Semi-supervised Object Detection with Object-wise Contrastive learning and regression uncertainty

Honggyu Choi[1], Zhixiang Chen[2], Xuepeng Shi[3], Tae-Kyun (T-K) Kim[1,3]

[1]KAIST, [2]The University of Sheffield, [3]Imperial College London

BMVC 2022

1. Introduction

Semi-Supervised Object Detection

- **Aims to boost detection performance by leveraging extra unlabeled data**
- **Since the pseudo-labels are noisy, pseudo-label filtering is crucial**

Previous Works & Limitations

- **Teacher-Student Framework** [1]

  → Teacher network generates pseudo-labels for unlabeled data to assist the training of a student network

- **Previous Works …**

  → Adopted classification score to select pseudo-labels with confidence higher than a pre-defined threshold

  → Heuristic designs to measure the localization quality

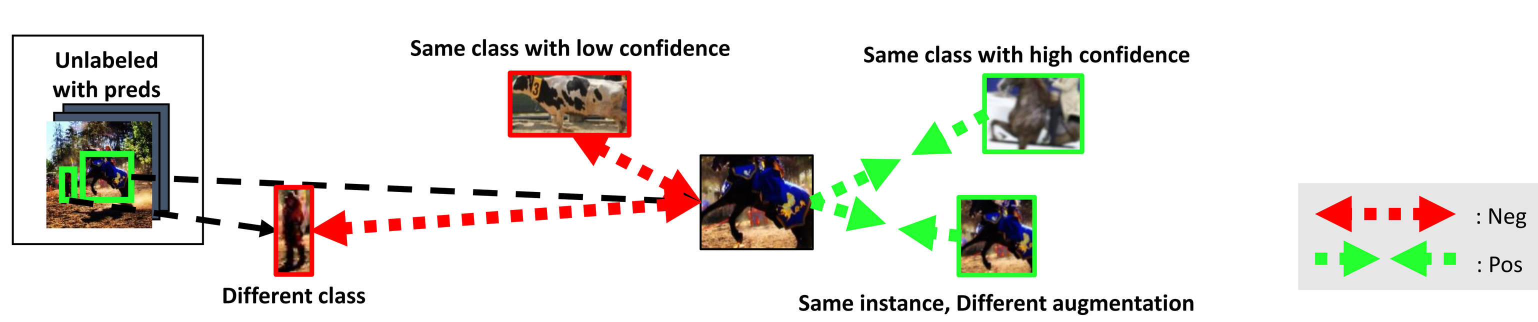
- **Existing classification scores and localization optimization is suboptimal**

  → Lack of labeled data makes **classification scores** less discriminative,

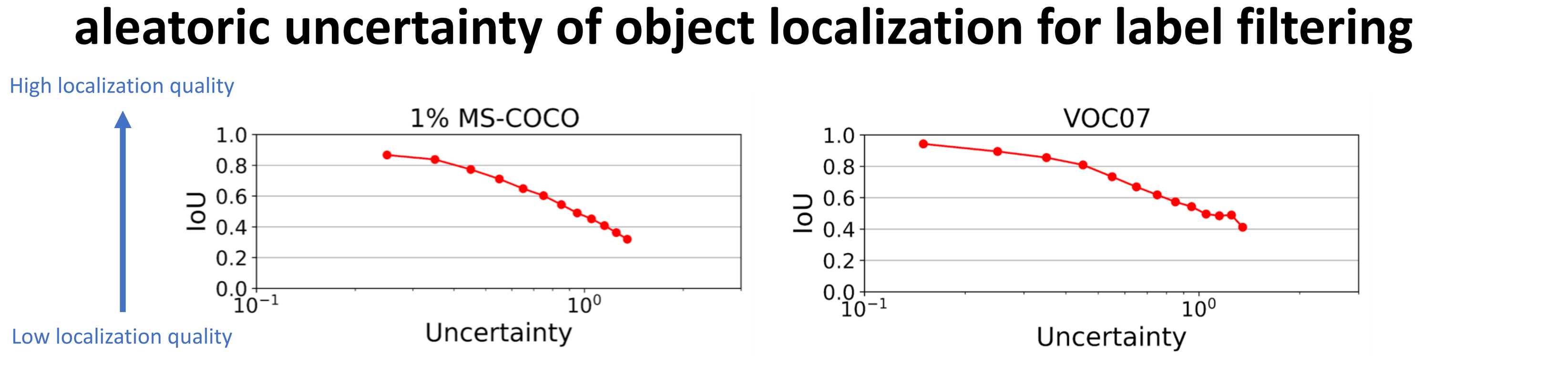
  → Measurement of pseudo-labels' **localization quality** is less investigated

Key Ideas

- **Object-wise Contrastive Learning (OCL) enhances the discriminativeness of the classification score**

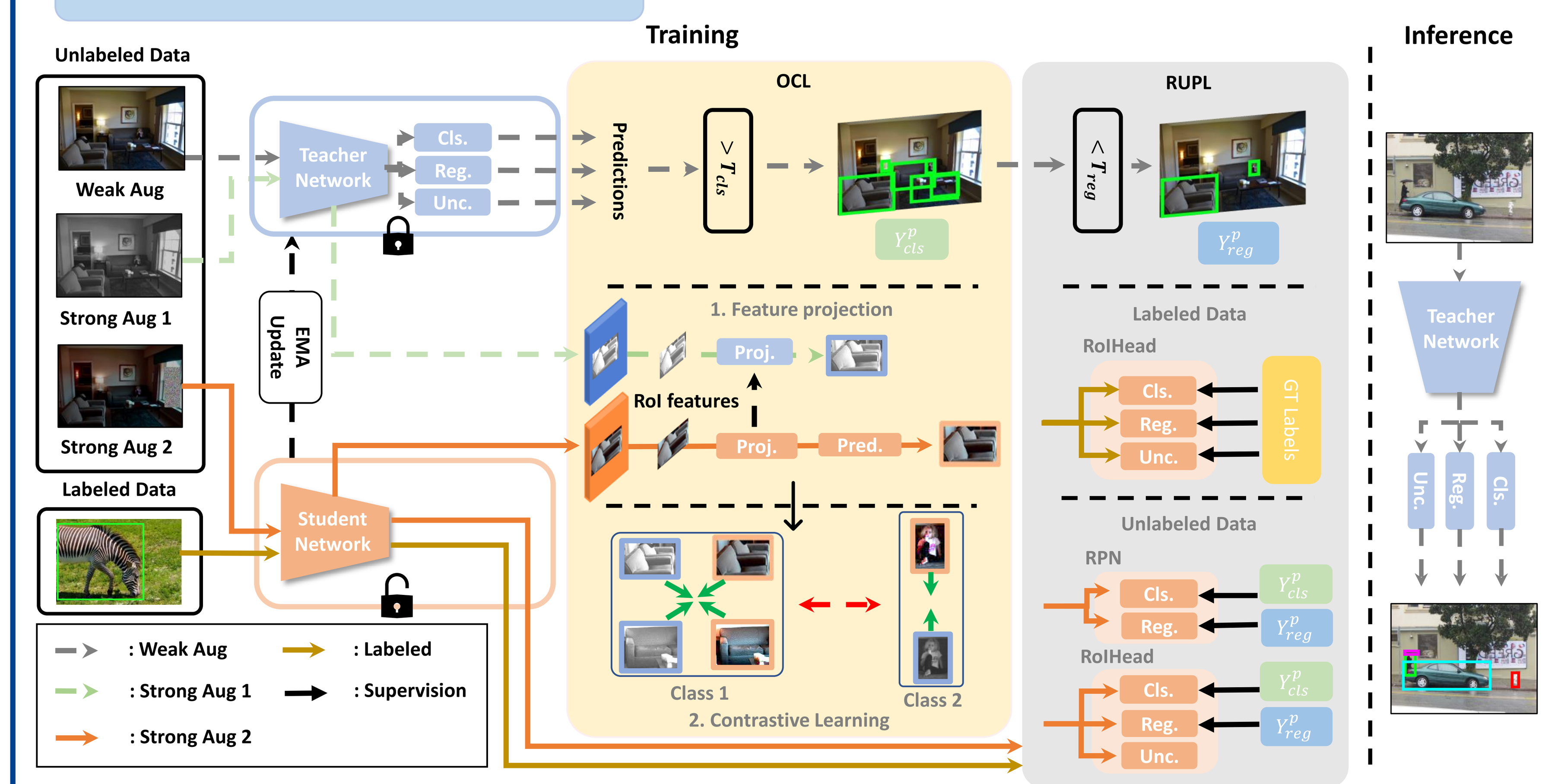


- **Regression-Uncertainty-guided Pseudo-Labeling (RUPL) models aleatoric uncertainty of object localization for label filtering**



2. Proposed Method

Framework Overview

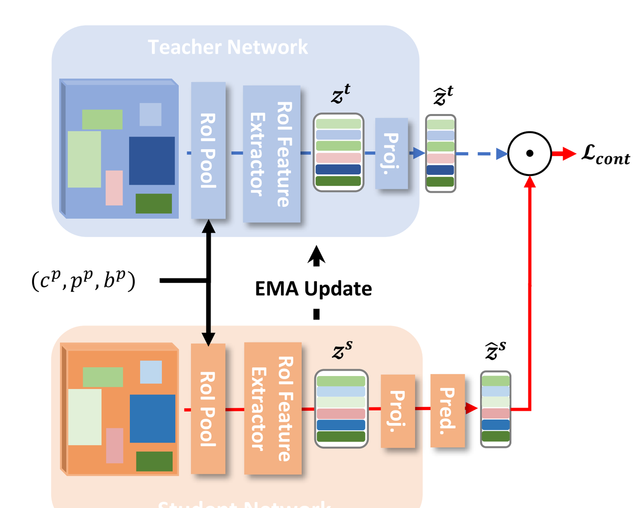


Object-wise Contrastive Learning (OCL)

- **Object-wise Contrastive Loss**

$$\mathcal{L}_{cont} = -\frac{1}{\mathcal{N}_O}\sum_{n=1}^{\mathcal{N}_O}\frac{1}{\sum_{m=1}^{\mathcal{N}_O}\mathbb{1}_{(w_{nm}>0)}}\sum_{m=1}^{\mathcal{N}_O} w_{nm}\log\frac{\exp(\hat{z}_n^s\cdot\hat{z}_m^t/\tau)}{\sum_{l}^{\mathcal{N}_O}\mathbb{1}_{(l\neq n)}\exp(\hat{z}_n^s\cdot\hat{z}_l^t/\tau)}$$

$$w_{nm} = \begin{cases} 1 & \text{if } n = m \\ p_n^p\cdot p_m^p & \text{if } \left[\begin{array}{c} n\neq m \\ c_n^p = c_m^p \\ P_n^p,\, P_m^p > T_{cont}, \end{array}\right. \\ 0 & \text{otherwise.} \end{cases}$$



- **Pseudo-label for Classification**

$$\{\boldsymbol{X}_{cls}^p, \boldsymbol{Y}_{cls}^p\} = \{(\boldsymbol{x}_i^p, \boldsymbol{y}_i^p)|p_{i,\cdot}^p > T_{cls}\}_{i=1}^{N_u}$$

Regression-Uncertainty-guided Pseudo-labeling (RUPL)

- **Uncertainty-aware Regression Loss** [2]

$$\mathcal{L}_{roi,reg} = \frac{smoothL_1(\hat{t},t)}{\sigma^2} + \lambda_{unc}\log(\sigma^2)$$

- **Pseudo-label for Regression**

$$\{\boldsymbol{X}_{reg}^p, \boldsymbol{Y}_{reg}^p\} = \{(\boldsymbol{x}_i^p, \boldsymbol{y}_i^p)|p_{i,\cdot}^p > T_{cls} \text{ and } \bar{\sigma}_{i,\cdot}^p < T_{reg}\}_{i=1}^{N_u}$$

3. Experiments

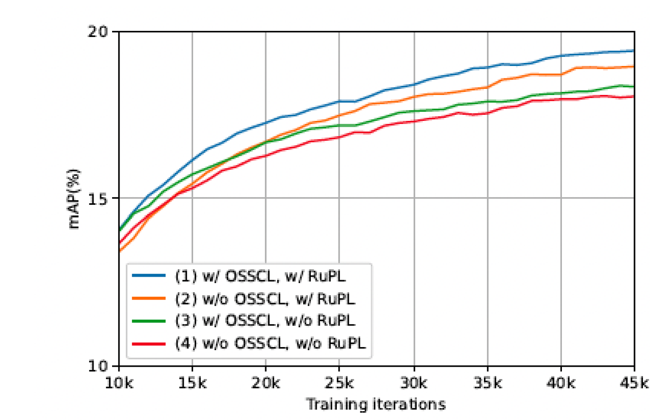
Main Results

Methods	VOC	VOC + coco20cls
Baseline [2]	48.69	50.34
Ours	**52.04**	**53.88**

Methods	COCO 1%	COCO 5%	COCO 10%	Additional	coco-35k
Baseline [2]	20.75	28.27	31.50	41.30	36.36
Ours	**21.63**	**30.66**	**33.53**	**41.89**	**37.13**

Ablation Study

	OCL	RUPL	$AP_{50:95}$	AP_{50}	AP_{75}
(1)	√	√	**19.42**	34.65	**19.36**
(2)		√	18.95	33.82	19.07
(3)	√		18.34	**35.21**	17.16
(4)			18.05	34.45	17.09



Method	$AP_{50:95}$
Box jittering [4]	18.15
Predicted IoU [5]	18.85
Aleatoric uncertainty [3] (Ours)	**18.95**

Method	$AP_{50:95}$
w/o CL	18.05
Self-sup CL	18.24
OCL (Ours)	**18.34**

4. Conclusion

- Propose a two-step pseudo-label filtering for SSOD
- Deal with both classification and regression heads
- OCL enhances discriminativeness of classification score
- RUPL learns regression uncertainty to measure the localization quality
- Achieve remarkable performance gain against our baseline and show competitive results compared to other SOTA

References

[1] A.Tarvainen, *et al*. Mean teachers are better role models: Weight averaged consistency targets improve semi-supervised deep learning results. In NIPS, 2017.

[2] Y.Liu, *et al.* Unbiased teacher for Semi-Supervised Object Detection. In ICLR, 2021

[3] A.Kendall, *et al.* What uncertainties do we need in bayesian deep learning for computer vision? In NIPS, 2017.

[4] M.Xu, *et al*. End-to-end semi-supervised object detection with soft teacher. In ICCV, 2021.

[5] H.Wang, *et al*. 3dioumatch:Leveraging iou prediction for semi-supervised 3d object detection. In CVPR, 2021.