14.1±0.9	27.5±1.6	12.7±1.0	15.1±1.3	27.2±2.1	**14.4±1.5**
	FSCE (Ours)	**24.2±1.8**	**44.0±2.6**	**23.6±2.6**	**15.1±2.1**	**30.6±3.5**	**12.8±2.4**	**17.0±1.8**	**33.4±3.4**	14.3±1.9
3	FRCN-baseline	13.7±1.0	21.6±1.6	14.8±1.1	9.8±0.9	17.4±1.6	9.4±1.0	11.1±0.9	19.0±1.5	11.2±1.0
	FSRW [2]	16.8±0.9	29.8±1.6	16.5±1.0	12.5±0.7	25.1±1.4	10.4±0.7	14.2±0.7	26.8±1.4	13.1±0.7
	TFA w/ fc [1]	22.6±1.2	40.4±1.7	22.4±1.7	15.2±0.8	30.5±1.5	13.1±0.8	18.1±1.0	34.7±1.6	16.2±1.3
	TFA w/ cos [1]	25.4±0.9	42.1±1.5	**27.0±1.2**	16.0±0.8	30.9±1.6	14.4±0.9	18.9±1.0	34.3±1.7	**18.1±1.4**
	FSCE (Ours)	**25.4±1.4**	**46.9±2.5**	24.3±1.9	**18.1±1.6**	**38.4±2.4**	**14.5±1.9**	**19.3±1.4**	**39.5±3.1**	16.0±1.3
5	FRCN-baseline	17.9±1.1	28.0±1.7	19.2±1.3	12.4±0.9	21.9±1.5	12.1±0.9	14.0±0.9	23.9±1.7	13.7±0.9
	FSRW [2]	20.6±0.8	36.5±1.4	20.0±0.9	15.7±0.8	31.4±1.5	13.3±0.9	18.0±0.7	33.8±1.4	16.5±0.8
	TFA w/ fc [1]	25.9±1.0	46.7±1.4	25.3±1.2	17.5±0.7	34.6±1.1	15.5±0.9	21.4±0.9	40.8±1.3	19.4±1.0
	TFA w/ cos [1]	28.9±0.9	47.9±1.2	30.6±1.0	17.8±0.8	34.1±1.4	16.2±1.0	22.8±0.9	40.8±1.4	22.1±1.1
	FSCE (Ours)	**30.7±1.2**	**52.9±1.9**	**31.3±1.4**	**22.0±0.7**	**43.0±1.4**	**19.8±0.9**	**25.5±1.5**	**47.3±2.5**	**23.7±1.7**
10	FRCN-baseline	22.7±0.9	35.6±1.5	24.4±1.0	17.0±0.8	29.8±1.4	16.7±0.9	18.4±0.8	31.0±1.2	18.7±1.0
	TFA w/ fc [1]	29.3±0.7	52.0±1.1	29.0±0.9	20.2±0.5	39.7±0.9	18.0±0.7	23.3±0.8	44.6±1.1	21.0±1.2
	TFA w/ cos [1]	32.0±0.6	52.8±1.0	33.7±0.7	20.8±0.6	39.5±1.1	19.2±0.6	25.4±0.7	45.6±1.1	24.7±1.1
	FSCE (Ours)	**34.0±1.0**	**58.7±1.7**	**35.0±1.3**	**25.2±0.9**	**48.5±1.7**	**23.1±1.0**	**29.2±1.4**	**54.0±1.9**	**27.5±1.5**

Table 1. The averaged few-shot object detection performance on PASCAL VOC. For each metric, we report the mean and 95% confidence interval over 10 random runs.

# Shots	Method	Novel Average Precision					
		nAP	nAP50	nAP75	nAPs	nAPm	nAPl
10	MetaDet [3]	7.1±n/a	14.6±n/a	6.1±n/a	1.0±n/a	4.1±n/a	12.2±n/a
	Meta R-CNN [4]	8.7±n/a	19.1±n/a	6.6±n/a	2.3±n/a	7.7±n/a	14.0±n/a
	TFA w/ fc [1]	9.1±0.4	17.3±0.6	8.5±0.4	-	-	-
	TFA w/ cos [1]	9.1±0.4	17.1±0.7	8.8±0.5	-	-	-
	FSCE (Ours)	**11.1±0.2**	**23.0±0.4**	**9.8±0.3**	**3.5±0.4**	**10.9±0.3**	**16.6±0.5**
30	MetaDet [3]	11.3±n/a	21.7±n/a	8.1±n/a	1.1±n/a	6.2±n/a	17.3±n/a
	Meta R-CNN [4]	12.4±n/a	25.3±n/a	10.8±n/a	2.8±n/a	11.6±n/a	19.0±n/a
	TFA w/ fc [1]	12.0±0.4	22.2±0.6	11.8±0.4	-	-	-
	TFA w/ cos [1]	12.1±0.4	22.0±0.7	12.0±0.5	-	-	-
	FSCE (Ours)	**15.3±0.3**	**29.0±0.5**	**14.2±0.3**	**5.2±0.5**	**15.4±0.4**	**22.6±0.8**

Table 2. The averaged few-shot object detection performance on COCO. For each metric, we report the mean and 95% confidence interval over 10 random runs. The "n/a" here indicates the confidence intervals are not reported by the authors.