Consistency-based Semi-supervised learning for Object Detection

Authors: Jisoo Jeong, Seungeui Lee, Jeesoo Kim, Nojun Kwak

Venue: Advances in Neural Information Processing Systems 32 (NIPS 2019)

Presenter: Ishan Dave



Outline of this presentation

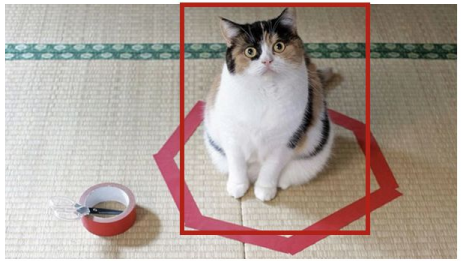
- Understanding the problem and terminologies
- Existing Approaches to solve the problem
- Inspiration of the proposed work
- Proposed method
- Technical Contribution
- Results
- Conclusion

Understanding the problem and terminologies: **Object Detector**



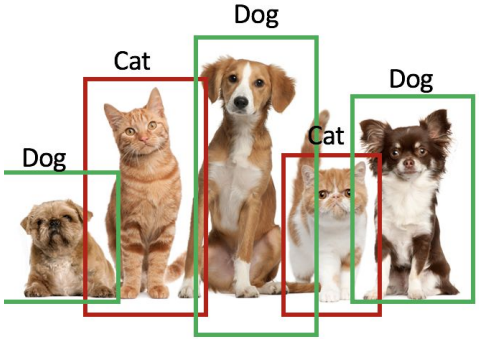
Is this image of Cat or not?

Image classification problem



Where is Cat?

Classification with localization probl

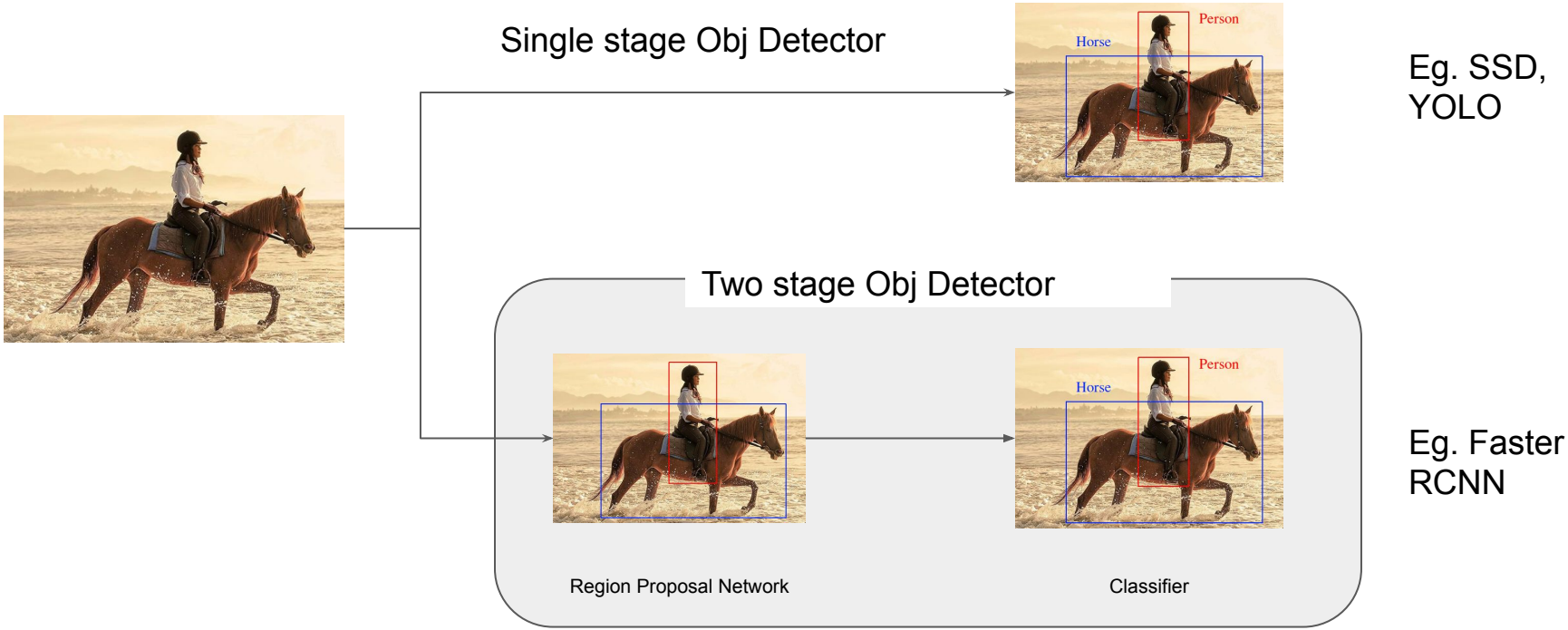


Which animals are there in image and where?

Object detection problem

Image Courtesy: https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad

Type of object detector: quick review



Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).
Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

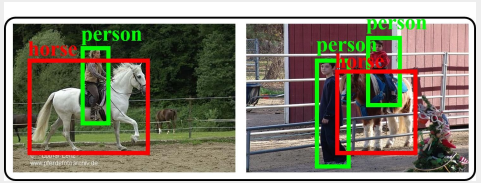
Understanding the problem and terminologies: **Motivation**



Object Detector Model

Classification and Localization Loss

Supervised Learning



Object Detector Model

Classification and Localization Loss for **labeled data**

But what losses for **Unlabled data**?

Semi-Supervised Learning

Any questions till this point?

- ✓ Understanding the problem and terminologies
  - ✓ Object Detector
  - ✓ Different Level of Supervision
  - ✓ Motivation
- ❑ **Existing Approaches to solve the problem**

A common approach in Semi-supervised learning

"Self-Training approach"

Training Mode

Labeled data



Object Detector Model

Classification and Localization Loss

Inference Mode



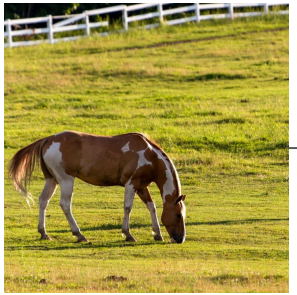
TRAINED Object Detector Model

Prediction
"Horse" at [245, 240, 50, 50]
with **0.93 confidence**

Confidence threshold

"Pseudo Labels"

Consistency regularization

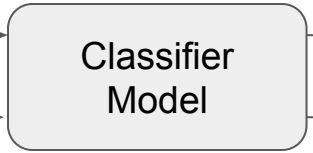


Original (x)



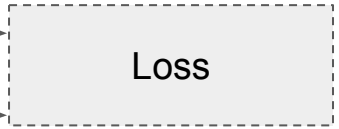
Perturbed (x')

State-of-the-Art in Semi-Supervised Classification!



f(x)

f(x')



Consistency regularization



Original (x)

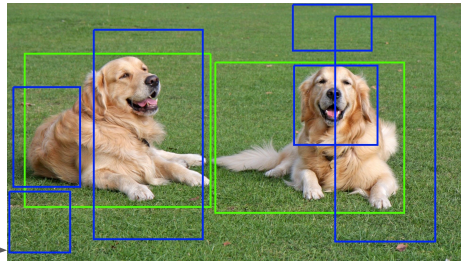


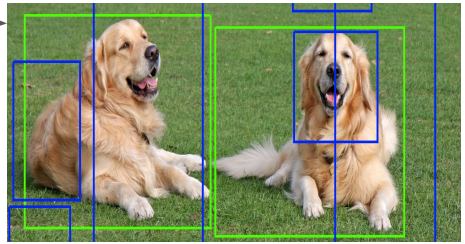
Perturbed (x')

Localization Model

f(x)

f(x')

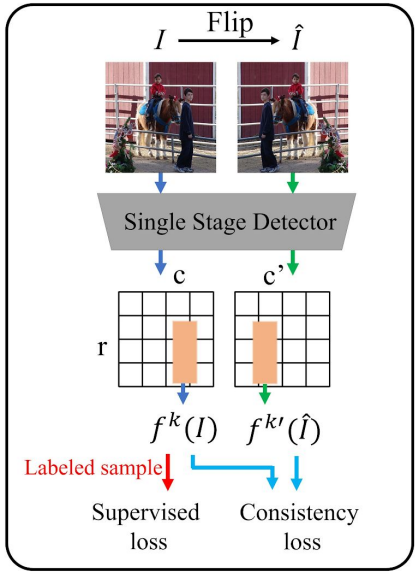




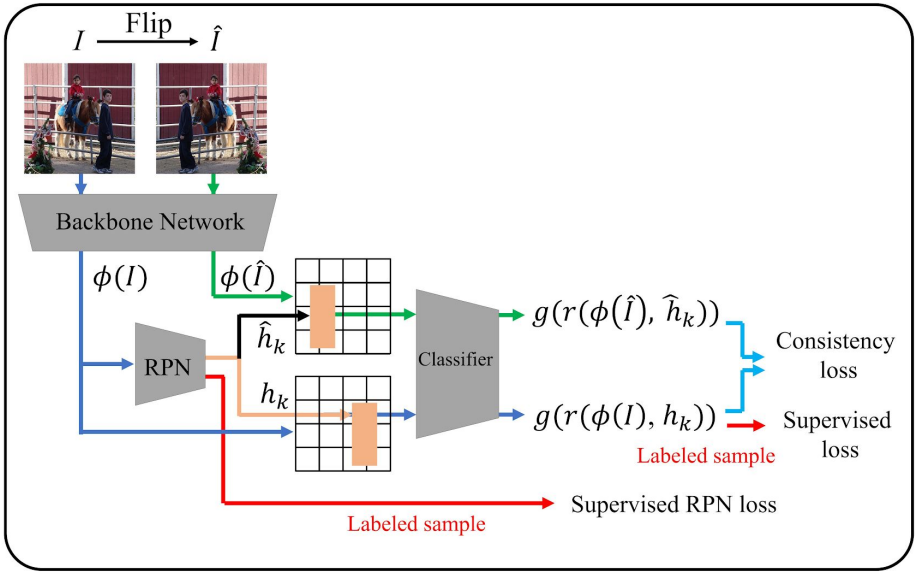
Hard to find one to one correspondence!

Image courtesy https://www.petmoo.com/dogs/golden-retriever-puppies/

Proposed Work “CSD”

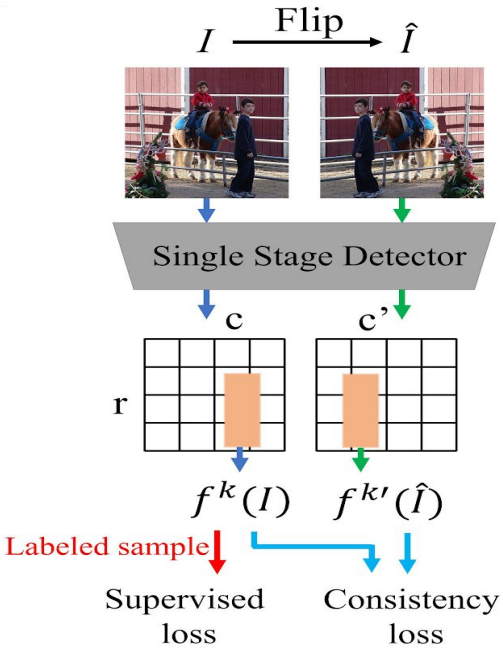


(a) single stage detector

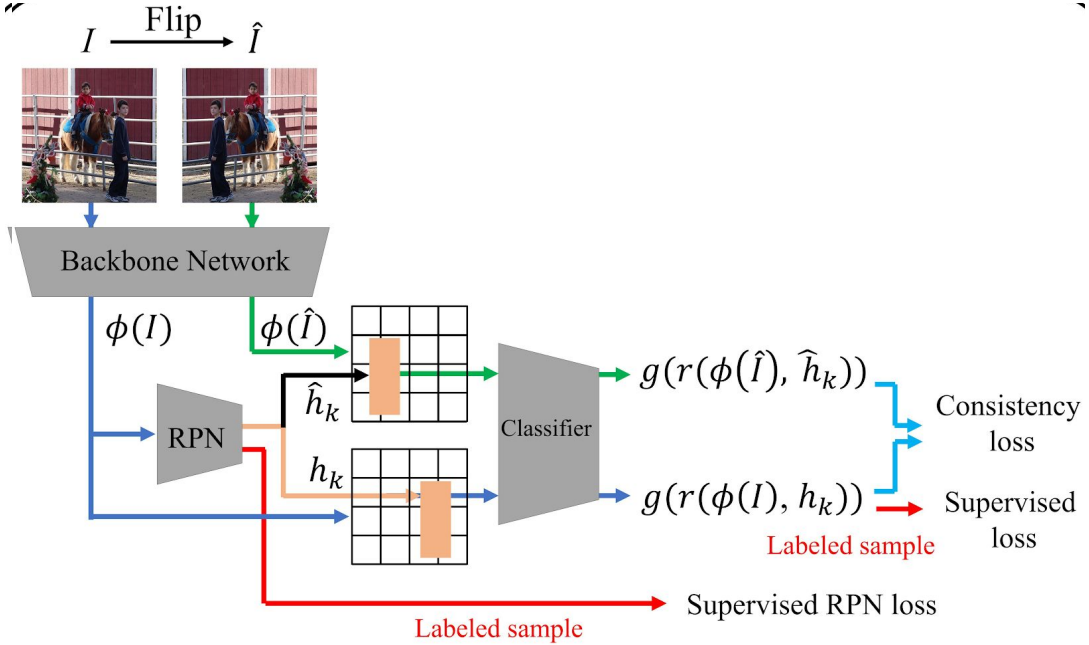


(b) two stage detector

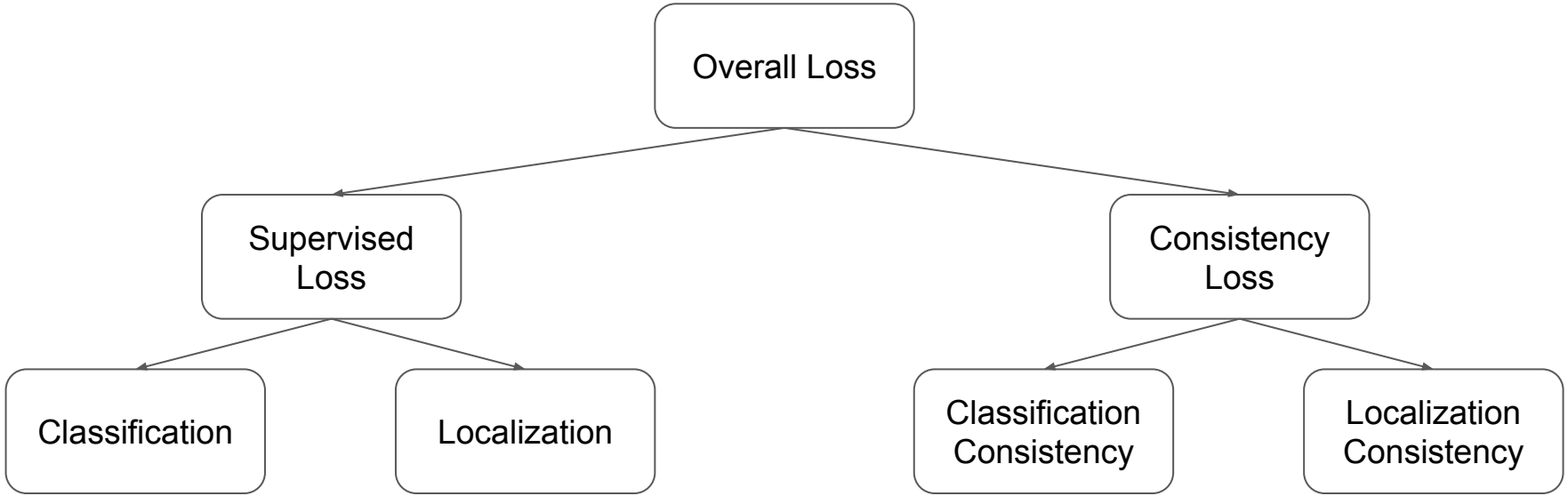
CSD-**Single-Stage Detector**



CSD-**Two-Stage Detector**



Loss function



Consistency Loss

Classification

For a pair of Bounding Box

$$l_{con_cls}(f^k_{cls}(I), f^{k'}_{cls}(\hat{I})) = JS(f^k_{cls}(I), f^{k'}_{cls}(\hat{I}))$$

Average classification loss for all Bounding box pair

$$\mathcal{L}_{con-c} = \mathbb{E}_k[l_{con_cls}(f^k_{cls}(I), f^{k'}_{cls}(\hat{I}))]$$

Localization

Prediction of Localization:

$$f^k_{loc}(I) = [\Delta cx, \Delta cy, \Delta w, \Delta h]$$

$$l_{con_loc}(f^k_{loc}(I), f^{k'}_{loc}(\hat{I})) = \frac{1}{4}(\|\Delta cx^k - (-\Delta \hat{cx}^{k'})\|^2 + \|\Delta cy^k - \Delta \hat{cy}^{k'}\|^2 + \|\Delta w^k - \Delta \hat{w}^{k'}\|^2 + \|\Delta h^k - \Delta \hat{h}^{k'}\|^2)$$

Average classification loss for all Bounding box pair

$$\mathcal{L}_{con-l} = \mathbb{E}_k[l_{con_loc}(f^k_{loc}(I), f^{k'}_{loc}(\hat{I}))]$$

Total Consistency Loss $\mathcal{L}_{con} = \mathcal{L}_{con-c} + \mathcal{L}_{con-l}$

Overall loss for Object Detector

$$\mathcal{L} = \mathcal{L}_c + \mathcal{L}_l + w(t) \cdot \mathcal{L}_{con}$$

Weight Scheduling for “Ramp-up and Ramp-down Strategies”

Background Elimination

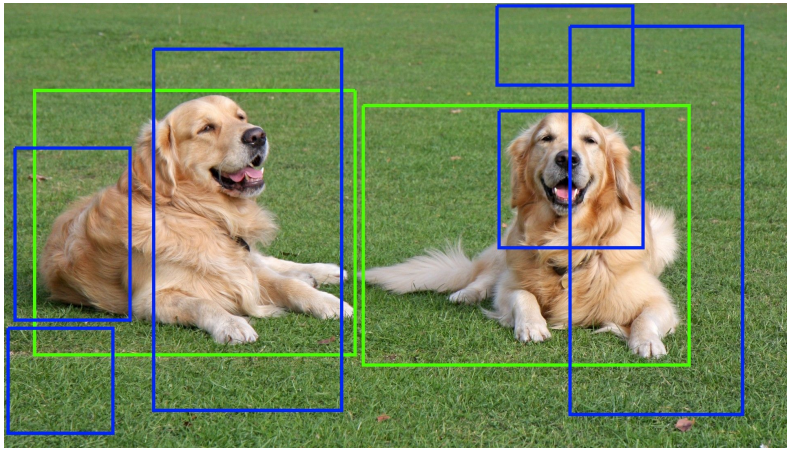


Image courtesy:
https://www.learnopencv.com/wp-content/uploads/2017/10/object-recognition-false-positives-true-positives.jpg

Consistency loss computed with all candidates will be easily dominated by background objects

$$m^k = \begin{cases} 1, & \text{if } \operatorname{argmax}(f^k_{cls}(I)) \neq background \\ 0, & \text{otherwise.} \end{cases}$$

Example:
For background object (in blue) the detector predicts some prediction value:
"Background" as "Dog": 0.3, 0.2, 0.3, 0.1, 0.2----> should be '0'
Loss: 0.3+0.2+0.3+0.1+0.2= **1.1**

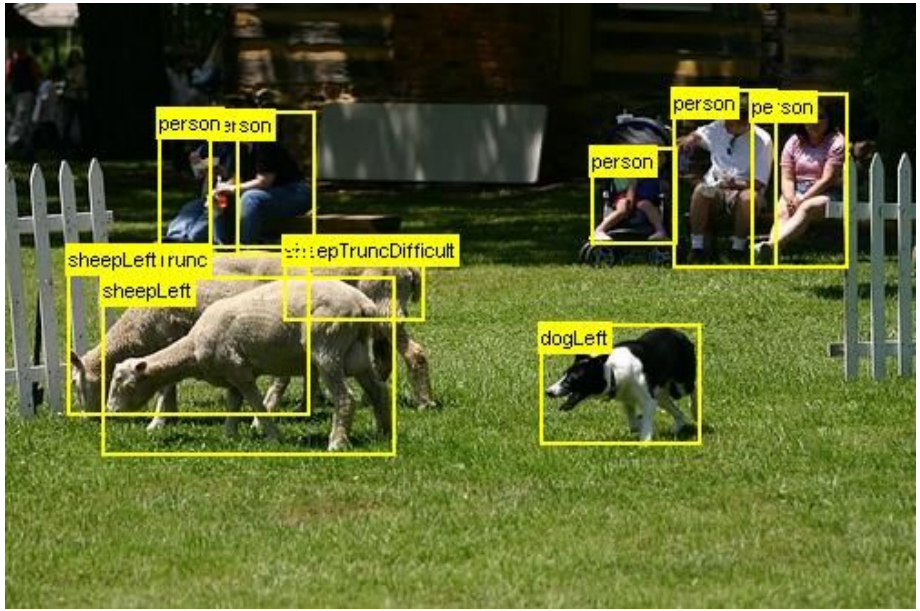
"Dog" as "Dog": 0.9, 0.8---->Should be '1'
Loss:0.1+0.2= **0.3**

Experiments

Datasets:

1. PASCAL VOC: 20 Classes
   a. VOC07- 5k images
   b. VOC12- 11.5k images

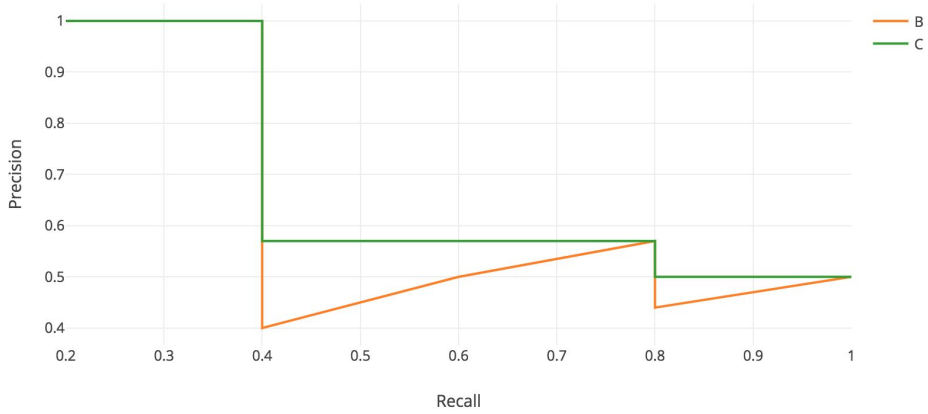
2. MSCOCO: 80 Classes



Experiments

Metric:

Mean Average Precision (mAP): AUC of Precision-Recall curve and take average across classes



Results

Labeled data	Unlabeled data	Consistency		Background Elimination	mAP (%)		
		cls	loc		SSD 300	SSD 512	R-FCN
VOC07	-	-	-	-	68.0*/70.2	71.6*/73.3	73.9
VOC0712	-	-	-	-	74.3*/77.2	76.8*/79.6	79.5*/79.4
VOC07	VOC12	✓	-	-	71.6 (1.4)	74.6 (1.3)	74.0 (0.1)
		-	✓	-	72.2 (2.0)	74.6 (1.3)	73.9 (0.0)
		✓	✓	-	72.0 (1.8)	74.8 (1.5)	74.0 (0.1)
VOC07	VOC12	✓	-	✓	71.7 (1.5)	75.4 (2.1)	74.5 (0.6)
		-	✓	✓	71.9 (1.7)	75.2 (1.9)	74.4 (0.5)
		✓	✓	✓	72.3 (2.1)	75.8 (2.5)	74.7 (0.8)

1. Using CSD method improves the mAP
2. Combined classification and localization consistency provides better mAP than the individual
3. Combined BE, classification and localization consistency provides the best mAP
4. Improvement in 2-stage object detector is lesser than that of 1-stage object detector

Results-Effect of Unlabeled data with different distribution

20 Classes

Labeled data	Unlabeled data	CSD Method (mAP)		
		SSD300	SSD512	R-FCN
VOC07	-	70.2	73.3	73.9
VOC07	VOC12	72.3	75.8	74.7
	VOC12+COCO[§]	71.7	75.1	74.9
	VOC12+COCO[†]	72.6	75.9	75.1

Results: Consistency loss without unlabeled data

Labeled data	Unlabeled data	Consistency		Eliminating Background	Method (mAP) SSD300
		cls	loc		
VOC07	-	-	-	-	68.0*/70.2
VOC0712	-	-	-	-	74.3*/77.2
VOC07	-	✓	-	-	69.4
		-	✓	-	69.9
		✓	✓	-	69.7
VOC07	-	✓	-	✓	70.2
		-	✓	✓	69.8
		✓	✓	✓	69.3

Limitations

1. Not significant improvement on 2-stage Object Detector (~0.8%)
2. Unlabeled data should have a similar distribution to the labeled data

Conclusion

Using Semi-supervised technique- Consistency regularization this paper improves the performance of the object detector using unlabeled data