Tencent YouTu Lab

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

MixTeacher: Mining Promising Labels with Mixed Scale Teacher for Semi-Supervised Object Detection

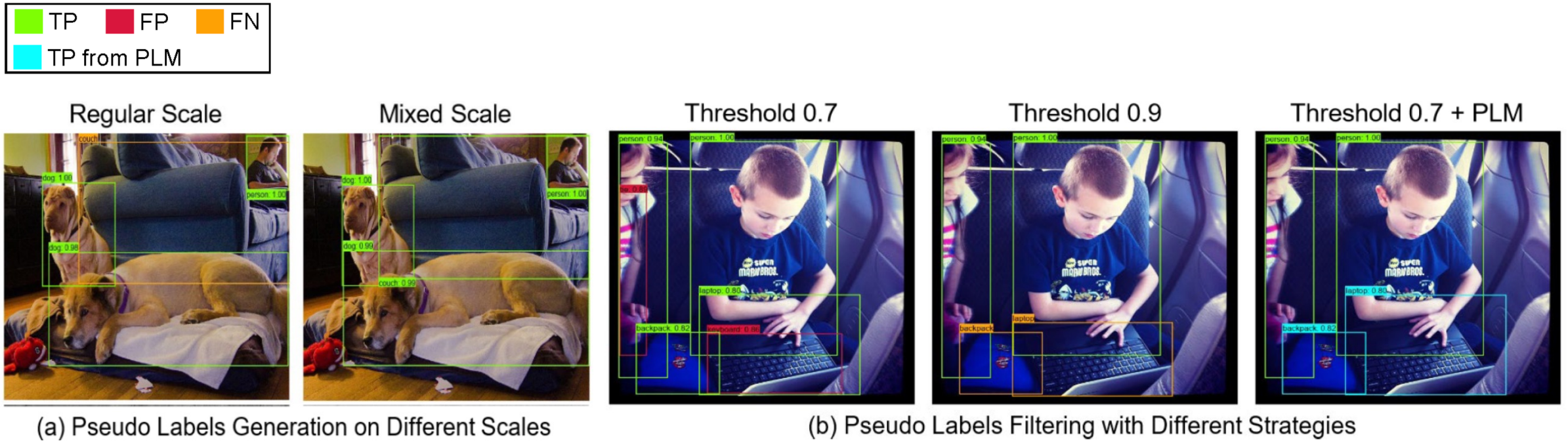
Liang Liu[1], Boshen Zhang[1], Jiangning Zhang[1], Wuhao Zhang[1], Zhenye Gan[1],
Guanzhong Tian[3], Wenbing Zhu[4], Yabiao Wang[1*], Chengjie Wang[1,2*],
[1.] Youtu Lab, Tencent [2.] Shanghai Jiao Tong University
[3.] Ningbo Research Institute, Zhejiang University [4.] Rongcheer Co., Ltd.
(* indicates Corresponding authors)

Poster: TUE-PM-310

MixTeacher Overview

Tencent YouTu Lab

- MixTeacher is a framework to address the challenge of scale variation in semi-supervised object detection.
- MixTeacher generates high-quality initial pseudo labels using a mixed-scale feature pyramid.
- MixTeacher mines low-confidence pseudo labels according to the score promotion of predictions across scales.
- MixTeacher achieves SoTA on various benchmarks under SSOD settings.



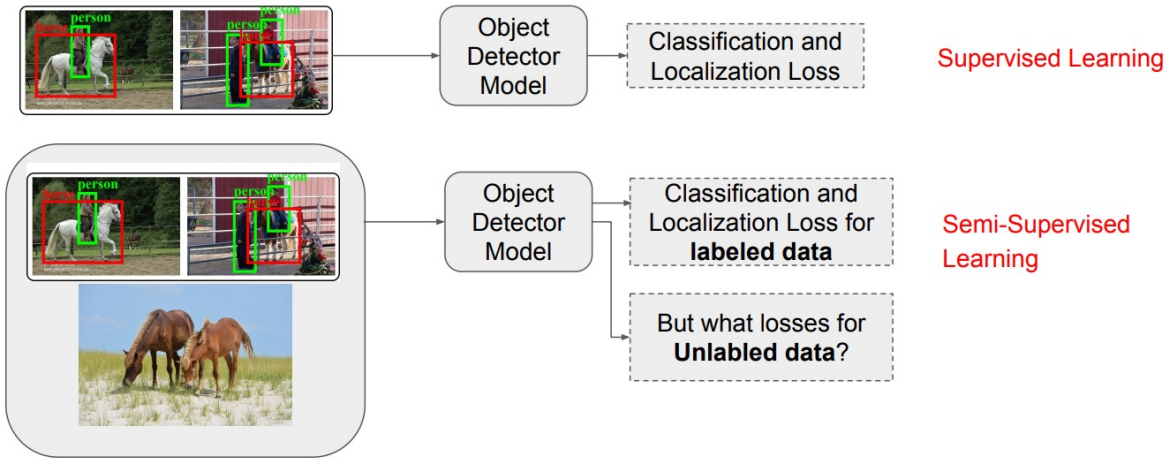
Background: Semi-Supervised Object Detection

Tencent YouTu Lab

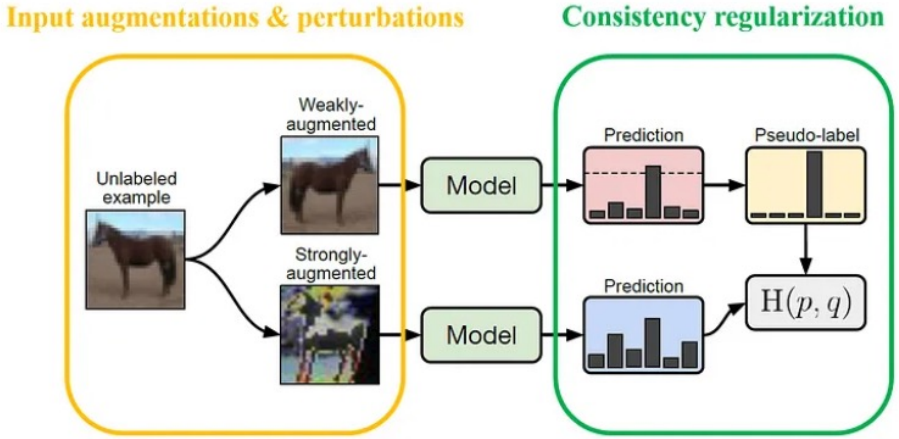
Small amount of labeled images +
Large amount of unlabeled images

⇩

Semi-Supervised Learning

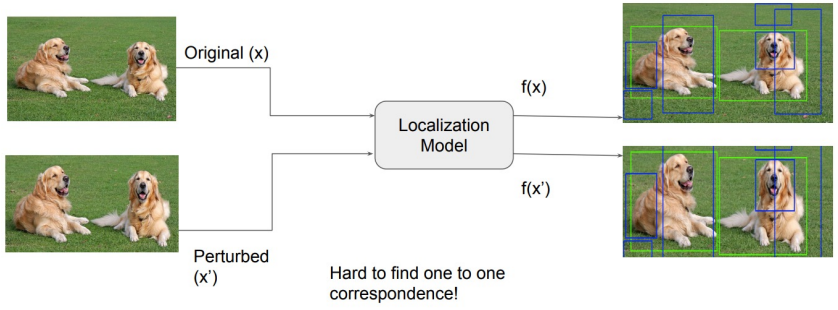


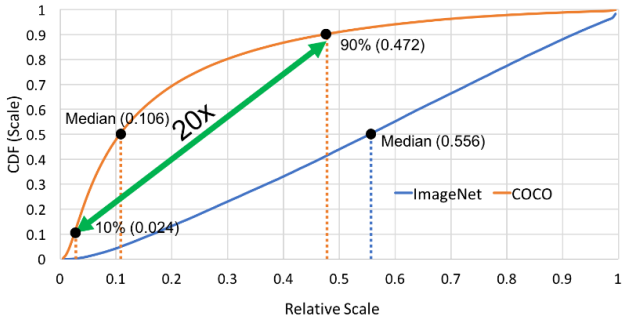
Semi-supervise in Image Classification



☺ Clear correspondence between the predictions in different views

Semi-supervise in Object Detection





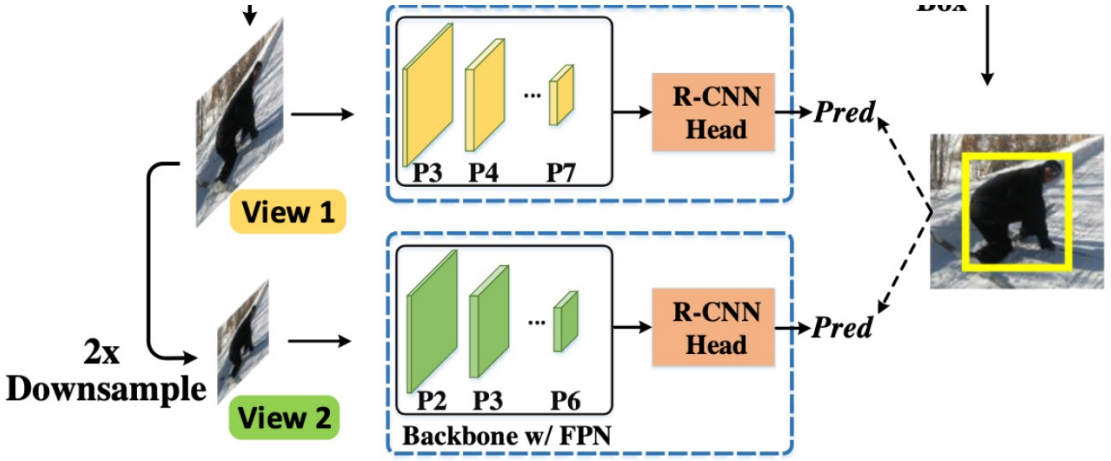
☹ Clear correspondence between the predictions in different views

☹ Object instance scale changes dramatically in detection task

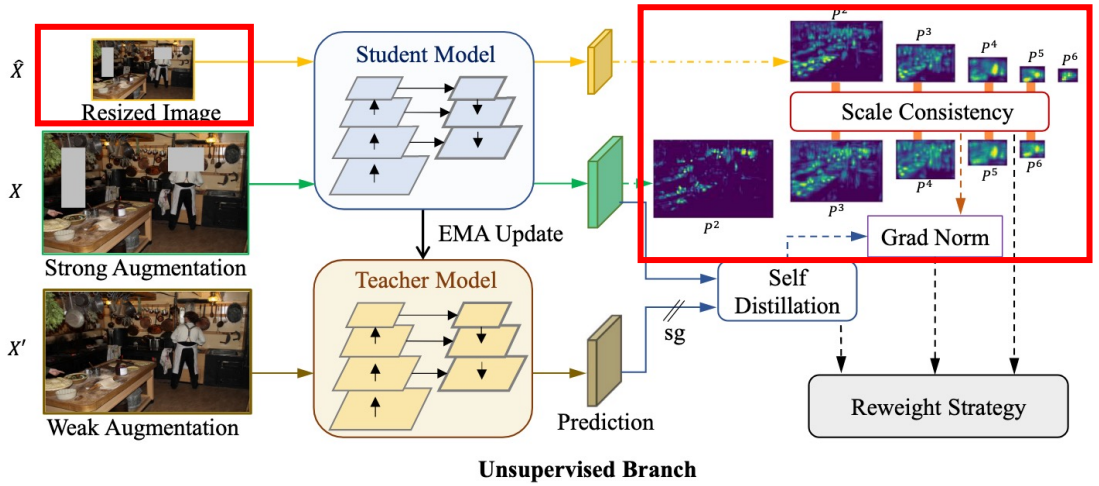
Existing Observation: Additional Scales Helps SSOD

Tencent YouTu Lab

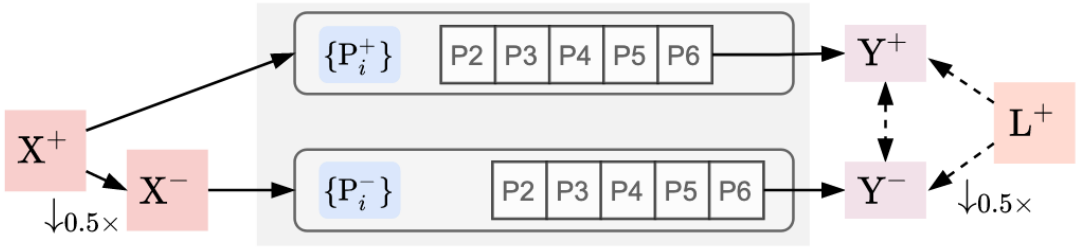
Additional down-sampling scale input helps SSOD



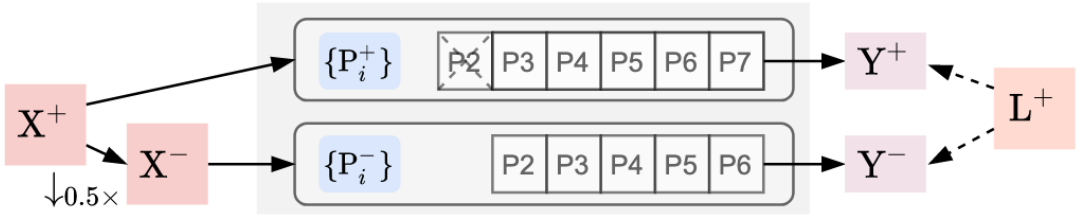
[PseCo, Li et al, ECCV 2022]



[SED, Guo et al, CVPR 2022]



(a) Scale Consistency Regularization in SED



(b) Multiview Scale-Invariant Learning in PseCo

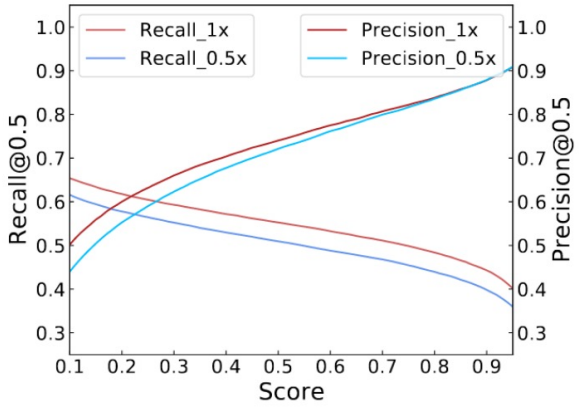
Previous methods adopt pseudo labels from large scale inputs and take additional scales as regularization.

Our Motivations: Regular Scale Pseudo Labels Are Not Always Reliable

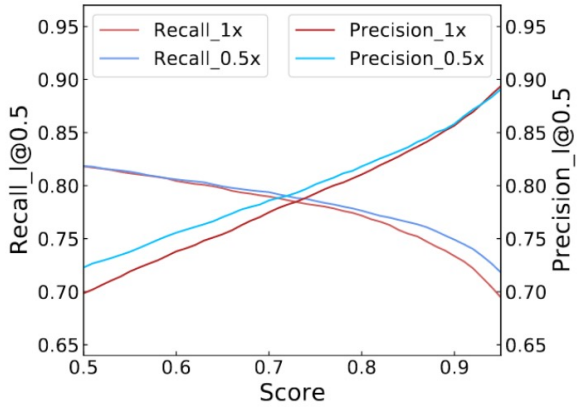
Tencent YouTu Lab

Input size significantly affects pseudo-labels in SSOD

Small scale input generates better predictions on large objects.



(a) Precision and recall for all objects



(b) Precision and recall for large objects

Model test on different scales

P^+: regular scale

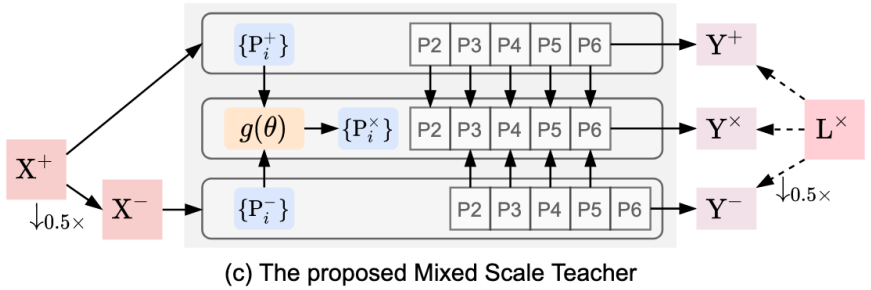
P^-: 0.5x down-sample scale

	mAP	AP_s	AP_m	AP_l
Test on $\mathbb{P}^-$	33.2	14.6	36.1	50.0
Test on $\mathbb{P}^+$	36.7	**21.8**	39.2	48.6

Method: Mixed Scale Teacher

Tencent YouTu Lab

Pseudo Labels from Mixed Scale

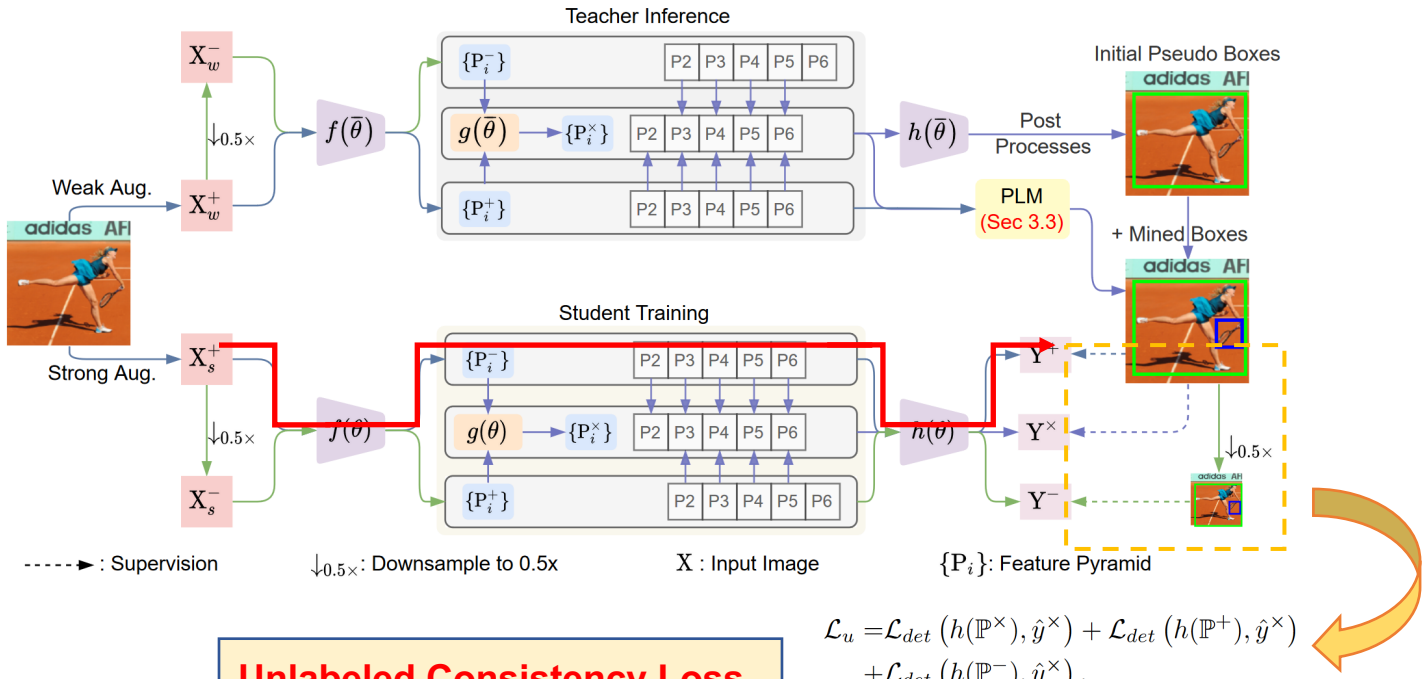


(c) The proposed Mixed Scale Teacher

$$\gamma = \sigma\left(g\left(P_i^+, P_{i-1}^-|\theta_g\right)\right),$$

$$P_i^\times = \gamma P_i^+ + (1-\gamma)P_{i-1}^-.$$

Mixed Scale Feature Pyramid



$$\mathcal{L}_u = \mathcal{L}_{det}\left(h(\mathbb{P}^\times), \hat{y}^\times\right) + \mathcal{L}_{det}\left(h(\mathbb{P}^+), \hat{y}^\times\right) + \mathcal{L}_{det}\left(h(\mathbb{P}^-), \hat{y}^\times\right),$$

Unlabeled Consistency Loss

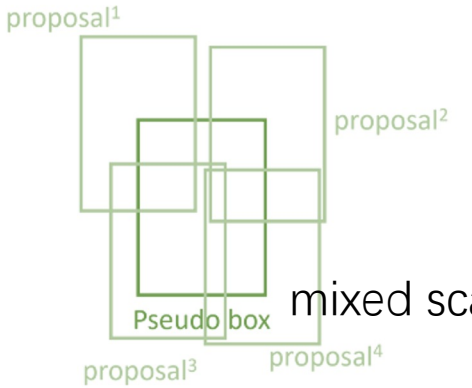
☺ **Efficient Training.** Only an additional module of two layer MLPs is insert into the original detector.

☺ **Fair Testing.** During testing, all parts not present in the original detector are discarded.

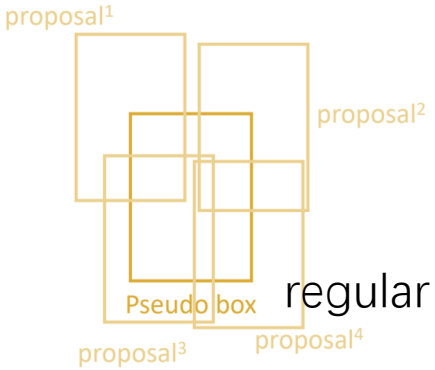
Method: Promising Labels Mining

Tencent YouTu Lab

Pseudo Labels from Lower Score Predictions :



mixed scale



regular scale

Mean score promotion for a pseudo box across scales

$$\Delta q = \frac{1}{K}\left(\sum_{i=1}^{K} h_c(b_i, \mathbb{P}^{\times}|\bar{\theta}) - \sum_{i=1}^{K} h_c(b_i, \mathbb{P}^{+}|\bar{\theta})\right)$$

Procedure:

1. Construct a bag of proposals for each low-confidence candidate.
2. Compute the mean score for mixed scale and regular scale.
3. Compute the gap for two scales as the mean score promotion.

Ablation Study:

Feature Scales $\mathbb{P}^+$	$\mathbb{P}^-$	$\mathbb{P}^\times$	MST	PLM	mAP	AP_{50}	AP_{75}
					26.8	45.1	28.4
✓					33.9 (+7.1)	54.0	37.0
✓	✓				34.7 (+7.9)	54.7	37.8
✓	✓	✓			34.4 (+7.4)	54.2	37.2
✓	✓	✓	✓		36.2 (+9.4)	56.5	39.5
✓	✓	✓	✓	✓	**36.7 (+9.9)**	**57.0**	**39.7**

Table 5. Analysis of various components of proposed approach. MST indicates generating pseudo labels from the mixed scale feature pyramid. PLM indicates promising labels mining.

Results of MixTeacher

Tencent YouTu Lab

Experimental comparison: Compared with 10+ methods on VOC 07, 12 and COCO 14 data sets, the best mAP has been achieved.

COCO: Gains over baseline > 10 mAP

	COCO Partially Labeled				COCO Additional
	1%	2%	5%	10%	100 %
Supervised Baseline	12.15±0.27	16.65±0.18	21.45±0.16	27.10±0.07	40.9
STAC [35]	13.97±0.35	18.25±0.25	24.38±0.12	28.64±0.21	39.5 $\xrightarrow{-0.3}$ 39.2
SED [10]	[illegible]	[illegible]	[illegible]	[illegible]	40.2 $\xrightarrow{+3.2}$ 43.4
Unbiased Teacher [25]*	20.75±0.12	24.30±0.07	28.27±0.11	31.50±0.10	40.2 $\xrightarrow{+1.1}$ 41.3
Soft Teacher [40]†	20.46±0.39	-	30.74±0.08	34.04±0.14	40.9 $\xrightarrow{+3.6}$ 44.5
LabelMatching [3]*	**25.81±0.28**	-	32.70±0.18	35.49±0.17	40.3 $\xrightarrow{+5.0}$ 45.3
PseCo [17]†	22.43±0.36	27.77±0.18	32.50±0.08	36.06±0.24	41.0 $\xrightarrow{+5.1}$ **46.1**
DTG-SSOD [16]†	21.27±0.12	26.84±0.25	31.90±0.08	35.92±0.26	40.9 $\xrightarrow{+4.8}$ 45.7
Unbiased Teacher v2 [26]*	25.40±0.36	28.37±0.03	31.85±0.09	35.08±0.02	40.9 $\xrightarrow{+3.9}$ 44.8
MixTeacher (Ours)†	25.16±0.26	**29.11±0.21**	**34.06±0.13**	**36.72±0.16**	40.9 $\xrightarrow{+4.8}$ 45.7

VOC: Gains over baseline > 13 mAP

	AP_{50}	$AP_{50:95}$
Supervised [25]	72.63	42.13
STAC [35]	77.45 (+4.82)	44.64 (+2.51)
Humble Teacher [36]	80.94 (+8.31)	53.04 (+10.91)
Rethinking Pse [18]	79.00 (+6.37)	54.60 (+12.47)
LabelMatching [3]	85.48 (+12.85)	55.11 (+12.98)
MixTeacher (Ours)	**85.85** (+13.22)	**56.25** (+14.12)

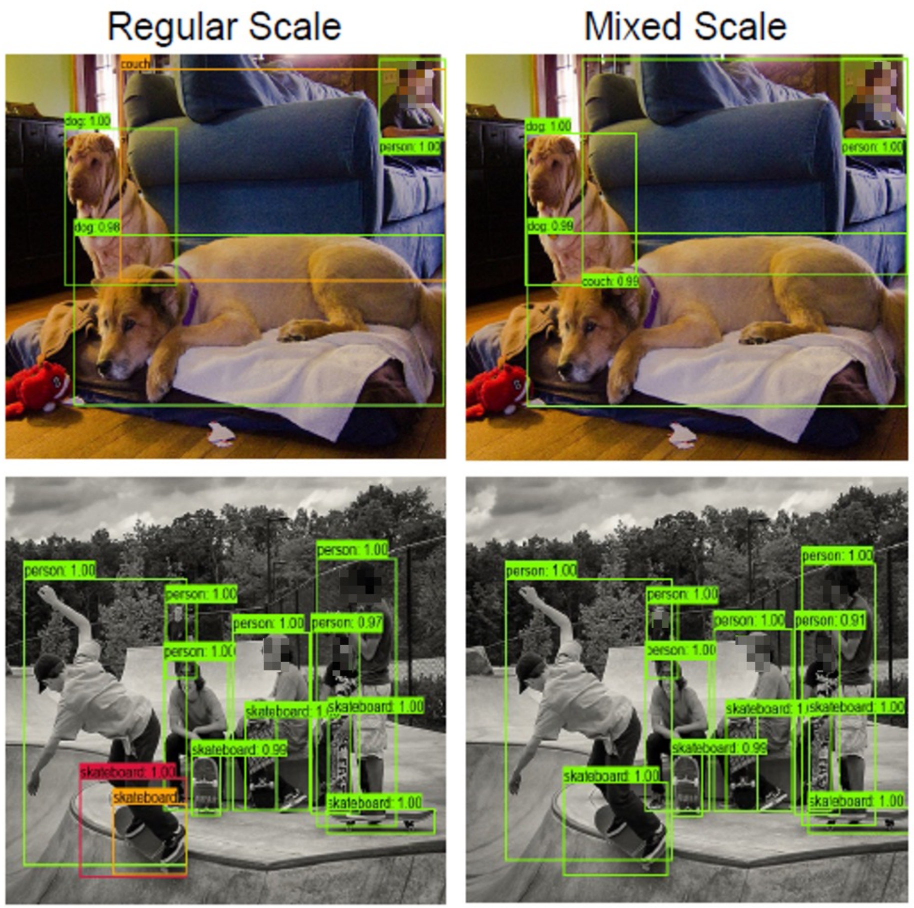
Table 3. Experimental results on the VOC Additional setting.

	AP_{50}	$AP_{50:95}$
Supervised [25]	72.63	42.13
STAC [35]	79.08 (+6.45)	46.01 (+3.88)
Humble Teacher [36]	81.29 (+8.66)	54.41 (+12.28)
Rethinking Pse [18]	79.60 (+6.79)	56.10 (+13.97)
LabelMatching† [3]	85.81 (+13.18)	55.50 (+13.37)
MixTeacher (Ours)	**86.58** (+13.95)	**56.83** (+14.70)

Table 4. Experimental results on the VOC Mixture setting.

Mixed Scale Teacher with FCOS

	1%	2%	5%	10%
Dense Teacher [46]	22.38	27.20	33.01	**37.13**
Unbiased Teacher v2 [26]	22.71	26.03	30.08	32.61
MixTeacher (Ours)	**23.83**	**27.88**	**33.42**	36.95



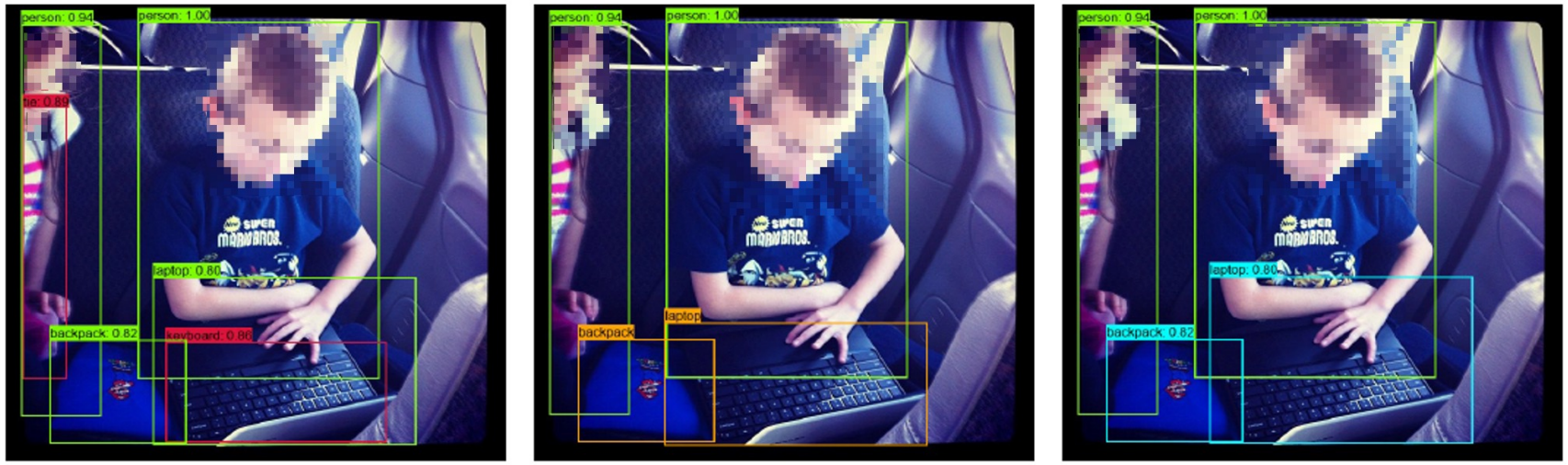
(a) Pseudo Labels Generation on Different Scales

Pseudo-labels from mixed scales are more accurate than regular scales

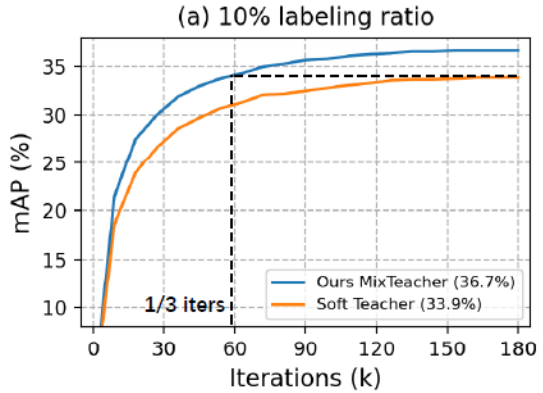
More Results of MixTeacher:

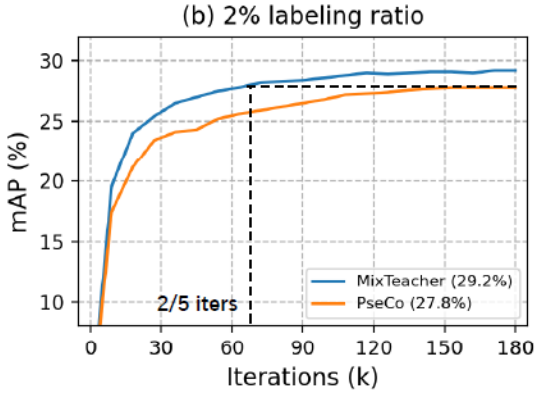
Tencent YouTu Lab

Threshold 0.7 | Threshold 0.9 | Threshold 0.7 + PLM



Pseudo Labels Visualization.
Green Boxes: TP, Red Boxes: FP, Orange Boxes: FP, Blue Boxes: Correct Mined Boxes





Achieves Baseline Performance with 1/3 Iterations

Achieves SOTA Performance with 2/5 Iterations

Tencent YouTu Lab

MixTeacher: Mining Promising Labels with Mixed Scale Teacher for Semi-Supervised Object Detection

Thanks

Poster: TUE-PM-310