南京航空航天大學
Nanjing University of Aeronautics and Astronautics

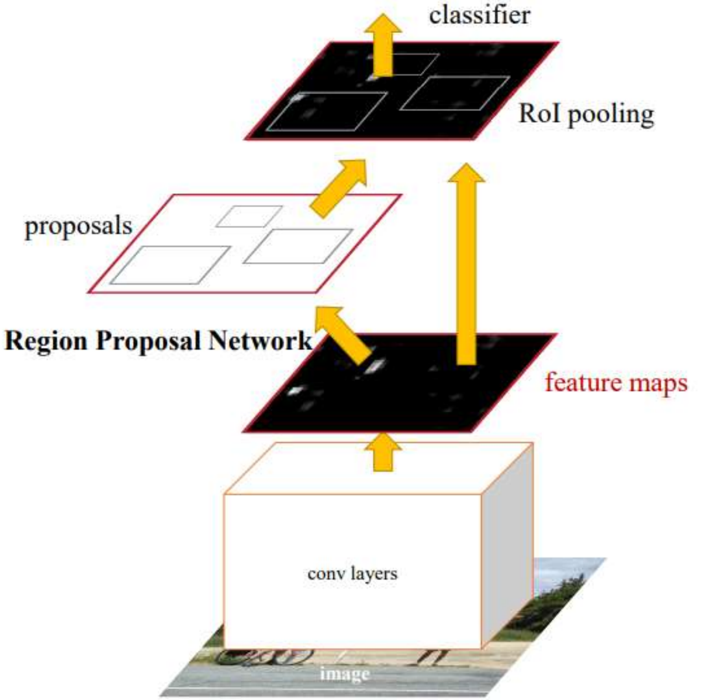
Label Matching Semi-Supervised Object Detection

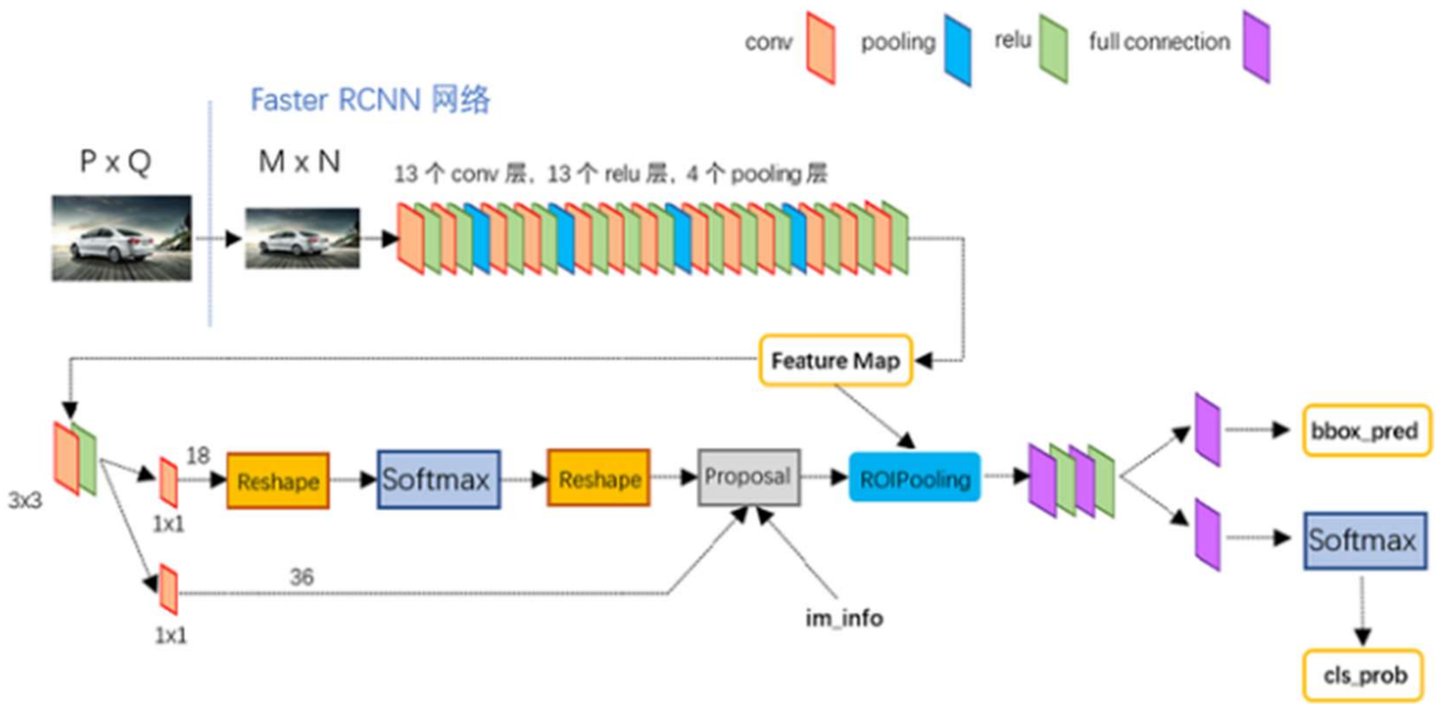
Binbin Chen[2], Weijie Chen[1,2,†], Shicai Yang[2], Yunyi Xuan[2], Jie Song[1]
Di Xie[2], Shiliang Pu[2], Mingli Song[1], Yueting Zhuang[1,†]
[1]Zhejiang University, [2]Hikvision Research Institute
{chenbinbin8,chenweijie5,yangshicai,xuanyunyi,xiedi,pushiliang.hri}@hikvision.com
{sjie,songml,yzhuang}@zju.edu.cn

CVPR 2022

1 Preface

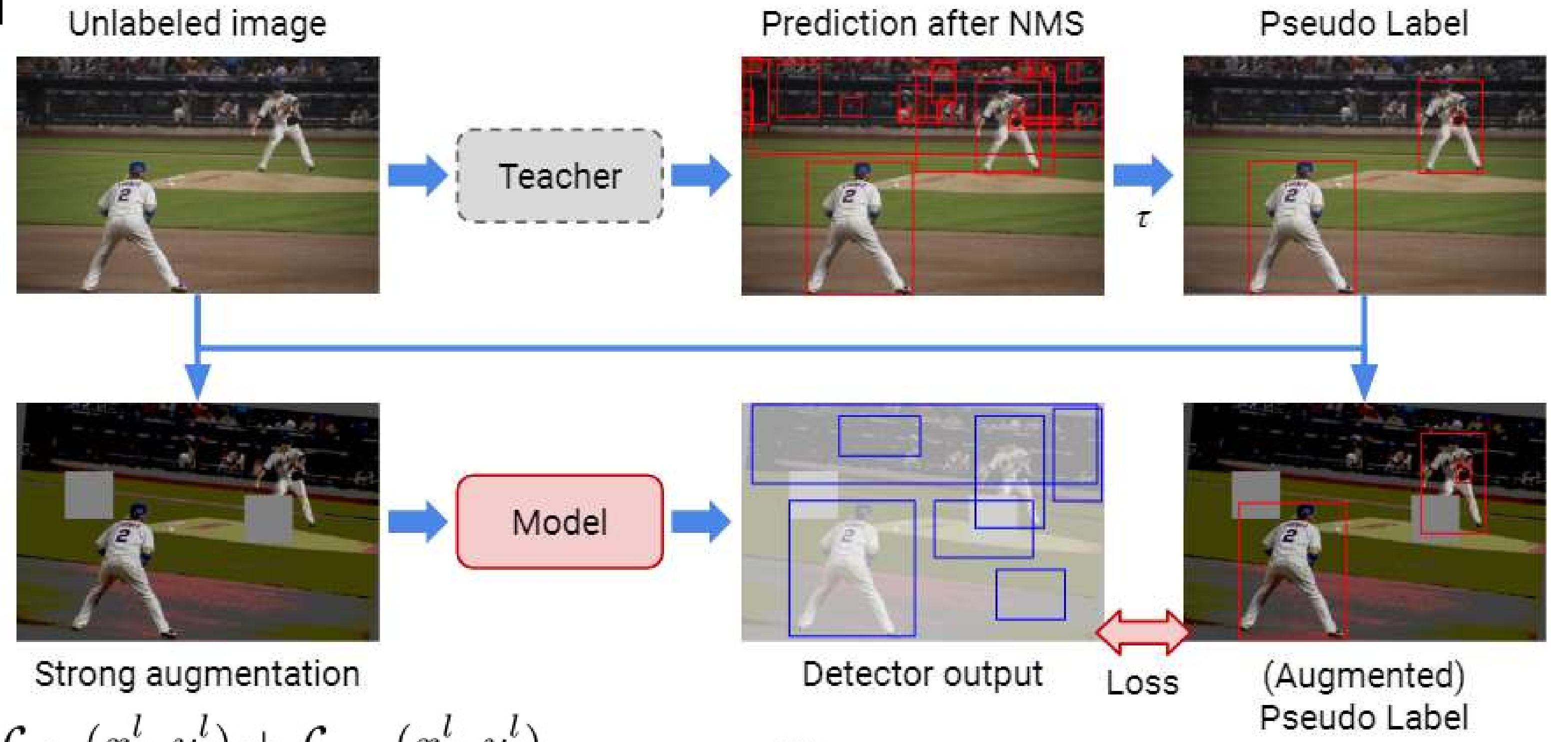
Faster RCNN





1 Preface

A Simple Semi-Supervised Learning Framework for Object Detection



$$\mathcal{L}_l = \sum_i \mathcal{L}_{cls}(x_i^l, y_i^l) + \mathcal{L}_{reg}(x_i^l, y_i^l), \quad (1)$$

$$\mathcal{L}_u = \sum_i \mathcal{L}_{cls}(x_i^u, y_i^u) + \mathcal{L}_{reg}(x_i^u, y_i^u), \quad (2)$$

$$\mathcal{L}_{total} = \mathcal{L}_l + \lambda \mathcal{L}_u, \quad (3)$$

2 Problem

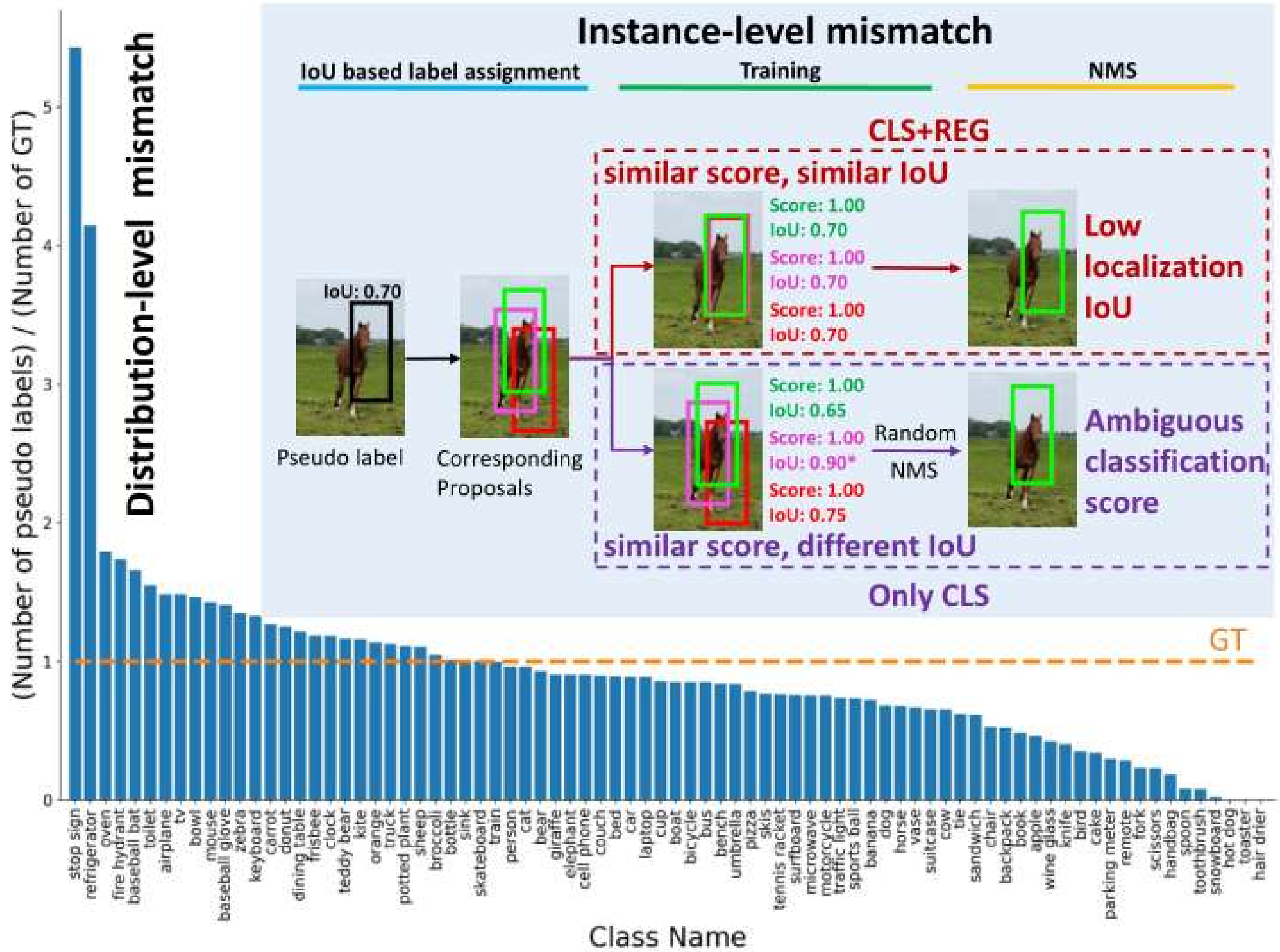


Figure 1. Label mismatch problems on the MS-COCO dataset. 1) **Distribution-level mismatch**: there exists a bias between the pseudo labels produced by the single confidence threshold and the ground truth labels (GT) during self-training, as shown in the relation of the blue bar and the orange dotted line. 2) **Instance-level mismatch**: there are two kinds of training patterns for the unlabeled data in the previous SSOD frameworks. One is the same as supervised learning, using both classification and box regression for optimization, which will overfit the poor-quality pseudo labels and result in low localization accuracy. To avoid incorrect box regression, another one merely exploits a classification objective [22], which will bring ambiguity due to the similar classification scores to confuse the post-processing of Non-Maximum-Suppression (NMS).

3 Method

南京航空航天大学
Nanjing University of Aeronautics and Astronautics

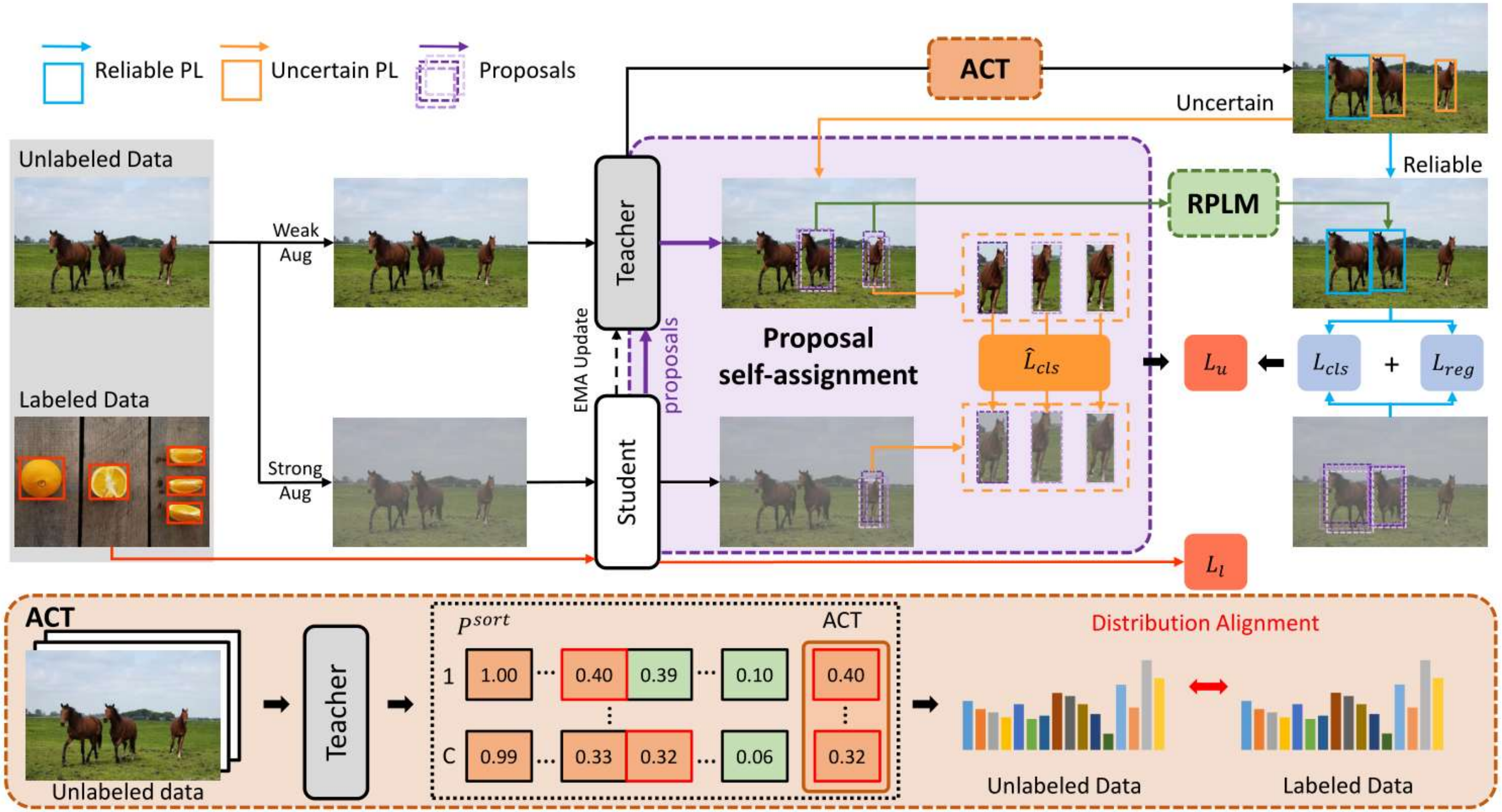


Figure 2. An overview of LabelMatch framework. Labeled data: only applied to the student with a supervised loss. Unlabeled data: annotated by the teacher to get pseudo labels (PL) according to the *adaptive label-distribution-aware confidence thresholds* (ACT), which are then split into reliable ones and uncertain ones for separated optimization. Reliable pseudo labels directly follow the IoU based assignment strategy, acting as hard labels to train the student model. As for uncertain labels, the *proposal self-assignment* method guides the student training with the supervision provided by the corresponding proposal prediction in the teacher. Besides, a *reliable pseudo label mining* (RPLM) strategy is utilized to convert the high-quality uncertain pseudo labels into reliable ones as the training goes on.

3.1 Re-distribution Mean Teacher

$$\underset{t_1,\ldots,t_C}{\operatorname{argmin}} \quad D_{KL}([\underbrace{r_1^l,\ldots,r_C^l}_{f-f}, \underbrace{r_f^l}_{f-b}], [\underbrace{r_1^u,\ldots,r_C^u}_{f-f}, \underbrace{r_f^u}_{f-b}])$$

$$\text{s.t.} \quad r_c^l = \frac{n_c^l}{\sum_{i=1}^{C} n_i^l},$$

$$r_f^l = \frac{\sum_{i=1}^{C} n_i^l}{N_l}, \qquad (4)$$

$$r_c^u = \frac{\sum_{j=1}^{N_u} \sum (P_c^j > t_c)}{\sum_{i=1}^{C} (\sum_{j=1}^{N_u} \sum (P_i^j > t_i))},$$

$$r_f^u = \frac{\sum_{i=1}^{C} (\sum_{j=1}^{N_u} \sum (P_i^j > t_i))}{N_u},$$

$$t_c = P_c^{sort}[n_c^l \cdot \frac{N_u}{N_l}], \qquad (5)$$

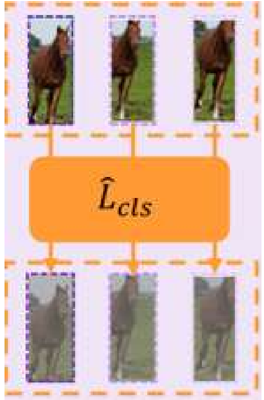
3.2 Proposal Self-Assignment

It is worth noting that the quality of pseudo labels cannot be guaranteed, especially at the early beginning of self-training.

$$t_c^r = P_c^{sort}[\alpha\% \cdot n_c^l \cdot \frac{N_u}{N_l}], \qquad (6)$$

Obviously, the uncertain pseudo labels potentially lead to low localization accuracy. To avoid poor box regression, Unbiased Teacher(ICLR 2021) removes the box regression loss for the unlabeled data, yet resulting in ambiguity in label assignment as shown in Fig. 1.

$$\hat{\mathcal{L}}_{cls} = \sum_{i=1}^{n_p} \sum_{c=1}^{C} -p_{i,c}^t \log p_{i,c}^s, \qquad (7)$$



$$\mathcal{L}_u = \sum_i \mathcal{L}_{cls}(x_i^u, y_i^{ur}) + \mathcal{L}_{reg}(x_i^u, y_i^{ur}) + \hat{\mathcal{L}}_{cls}(x_i^u, y_i^{uu}), \qquad (8)$$

3.3 Reliable Pseudo Label Mining

To benefit from the continuously evolved teacher model and encourage the cycle positive feedback during self-training, we present a reliable pseudo label mining strategy to convert the high-quality uncertain pseudo labels into reliable ones.

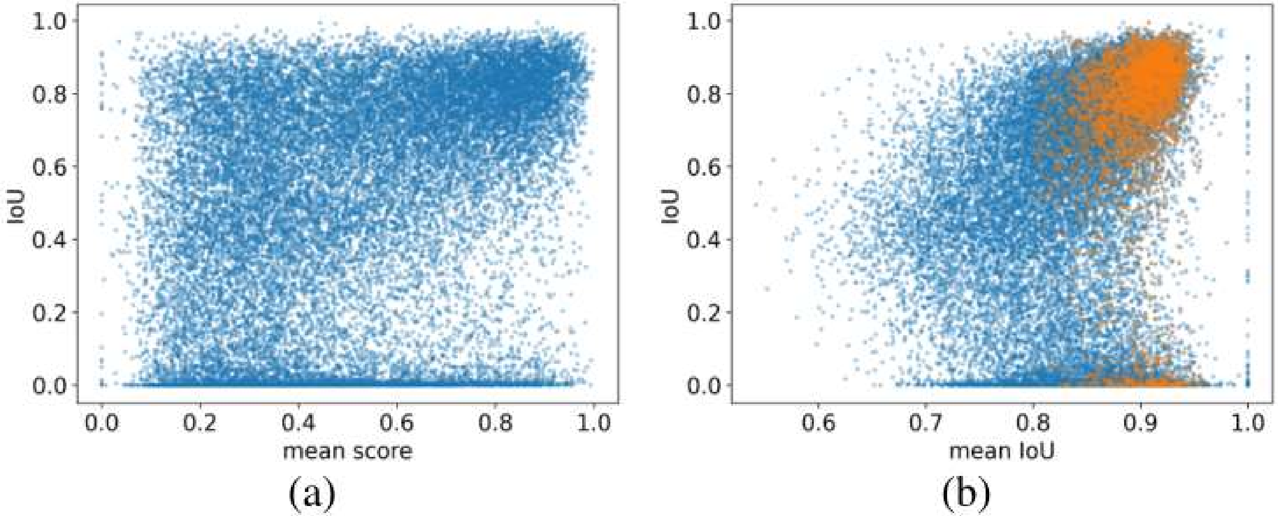


Figure 3. 5k images are selected to estimate the quality of pseudo labels. (a) the correlation between the IoU with ground truth and *mean score*. (b) the correlation between the IoU with ground truth and *mean IoU*. The orange points represent the predictions with the *mean score* larger than 0.8 and *mean IoU* larger than 0.8

4 Experiments

- COCO-standard: 1%, 5%, 10% images of train2017 set are sampled as the labeled training data and the remaining ones as the unlabeled data.

- COCO-additional: we use the entire train2017 set as the labeled data and the additional COCO2017-unlabeled set as the unlabeled data.

- VOC: we use VOC07 trainval set as the labeled data and the VOC12 trainval set as the unlabeled data.

- The validation sets in the COCO setting and VOC setting are COCO val2017 and VOC07 test set, respectively.

4 Experiments

	Threshold	1%	5%	10%
Supervised [22]	-	9.05 ± 0.16	18.47 ± 0.22	23.86 ± 0.81
STAC [31]	0.9	13.97 ± 0.35 (+4.92)	24.38 ± 0.12 (+5.91)	28.64 ± 0.21 (+4.78)
ISMT [38]	0.9	18.88 ± 0.74 (+9.83)	26.37 ± 0.24 (+7.90)	30.53 ± 0.52 (+6.67)
Instant Teaching [40]	0.9	18.05 ± 0.15 (+9.00)	26.75 ± 0.05 (+8.28)	30.40 ± 0.05 (+6.54)
Unbiased Teacher [22]	0.7	20.75 ± 0.12 (+11.70)	28.27 ± 0.11 (+9.80)	31.50 ± 0.10 (+7.64)
Soft Teacher [37]	0.9	20.46 ± 0.39 (+11.41)	30.74 ± 0.08 (+12.27)	34.04 ± 0.14 (+10.18)
LabelMatch (Ours)	ACT	$\mathbf{25.81 \pm 0.28}$ (+16.76)	$\mathbf{32.70 \pm 0.18}$ (+14.23)	$\mathbf{35.49 \pm 0.17}$ (+11.63)

Table 1. Experimental results on COCO-standard ($AP_{50:95}$). All the results are the average of all 5 folds.

	Iterations	$AP_{50:95}$
STAC [31]	540k	$39.5 \xrightarrow{-0.3} 39.2$
Unbiased Teacher [22]	270k	$40.2 \xrightarrow{+1.1} 41.3$
Soft Teacher [37]	370k	$40.9 \xrightarrow{+3.6} 44.5$
LabelMatch (Ours)	540k	$40.3 \xrightarrow{+5.0} \mathbf{45.3}$

Table 2. Experimental results on COCO-additional.

	AP_{50}	$AP_{50:95}$
Supervised [22]	72.63	42.13
STAC [31]	77.45 (+4.82)	44.64 (+2.51)
ISMT [38]	77.23 (+4.60)	46.23 (+4.10)
Instant Teaching [40]	79.20 (+6.57)	50.00 (+7.87)
Unbiased Teacher [22]	77.37 (+4.74)	48.69 (+6.56)
LabelMatch (Ours)	**85.48** (+12.85)	**55.11** (+12.98)

Table 3. Experimental results on VOC.

4 Experiments——Ablation

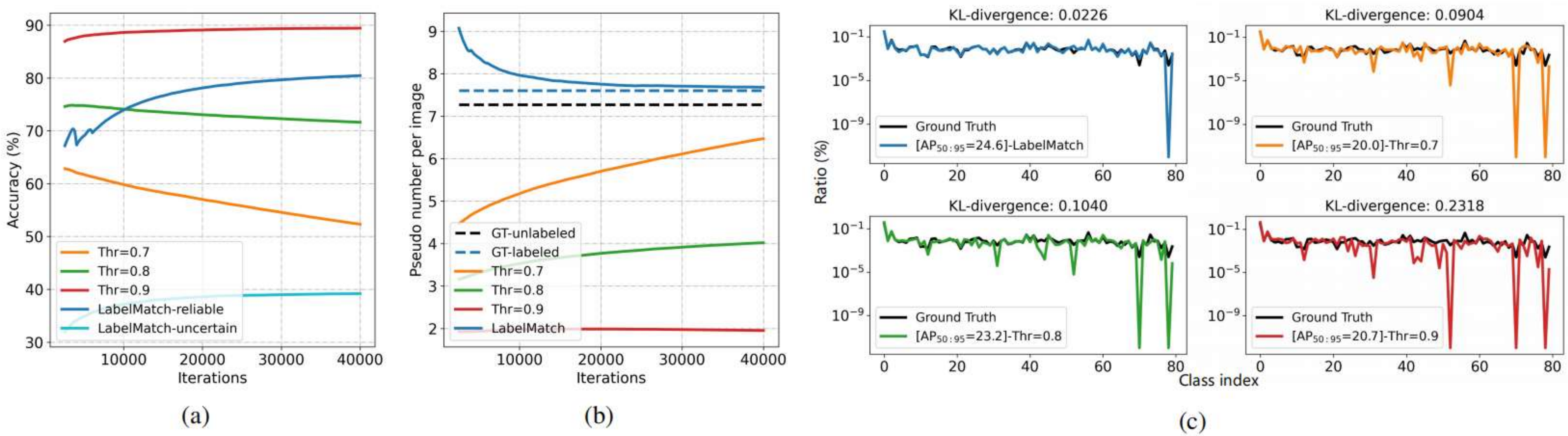
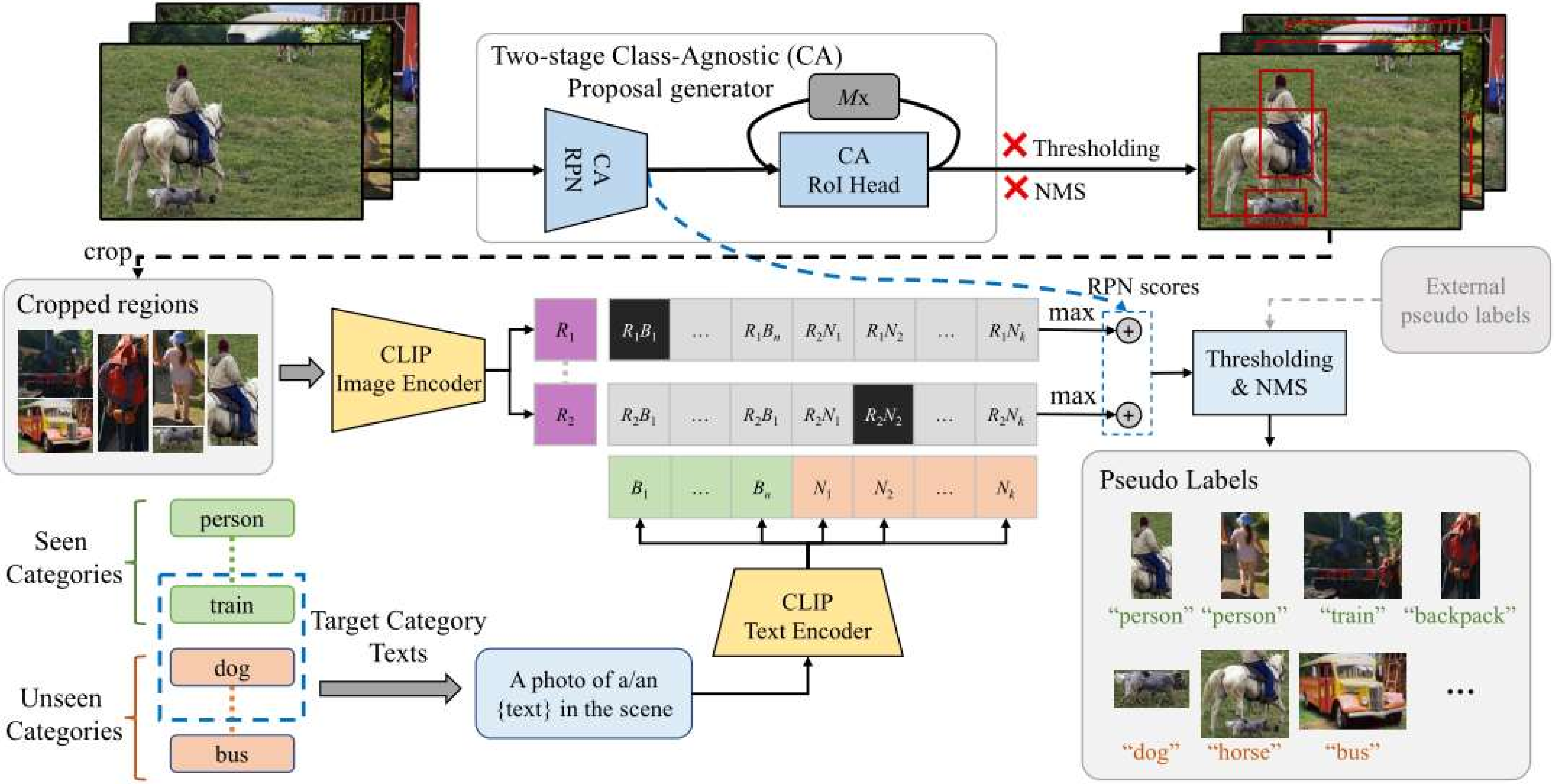


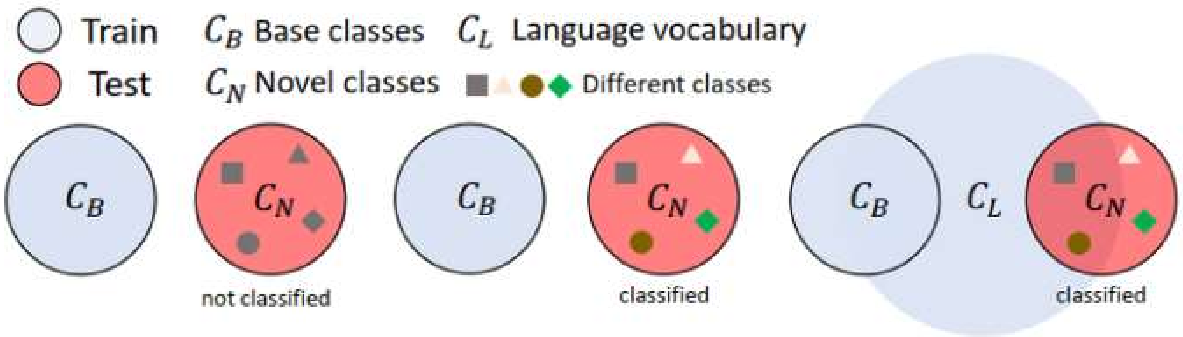
Figure 4. Ablation study about the quality of pseudo labels. (a) the accuracy of pseudo labels in the training phase. (Note: the pseudo labels with IoU overlapping the ground truth greater than 0.5 are regarded as true positives) (b) The average number of pseudo labels per image in the training phase. (c) KL divergence of the class distribution between the pseudo labels and the ground truth in the training phase.

5 Another

Exploiting Unlabeled Data with Vision and Language Models for Object Detection(ECCV2022)



5 Another



(a) Open-Set/Open World/OOD (b) Zero-Shot (c) Open Vocabulary

Table 1. Evaluations for open vocabulary detection on the COCO 2017 [32]. RegionCLIP* indicates a model without refinement using image-caption pairs.

Method	Training Source	Novel AP	Base AP	Overall AP
Bansal *et al.* [4]	instance-level labels in $\mathcal{S}_B$	0.31	29.2	24.9
Zhu *et al.* [63]		3.41	13.8	13.0
Rahman *et al.* [40]		4.12	35.9	27.9
OVR-CNN [56]	image-caption pairs in $\mathcal{S}_B \cup \mathcal{S}_N$ instance-level labels in $\mathcal{S}_B$	22.8	46.0	39.9
Gao *et al.* [14]	raw image-text pairs via Internet image-caption pairs in $\mathcal{S}_B \cup \mathcal{S}_N$ instance-level labels in $\mathcal{S}_B$	30.8	46.1	42.1
RegionCLIP [59]		31.4	57.1	50.4
RegionCLIP* [59]	raw image-text pairs via Internet instance-level labels in $\mathcal{S}_B$	14.2	52.8	42.7
ViLD [16]		27.6	59.5	51.3
VL-PLM (Ours)		**34.4**	**60.2**	**53.5**

Table 3. Evaluation of pseudo labels for semi-supervised object detection on COCO [32].

Methods	1% COCO	2% COCO	5% COCO	10% COCO
Supervised	9.25	12.70	17.71	22.10
Supervised+PLs	11.18	14.88	21.20	25.98
Supervised+VL-PLM	**15.35**	**18.60**	**23.70**	**27.23**
STAC [46]	13.97	18.25	24.38	28.64
STAC+VL-PLM	**17.71**	**21.20**	**26.21**	**29.61**

南京航空航天大學

Nanjing University of Aeronautics and Astronautics

Thanks for Listening