JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Ambiguity-Resistant Semi-Supervised Learning for Dense Object Detection

Chang Liu[1], Weiming Zhang[2], Xiangru Lin[2], Wei Zhang[2], Xiao Tan[2], Junyu Han[2], Xiaomao Li[1,3], Errui Ding[2], Jingdong Wang[2]

[1] Shanghai University, [2] Baidu VIS, [3] Shanghai Artificial Intelligence Laboratory

Paper：https://arxiv.org/abs/2303.14960

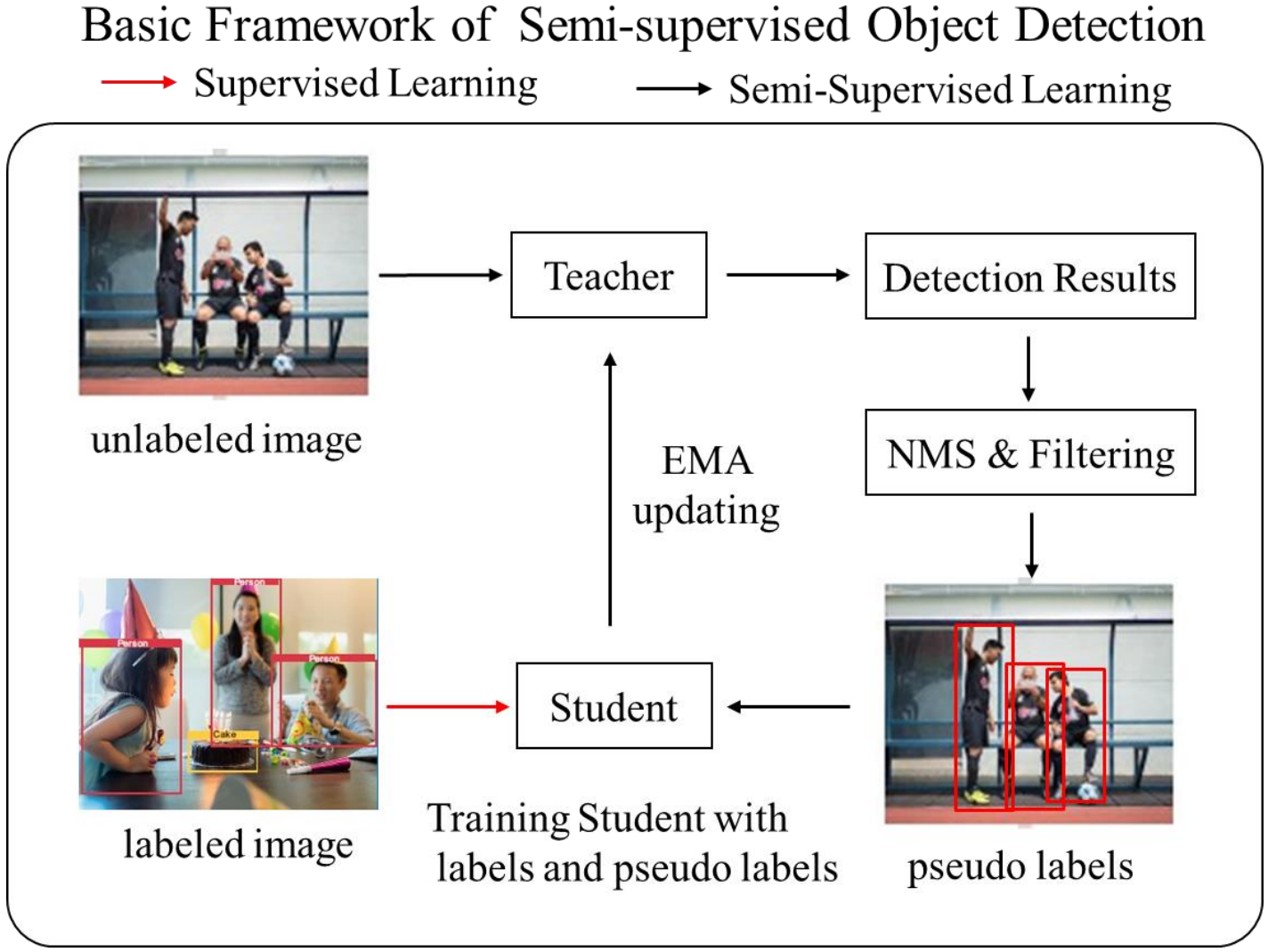
Code：https://github.com/PaddlePaddle/PaddleDetection

上海大学 Shanghai University

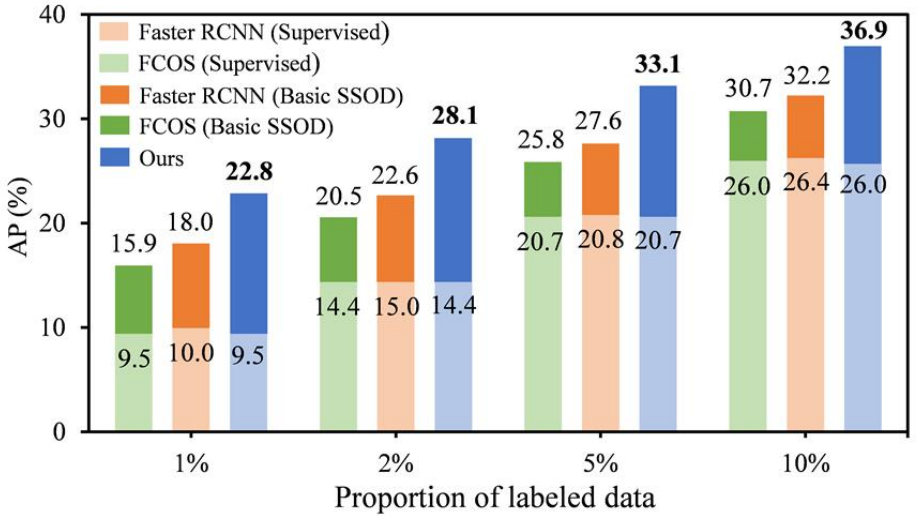
Baidu 百度

Motivation

Basic Framework of Semi-supervised Object Detection



SSOD Performance



Under this basic SSOD pipeline, FCOS achieves a relatively limited improvement compared with Faster RCNN.

The root lies in the selection and assignment ambiguity of pseudo lables.

Analysis: Selection Ambiguity

Selection ambiguity of pseudo labels：

The mismatch between classification scores and localization quality affects the selection of high-quality pseudo labels, suppressing the semi-supervised performance.

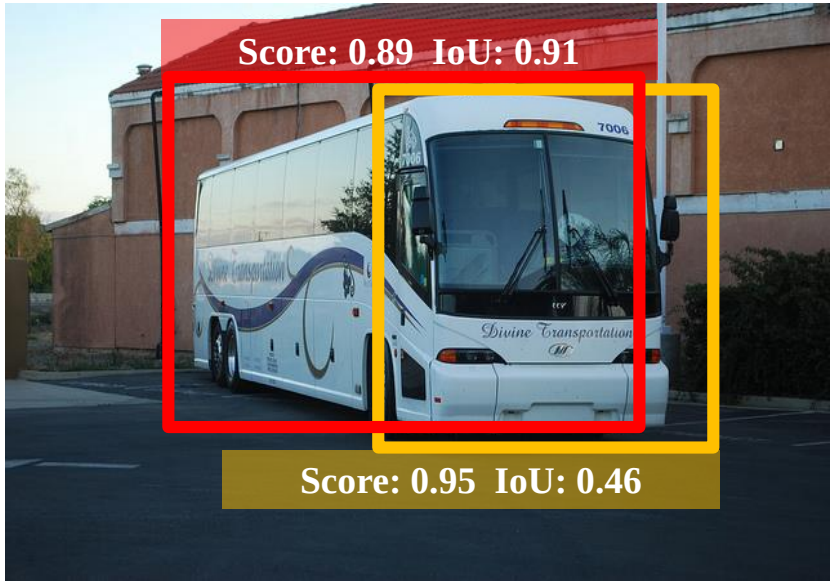


Table 1. Comparison on pseudo labels predicted by Faster RCNN and FCOS. 'vanilla FCOS' denotes the FCOS without the centerness branch. 'Top-5 IoU' represents the mean IoU of top-5 detection results based on classification scores in each image. 'PCC' represents the Pearson Correlation Coefficient between the normalized classification scores and localization quality.

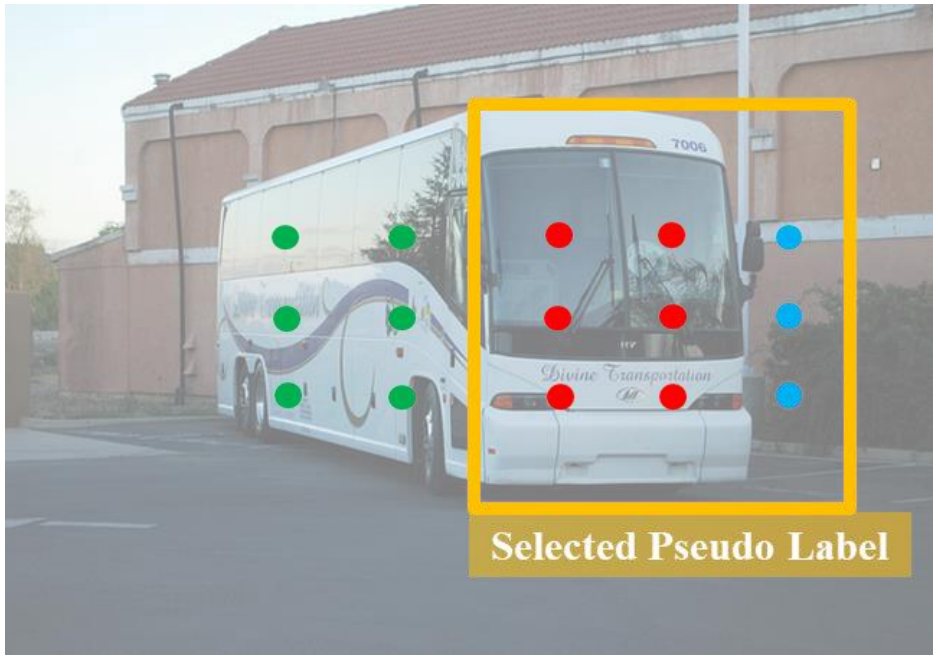
Method	AP	Mean IoU	Top-5 IoU	PCC
Faster RCNN	26.4	0.348	0.641	0.439
vanilla FCOS	25.2	0.369	0.585	0.235
FCOS	26.0	0.369	0.593	0.279

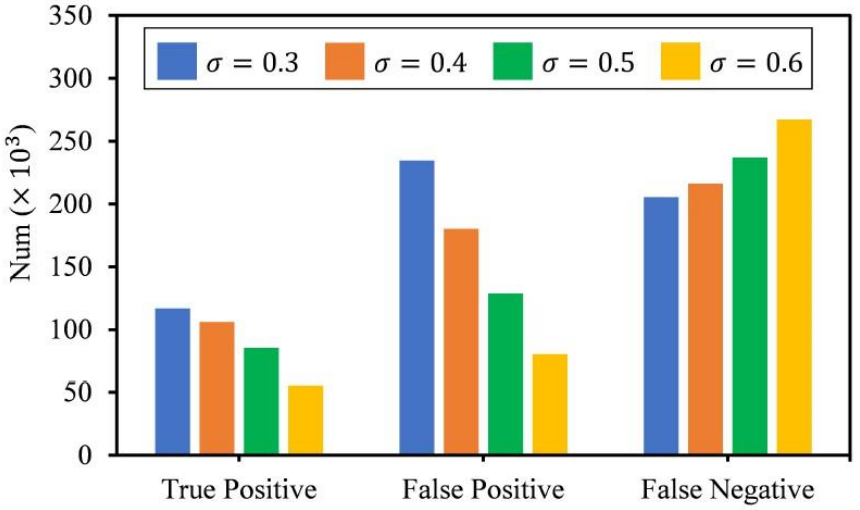
Analysis: Assignment Ambiguity

Assignment ambiguity of pseudo lables：

The box-based assignment is naturally not robust to inaccurate pseudo boxes and missed objects, generating many false negatives and false positives.

● True Positive ● False Negative ● False Positive





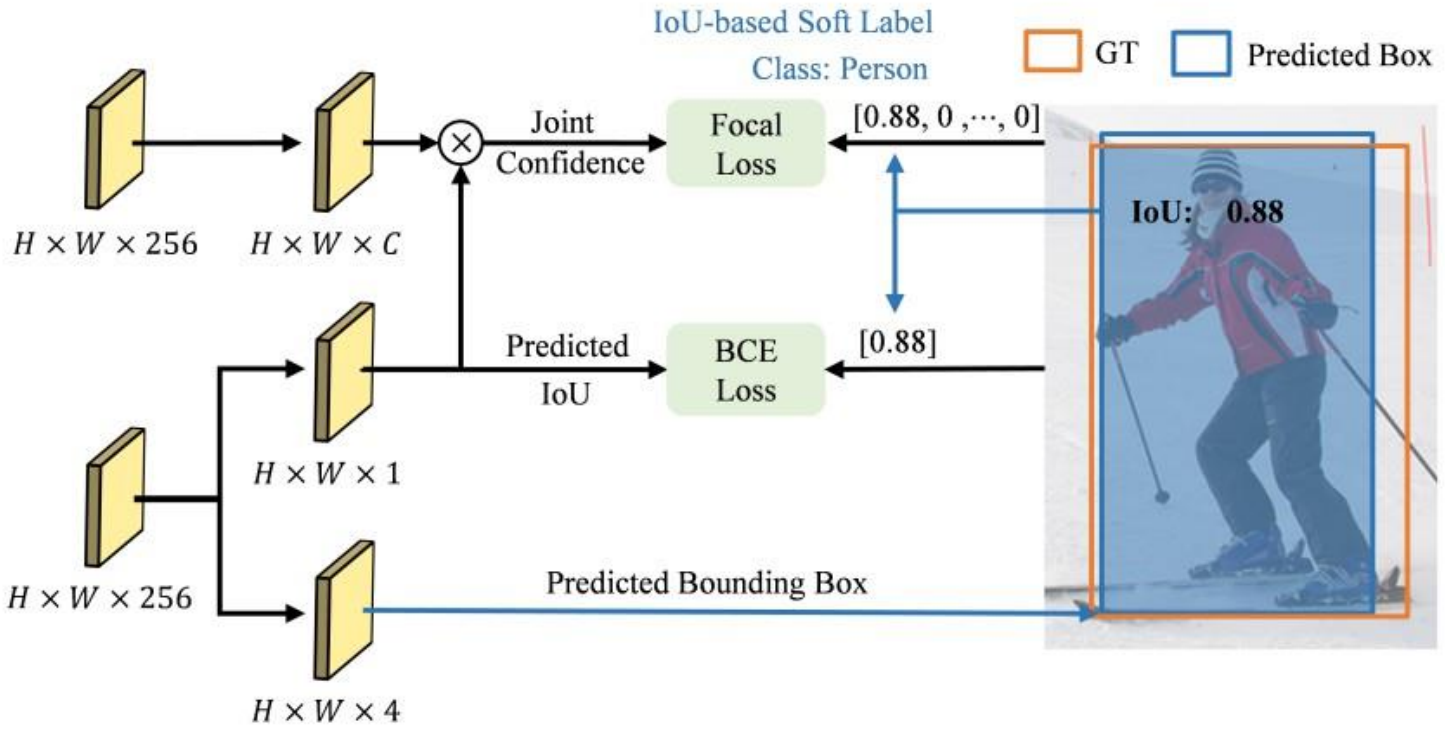
Thresh	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$	$\sigma = 0.6$
AP (%)	28.7	29.8	30.7	29.5

Figure 3. Investigation on the assignment ambiguity of FCOS under different filtering thresholds σ. The assignment results are obtained based on selected pseudo labels.

Method: Joint-Confidence Estimation

Core idea：

To mitigate the selection ambiguity， JCE aims to predict the joint confidence of the classification and localization for pseudo-label selection.



Joint Confidence $\hat{S}$:

$$\hat{S} = \hat{S}_{cls} * \hat{S}_{iou}$$

Learning Target S:

$$S = \begin{cases} \{0, \cdots, IoU, \cdots, 0\}, & \text{Labeled} \\ \{0, \cdots, Max(\hat{S}_t), \cdots, 0\}, & \text{Unlabeled} \end{cases}$$

Classification Loss L_{cls}:

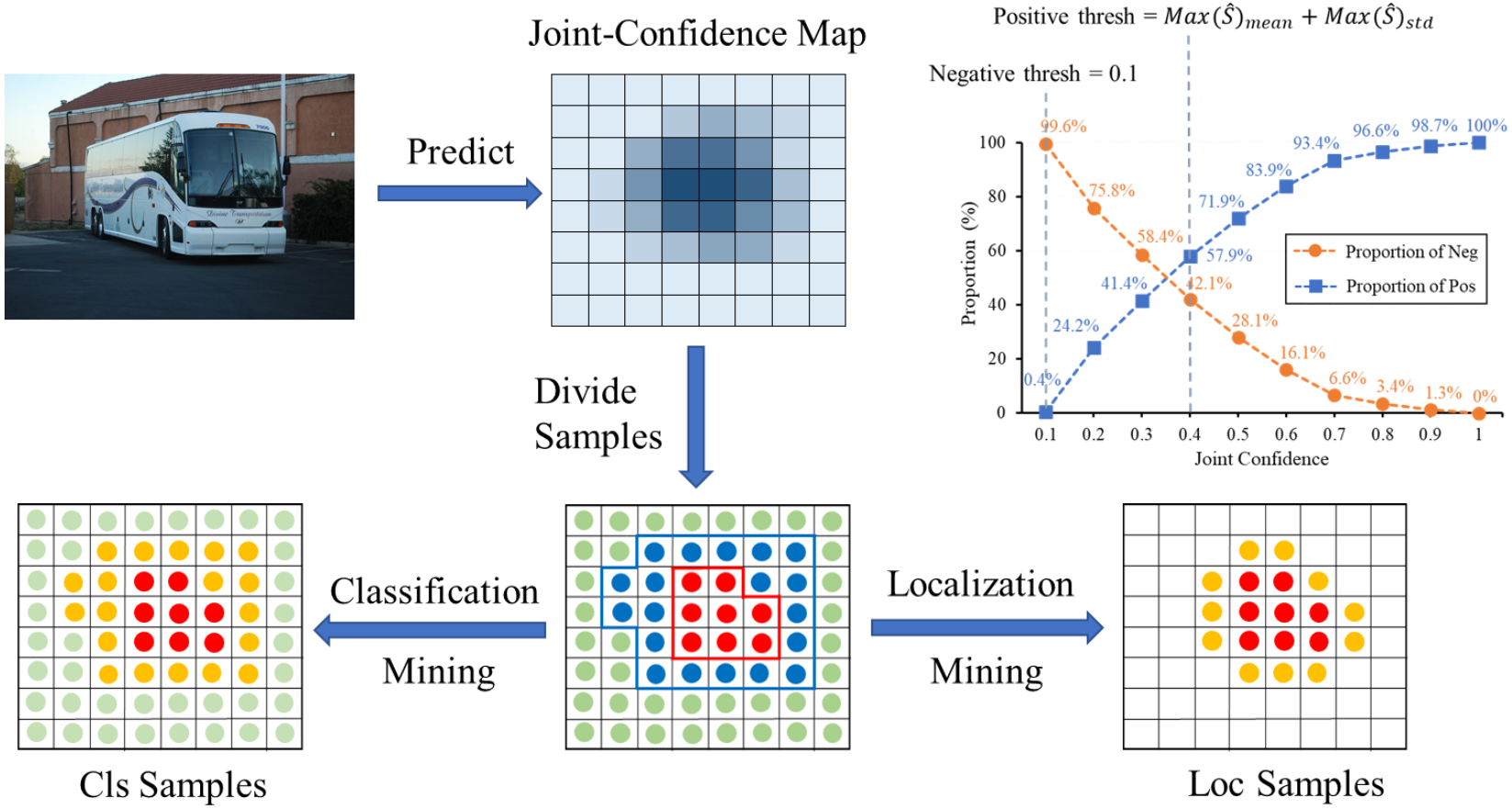
$$L_{cls} = FocalLoss(\hat{S}, S)$$

Method: Task-Separation Assignment

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Core idea：

To alleviate the assignment ambiguity, TSA assigns labels based on pixel-level predictions rather than unreliable pseudo boxes, and further exploits potential positives for the classification and localization task separately.



- Positives
- Negatives
- Ambiguous Candidates
- Potential Positives

Divide Samples:

Employ neg and pos threshes to divide samples into negatives, positives, and ambiguous candidates.

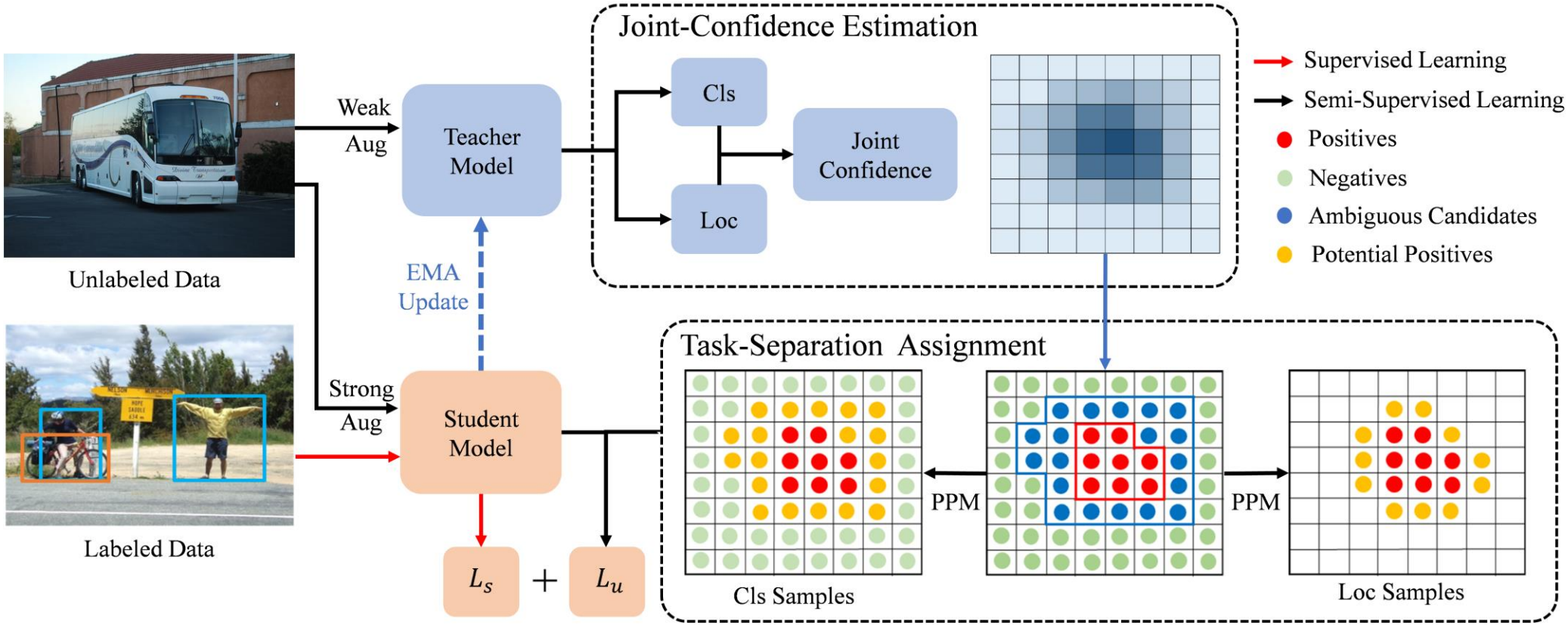
Classification Mining:

All candidates participate in the consistency learning to mimic the classification responses of the teacher.

Localization Mining:

Select potential positives according to their similarity with positives.

Method: Framework of ARSL



Experiments: Comparison with SOTA

Table 2. Experimental results on COCO-Standard. Two-stage detectors employ Faster RCNN as the baseline, while FCOS is used for one-stage detectors. * and † denotes the additional patch-shuffle and large scale jittering augmentation respectively.

Methods	Reference	COCO-Standard			
		1%	2%	5%	10%
Faster RCNN [23] (Supervised)	-	10.02 ± 0.38	15.04 ± 0.31	20.82 ± 0.13	26.44 ± 0.11
STAC [27]	arXiv20	13.97 ± 0.35	18.25 ± 0.25	24.38 ± 0.12	28.64 ± 0.21
ISMT [34]	CVPR21	18.88 ± 0.74	22.43 ± 0.56	26.37 ± 0.24	30.53 ± 0.52
Humble Teacher [28]	CVPR21	16.96 ± 0.38	21.72 ± 0.24	27.70 ± 0.15	31.61 ± 0.28
Unbiased Teacher [19]	ICLR21	20.75 ± 0.12	24.30 ± 0.07	28.27 ± 0.11	31.50 ± 0.10
Active Teacher [21]	CVPR22	22.20	24.99	30.07	32.58
Unbiased Teacher V2 [20]	CVPR22	21.84 ± 0.13	26.14 ± 0.01	30.06 ± 0.14	33.50 ± 0.03
Soft Teacher† [33]	ICCV21	20.46 ± 0.39	-	30.74 ± 0.08	34.04 ± 0.14
PseCo [13]	ECCV22	22.43 ± 0.36	27.77 ± 0.18	32.50 ± 0.08	36.06 ± 0.24
FCOS [30] (Supervised)	-	9.05 ± 0.31	14.40 ± 0.28	20.69 ± 0.22	26.01 ± 0.15
Unbiased Teacher V2 [20]	CVPR22	22.71 ± 0.42	26.03 ± 0.12	30.08 ± 0.04	32.61 ± 0.03
Dense Teacher [36]	ECCV22	19.64 ± 0.34	25.39 ± 0.13	30.83 ± 0.21	35.11 ± 0.13
DSL* [3]	CVPR22	22.03 ± 0.28	25.19 ± 0.37	30.87 ± 0.24	36.22 ± 0.18
ARSL (FCOS)	-	22.82 ± 0.26	28.11 ± 0.19	33.14 ± 0.12	36.90 ± 0.03
ARSL† (FCOS)	-	25.36 ± 0.32	29.08 ± 0.21	34.45 ± 0.16	38.50 ± 0.05
ARSL† (RetinaNet)	-	25.16 ± 0.25	28.68 ± 0.24	34.30 ± 0.21	38.42 ± 0.03

Experiments: Ablation Studies

Table 5. The impacts of components on detection performance. JCE, TSA indicate the proposed Joint-Confidence Estimation and Task-Separation Assignment.

Methods	AP	AP_{50}	AP_{75}
FCOS (Supervised)	26.0	43.6	26.7
FCOS (Semi-Supervised)	30.7	47.1	32.4
+ JCE	34.7	52.4	37.3
+ TSA (w/o mining)	35.6	54.3	38.1
+ TSA (w/ mining)	36.9	55.4	39.6

Table 6. Ablation studies on Joint-Confidence Learning. 'United Supervision' indicates the joint training of the IoU-prediction and classification task. 'Specific targets' denotes that the classification targets of unlabeled data is set as max responses of the teacher.

Strategies of JCE	AP
baseline	30.7
+ IoU prediction	32.0(+1.3)
+ United supervision	34.2(+2.2)
+ Specific targets for unlabeled data	34.7(+0.5)

Experiments: Ambiguity Mitigation

Table 8. Selection Ambiguity Mitigation. 'T-Head' denotes the task-aligned head in TOOD and QFL is the quality focal loss in GFL. The metrics follow the settings presented in Sec. 3.2. The statistics are calculated by the final model of 10% split on validation set.

Methods	Top-5 IoU	PCC	AP
FCOS	0.614	0.299	30.7
FCOS w/ T-head [5]	0.632	0.361	31.9
FCOS w/ QFL [15]	0.628	0.353	32.3
FCOS w/ JCE	0.656	0.395	34.7

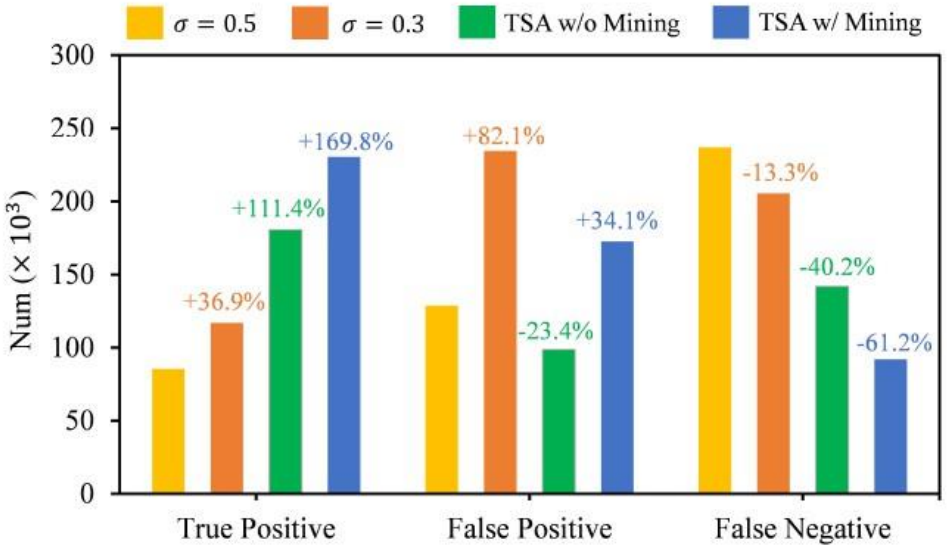


Figure 6. Mitigation of Assignment Ambiguity. σ indicates the filtering threshold of pseudo boxes. The statistics are counted on the COCO validation set.

Thanks!