Frustratingly Simple Few-Shot Object Detection
Supplementary Material

Xin Wang [*1] **Thomas E. Huang** [*2] **Trevor Darrell** [1] **Joseph E. Gonzalez** [1] **Fisher Yu** [1]

A. Generalized Object Detection Benchmarks

We present the full benchmark results of PASCAL VOC (Table 1) and COCO (Table 2) on the revised benchmark used in this work. We report the average AP, AP50 and AP75 for all the classes, base classes only, and novel classes only in the tables. For each evaluation metric, we report the average value of n repeated runs with different groups of randomly sampled training shots (30 for PASCAL VOC and 10 for COCO) as well as the 95% confidence interval estimate of the mean values. The 95% confidence interval is calculated by

$$95\%\ CI = 1.96 \cdot \frac{s}{\sqrt{n}}, \tag{1}$$

where 1.96 is the Z-value, s is the standard deviation, and n is the number of repeated runs.

We compare two of our methods, one using a FC-based classifier (`TFA w/fc`) and one using a cosine similarity based classifier (`TFA w/cos`). We also compare against a fine-tuning baseline `FRCN+ft-full` and against FSRW (Kang et al., 2019) using their released code on PASCAL VOC shown in Table 1.

As shown in Table 1, `TFA w/cos` is able to significantly outperform `TFA w/fc` in overall AP across most splits and shots. We observe that using a cosine similarity based classifier can achieve much higher accuracy on base classes, especially in higher shots. On split 1 and 3, `TFA w/cos` is able to outperform `TFA w/fc` by over 3 points on bAP on 5 and 10 shots. Across all shots in split 1, `TFA w/cos` consistently outperforms `TFA w/fc` on nAP75 by over 2 points in the novel classes.

Moreover, the AP of our models is usually over 10 points higher than that of `FRCN+ft-full` and FSRW on all settings. Note that FSRW uses YOLOv2 as the base object detector, while we are using Faster R-CNN. Wang et al. (2019) shows that there are only about 2 points of difference when using a one or two-stage detector. Therefore, our improvements should still be significant despite the difference in the base detector.

We evaluate on COCO over six different number of shots $K = 1, 2, 3, 5, 10, 30$ shown in Table 2. Although the differences are less significant than on PASCAL VOC, similar observations can be made about accuracy on base classes and novel classes.

B. Performance over Multiple Runs

In our revised benchmark, we adopt n repeated runs with different randomly sampled training shots to increase the reliability of the benchmark. In our experiments, we adopt $n = 30$ for PASCAL VOC and $n = 10$ for COCO.

In this section, we provide plots of cumulative means with 95% confidence intervals of the repeated runs to show that the selected value of n is sufficient to provide statistically stable results.

We plot the model performance measured by AP, AP50 and AP75 of up to 40 random groups of training shots across all three splits in Figure 1. For COCO, we plot up to 10 random groups of training shots in Figure 2.

As we can observe from both Figure 1 and Figure 2, after around 30 runs on PASCAL VOC and 8 runs on COCO, the means and variances stabilize and our selected values of n are sufficient to obtain stable estimates of the model performances and reliable comparisons across different methods.

We also observe that the average value across multiple runs is consistently lower than that on the first run, especially in the one-shot case. For example, the average AP50 across 40 runs is around 15 points lower than the AP50 on the first run in the 1-shot case on split 1 on PASCAL VOC. This indicates that the accuracies on the first run, adopted by the previous work (Kang et al., 2019; Yan et al., 2019; Wang et al., 2019), often overestimate the actual performance and thus lead to unreliable comparison between different approaches.

[*]Equal contribution [1]EECS, UC Berkeley [2]EECS, University of Michigan. Correspondence to: Xin Wang <xinw@berkeley.edu>, Thomas E. Huang <thomaseh@umich.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

Table 1. Generalized object detection benchmarks on PASCAL VOC. For each metric, we report the average and 95% confidence interval computed over 30 random samples.

Split	# shots	Method	Overall			Base class			Novel class		
			AP	AP50	AP75	bAP	bAP50	bAP75	nAP	nAP50	nAP75
Split 1	1	FSRW (Kang et al., 2019)	27.6 ± 0.5	50.8 ± 0.9	26.5 ± 0.6	34.1 ± 0.5	62.9 ± 0.9	32.6 ± 0.5	8.0 ± 1.0	14.2 ± 1.7	7.9 ± 1.1
		FRCN+ft-full	30.2±0.6	49.4±0.7	32.2±0.9	38.2±0.8	62.6±1.0	40.8±1.1	6.0±0.7	9.9±1.2	6.3±0.8
		TFA w/fc	39.6±0.5	63.5±0.7	43.2±0.7	48.7±0.7	77.1±0.7	53.7±1.0	12.2±1.6	22.9±2.5	11.6±1.9
		TFA w/cos	40.6±0.5	64.5±0.6	44.7±0.6	49.4±0.4	77.6±0.2	54.8±0.5	14.2±1.4	25.3±2.2	14.2±1.8
	2	FSRW (Kang et al., 2019)	28.7±0.4	52.2±0.6	27.7±0.5	33.9±0.4	61.8±0.5	32.7±0.5	13.2±1.0	23.6±1.7	12.7±1.1
		FRCN+ft-full	30.5±0.6	49.4±0.8	32.6±0.7	37.3±0.7	60.7±1.0	40.1±0.9	9.9±0.9	15.6±1.4	10.3±1.0
		TFA w/fc	40.5±0.5	65.5±0.7	43.8±0.7	47.8±0.7	75.8±0.7	52.2±1.0	18.9±1.5	34.5±2.4	18.4±1.9
		TFA w/cos	42.6±0.3	67.1±0.4	47.0±0.4	49.6±0.3	77.3±0.2	55.0±0.4	21.7±1.0	36.4±1.6	22.8±1.3
	3	FSRW (Kang et al., 2019)	29.5±0.3	53.3±0.6	28.6±0.4	33.8±0.3	61.2±0.6	32.7±0.4	16.8±0.9	29.8±1.6	16.5±1.0
		FRCN+ft-full	31.8±0.5	51.4±0.8	34.2±0.6	37.9±0.5	61.3±0.7	40.7±0.6	13.7±1.0	21.6±1.6	14.8±1.1
		TFA w/fc	41.8±0.9	67.1±0.9	45.4±1.2	48.2±0.9	76.0±0.9	53.1±1.2	22.6±1.2	40.4±1.7	22.4±1.7
		TFA w/cos	43.7±0.3	68.5±0.4	48.3±0.4	49.8±0.3	77.3±0.2	55.4±0.4	25.4±0.9	42.1±1.5	27.0±1.2
	5	FSRW (Kang et al., 2019)	30.4±0.3	54.6±0.5	29.6±0.4	33.7±0.3	60.7±0.4	32.8±0.4	20.6±0.8	36.5±1.4	20.0±0.9
		FRCN+ft-full	32.7±0.5	52.5±0.8	35.0±0.6	37.6±0.4	60.6±0.6	40.3±0.5	17.9±1.1	28.0±1.7	19.2±1.3
		TFA w/fc	41.9±0.6	68.0±0.7	45.0±0.8	47.2±0.6	75.1±0.6	51.5±0.8	25.9±1.0	46.7±1.4	25.3±1.2
		TFA w/cos	44.8±0.3	70.1±0.4	49.4±0.4	50.1±0.2	77.4±0.3	55.6±0.3	28.9±0.8	47.9±1.2	30.6±1.0
	10	FRCN+ft-full	33.3±0.4	53.8±0.6	35.5±0.4	36.8±0.4	59.8±0.6	39.2±0.4	22.7±0.9	35.6±1.5	24.4±1.0
		TFA w/fc	42.8±0.3	69.5±0.4	46.0±0.4	47.3±0.3	75.4±0.3	51.6±0.4	29.3±0.7	52.0±1.1	29.0±0.9
		TFA w/cos	45.8±0.2	71.3±0.3	50.4±0.3	50.4±0.2	77.5±0.2	55.9±0.3	32.0±0.6	52.8±1.0	33.7±0.7
Split 2	1	FSRW (Kang et al., 2019)	28.4±0.5	51.7±0.9	27.3±0.6	35.7±0.5	64.8±0.9	34.6±0.7	6.3±0.9	12.3±1.9	5.5±0.7
		FRCN+ft-full	30.3±0.5	49.7±0.5	32.3±0.7	38.8±0.6	63.2±0.7	41.6±0.9	5.0±0.6	9.4±1.2	4.5±0.7
		TFA w/fc	36.2±0.8	59.6±0.9	38.7±1.0	45.6±0.9	73.8±0.9	49.4±1.2	8.1±1.2	16.9±2.3	6.6±1.1
		TFA w/cos	36.7±0.6	59.9±0.8	39.3±0.8	45.9±0.7	73.8±0.8	49.8±1.1	9.0±1.2	18.3±2.4	7.8±1.2
	2	FSRW (Kang et al., 2019)	29.4±0.3	53.1±0.6	28.5±0.4	35.8±0.4	64.2±0.6	35.1±0.5	9.9±0.7	19.6±1.3	8.8±0.6
		FRCN+ft-full	30.7±0.5	49.7±0.7	32.9±0.6	38.4±0.5	61.6±0.7	41.4±0.7	7.7±0.8	13.8±1.4	7.4±0.8
		TFA w/fc	38.5±0.5	62.8±0.6	41.2±0.6	46.9±0.5	74.9±0.5	51.2±0.7	13.1±1.0	26.4±1.9	11.3±1.1
		TFA w/cos	39.0±0.4	63.0±0.5	42.1±0.6	47.3±0.4	74.9±0.4	51.9±0.7	14.1±0.9	27.5±1.6	12.7±1.0
	3	FSRW (Kang et al., 2019)	29.9±0.3	53.9±0.4	29.0±0.4	35.7±0.3	63.5±0.4	35.1±0.4	12.5±0.7	25.1±1.4	10.4±0.7
		FRCN+ft-full	31.1±0.3	50.1±0.5	33.2±0.5	38.1±0.4	61.0±0.6	41.2±0.5	9.8±0.9	17.4±1.6	9.4±1.0
		TFA w/fc	39.4±0.4	64.2±0.5	42.0±0.5	47.5±0.4	75.4±0.5	51.7±0.6	15.2±0.8	30.5±1.5	13.1±0.8
		TFA w/cos	40.1±0.3	64.5±0.5	43.3±0.4	48.1±0.3	75.6±0.4	52.9±0.5	16.0±0.8	30.9±1.6	14.4±0.9
	5	FSRW (Kang et al., 2019)	30.4±0.4	54.6±0.5	29.5±0.5	35.3±0.3	62.4±0.4	34.9±0.5	15.7±0.8	31.4±1.5	13.3±0.9
		FRCN+ft-full	31.5±0.3	50.8±0.7	33.6±0.4	37.9±0.4	60.4±0.6	40.8±0.5	12.4±0.9	21.9±1.5	12.1±0.9
		TFA w/fc	40.0±0.4	65.1±0.5	42.6±0.5	47.5±0.4	75.3±0.5	51.6±0.5	17.5±0.7	34.6±1.1	15.5±0.9
		TFA w/cos	40.9±0.4	65.7±0.5	44.1±0.5	48.6±0.4	76.2±0.4	53.3±0.5	17.8±0.8	34.1±1.4	16.2±1.0
	10	FRCN+ft-full	32.2±0.3	52.3±0.4	34.1±0.4	37.2±0.3	59.8±0.4	39.9±0.4	17.0±0.8	29.8±1.4	16.7±0.9
		TFA w/fc	41.3±0.2	67.0±0.3	44.0±0.3	48.3±0.2	76.1±0.3	52.7±0.4	20.2±0.5	39.7±0.9	18.0±0.7
		TFA w/cos	42.3±0.3	67.6±0.4	45.7±0.3	49.4±0.2	76.9±0.3	54.5±0.3	20.8±0.6	39.5±1.1	19.2±0.6
Split 3	1	FSRW (Kang et al., 2019)	27.5±0.6	50.0±1.0	26.8±0.7	34.5±0.7	62.5±1.2	33.5±0.7	6.7±1.0	12.5±1.6	6.4±1.0
		FRCN+ft-full	30.8±0.6	49.8±0.8	32.9±0.8	39.6±0.8	63.7±1.0	42.5±0.9	4.5±0.7	8.1±1.3	4.2±0.7
		TFA w/fc	39.0±0.6	62.3±0.7	42.1±0.8	49.5±0.8	77.8±0.8	54.0±1.0	7.8±1.1	15.7±2.1	6.5±1.0
		TFA w/cos	40.1±0.3	63.5±0.6	43.6±0.5	50.2±0.4	78.7±0.2	55.1±0.5	9.6±1.1	17.9±2.0	9.1±1.2
	2	FSRW (Kang et al., 2019)	28.7±0.4	51.8±0.7	28.1±0.5	34.5±0.4	62.0±0.7	34.0±0.5	11.3±0.7	21.3±1.0	10.6±0.8
		FRCN+ft-full	31.3±0.5	50.2±0.9	33.5±0.6	39.1±0.5	62.4±0.9	42.0±0.7	8.0±0.8	13.9±1.4	7.9±0.9
		TFA w/fc	41.1±0.6	65.1±0.7	44.3±0.7	50.1±0.7	77.7±0.7	54.8±0.9	14.2±1.2	27.2±2.0	12.6±1.3
		TFA w/cos	41.8±0.4	65.6±0.6	45.3±0.4	50.7±0.3	78.4±0.2	55.6±0.4	15.1±1.3	27.2±2.1	14.4±1.5
	3	FSRW (Kang et al., 2019)	29.2±0.4	52.7±0.6	28.5±0.4	34.2±0.3	61.3±0.6	33.6±0.4	14.2±0.7	26.8±1.4	13.1±0.7
		FRCN+ft-full	32.1±0.5	51.3±0.8	34.3±0.6	39.1±0.5	62.1±0.7	42.1±0.6	11.1±0.9	19.0±1.5	11.2±1.0
		TFA w/fc	40.4±0.5	65.4±0.7	43.1±0.7	47.8±0.5	75.6±0.5	52.1±0.7	18.1±1.0	34.7±1.6	16.2±1.3
		TFA w/cos	43.1±0.4	67.5±0.5	46.7±0.5	51.1±0.3	78.6±0.2	56.3±0.4	18.9±1.1	34.3±1.7	18.1±1.4
	5	FSRW (Kang et al., 2019)	30.1±0.3	53.8±0.5	29.3±0.4	34.1±0.3	60.5±0.4	33.6±0.4	18.0±0.7	33.8±1.4	16.5±0.8
		FRCN+ft-full	32.4±0.5	51.7±0.8	34.4±0.6	38.5±0.5	61.0±0.7	41.3±0.6	14.0±0.9	23.9±1.7	13.7±0.9
		TFA w/fc	41.3±0.5	67.1±0.6	44.0±0.6	48.0±0.5	75.8±0.5	52.2±0.6	21.4±0.9	40.8±1.3	19.4±1.0
		TFA w/cos	44.1±0.3	69.1±0.4	47.8±0.4	51.3±0.2	78.5±0.3	56.4±0.3	22.8±0.9	40.8±1.4	22.1±1.1
	10	FRCN+ft-full	33.1±0.5	53.1±0.7	35.2±0.5	38.0±0.5	60.5±0.7	40.7±0.6	18.4±0.8	31.0±1.2	18.7±1.0
		TFA w/fc	42.2±0.4	68.3±0.5	44.9±0.6	48.5±0.4	76.2±0.4	52.9±0.5	23.3±0.8	44.6±1.1	21.0±1.2
		TFA w/cos	45.0±0.3	70.3±0.4	48.9±0.4	51.6±0.2	78.6±0.2	57.0±0.3	25.4±0.7	45.6±1.1	24.7±1.1

Table 2. Generalized object detection benchmarks on COCO. For each metric, we report the average and 95% confidence interval computed over 10 random samples.

# shots	Method	Overall						Base class			Novel class		
		AP	AP50	AP75	APs	APm	APl	bAP	bAP50	bAP75	nAP	nAP50	nAP75
1	FRCN+ft-full	16.2±0.9	25.8±1.2	17.6±1.0	7.2±0.6	17.9±1.0	23.1±1.1	21.0±1.2	33.3±1.7	23.0±1.4	1.7±0.2	3.3±0.3	1.6±0.2
	TFA w/fc	24.0±0.5	38.9±0.5	25.8±0.6	13.8±0.4	26.6±0.4	32.0±0.6	31.5±0.5	50.7±0.6	33.9±0.8	1.6±0.4	3.4±0.6	1.3±0.4
	TFA w/cos	24.4±0.6	39.8±0.8	26.1±0.8	14.7±0.7	26.8±0.5	31.4±0.7	31.9±0.7	51.8±0.9	34.3±0.9	1.9±0.4	3.8±0.6	1.7±0.5
2	FRCN+ft-full	15.8±0.7	25.0±1.1	17.3±0.7	6.6±0.6	17.2±0.8	23.5±0.7	20.0±0.9	31.4±1.5	22.2±1.0	3.1±0.3	6.1±0.6	2.9±0.3
	TFA w/fc	24.5±0.4	39.3±0.6	26.5±0.5	13.9±0.3	27.1±0.5	32.7±0.7	31.4±0.5	49.8±0.7	34.3±0.6	3.8±0.5	7.8±0.8	3.2±0.6
	TFA w/cos	24.9±0.6	40.1±0.9	27.0±0.7	14.9±0.7	27.3±0.6	32.3±0.6	31.9±0.7	50.8±1.1	34.8±0.8	3.9±0.4	7.8±0.7	3.6±0.6
3	FRCN+ft-full	15.0±0.7	23.9±1.2	16.4±0.7	6.0±0.6	16.1±0.9	22.6±0.9	18.8±0.9	29.5±1.5	20.7±0.9	3.7±0.4	7.1±0.8	3.5±0.4
	TFA w/fc	24.9±0.5	39.7±0.7	27.1±0.6	14.1±0.4	27.5±0.6	33.4±0.8	31.5±0.6	49.6±0.7	34.6±0.7	5.0±0.5	9.9±1.0	4.6±0.6
	TFA w/cos	25.3±0.6	40.4±1.0	27.6±0.7	14.8±0.7	27.7±0.6	33.1±0.7	32.0±0.7	50.5±1.0	35.1±0.7	5.1±0.6	9.9±0.9	4.8±0.6
5	FRCN+ft-full	14.4±0.8	23.0±1.3	15.6±0.8	5.6±0.4	15.2±1.0	21.9±1.1	17.6±0.9	27.8±1.5	19.3±1.0	4.6±0.5	8.7±1.0	4.4±0.6
	TFA w/fc	25.6±0.5	40.7±0.8	28.0±0.5	14.3±0.4	28.2±0.6	34.4±0.6	31.8±0.5	49.8±0.7	35.2±0.5	6.9±0.7	13.4±1.2	6.3±0.8
	TFA w/cos	25.9±0.6	41.2±0.9	28.4±0.6	15.0±0.6	28.3±0.5	34.1±0.6	32.3±0.6	50.5±0.9	35.6±0.6	7.0±0.7	13.3±1.2	6.5±0.7
10	FRCN+ft-full	13.4±1.0	21.8±1.7	14.5±0.9	5.1±0.4	14.3±1.2	20.1±1.5	16.1±1.0	25.7±1.8	17.5±1.0	5.5±0.9	10.0±1.6	5.5±0.9
	TFA w/fc	26.2±0.5	41.8±0.7	28.6±0.5	14.5±0.3	29.0±0.5	35.2±0.6	32.0±0.5	49.9±0.7	35.3±0.6	9.1±0.5	17.3±1.0	8.5±0.5
	TFA w/cos	26.6±0.5	42.2±0.8	29.0±0.6	15.0±0.5	29.1±0.4	35.2±0.5	32.4±0.6	50.6±0.9	35.7±0.7	9.1±0.5	17.1±1.1	8.8±0.5
30	FRCN+ft-full	13.5±1.0	21.8±1.9	14.5±1.0	5.1±0.3	14.6±1.2	19.9±2.0	15.6±1.0	24.8±1.8	16.9±1.0	7.4±1.1	13.1±2.1	7.4±1.0
	TFA w/fc	28.4±0.3	44.4±0.6	31.2±0.3	15.7±0.3	31.2±0.3	38.6±0.4	33.8±0.3	51.8±0.6	37.6±0.4	12.0±0.4	22.2±0.6	11.8±0.4
	TFA w/cos	28.7±0.4	44.7±0.7	31.5±0.4	16.1±0.4	31.2±0.3	38.4±0.4	34.2±0.4	52.3±0.7	38.0±0.4	12.1±0.4	22.0±0.7	12.0±0.5

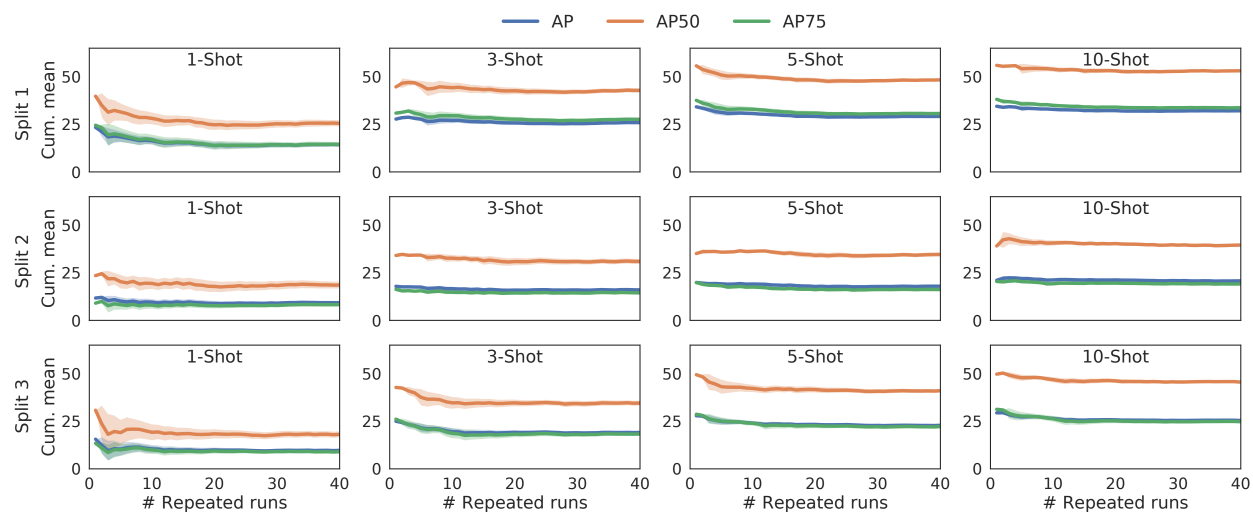

Figure 1. Cumulative means with 95% confidence intervals across 40 repeated runs, computed on the novel classes of all three splits of PASCAL VOC. The means and variances become stable after around 30 runs.

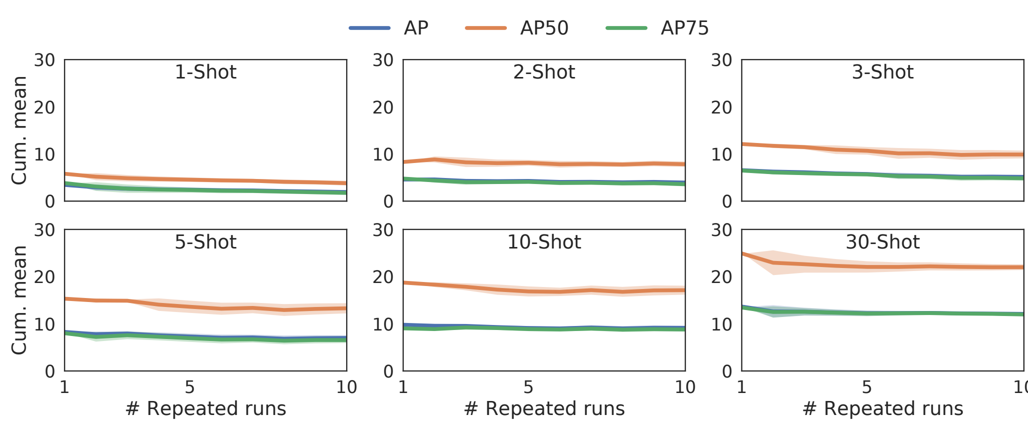

Figure 2. Cumulative means with 95% confidence intervals across 10 repeated runs, computed on the novel classes of COCO.

References

Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., and Darrell, T. Few-shot object detection via feature reweighting. In *ICCV*, 2019.

Wang, Y.-X., Ramanan, D., and Hebert, M. Meta-learning to detect rare objects. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9925–9934, 2019.

Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., and Lin, L. Meta r-cnn: Towards general solver for instance-level low-shot learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9577–9586, 2019.