PseCo: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection

2022-10-18 | Jihyun Lee

PseCo: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection

Gang Li[1,2], Xiang Li[1⋆], Yujie Wang[2], Yichao Wu[2], Ding Liang[2], and Shanshan Zhang[1⋆]

[1] Nanjing University of Science and Technology
[2] SenseTime Research
{gang.li, xiang.li.implus, shanshan.zhang}@njust.edu.cn
{wangyujie,wuyichao,liangding}@sensetime.com

https://arxiv.org/pdf/2203.16317.pdf
https://github.com/ligang-cs/PseCo

http://arxiv.org › cs

PseCo: Pseudo Labeling and Consistency Training for Semi ...

G Li 저술 · 2022 · 1회 인용 — **PseCo**: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection. Authors:Gang Li, Xiang Li, Yujie Wang, Yichao Wu, Ding ...

Comments: ECCV 2022
Subjects: **Computer Vision and Pattern Recognition (cs.CV)**; Artificial Intelligence (cs.AI)
Cite as: arXiv:2203.16317 **[cs.CV]**
(or arXiv:2203.16317v2 **[cs.CV]** for this version)
https://doi.org/10.48550/arXiv.2203.16317

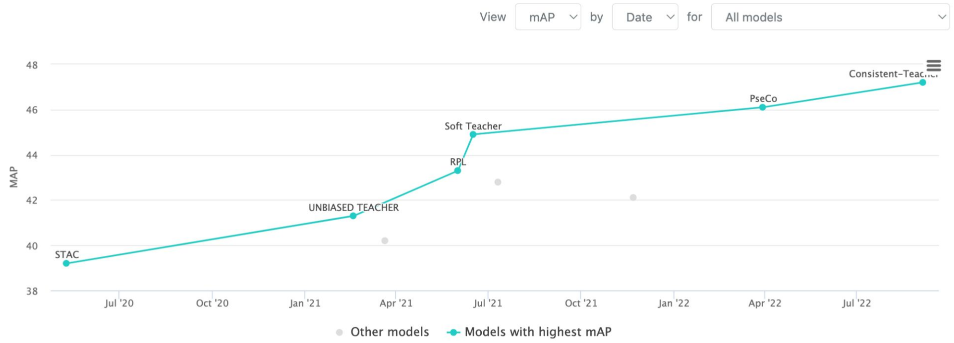
Content

- Introduction
- Related Work
- Method
  - Prediction-guided Label Assignment (PLA)
  - Positive-proposal Consistency Voting (PCV)
  - Multi-view Scale-invariant Learning (MSL)
- Experiments
- Conclusion

- **Pseco: Pse**udo Labeling and **Co**nsistency Training for Semi-Supervised Object Detection
  - Semi-supervised Object detection (SSOD) // c.f Semi-supervised Image Classification (SSIC)
  - 기존 Semi-supervised learning의 기본적인 2가지 Approach: Pseudo labeling과 Consistency Learning 관점에서 문제점을 개선
  - SSOD 특성상 발생되는 Noisy Pseudo Box 에 대해 모델이 덜 Misleading 되지않도록 3가지 Method 제안
    - Prediction-Guided Label Assignment(PLA)
    - Positive-proposal Consistency Voting (PCV)
    - Multi-view Scale-invariant Learning (MSL)
  - COCO benchmark - 2022 SOTA
    - 1,5,10 % labeling ratio
    - full dataset + 2017 unlabel COCO data (5% mAP 성능향상)

Semi-Supervised Object Detection on COCO 100% labeled data

Leaderboard Dataset



Filter: untagged

Edit Leaderboard

Rank	Model	mAP	Extra Training Data	Paper	Code	Result	Year	Tags
1	Consistent-Teacher	47.20	✓	Consistent Targets Provide Better Supervision in Semi-supervised Object Detection			2022	
2	PseCo	46.1	×	PseCo: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection			2022	

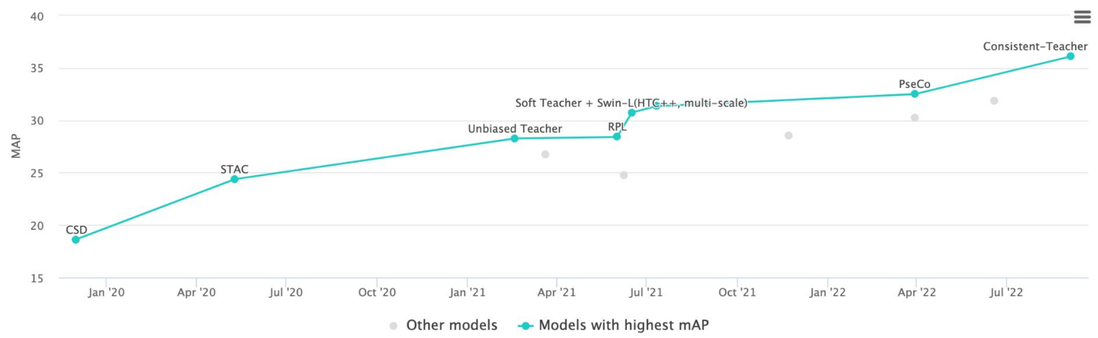
Introduction

- Semi-Supervised Object detection task

Semi-Supervised Object Detection on COCO 5% labeled data

Leaderboard Dataset

View mAP by Date



Filter: untagged

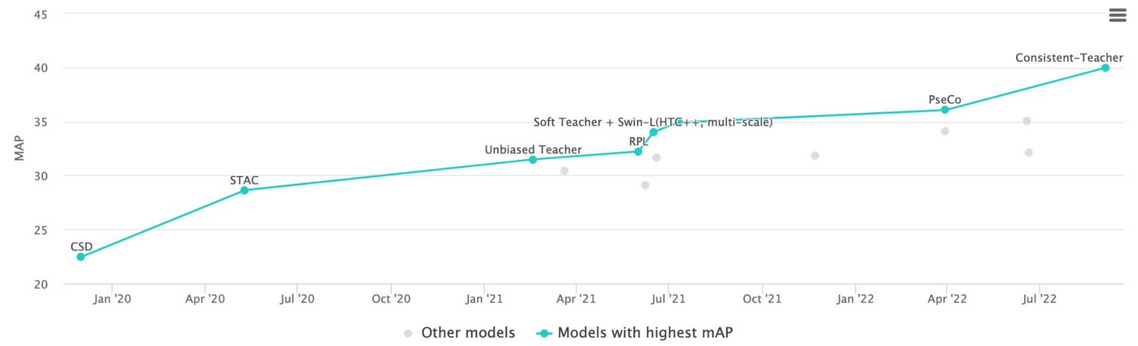
Edit Leaderboard

Rank	Model	mAP	Paper	Code	Result	Year	Tags
1	**Consistent-Teacher**	36.1	Consistent Targets Provide Better Supervision in Semi-supervised Object Detection			2022	
2	**PseCo**	32.5	PseCo: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection			2022	

Semi-Supervised Object Detection on COCO 10% labeled data

Leaderboard Dataset

View mAP by Date



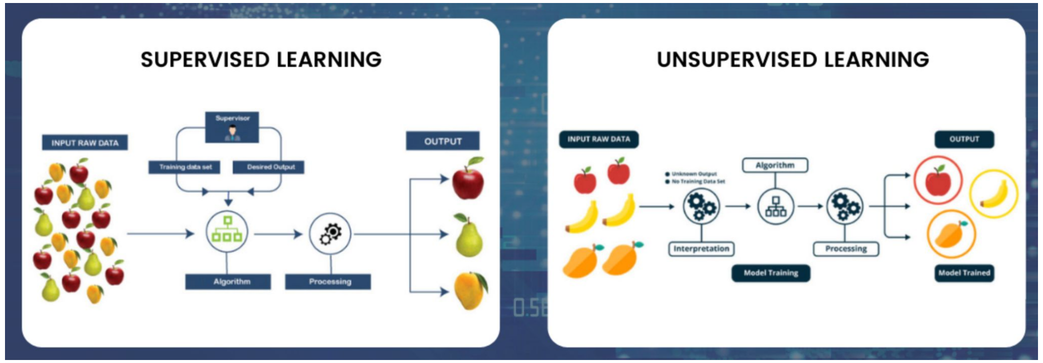
Filter: untagged

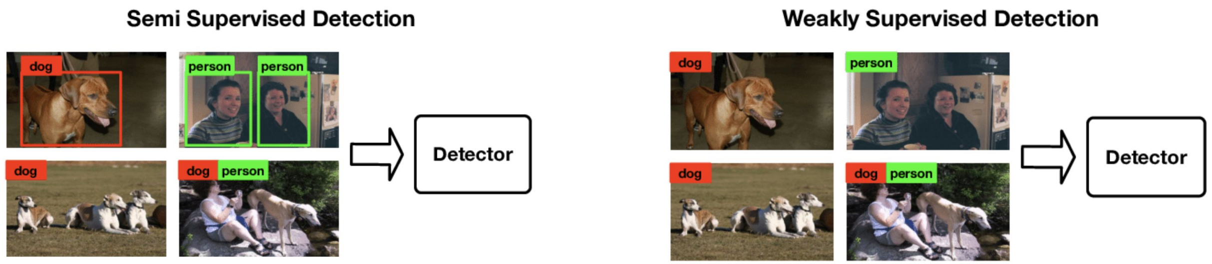
Edit Leaderboard

Rank	Model	mAP	Paper	Code	Result	Year	Tags
1	**Consistent-Teacher**	40.0	Consistent Targets Provide Better Supervision in Semi-supervised Object Detection			2022	
2	**PseCo**	36.06	PseCo: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection			2022	

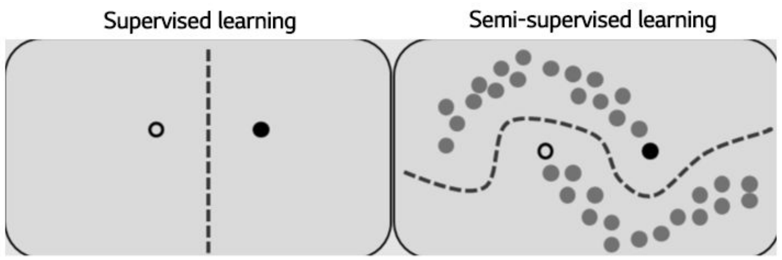
Introduction

- **Label 종류에 따른 학습방법**
  - Supervised Learning
    - 모든 이미지와 annotation 존재
  - Unsupervised Learning
    - 이미지만 존재
  - **Semi-Supervised Learning**
    - **소량의 Label 데이터와 대량의 Unlabel 데이터 ⇒ 데이터의 annotation cost 가 큰 task에 적용가능**
  - Weakly Supervised Learning
    - 타겟으로 하는 task의 annotation이 아닌 다른 label을 활용, i.e, object label 로만 localization을 수행





- **Semi-Supervised Learning**
  - 소량의 Label 데이터와 대량의 Unlabel 데이터 ⇒ 데이터의 annotation cost 가 큰 task에 적용가능
  - **소량의 Label로 Supervised Learning → 대량의 Unlabel에 대해서 Unsupervised Learning**
  - **주요목적: labeled data가 만들어놓은 decision boundary를 더 정교하고 신뢰성있게 변형시켜주는 것 (일반화, 성능향상)**



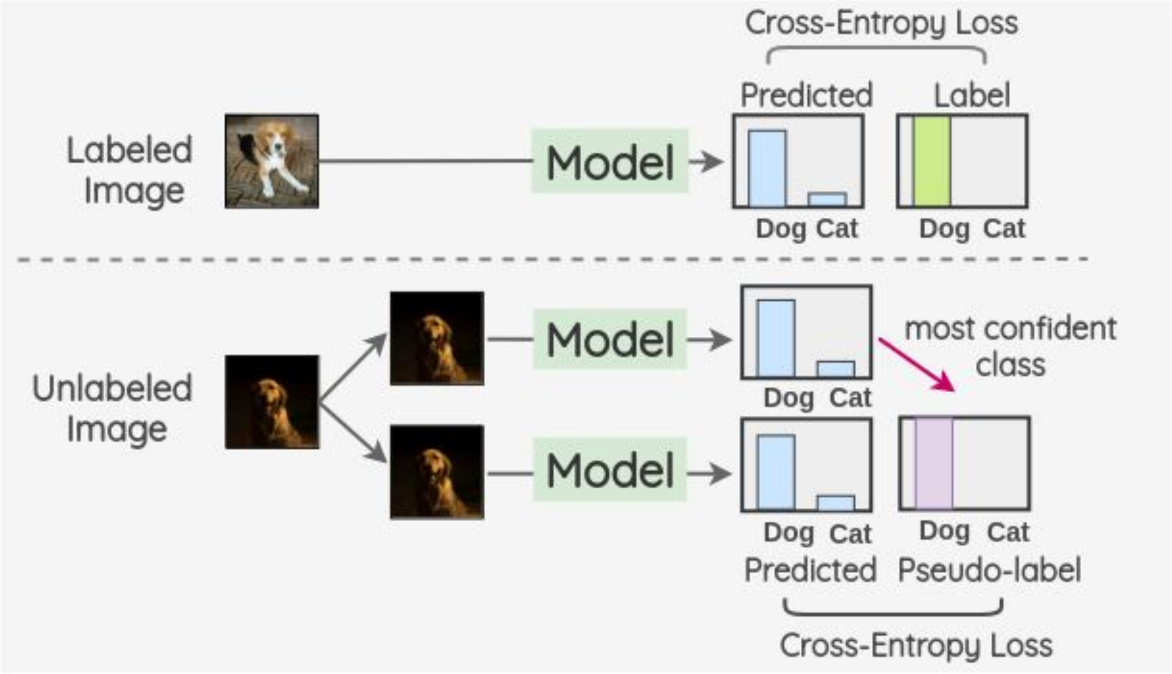
- **Unlabeled data를 사용함으로써 기존의 supervised learning model과 비교하여 추가적으로 고려해야 할 사항**
  - 어떻게 Unlabeled data의 loss를 계산할 것인가?
  - 계산된 unsupervised loss를 어떻게 supervised loss와 결합시켜 성능을 향상시킬 것인가?

⇒ SSL은 unlabled 된 데이터의 unsupervised loss를 어떻게 측정하고, 언제 어떻게 학습과정에 활용하는 가에 따라 그 종류가 달라짐.

- 기본적인 **SSL Concept**
  - **Consistency Regularization**
  - **Self-Training (Pseudo-Label)**

Introduction

- 기존 Semi-Supervised Learning 의 문제점
    - Pseudo Labeling
        - Self-Training - Pseudo Labeling
        - Self-Training: 학습과정 중간에 발생하는 결과치들을 바탕으로 스스로 학습을 하는 방법론
        - 가장 대표적인 Pseudo Labeling 방법

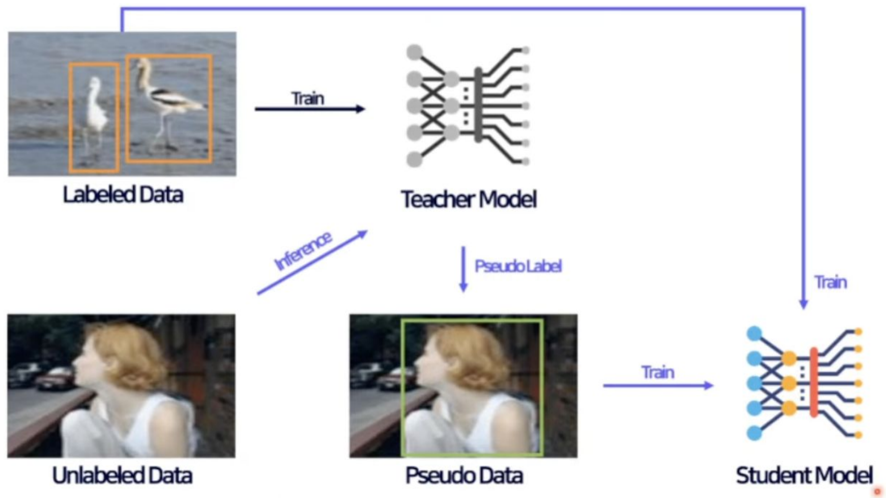


$$L = L_l + \alpha_t \times L_u$$

- 기존 Semi-Supervised Learning 의 문제점
  - Pseudo Labeling
    - Label 데이터로 teacher model 학습
    - Unlabel 데이터를 teacher model 에 inference 하여 pseudo labeling
    - 생성한 pseudo label 로 student model 학습
  - 문제점
    - classification 에 집중된 방법
    - Pseudo Labeling 에서 classification 의 성능이 bounding box 의 quality 를 보장하지 않음

IoU 가 0.9 일 때, precision 이 31%

잘못된 pseudo box 는
잘못된 label assignment 야기



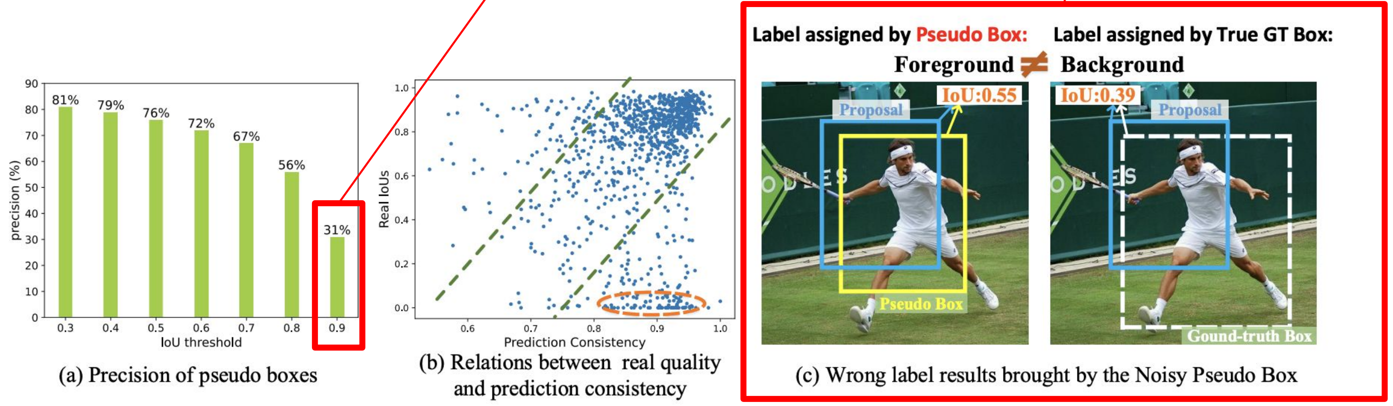
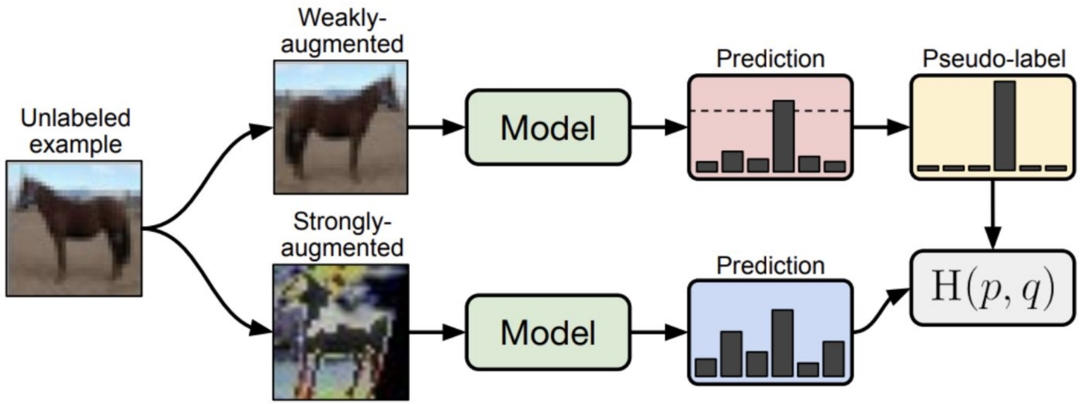


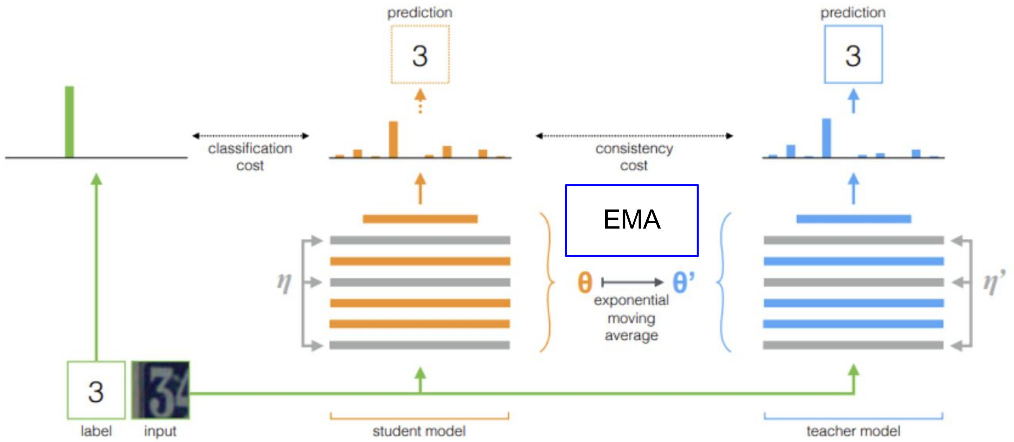
Fig. 1: (a) The precision of pseudo boxes under various IoU thresholds. (b) The scatter diagram of the relation between the prediction consistency and their true localization quality. Some dots falling in the orange ellipse are caused by annotation errors. We show some examples in Fig. 5. (c) One specific example to demonstrate that noisy pseudo boxes will mislead label assignment.

- 기존 Semi-Supervised Learning 의 문제점
  - Consistency Training
    - 모델이 데이터의 작은 변화에 민감하지 않게 (robust) 만드는 방법. 작은 변화란, 사람이 봤을 때 label 에 큰 영향을 주지 않을 정도의 noise
    - Label 데이터는 학습시키고,
    - Unlabel 데이터는 augmentation 을 시킨 후, 모델에 입력
    - 2개의 augmentation 된 unlabel 데이터는 동일한 classification 결과를 보여야 함 -> KL loss 를 통해 학습
  - 문제점
    - Object detection task 속성을 고려하지 않음
    - Object detection 의 경우 scale invariant 해야 함 (Label-Level consistency 만 고려)
    - Small object 의 성능을 높이기 위해 당연히 feature 피라미드(FPN)를 사용해야 함



FixMatch Consistency Training 예시

- Mean teacher
    - [“Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results” Tarvainen et al., Advances in neural information processing systems 30 (2017)]
    - Teacher 모델을 student 모델의 weight 를 활용함
    - 즉, EMA (Exponential Moving Average; 오래된 데이터에 대한 가중치는 지수적으로 감소하지만, 0이 되지는 않게끔 계산하는 방법) 을 이용해서 student 모델의 weight 를 teacher 모델의 weight 로 사용함



- Soft teacher

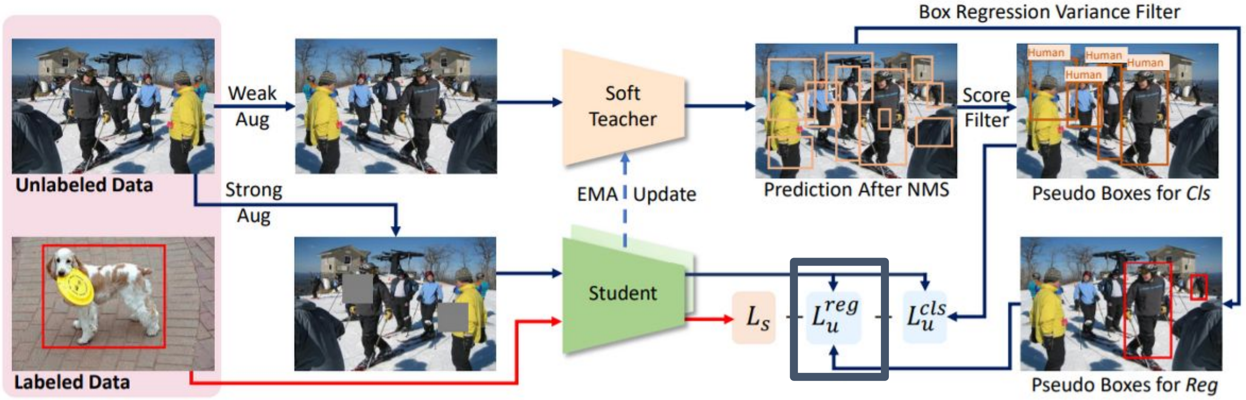


Figure 2. The overview of the end-to-end pseudo-labeling framework for semi-supervised object detection. Unlabeled images and labeled images form the training data batch. In each training iteration, a soft teacher is applied to perform pseudo-labeling on weak augmented unlabeled images on the fly. Two sets of pseudo boxes are produced: one is used for classification branch by filtering boxes according to the foreground score, and the other is used for box regression branch by filtering boxes according to box regression variance. The teacher model is updated by student model via exponential mean average (EMA) manner. The final loss is the sum of supervised detection loss L_s and unsupervised detection loss L_u.

- Architecture
  - 기본 구조는 soft teacher 와 동일
  - **Pseudo box 에 의한 regression loss 계산 제거** (왜냐하면, pseudo label 에 의한 bounding box 는 quality issue 로 학습 과정을 불안하게 함으로)
  - 3가지 방법
    - Prediction-guided Label Assignment (PLA)
    - Positive-proposal Consistency Voting (PCV)
    - Multi-view Scale-invariant Learning (MSL)

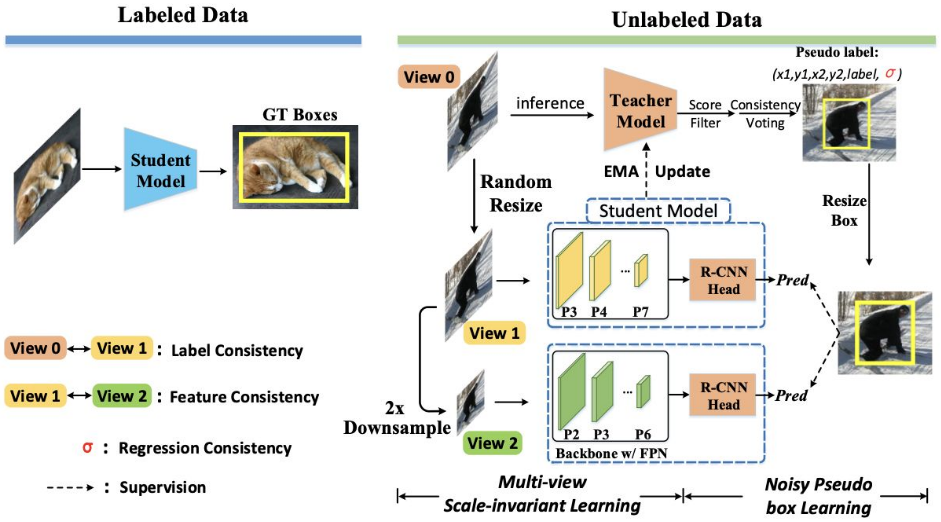
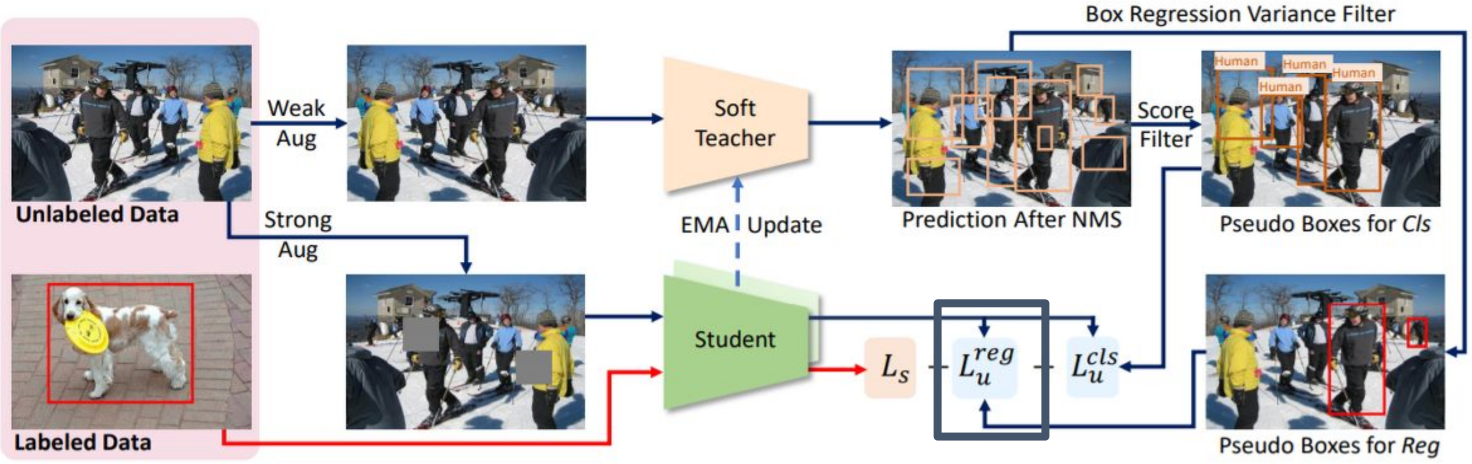
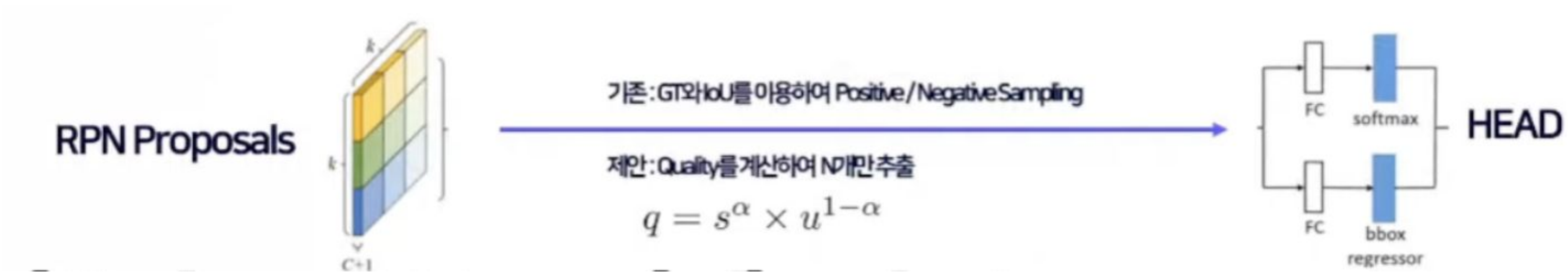


Fig. 2: The framework of our PseCo. Each training batch consists of both labeled and unlabeled images. On the unlabeled images, the student model trains on view V_1 and V_2 at the same time, taking the same pseudo boxes as supervisions. View V_0 refers to input images for the teacher model.

- **Prediction-guided Label Assignment (PLA)**
  - Baseline model 은 faster-rcnn
  - 기존의 방법들은 teacher model 의 NMS 결과를 pseudo label 로 사용
  - Label assignment 문제를 해결하고, Pseudo Label 에 의한 regression loss 제거를 대응하여 auxiliary information 전달
  - Teacher model 의 RPN 결과를 student model 과 공유하여 guidance 전달





- **Positive-proposal Consistency Voting (PCV)**
  - 기존의 classification score 와 objectiveness 로 scoring 하는 것이 아닌, **localization 도 잘 반영되어야 함**
  - **Regression consistency 제안**
    - Localization 성능을 네트워크에 반영

$$\sigma^j = \frac{\sum_{i=1}^{N} u_i^j}{N}, \qquad \mathcal{L}_{reg}^u = \frac{1}{MN}\sum_{j=1}^{M}\sigma^j\sum_{i=1}^{N}|reg_i^j - \hat{reg}_i^j|,$$

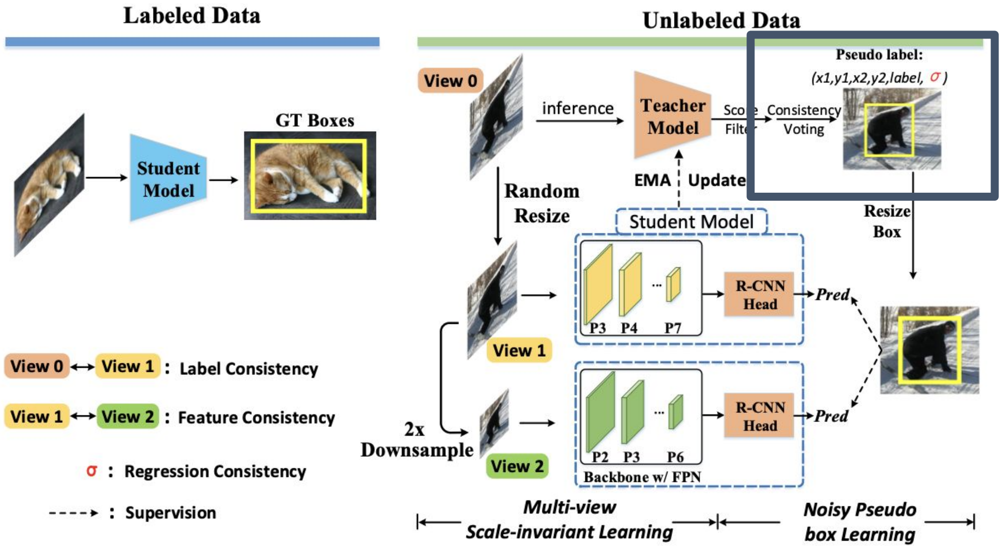
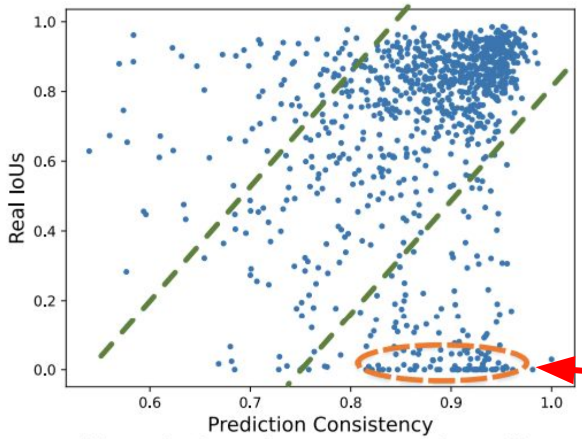
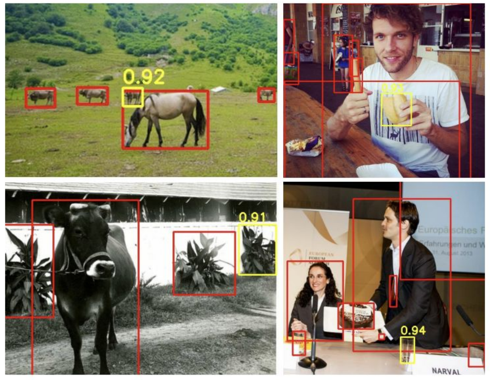


Fig. 2: The framework of our PseCo. Each training batch consists of both labeled and unlabeled images. On the unlabeled images, the student model trains on view V_1 and V_2 at the same time, taking the same pseudo boxes as supervisions. View V_0 refers to input images for the teacher model.



(b) Relations between real quality and prediction consistency

Annotation error



(a) Pseudo boxes produced by the teacher model

Method

- **Multi-view Scale-invariant Learning (MSL)**
    - Scale invariant 를 위하여 Label Level Consistency 와 feature level consistency 계산
    - **Label level consistency**
        - Random resize ratio 를 결합해 label alignment 진행
        - Random resize 를 적용한 이미지와 원본 이미지는 동일한 이미지이므로, 동일한 label 결과가 나타나야 함. 이를 통해서 consistency loss 계산
    - **Feature level consistency**
        - 같은 resolution 의 FPN stage feature 를 일치시키도록 학습
        - Object detection 의 scale invariant 해야 하는 속성을 반영하기 위해서, FPN 구조를 활용해 feature level consistency 제안

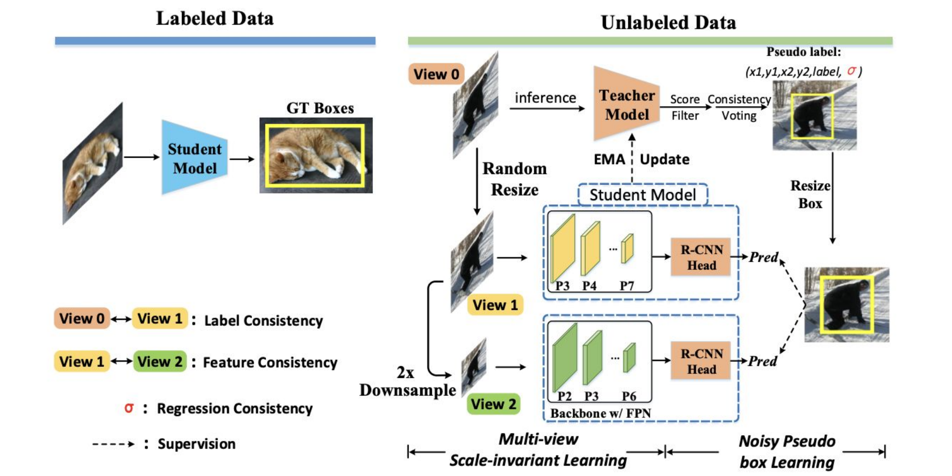
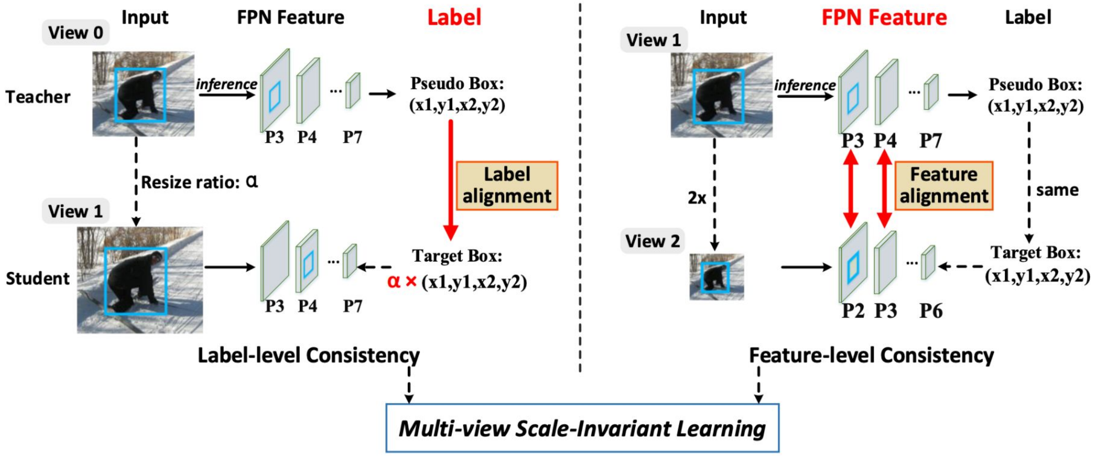


Fig. 2: The framework of our PseCo. Each training batch consists of both labeled and unlabeled images. On the unlabeled images, the student model trains on view V_1 and V_2 at the same time, taking the same pseudo boxes as supervisions. View V_0 refers to input images for the teacher model.



- **Results**
  - Partially Labeled Data: N% 만큼 MSCOCO train 2017 Labaled dataset 에서 random sampling (5-fold)
  - Fully labeled data: MSCOCO train2017 Data + MSCOCO train2017 unlabeled Data

Table 1: Comparisons with the state-of-the-art methods on val2017 set under the **Partially Labeled Data** and **Fully Labeled Data** settings.

Method	Partially Labeled Data				Fully Labeled Data
	1%	2%	5%	10%	
Supervised baseline	12.20±0.29	16.53±0.12	21.17±0.17	26.90±0.08	41.0
STAC [20]	13.97±0.35	18.25±0.25	24.38±0.12	28.64±0.21	39.5 $\xrightarrow{-0.3}$ 39.2
Humble Teacher [21]	16.96±0.35	21.74±0.24	27.70±0.15	31.61±0.28	37.6 $\xrightarrow{+4.8}$ 42.4
ISMT [27]	18.88±0.74	22.43±0.56	26.37±0.24	30.53±0.52	37.8 $\xrightarrow{+1.8}$ 39.6
Instant-Teaching [30]	18.05±0.15	22.45±0.15	26.75±0.05	30.40±0.05	37.6 $\xrightarrow{+2.6}$ 40.2
Unbiased Teacher [14]	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 [26]	20.46±0.39	-	30.74±0.08	34.04±0.14	40.9 $\xrightarrow{+3.6}$ 44.5
PseCo (ours)	**22.43**±0.36	**27.77**±0.18	**32.50**±0.08	**36.06**±0.24	41.0 $\xrightarrow{+5.1}$ **46.1**

- **Results**
  - Supervised baseline 에 비해 성능이 좋음

(b) Detection results of supervised baseline

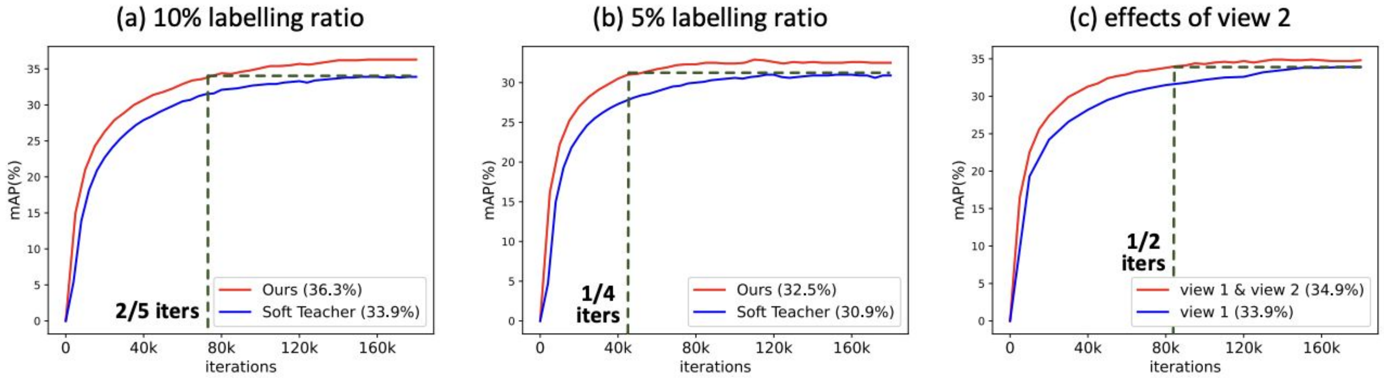
(c) Detection results of our method

  - 각각의 기능들에 대한 효과

Table 2: Ablation studies on each component of our method. MSL represents Multi-view Scale-invariant Learning; NPL represents Noisy Pseudo box Learning. In MSL, V_1 and V_2 are constructed for label- and feature-level consistency, respectively. In NPL, PCV and PLA stand for Positive-proposal Consistency Voting and Prediction-guided Label Assignment, respectively.

MSL		NPL		mAP	AP_{50}	AP_{75}
V_1	V_2	PCV	PLA			
				26.8	44.9	28.4
✓				33.9(+7.1)	55.2	36.0
✓	✓			34.9(+8.1)	56.3	37.1
✓		✓		34.8(+8.0)	55.1	37.4
✓		✓	✓	35.7(+8.9)	56.4	38.4
✓	✓	✓		36.0(+9.2)	56.9	38.7
✓	✓	✓	✓	**36.3**(+9.5)	**57.2**	**39.2**

  - 빠른 convergence 확인 가능



Conclusion

- 새로운 SSOD framework (PseCO) 제안
  - Consists of Noisy Pseudo box Learning
  - Multi-view Scale-invariant Learning

- To validate the effectiveness of methods, extensive experiments are conducted on COCO benchmark
  - 기존의 Semi Supervised Learning 의 문제점을 해결하여 SoTA 달성
  - Both in accuracy and efficiency

End of the Document