#4557

Novel Instance Mining with Pseudo-Margin Evaluation for Few-Shot Object Detection

Weijie Liu[1], Chong Wang[1,2*], Shenghao Yu[1], Jun Wang[2], Jiafei Wu[3]

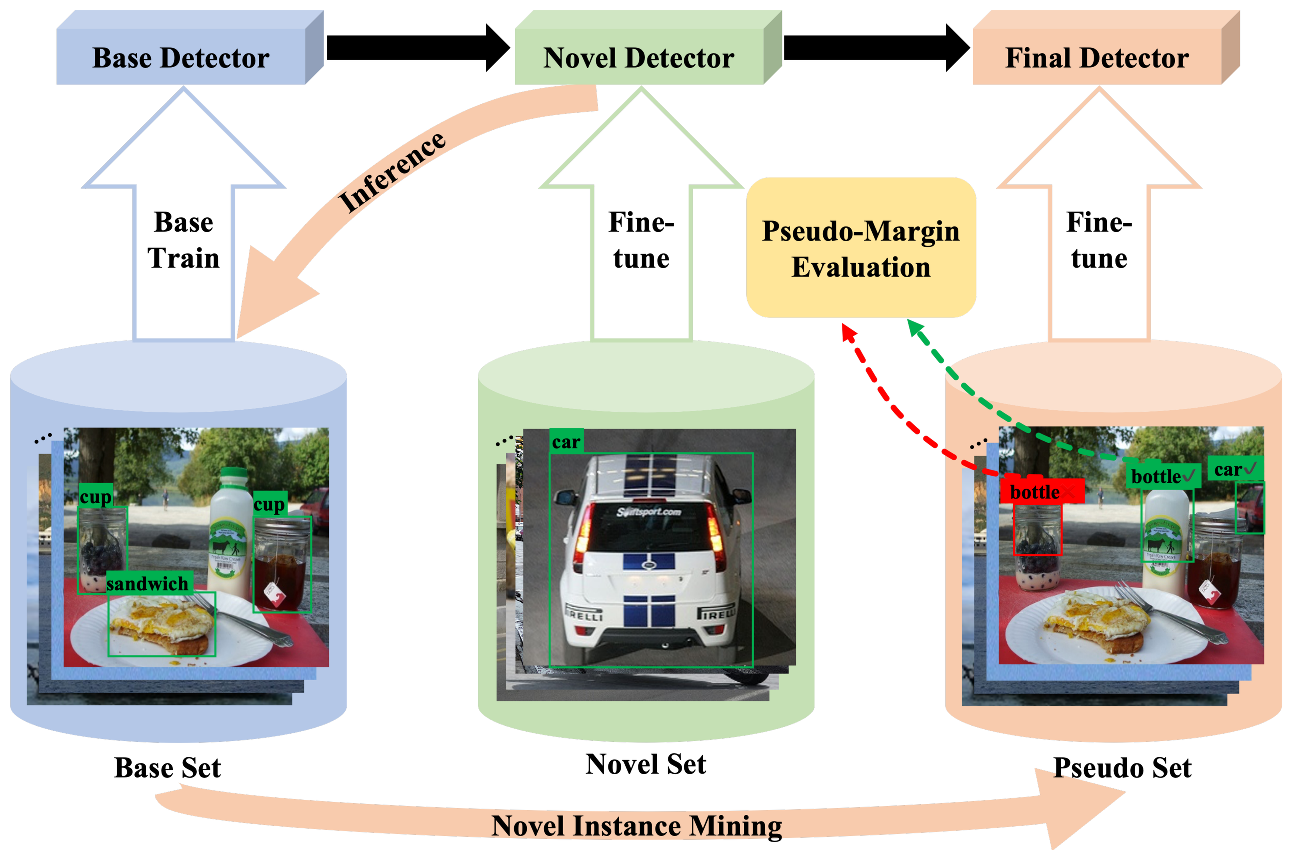
[1]Faculty of Electrical Engineering and Computer Science, Ningbo University, China

[2]School of Information and Control Engineering, China University of Mining and Technology, China

[3]SenseTime Research, China

Introduction

The unlabeled novel instances in the base set were untouched in previous works, which can be re-used to enhance the FSOD performance. Thus, a new instance mining model is proposed in this paper to excavate the novel samples from the base set. The detector is thus fine-tuned again by these additional free novel instances. Meanwhile, a novel pseudo-margin evaluation algorithm is designed to address the quality problem of pseudo-labels brought by those new novel instances.



A pseudo-margin evaluation model is proposed to introduce a new way to exploit the error-prone pseudo-labels by evaluating the uncertainty scores of both correct and incorrect pseudo-labels.

Novel Instance Mining

The key of our proposed model is to excavate more novel instances from the base set $D_b = \{(x_i, y_i^{(b)})\}_{i=1}^{N}$ for FSOD. the pseudo-label $\tilde{y}_i^{(n)}$ of novel instances for x_i can be defined as,

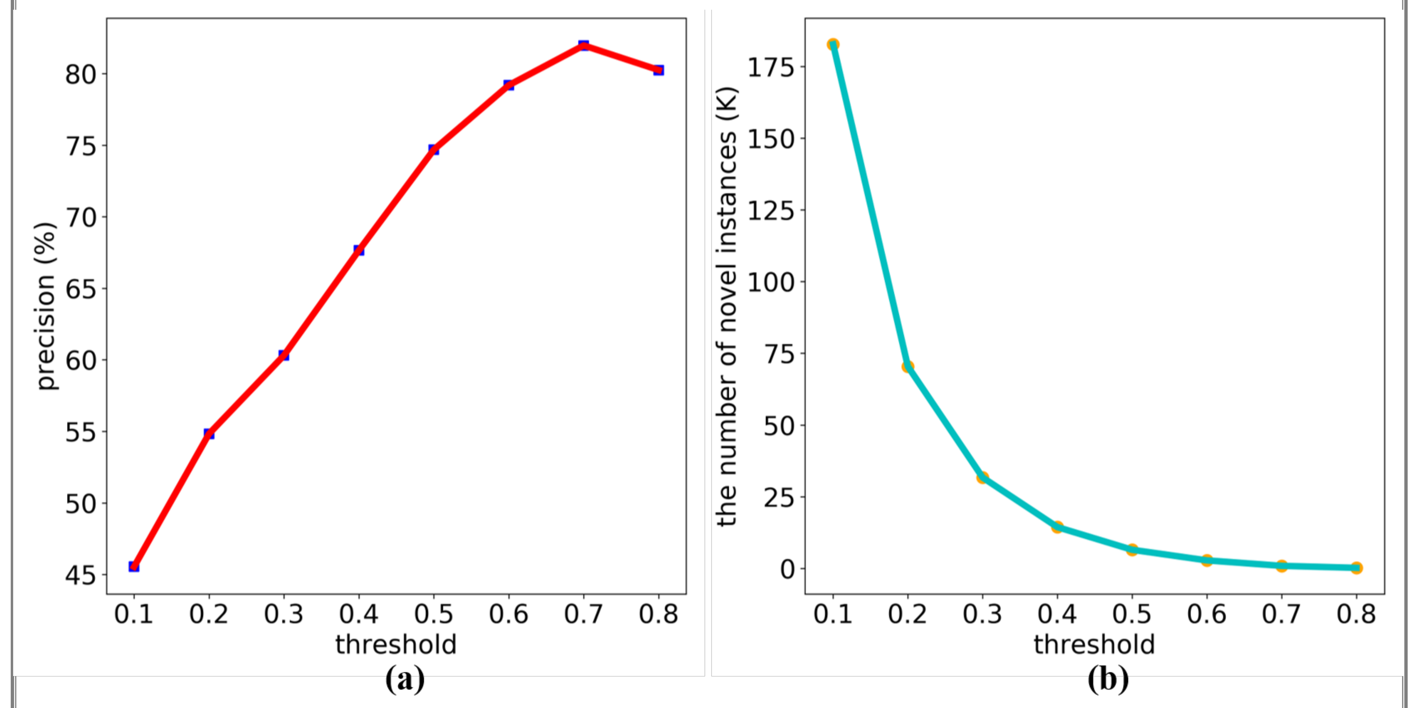
$$\tilde{y}_i^{(n)} = \left\{\left(\hat{c}_{i,k}^{(n)}, \hat{l}_{i,k}^{(n)}\right) | \hat{s}_{i,k}^{(n)} > \delta\right\}_{k=1}^{K_i}$$

Going through the whole dataset of D_b to collect N_p images containing pseudo-labels $\tilde{y}_i^{(n)}$, a new pseudo set D_p can then be formed as,

$$D_p = \left\{(x_i, \tilde{y}_i^{(n)} \cup y_i^{(b)})\right\}_{i=1}^{N_p}$$

To avoid class imbalance, original annotations $y_i^{(b)}$ of base classes are retained in D_p. It is also worth noting that the pseudo-label $\tilde{y}_i^{(n)}$ is not always true.

Pseudo-Margin Evaluation



A too small threshold δ will result numerous misclassified boxes, while a too large one can rarely find useful novel objects. Instead of selecting an empirical value of δ, the pseudo-margin evaluation (PME) is introduced. To estimate the quality of pseudo-labels, the uncertainty score $a_{i,j}^{(v)}$ can be evaluated by the cross-entropy,

$$a_{i,j}^{(v)} = Sigmoid(-\tilde{c}_{i,j}^{(v)} \log \hat{p}_{i,j})$$

The maximum and minimum of uncertainty scores are chosen in pairs to perform pseudo-margin evaluation,

$$L_{pme} = \sum_v max(0, \mu - max(A_v) + min(A_v))$$

The margin between them is used to guide the training of the box classifier, since a larger margin between true and false labels usually means better classification performance on the novel classes.

Loss Function

With the proposed PME loss, the learnable weights $\theta_f^{(p)}$ of prediction head is fine-tuned using D_p as,

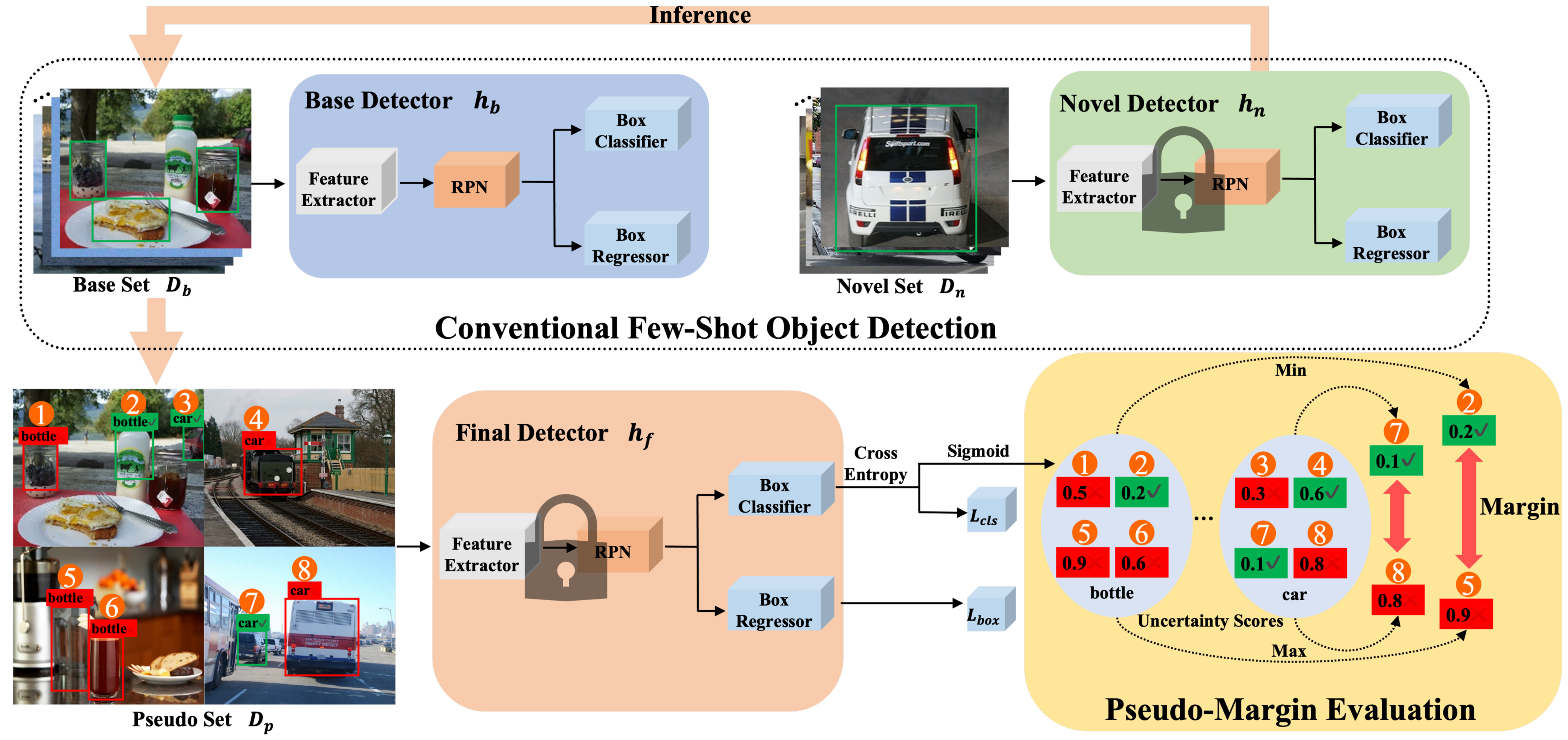
$$\theta_f^{(p)} \leftarrow \theta_f^{(p)} + \gamma \frac{\partial(L_{cls} + L_{box} + L_{pme})}{\theta_f^{(p)}}$$

Noting that D_p contains both true labels of base classes and pseudo-labels of novel classes. Hence, the classification loss L_{cls} is redefined as a weighted sum of the losses of base classes $L_{cls}^{(b)}$ and novel classes $L_{cls}^{(n)}$,

$$L_{cls} = \alpha L_{cls}^{(b)} + \beta L_{cls}^{(n)}$$

The regression loss L_{box} only consider the RoIs from base classes, which is defined as smooth L1 loss.

Network Architecture



The workflow of the proposed model includes three training stages, namely base training, novel set fine-tuning and pseudo set fine-tuning. The first two training stages are the same as the previous FSOD methods. In the pseudo set fine-tuning stage, the base set D_b is re-used to generate a new pseudo set D_p containing plausible novel instances. Such D_p is then utilized to fine-tune the network again to obtain the final detector $h_f(;\theta_f)$.

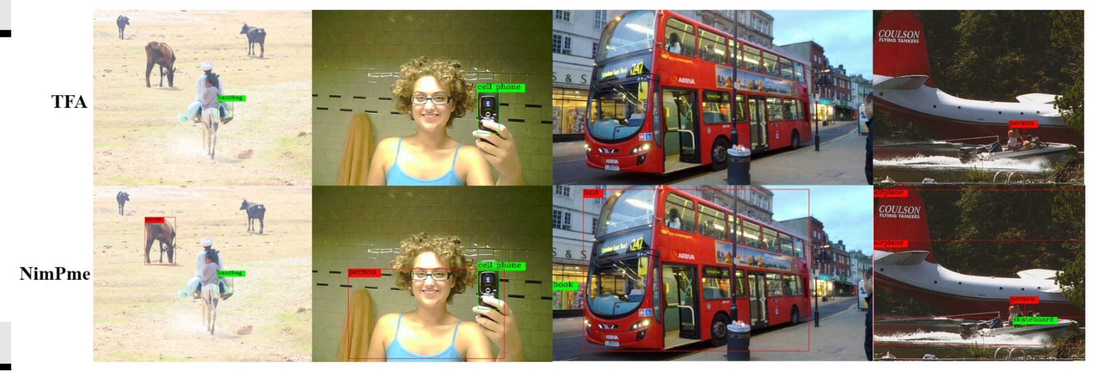
Experimental Results

Table 1: the mAP of novel classes on MS-COCO (%).

Shot	Method	AP	AP50	AP75
10	FSRW[17]	5.6	12.3	4.6
	MetaDet [18]	7.1	14.6	6.1
	Meta R-CNN[19]	8.7	19.1	6.6
	FRCN+ft-full [23]	9.2	17.0	9.2
	Retentive R-CNN [25]	10.5	-	-
	TFA w/fc [23]	10.0	19.2	9.2
	TFA w/cos [23]	10.0	19.1	9.3
	N-PME (Ours)	**10.6**	**21.1**	**9.4**
30	FSRW[17]	9.1	19.0	7.6
	MetaDet [18]	11.3	21.7	8.1
	Meta R-CNN[19]	12.4	25.3	10.8
	FRCN+ft-full [23]	12.5	23.0	12.0
	Retentive R-CNN [25]	13.8	-	-
	TFA w/fc [23]	13.4	24.7	13.2
	TFA w/cos [23]	13.7	24.9	13.4
	N-PME (Ours)	**14.1**	**26.5**	**13.6**

Table 2: Ablation study of 10-shot setting on MS-COCO (%).

Pseudo Set	PME	Novel Set	AP	AP50	AP75
			10.0	19.1	9.3
√			10.4	20.4	9.2
√	√		**10.6**	**21.1**	**9.4**
√	√	√	**10.6**	**21.1**	**9.4**



The N-PME achieves constantly higher accuracy than the baseline in term of AP, AP50 and AP75 in both of 10 and 30-shot settings. Compared to the significant improvement (2.0% and 1.6%) of AP50, the improvement of AP75 is relatively small (0.1% and 0.2%). The best guess is that it is limited by the accuracy of bounding box regression, since the novel instances are not used in the regression loss. This can also be observed in the visualization of detection results. as the result shown in Table 2, there is no difference from the one only using D_p. The best guess is that the detector has converged on the novel set with a small number of training samples. Thus, it could not improve the performance further.