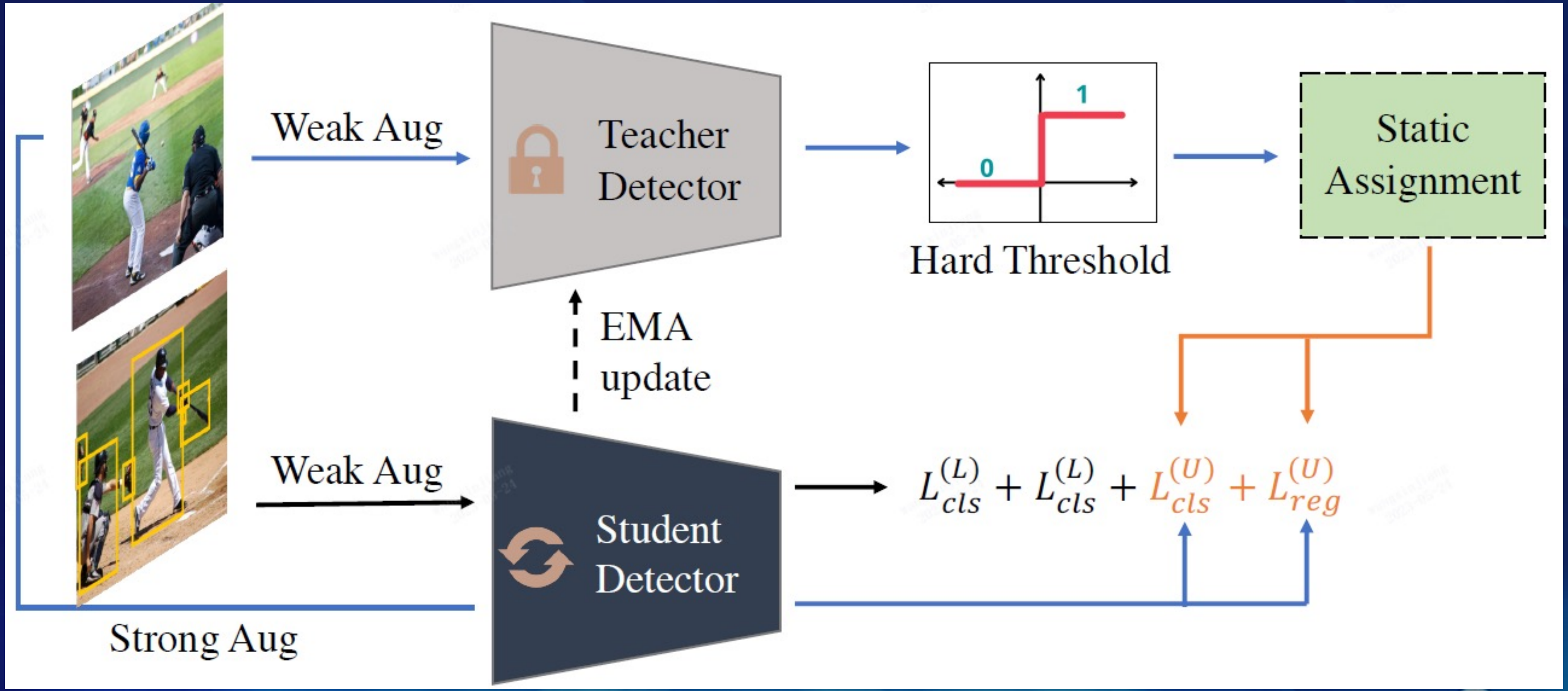
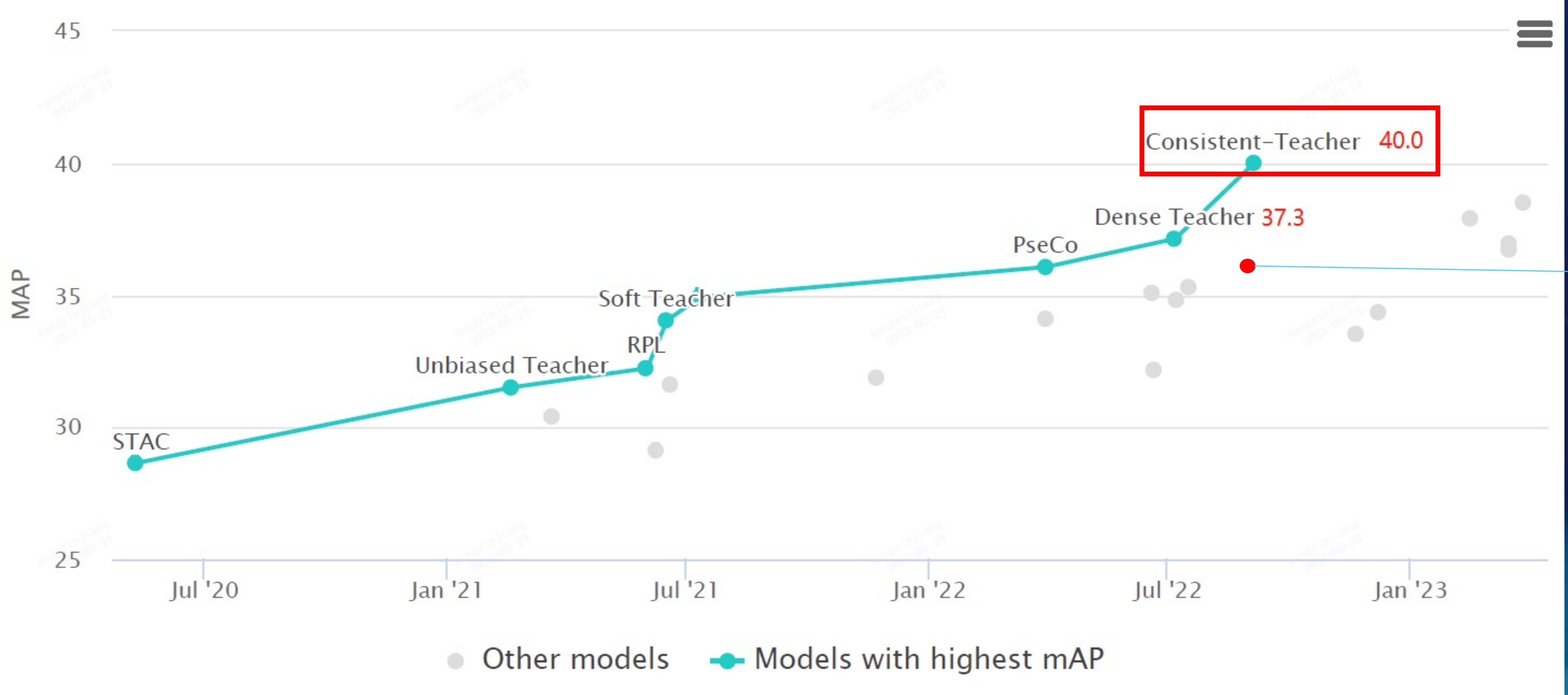
Consistent-Teacher: Towards Reducing Inconsistent Pseudo-targets in Semi-supervised Object Detection

Background: Traditional Semi-Supervised Detector Pipeline



Almost the same with classification

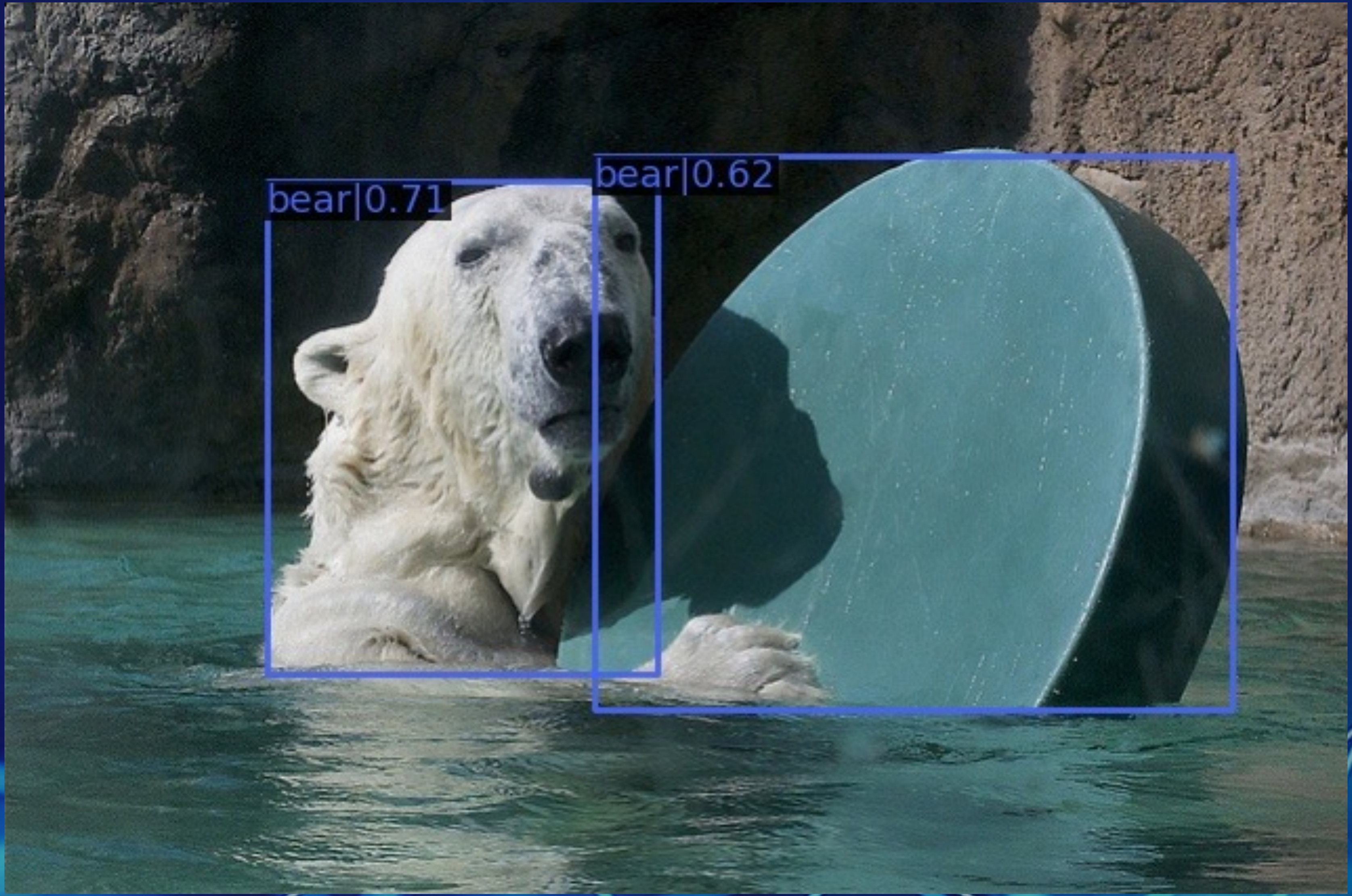
ConsistentTeacher: A SOTA Semi-Supervised Detector



Our Mean-Teacher Baseline

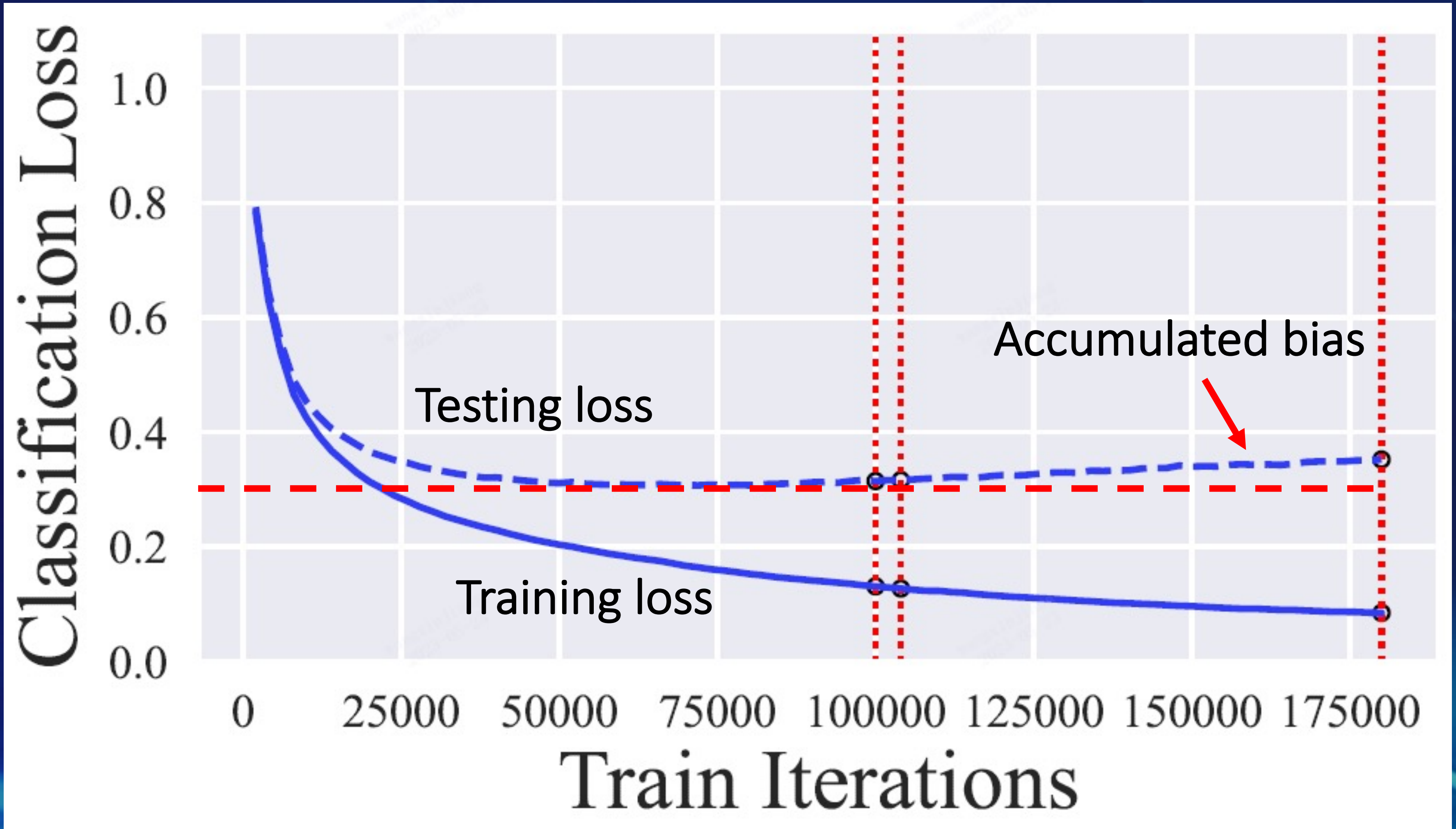
https://paperswithcode.com/sota/semi-supervised-object-detection-on-coco-10

Motivation: Pseudo-label evolvement in traditional SSOD



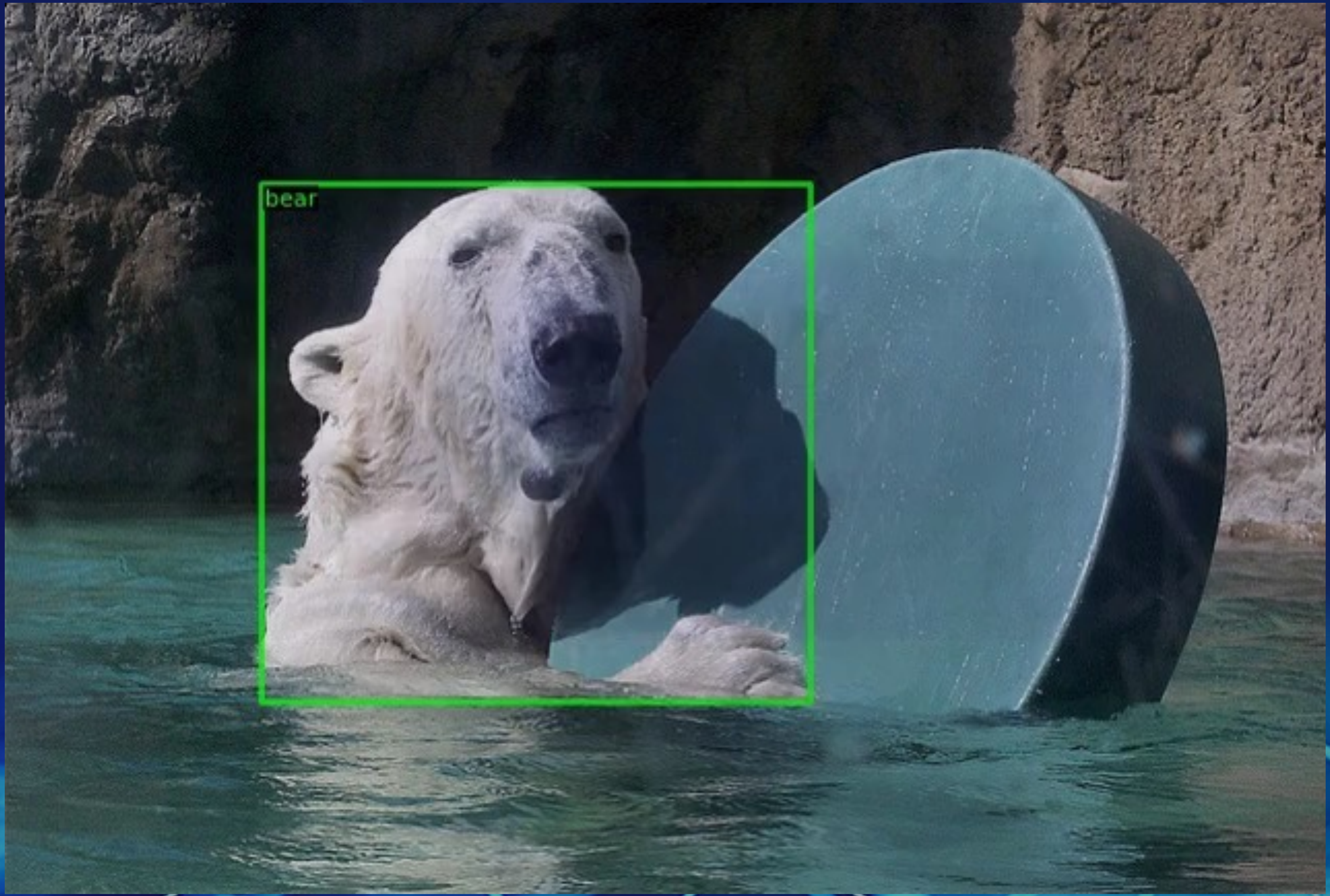
The metal board is surprisingly predicted as a bear by a traditional semi-supervised object detector

Motivation: Pseudo-label drifting in traditional SSOD

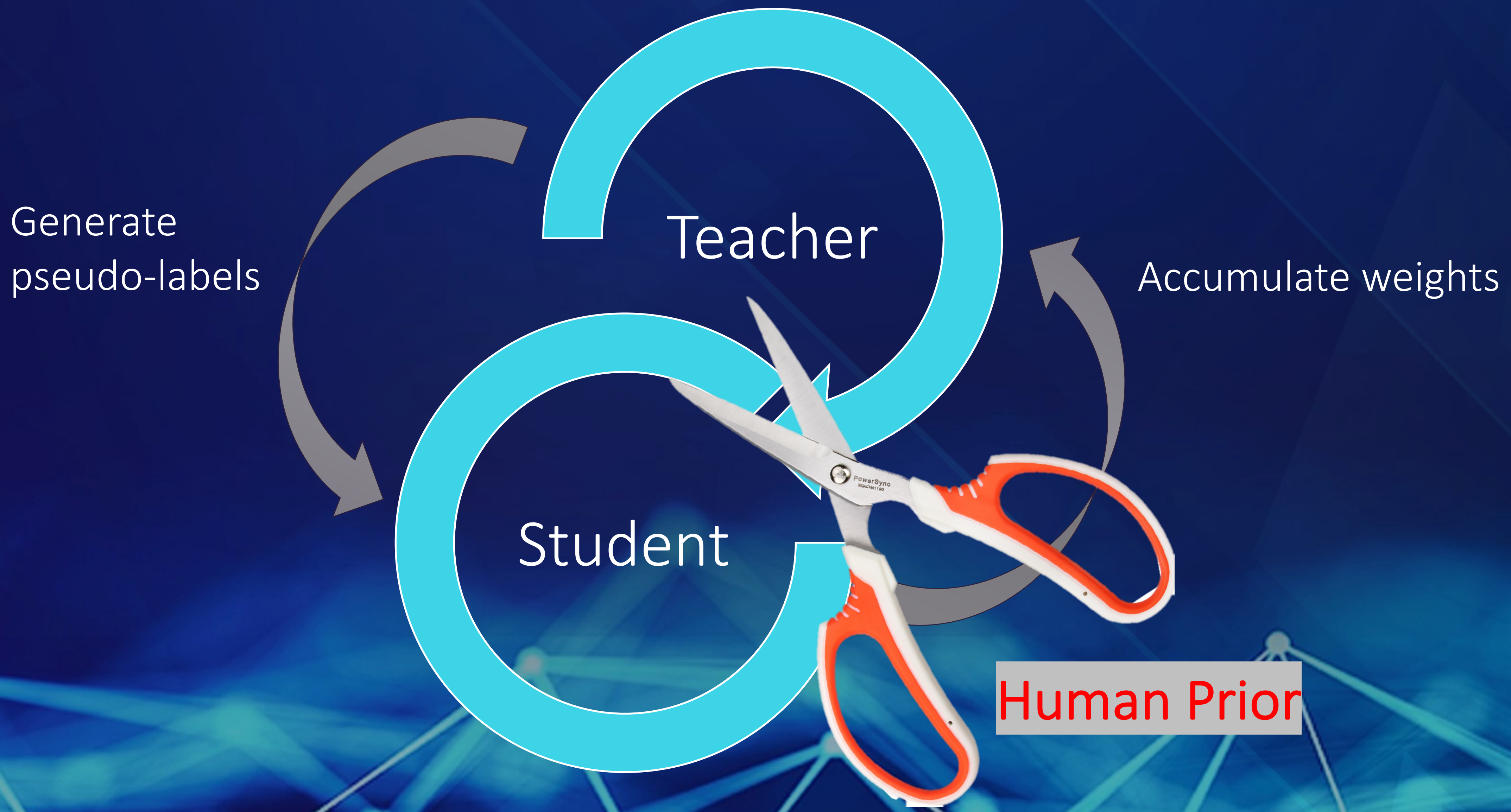


Training and testing loss on unlabeled images

Pseudo-label evolvement in traditional SSOD



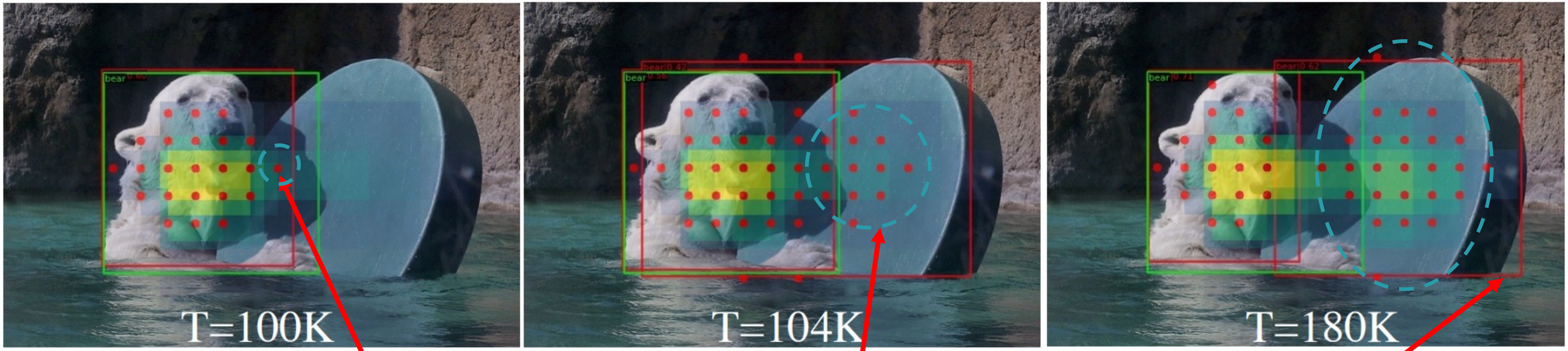
Reasons of pseudo-label drifting



Pseudo-label drifting in traditional SSOD

Pseudo bounding boxes

Ground truth bounding boxes

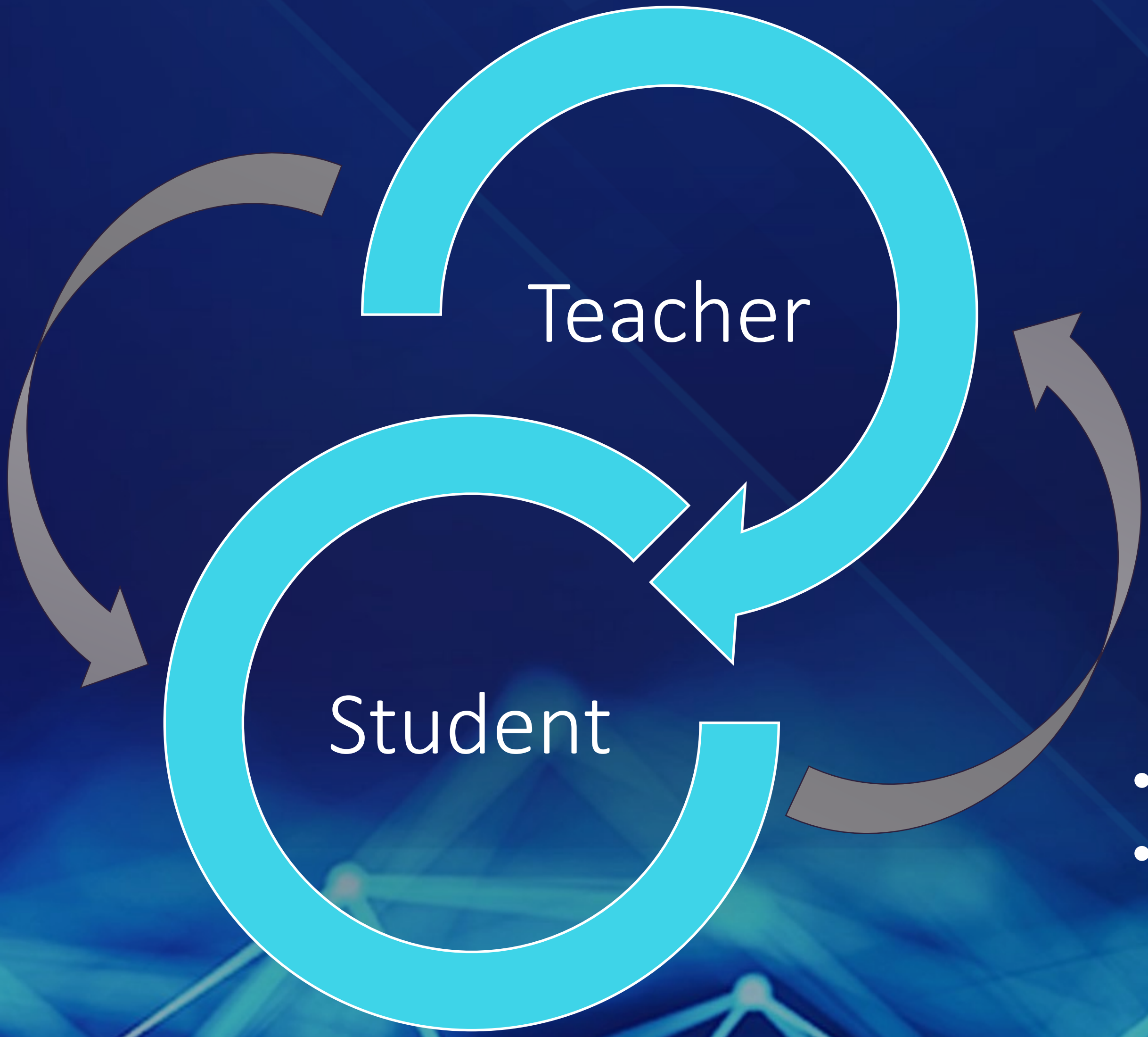


Suspicious assignment

Incorrect assignment

Biased prediction

Adaptive pseudo-label assignment



- Man thinks; God laughs.
- 治在无为

Adaptive pseudo-label assignment

Total Loss Function

$$\mathcal{L} = \frac{1}{N}\sum_i \left[\mathcal{L}_{cls}\left(f_s(T(\mathbf{x}_i^l)), \mathbf{y}_i^l\right) + \mathcal{L}_{reg}\left(f_s(T(\mathbf{x}_i^l)), \mathbf{y}_i^l\right)\right] + \lambda_u \frac{1}{M}\sum_j \left[\mathcal{L}_{cls}\left(f_s(T'(\mathbf{x}_j^u)), \hat{\mathbf{y}}_j^u\right) + \mathcal{L}_{reg}\left(f_s(T'(\mathbf{x}_j^u)), \hat{\mathbf{y}}_j^u\right)\right]$$

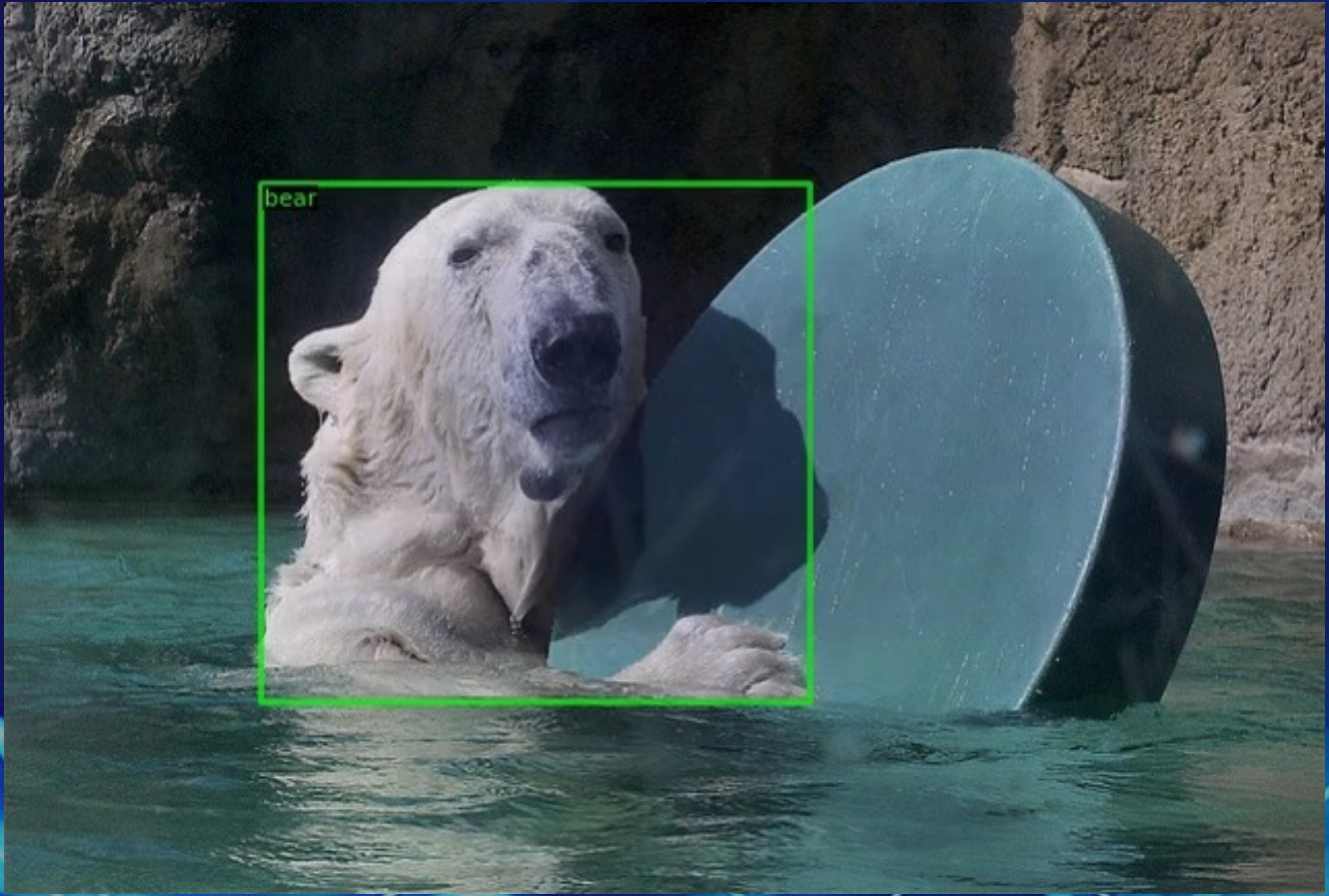
Adaptive assignment

$$\hat{c} = \operatorname*{argmin}_c \mathcal{L}(f_t(\mathbf{x}^u), c),$$

Adaptive assignment

$$\min_{a_1,\cdots,a_N} \sum_n^N \left[\mathcal{L}_{cls}\left(f_s(\mathbf{x}^u)_n, \hat{\mathbf{y}}_{a_n}^u\right) + \mathcal{L}_{reg}\left(f_s(\mathbf{x}^u)_n, \hat{\mathbf{y}}_{a_n}^u\right)\right]$$

Adaptive pseudo-label assignment



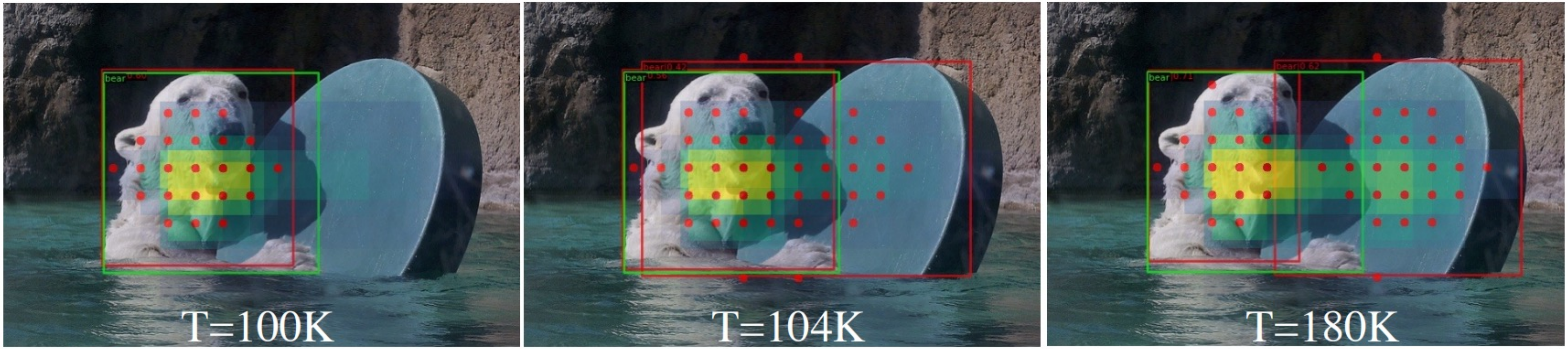
Assignment	$AP^{1\times}_{50:95}$	$AP^{10\%}_{50:95}$
IoU-based	38.4	35.50
our ASA	40.1 (+1.7)	38.50 (+3.0)

- Large improvement over the baseline
- Twice as much gain as in supervised learning

Pseudo-label drifting in traditional SSOD

Pseudo bounding boxes

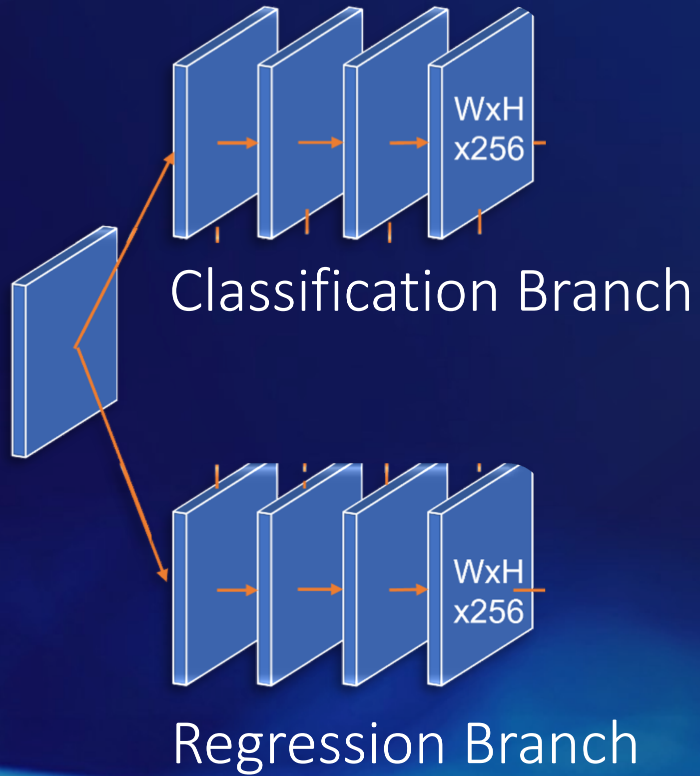
Ground truth bounding boxes

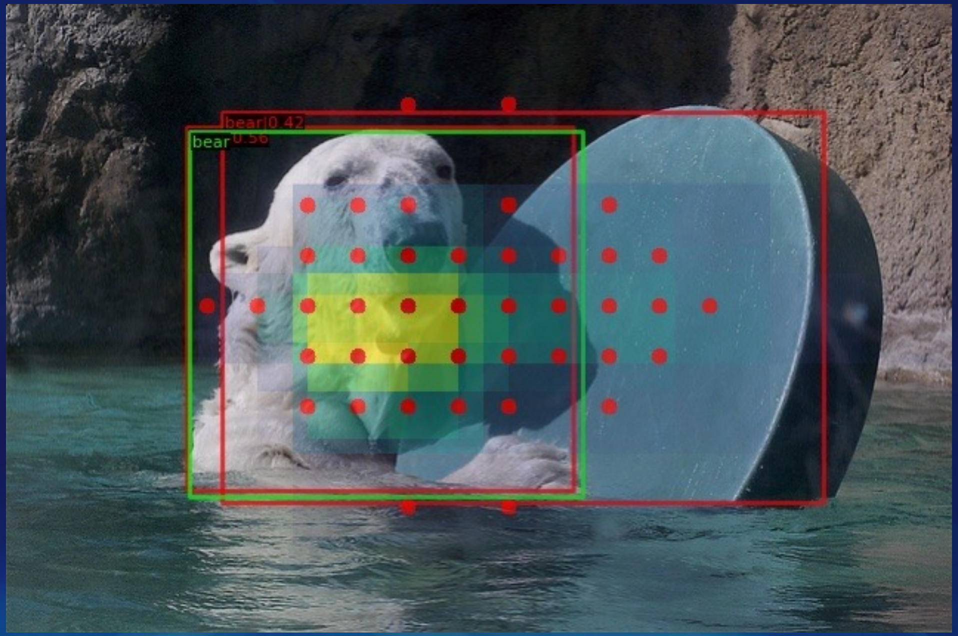


Strong Oscillation

Accumulated Bias

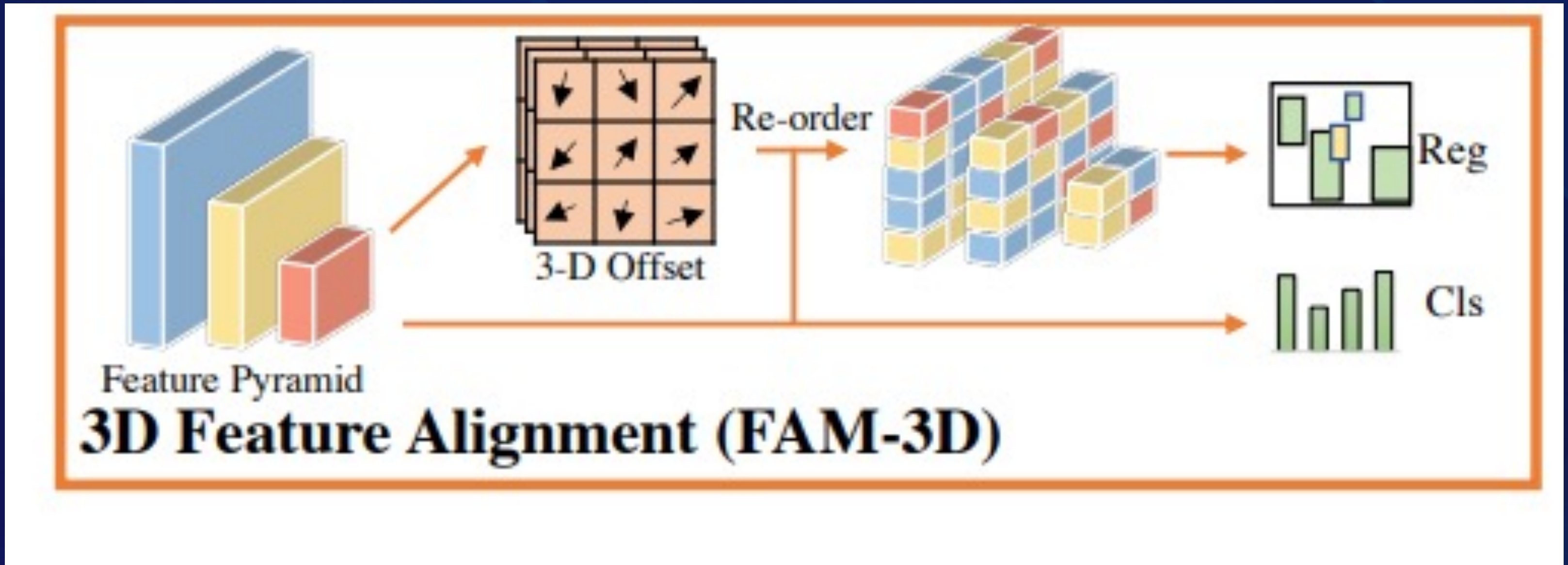
Task inconsistency in SSOD





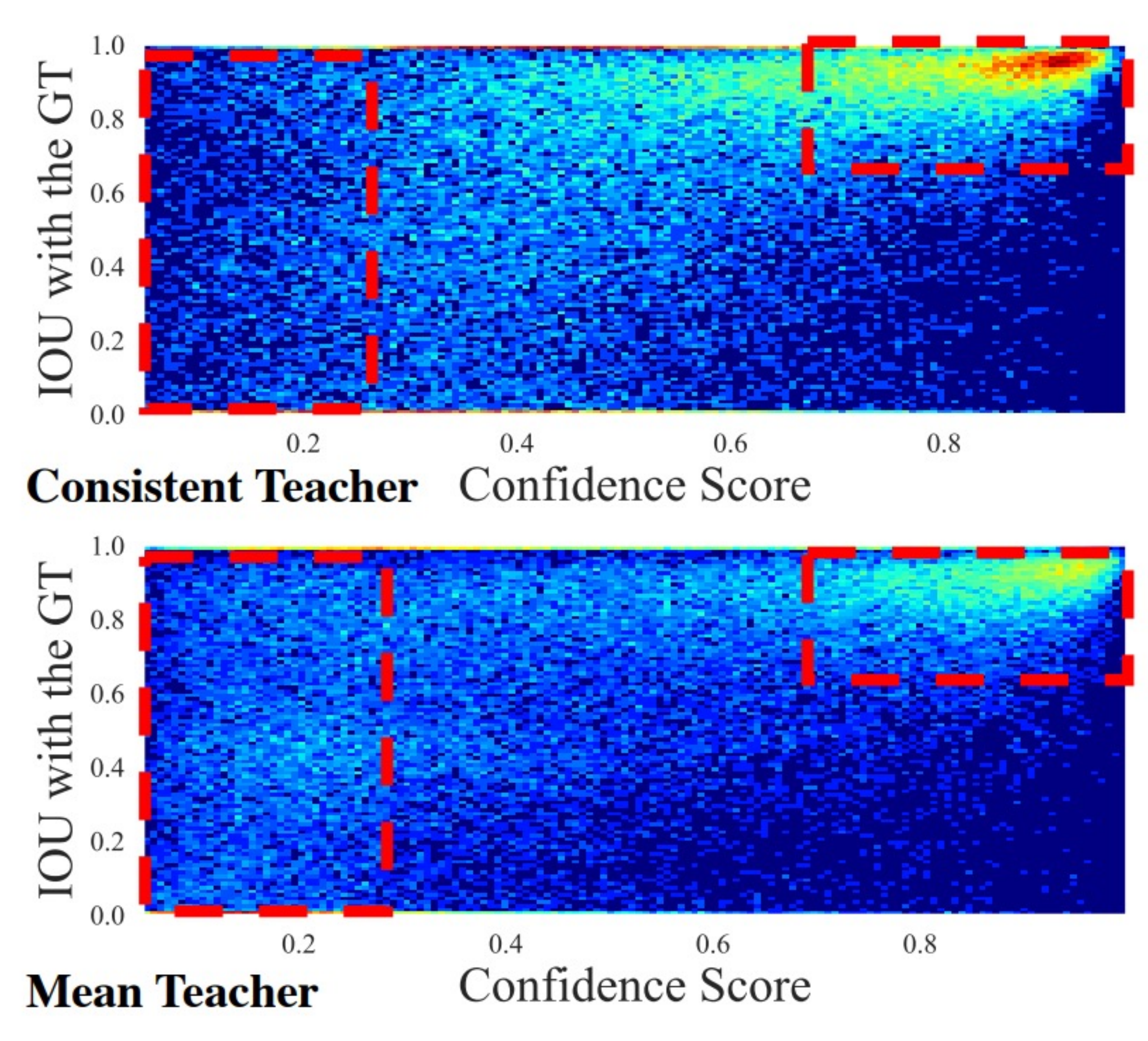
- Similar classification score
- Diverse bbox boundaries
- Independent classification & regression tasks

3D Feature Alignment



- Align Regression and Classification tasks
- Implicit solution of a pyramidal Rubik's Cube
- End-to-end optimization

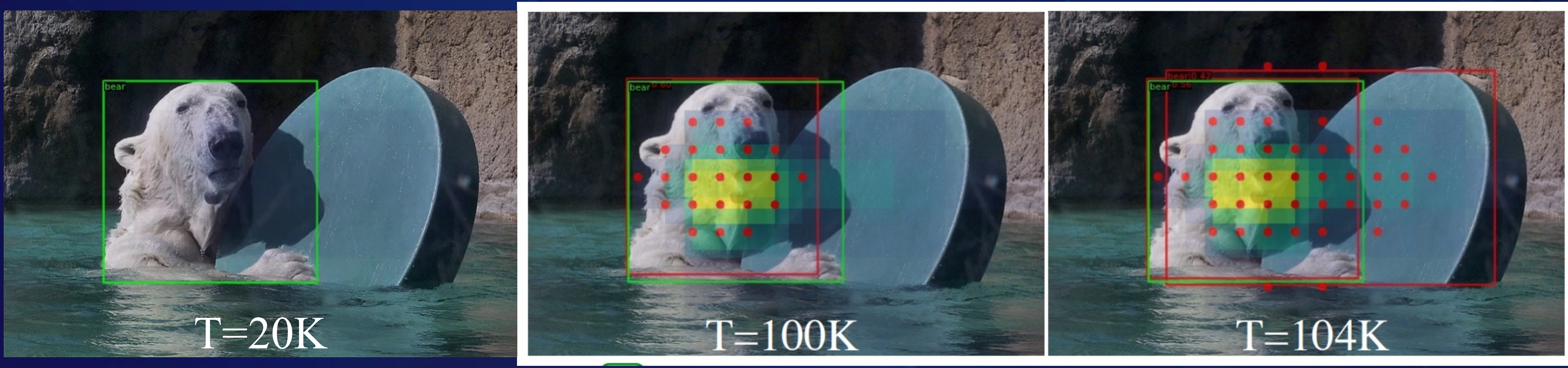
More aligned Cls-Reg tasks in SSOD



Method	FLOPs (G)	$AP^{1\times}_{50:95}$	$AP^{10\%}_{50:95}$
Ours w/o FAM	205.21	40.1	38.5
Ours w FAM-2D	205.70	40.4 (+0.3)	39.1 (+0.6)
Ours w FAM-3D	208.49	40.7 (+0.6)	39.5 (+1.0)

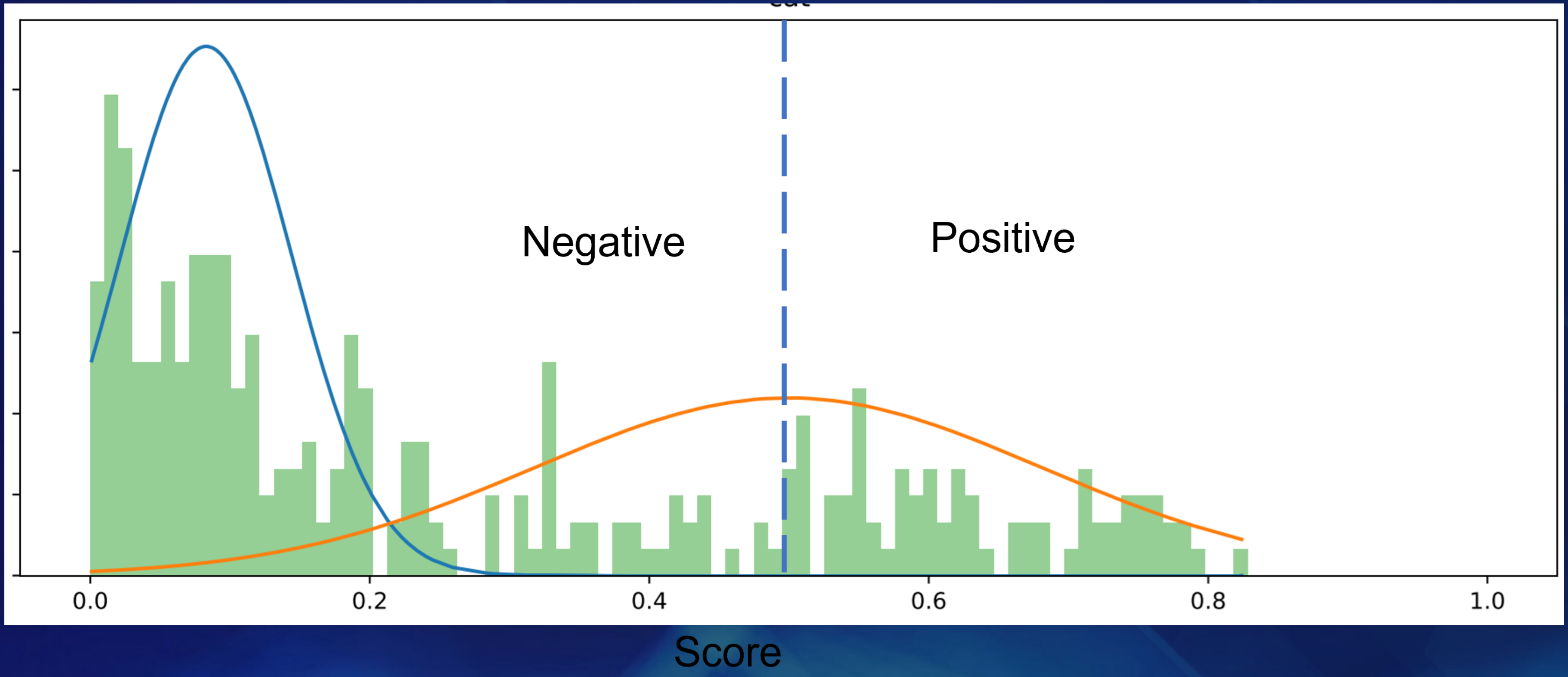
- More aligned cls & reg tasks
- Further improvement over FAM-2D
- Twice as much gain as in supervised learning

Score threshold inconsistency



- Temporally inconsistent target
- Sensitive to noise
- Underfit at first, overfit at last

Pseudo-label score distribution as GMM



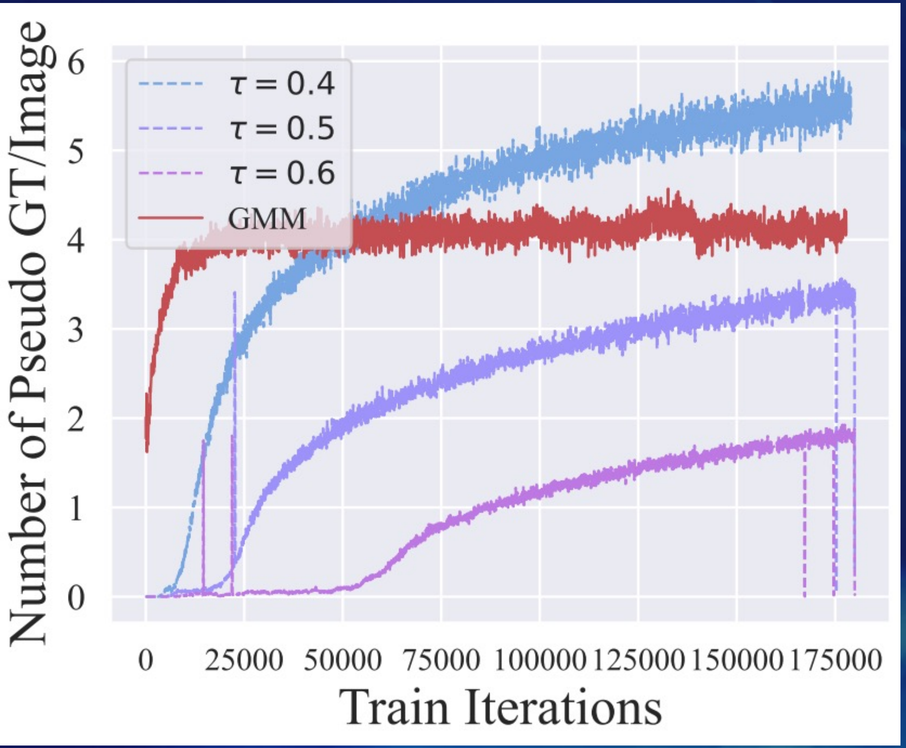
Score distribution as GMM

$$\mathcal{P}(s^c) = w_n^c \mathcal{N}(s^c | \mu_n^c, (\sigma_n^c)^2) + w_p^c \mathcal{N}(s^c | \mu_p^c, (\sigma_p^c)^2)$$

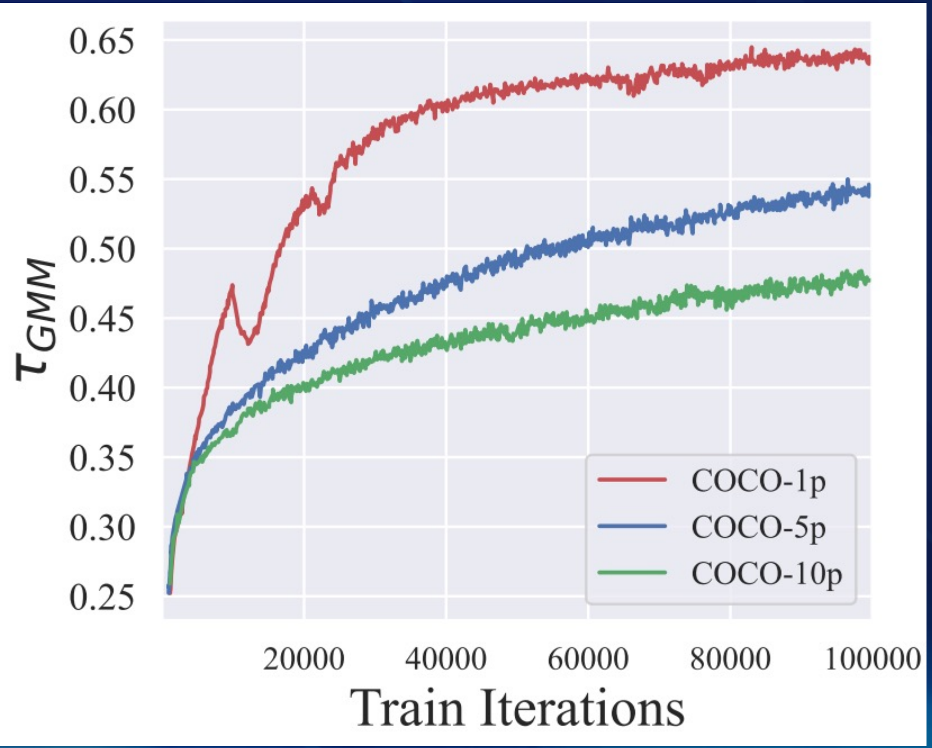
Dynamic score threshold

$$\tau^c = \underset{s^c}{\operatorname{argmax}} \, \mathcal{P}(pos | s^c, \mu_p^c, (\sigma_p^c)^2)$$

Number of pseudo labels in semi-supervised object detection

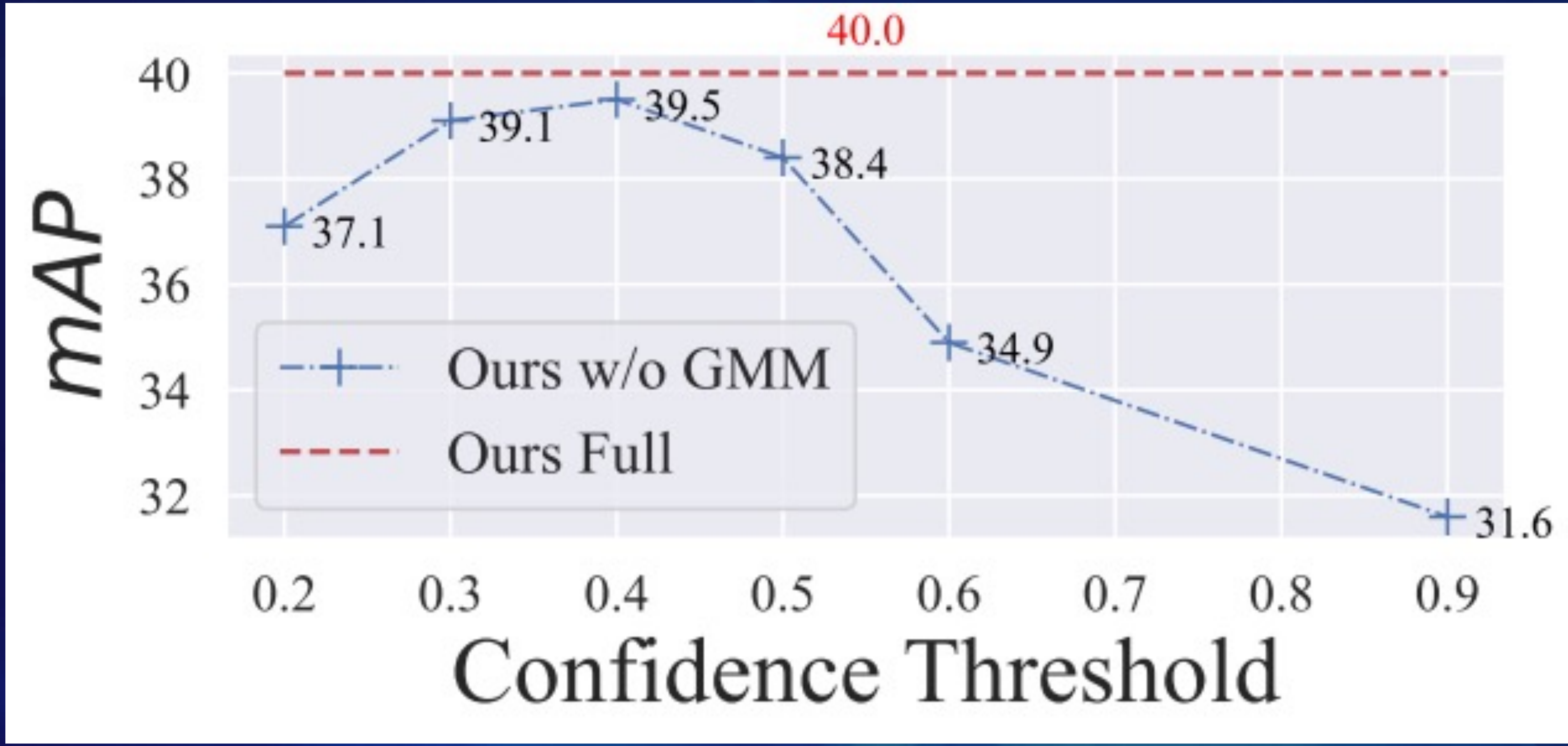


Number of pseudo labels



Score threshold dynamics

Performance of GMM



- Free from hyper-parameter tuning
- Increase over best hard threshold

State-of-the-art comparison on partial COCO

Method	1% COCO	2% COCO	5% COCO	10% COCO
Labeled Only	9.05	12.70	18.47	23.86
CSD	10.51	13.93	18.63	22.46
STAC	13.97	18.25	24.38	28.64
Instant Teaching	18.05	22.45	26.75	30.40
Humble teacher	16.96	21.72	27.70	31.61
Unbiased Teacher	20.75	24.30	28.27	31.50
Soft Teacher	20.46	-	30.74	34.04
ACRST	<u>26.07</u>	<u>28.69</u>	31.35	34.92
PseCo	22.43	27.77	32.50	36.06
Labeled Only	10.22	13.80	19.40	24.10
Unbiased Teacher v2	22.71	26.03	30.08	32.61
DSL	22.03	25.19	30.87	36.22
Dense Teacher	22.38	27.20	<u>33.01</u>	<u>37.13</u>
S4OD	20.10	-	30.00	32.90
Mean-Teacher	20.40	26.00	30.40	35.50
`Consistent-Teacher`	**25.30**	**30.40**	**36.10**	**40.00**

State-of-the-art comparison on additional COCO

Method	$AP_{50:95}$
CSD(3×)	$40.20 \xrightarrow{-1.38} 38.82$
STAC(6×)	$39.48 \xrightarrow{-0.27} 39.21$
Unbiased Teacher(3×)	$40.20 \xrightarrow{+1.10} 41.30$
ACRST(3×)	$40.20 \xrightarrow{+2.59} 42.79$
Soft Teacher(16×)	$40.90 \xrightarrow{+3.70} 44.50$
DSL(2×)	$40.20 \xrightarrow{+3.60} 43.80$
PseCo(8×)	$41.00 \xrightarrow{+5.10} 46.10$
Dense Teacher(8×)	$41.24 \xrightarrow{+4.88} 46.12$
Consistent-Teacher (8×)	$40.50 \xrightarrow{\mathbf{+7.20}} \mathbf{47.70}$

State-of-the-art comparison on partial VOC

Method	AP_{50}	$AP_{50:95}$
Labeled Only	72.63	42.13
CSD	74.70	-
STAC	77.45	44.64
ACRST	78.16	50.12
Instant Teaching	79.20	50.00
Humble Teacher	80.94	53.04
Unbiased Teacher	77.37	48.69
Unbiased Teacher v2	<u>81.29</u>	<u>56.87</u>
Mean-Teacher	77.02	53.61
`Consistent-Teacher`	**81.00**	**59.00**