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

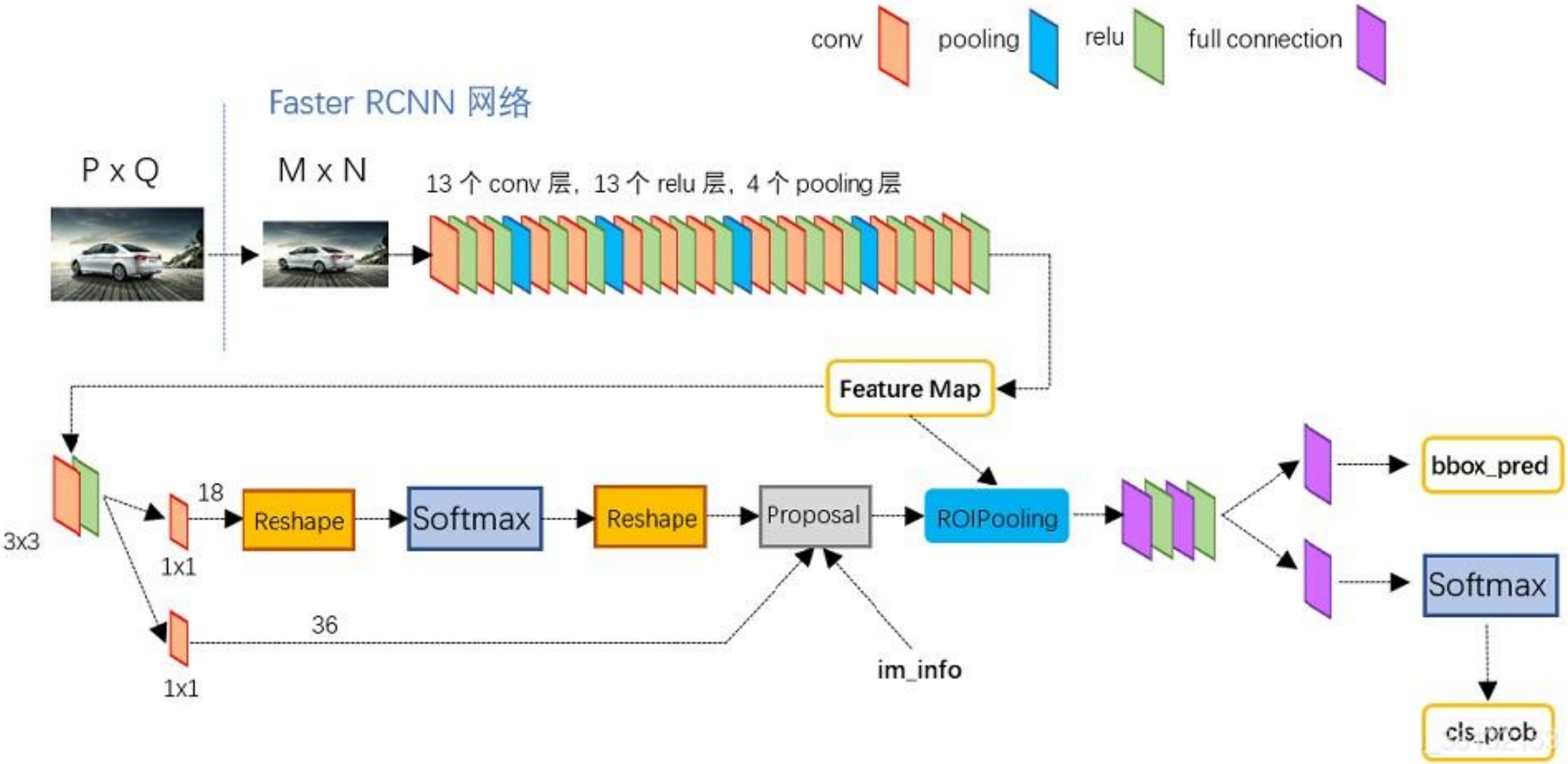
Humble Teachers Teach Better Students for Semi-Supervised Object Detection

Yihe Tang [†,*] Weifeng Chen [‡] Yijun Luo [‡] Yuting Zhang [‡]
[†] Carnegie Mellon University, [‡] Amazon Web Services
tangacademic@gmail.com {weifec,yijunl,yutingzh}@amazon.com

CVPR 2021

Contents

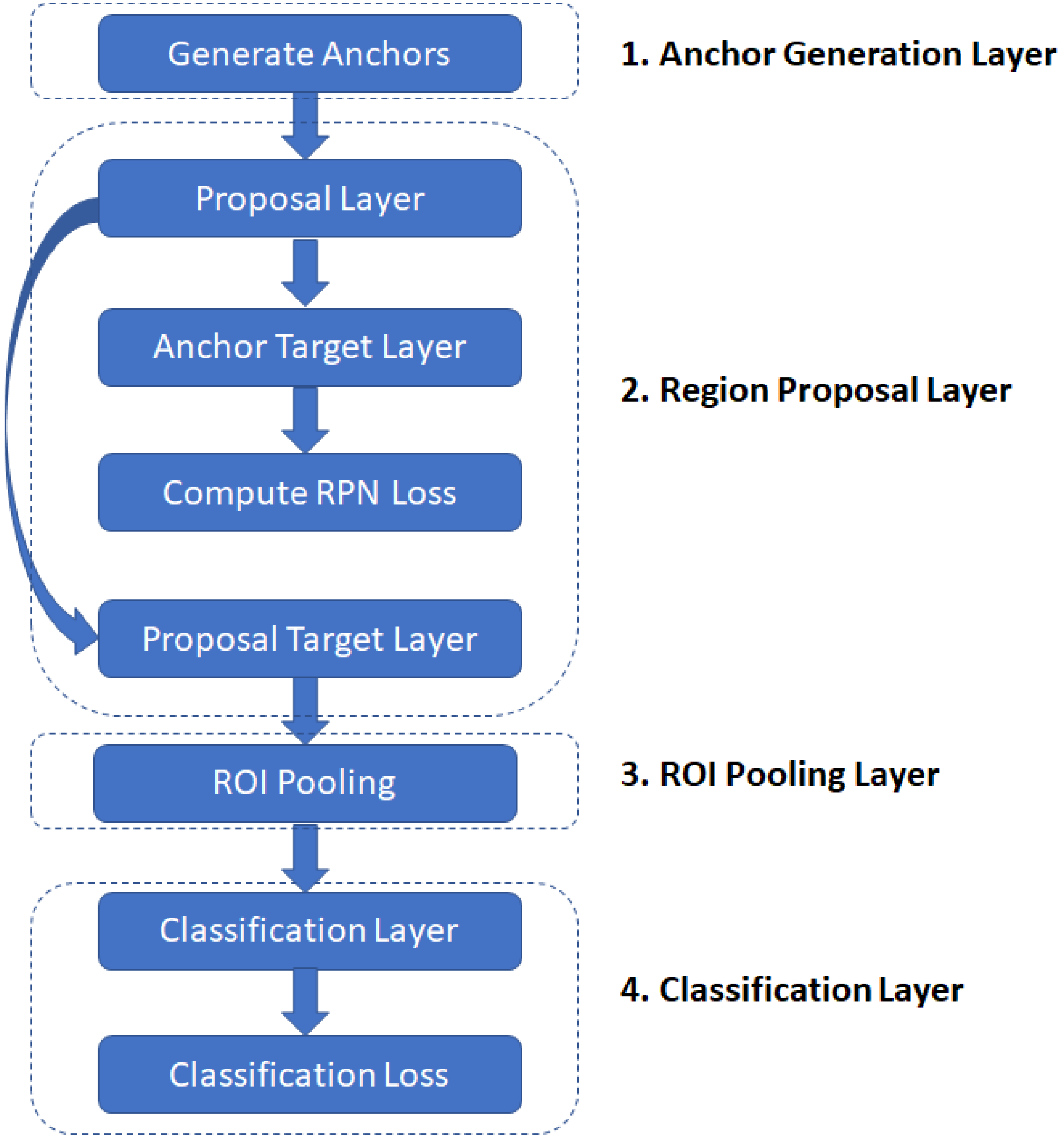
Two-stage object detectors



RPN (region proposal network)

ROI (region of interest)

Two-stage object detectors



RPN (region proposal network):
为第二阶段提供高质量的目标候选框

ROI (region of interest):
在rpn提供的proposal的基础上，筛选出第二阶段的训练样本，并提取相应的特征，用于组建第二阶段的训练网络

Dual Model Framework

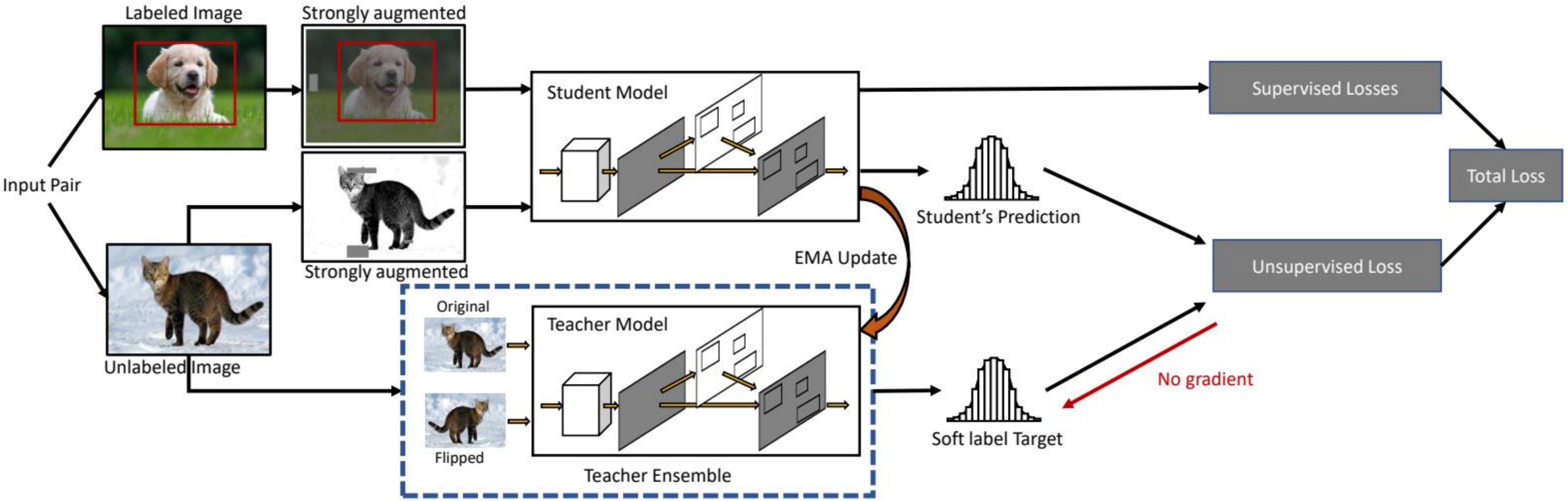


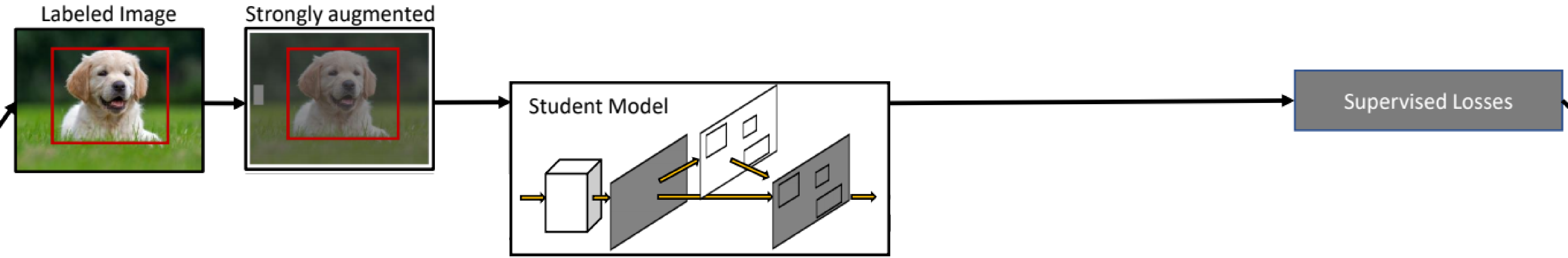
Figure 2: An overview of our Humble Teacher approach. The teacher model produces soft pseudo-labels for the student to learn from, and is updated via exponential moving average (EMA).

Overview & Supervised Branch

Semi-supervised: $$L = L_S + \frac{n_U}{n_S} \beta L_U$$

where n_U, n_S are the numbers of unlabeled and labeled images, and β is set to 0.5 by default.

Supervised loss: {classification loss, localization loss} x {RPN head, ROI head}

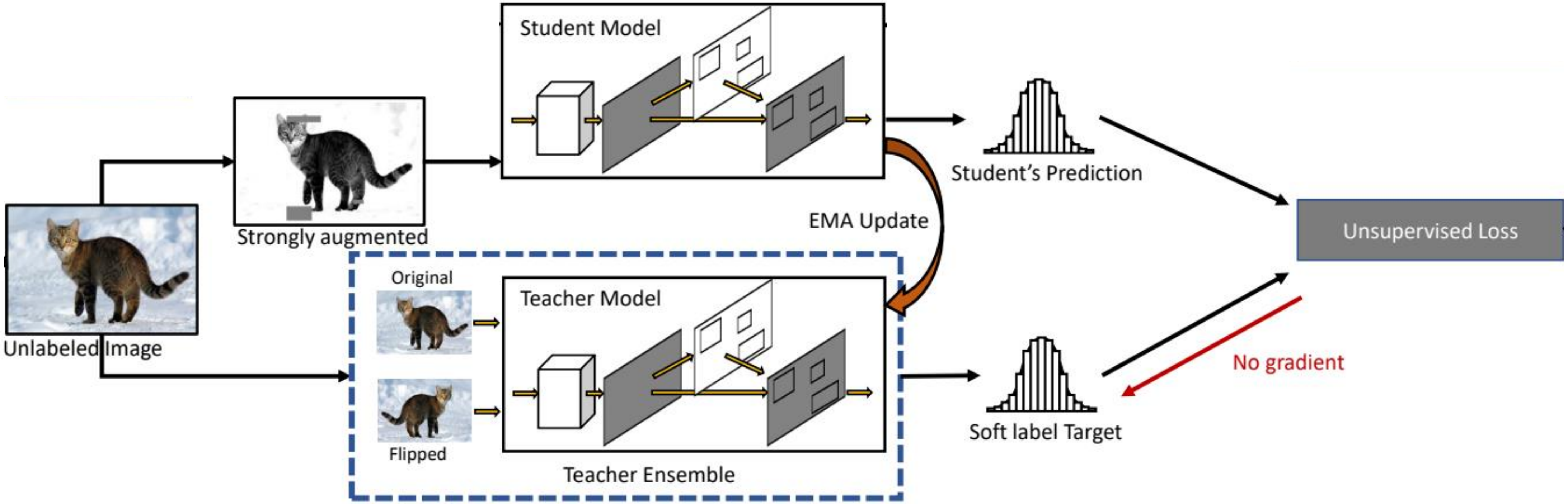


Augmentation:
- Weak(flip, resize)
- Strong [Based on weak] (color change, sharpness, contrast, Gaussian noise, cutout)

Without rotation/translation

$$L_S = L_{\text{cls}}^{\text{rpn}} + L_{\text{loc}}^{\text{rpn}} + L_{\text{cls}}^{\text{roi}} + L_{\text{loc}}^{\text{roi}}$$

Unsupervised Branch

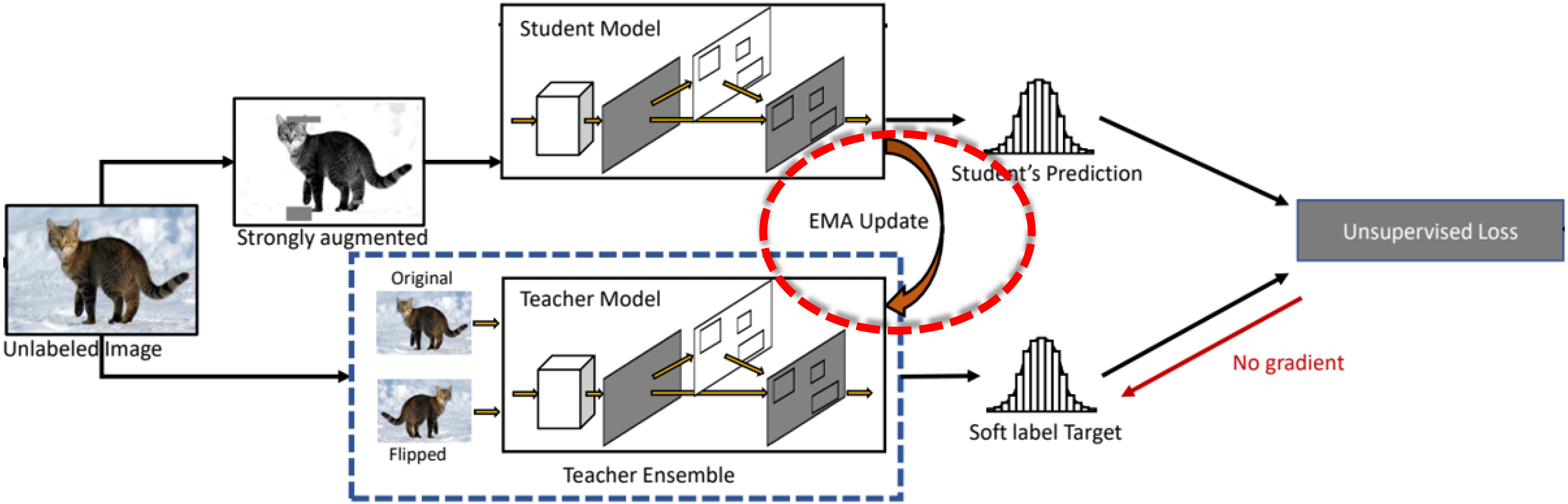


RPN & ROI: Classification loss: KLD Localization loss: L2
ROI: Feed proposals from teacher network to both networks

$$L_U^{\mathrm{rpn}} = \sum_{i \in S_A} D_{KL}(\mathbf{t}_{\mathrm{cls}}^{rpn,i} \| \mathbf{s}_{\mathrm{cls}}^{rpn,i}) + \| \mathbf{t}_{\mathrm{reg}}^{rpn,i} - \mathbf{s}_{\mathrm{reg}}^{rpn,i} \|_2$$

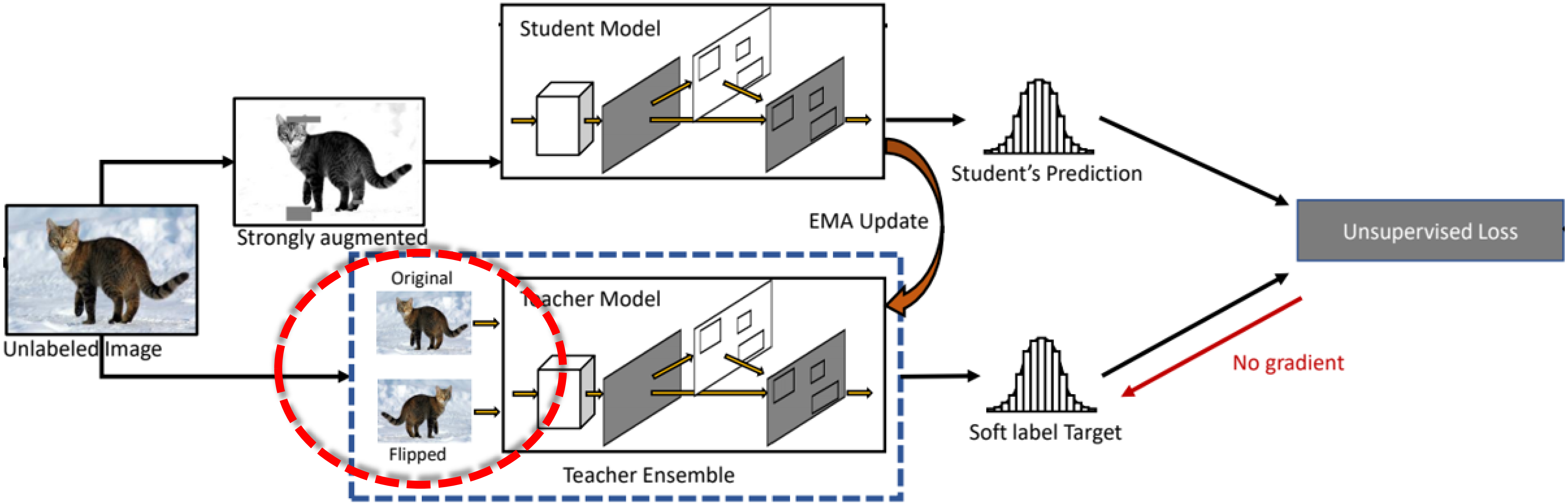
$$L_U^{\mathrm{roi}} = \Sigma_{i \in S_P} D_{KL}(\mathbf{t}_{\mathrm{cls}}^{roi,i} \| \mathbf{s}_{\mathrm{cls}}^{roi,i}) + \| \mathbf{t}_{\mathrm{reg}}^{roi,i} - \mathbf{s}_{\mathrm{reg}}^{roi,i} \|_2$$

Exponential Moving Average for the Teacher Model Update



$$W_{\text{teacher}} = \alpha W_{\text{teacher}} + (1 - \alpha) W_{\text{student}} \quad \text{where we set } \alpha = 0.999$$

Teacher Ensemble with Horizontal Flipping



$$f = \mathrm{ROIAlign}(f_B, P),$$

$$\hat{f} = \mathrm{ROIAlign}(\hat{f}_B, \hat{P}),$$

$$P_{\mathrm{cls}} = 0.5(C(f) + C(\hat{f})),$$

$$\sigma_{\mathrm{reg}} = 0.5(R(f) + T(R(\hat{f}))).$$

Experiments

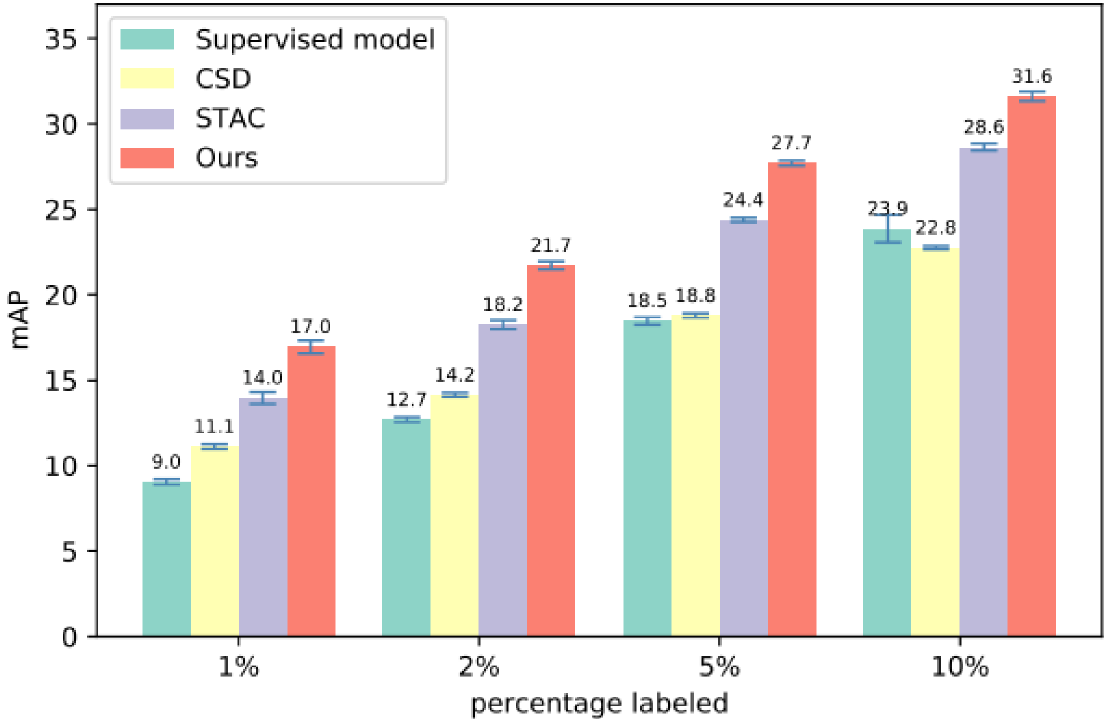


Figure 1: Comparing CSD [19], STAC [40], and our approach trained on full *MS-COCO train* 2017 with 1%, 2%, 5%, and 10% labeled over five runs using the splits in Sec. 4.1. Our approach consistently outperforms others.

Experiments

Model	Labeled Dataset	Unlabeled Dataset	AP50	AP
Supervised model	VOC07	N/A	76.3	42.60
Supervised model	VOC07 + VOC12	N/A	82.17	54.29
CSD‡	VOC07	VOC12	76.76	42.71
STAC [40]	VOC07	VOC12	77.45	44.64
Humble teacher (ours)	VOC07	VOC12	**80.94**	**53.04**
CSD‡	VOC07	VOC12 + MS-COCO20 (2017)	77.10	43.62
STAC [40]	VOC07	VOC12 + MS-COCO20 (2017)	79.08	46.01
Humble teacher (ours)	VOC07	VOC12 + MS-COCO20 (2017)	**81.29**	**54.41**

Table 1: Results on Pascal VOC, evaluated on the *VOC07 test* set. Our model consistently outperforms others in all experiment setups. CSD‡ is our ResNet-50-based re-implementation, which achieves better performance than the original CSD [19].

Experiments

Percentage labeled	1%	2%	5%	10%
Supervised model	9.05±0.16	12.70±0.15	18.47±0.22	23.86±0.81
CSD‡	11.12±0.15 (+2.07)	14.15±0.13 (+1.45)	18.79±0.13 (+0.32)	22.76±0.09 (−1.10)
STAC [40]	13.97±0.35 (+4.92)	18.25±0.25 (+5.55)	24.38±0.12 (+5.91)	28.64±0.21 (+4.78)
Humble teacher (ours)	**16.96±0.38** (+**7.91**)	**21.72±0.24** (+**9.02**)	**27.70±0.15** (+**9.23**)	**31.61±0.28** (+**7.74**)

Table 2: The mAP (50:95) results on *MS-COCO val 2017* by models trained on different percentage of labeled *MS-COCO train 2017*. All models are with the ResNet-50 backbone. CSD‡ is our re-implementation with better performance. Our method consistently outperforms others.

Experiments

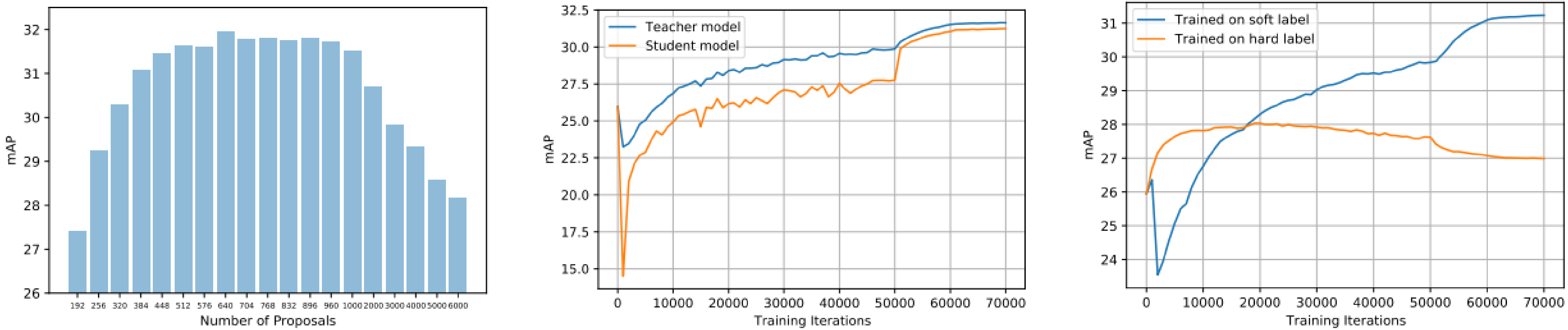
Model (Faster R-CNN with Resnet-50)	AP
Base supervised model	37.63
MOCOv2 + MS-COCO Unlabeled [7]	35.29
MOCOv2 + ImageNet-1M [7]	40.80
MOCOv2 + Instagram-1B [7]	41.10
Proposal learning [42]	38.4
CSD‡	38.52(+0.89)
STAC [40]	39.21(+1.58)
Humble teacher (ours)	**42.37(+4.74)**
Model (Cascade R-CNN with ResNet-152)	AP
Base supervised model	50.23
Humble teacher (ours)	**53.38 (+3.15)**

Table 3: The mAP (50:95) results on ***MS-COCO val 2017*** by models trained on *MS-COCO train 2017* + *MS-COCO unlabeled*. CSD‡ is with a ResNet-50 backbone.

Model (Cascade R-CNN with ResNet-152)	AP
Base supervised model	50.7
Humble teacher (ours)	**53.8 (+3.1)**

Table 4: The mAP (50:95) results on ***MS-COCO test-dev 2017*** by models trained on *MS-COCO train 2017* + *MS-COCO unlabeled*.

Ablation



(a) Comparison between models with different number of region proposals used in unsupervised loss. The student-teacher framework is jointly trained on the 10% labeled and 90% unlabeled *MS-COCO train 2017* split.

(b) Comparison between teacher and student performance on the 10% labeled *MS-COCO train 2017* setup. The student-teacher framework is jointly trained on the 10% labeled and 90% unlabeled *MS-COCO train 2017* split.

(c) Teacher models' performance on unlabeled data. Both models are trained on 10% labeled *MS-COCO train 2017* with the remaining 90% as unlabeled.

Figure 3: Ablation study on hyperparameters and hard/soft labels.

Ablation

Model	AP
No update	27.26±0.21
Copy weights from student to teacher every 10K iters	28.61±0.18
EMA update at every iter	**31.61±0.28**

Table 5: Comparison between different update rules on *MS-COCO train 2017* with 10% data labeled. The mean and standard deviation over five data splits are reported (the same five splits of *MS-COCO train 2017* as in Sec. 4.1).

Model	AP
With hard label	27.97±0.13
With soft label	**30.97±0.16**

Table 6: Comparison between training on soft label and hard label when 10% labeled *MS-COCO train 2017* is provided. The mean and standard deviation over five data splits are reported (the same five splits of *MS-COCO train 2017* described in Sec. 4.1).

Conclusion

1. Iteration-wise EMA teacher update
2. Soft label with a balanced number of teacher’ s region proposals
3. Data ensemble for the teacher

THANKS