Data-Uncertainty Guided Multi-Phase Learning for Semi-Supervised Object Detection

Zhenyu Wang   Yali Li Ye   Guo Lu Fang   Shengjin Wang

CVPR2021

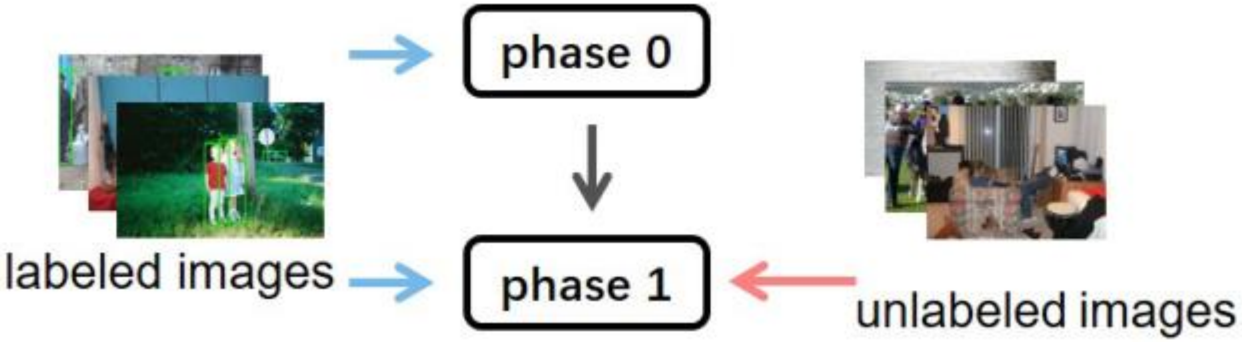
Motivation

◇ **weakly supervised object detection (WSOD)**: utilize large data with weak annotations, such as image labels, points.

◇ **semi-supervised object detection (SSOD)**: learn detectors with a small amount of box-level labeled images and large unlabeled images.

pseudo labels

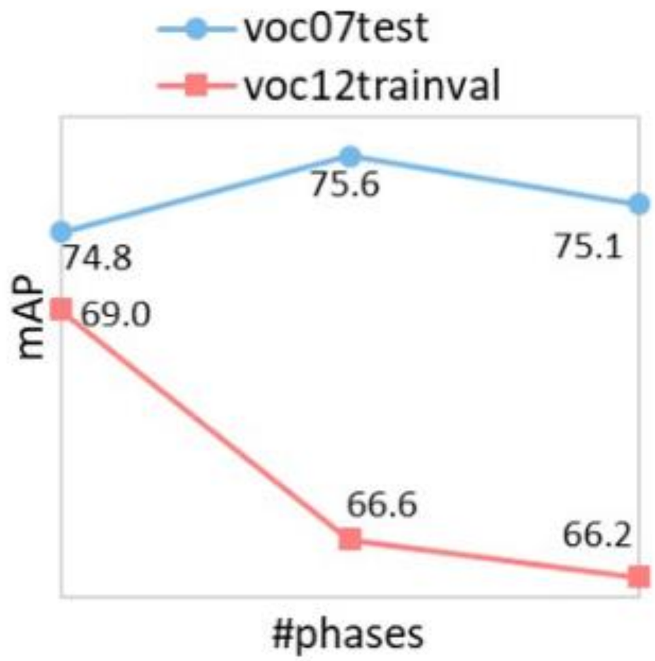
consistency



Motivation

When pseudo annotations are noisy with some false information, detection models are also able to learn to fit them. This fitting ability to incorrect annotations surpasses the representative learning for correct ones.

label noise overfitting problem



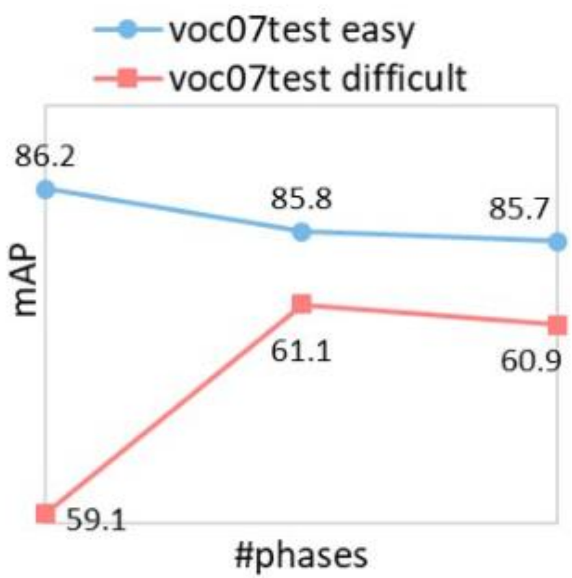
VOC 2007 trainval as labeled set
VOC 2012 trainval as unlabeled set

Label Noise Overfitting Problem

◇ image level

the model over-focuses on difficult images with more noise and ignores easy images

Image Uncertainty Guided Selection



◇ region level

some regions are similar to some existing objects but they are not highly overlapped with any positive instances.

Region Uncertainty Guided RoI Re-weighting

Image Uncertainty Guided Selection

the detected objects in image(i.e.pseudo labels): $\{(bb_{mn}, s_{mn})\}_{m=1}^{M}$

The average of all bounding boxes' scores inside an image measures the certainty degree of all annotations.

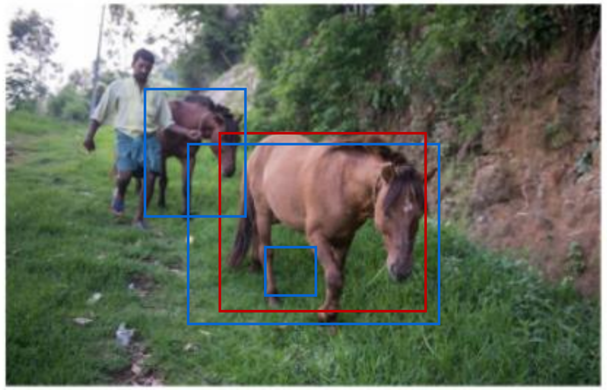
$$\overline{s}_m = \sum_{m=1}^{M} s_{mn}/M$$

Images with a small $\overline{s}_m$ are regarded as difficult ones and are filtered out in the first several phases.

Region Uncertainty Guided RoI Re-weighting

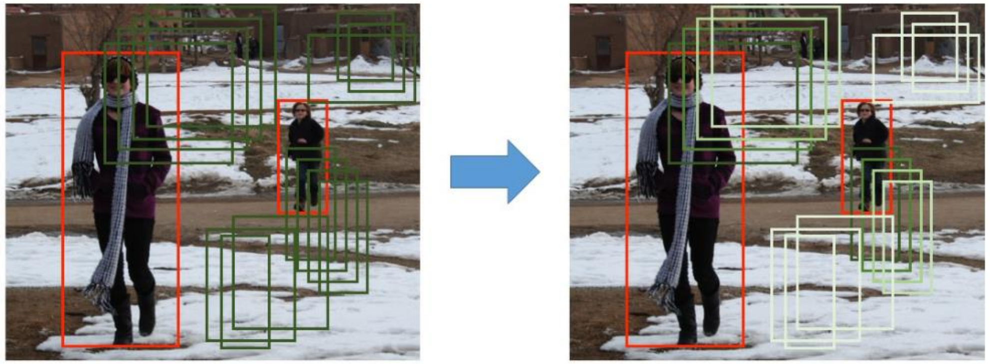
The strategy discovers uncertain regions and reduces their gradients by down-weighting to facilitate more accurate and certain regions standing out.

◇ overlap based uncertainty



$$w_i = a + (1 - a)e^{-be^{-c \cdot \mathrm{IoU}_i}}$$ [1]

Gompertz function:
$y(t) = ae^{-be^{-ct}}$



ROI Assignment

Gradient Weighting

[1] *Soft Sampling for Robust Object Detection*

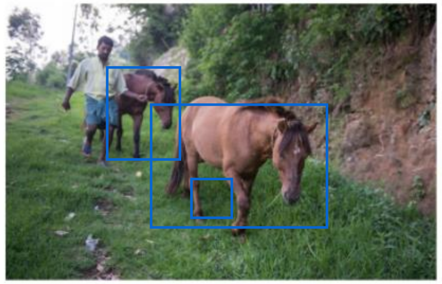
Region Uncertainty Guided RoI Re-weighting

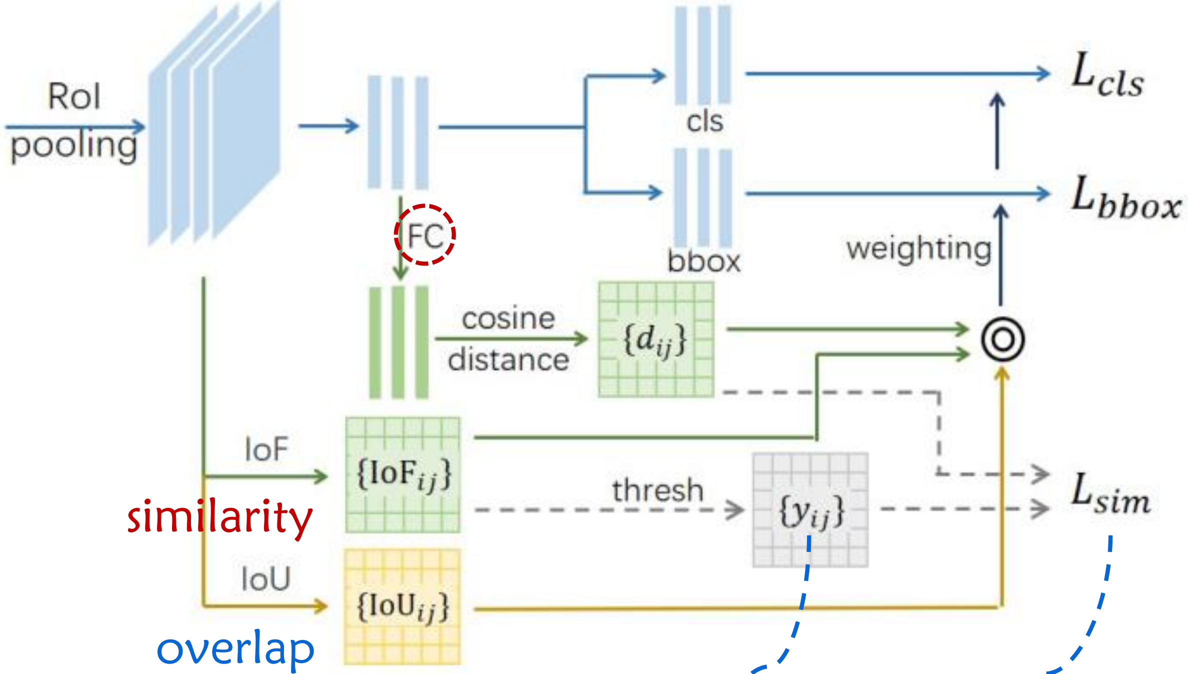
◇ similarity based uncertainty

$$D_i = 1 - \max_j d_{ij}(1 - \boxed{\mathrm{IoF_{ij}}})$$

Intersection over foreground

$$d_{ij} = \frac{|f_i^T f_j|}{\|f_i\|\|f_j\|}$$





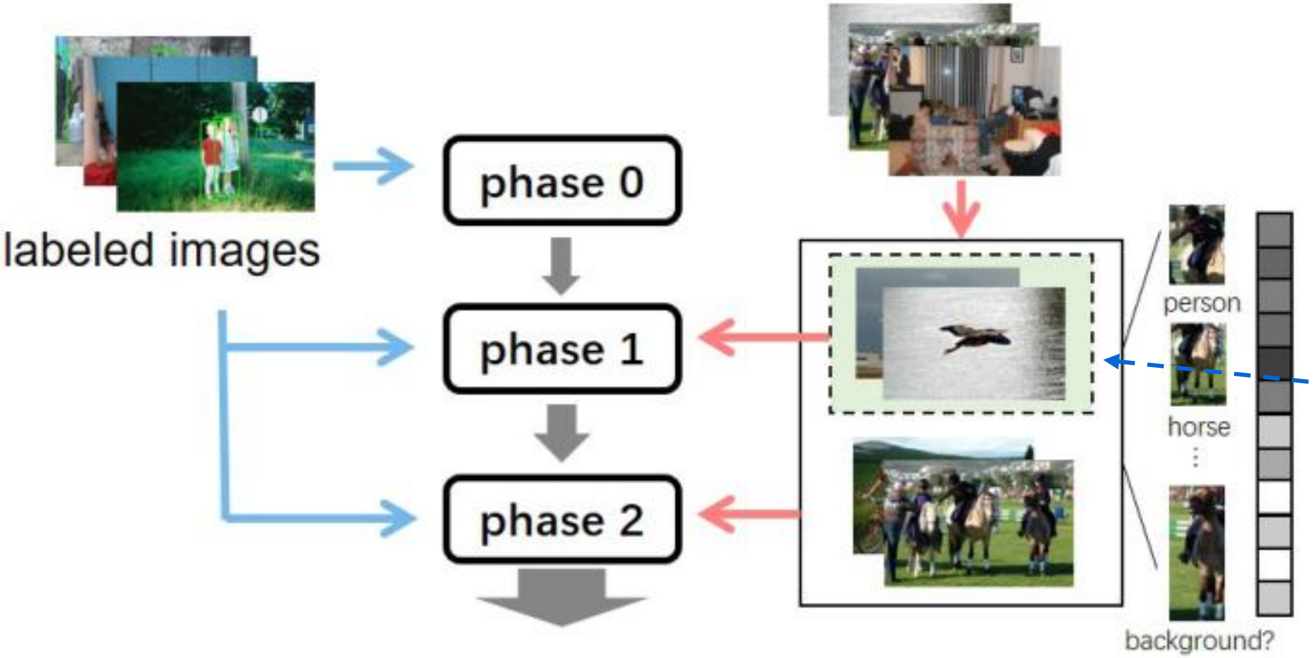
➡ ultimate uncertainty based weights

$$w_i = \underbrace{(a + (1-a)e^{-be^{-c_1 \cdot IoU_i}})}_{\text{overlap}}\underbrace{e^{-be^{-c_2 D_i}}}_{\text{similarity}}$$

$$y_{ij} = I(\mathrm{IoF_{ij}} > t)\ or\ I(\mathrm{IoF_{ji}} > t)$$

$$L_{sim} = y_{ij}(1 - d_{ij})^2 + (1 - y_{ij})d_{ij}^2$$

Multi-Phase Learning



Algorithm 1 The overall procedure for multi-phase SSOD learning.

Require:

The number of training phases, N

Training:

Train a FSOD model with all labeled data.

Set the initial easy data fraction: $k = 1/N$

for $i = 1; i <= N; i++$ **do**

1. Predict on unlabeled data with all current models.
2. Take the intersection for all current pseudo labels.
3. Select top k easy images from unlabeled images.
4. Train a SSOD model with labeled and easy unlabeled data.
5. $k = k + 1/N$

end for

Testing:

Ensemble testing results from all models to generate ultimate results.

Uncertainty Guided Multi-phase Learning

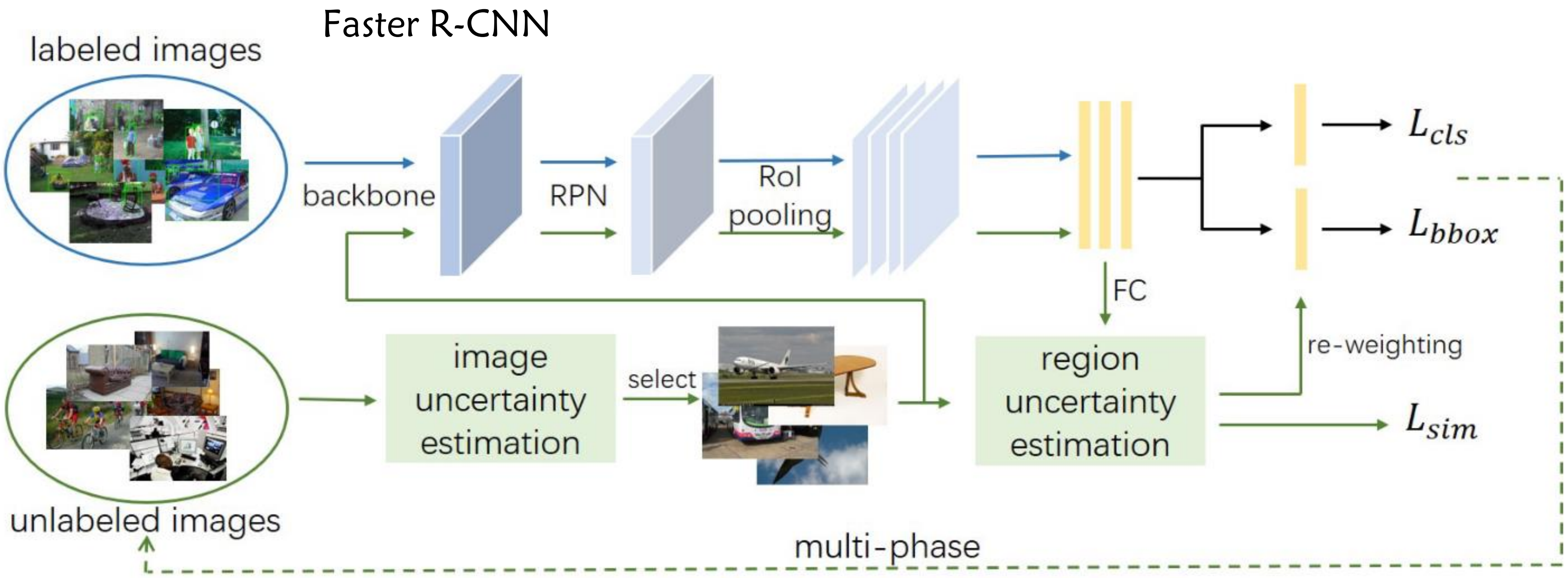


Figure 3: **The diagram of uncertainty guided multi-phase learning.** Multi-phase self-training is designed for unlabeled images flow in SSOD. Image uncertainty estimation and region uncertainty estimation guide the multi-phase SSOD learning.

Experiments

dataset: PASCAL VOC, MSCOCO

two-stage

one-stage

Table 1: **Semi-supervised Detection Results on PASCAL VOC 2007 test** *vs.* current SSOD methods and fully-supervised results trained on VOC07 or VOC0712. (L: labeled data, Un: unlabeled data.)

Model	Backbone	Method	L	Un	mAP
Faster RCNN	ResNet50	FS	VOC07	-	74.8
		Baseline	VOC07	VOC12	75.6
		DD [33]	VOC07	VOC12	76.0
		ours	VOC07	VOC12	**78.6**
		FS	VOC0712	-	81.2
SSD300	VGG16	FS	VOC07	-	70.2
		Baseline	VOC07	VOC12	71.8
		CSD [16]	VOC07	VOC12	72.3
		ISD [17]	VOC07	VOC12	73.3
		ours	VOC07	VOC12	**74.5**
		FS	VOC0712	-	77.2

Table 2: **Semi-supervised detection Results on COCO minival** *vs.* current SSOD and FSOD results. † denotes that the performance is obtained by the final model after the multi-phase learning without ensemble.

Backbone	Method	L	Un	AP	AP_{50}	AP_{75}
ResNet50	FS	co-35	-	31.3	52.0	33.0
	DD	co-35	co-80	33.1	53.3	35.4
	ours	co-35	co-80	**34.8**	**55.1**	**37.2**
	ours + DD	co-35	co-80	**35.2**	**55.7**	**37.6**
	FS	co-115	-	37.4	58.1	40.4
	DD	co-115	co-120	37.9	60.1	40.8
	PL [40]	co-115	co-120	38.4	59.7	41.7
	ours	co-115	co-120	**40.1**	**60.4**	**43.7**
	ours† + DD	co-115	co-120	38.9	59.4	42.3
	ours + DD	co-115	co-120	**40.3**	**61.0**	**43.9**
ResNet101	FS	co-115	-	39.4	60.1	43.1
	DD	co-115	co-120	40.1	62.1	43.5
	ours	co-115	co-120	**42.2**	**62.5**	**46.1**
	ours† + DD	co-115	co-120	41.2	61.5	44.9
	ours + DD	co-115	co-120	**42.3**	**62.7**	**46.3**

Ablation Study

Table 3: **Ablation Study on PASCAL VOC 2007 test.** (RR: RoI Re-weighting)

Model	L	Un	Two-Phase	RR	Ensemble	mAP
Faster RCNN	VOC07	-				74.8
	VOC07	VOC12				75.6
	VOC07	VOC12	✓			76.1
	VOC07	VOC12	✓	✓		77.4
	VOC07	VOC12	✓	✓	✓	**78.6**
SSD300	VOC07	-				70.2
	VOC07	VOC12				71.8
	VOC07	VOC12	✓			72.3
	VOC07	VOC12	✓		✓	**74.5**

We perform ablation study on PASCAL VOC to analyze the impacts of 1) multi-phase learning, 2) RoI Re-weighting strategy , 3) model ensembling during inference.

Ablation Study

◇ RoI Re-weighting Analysis

Table 4: **Effect of RoI Re-weighting on SSOD compared with Baseline and Soft Sampling.** $0 \sim 2$ is the ensemble result of model from phase 0 (FS model) to phase 2.

Phase	Baseline	Soft Sampling	RoI Re-weighting
0 (FS Model)	74.8	74.8	74.8
1	75.9	76.2	**76.6**
2	76.1	76.6	**77.4**
$0 \sim 2$	77.8	78.1	**78.6**

◇ Model Divergence Analysis

Table 5: **Detection results from different models.** $0 \sim 2$ indicates the ensemble result of model from phase 0 (fully-supervised model) to phase 2.

Phase	VOC07 Test	VOC07 Test (easy)	VOC07 Test (difficult)
0 (FS Model)	74.8	86.2	59.1
1	76.6	**86.7**	62.7
2	**77.4**	86.4	**63.8**
$1 \sim 2$	78.3	**87.3**	65.4
$0 \sim 2$	**78.6**	**87.3**	**66.2**

Discussion

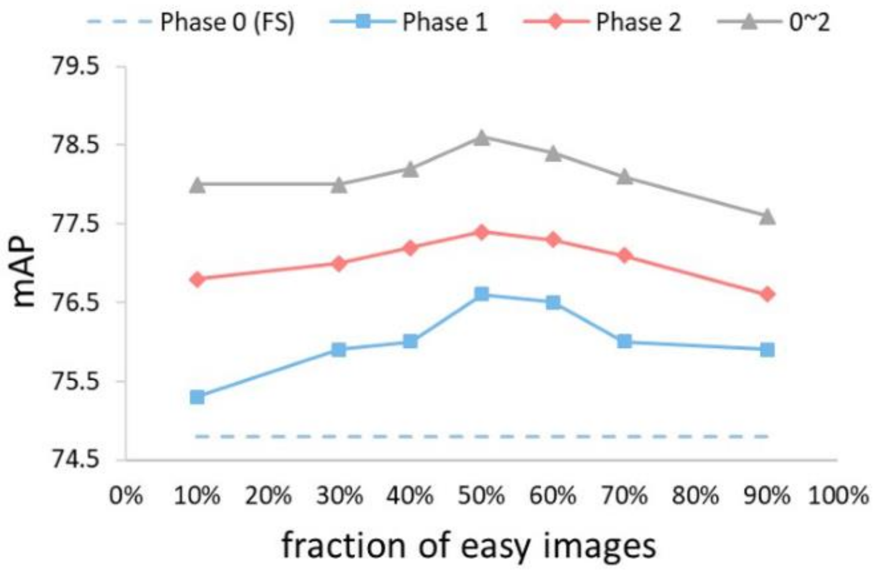


Figure 6: **Two-phase SSOD with different amount of easy data,** mAP reaches the peak when the ratio is 50%.

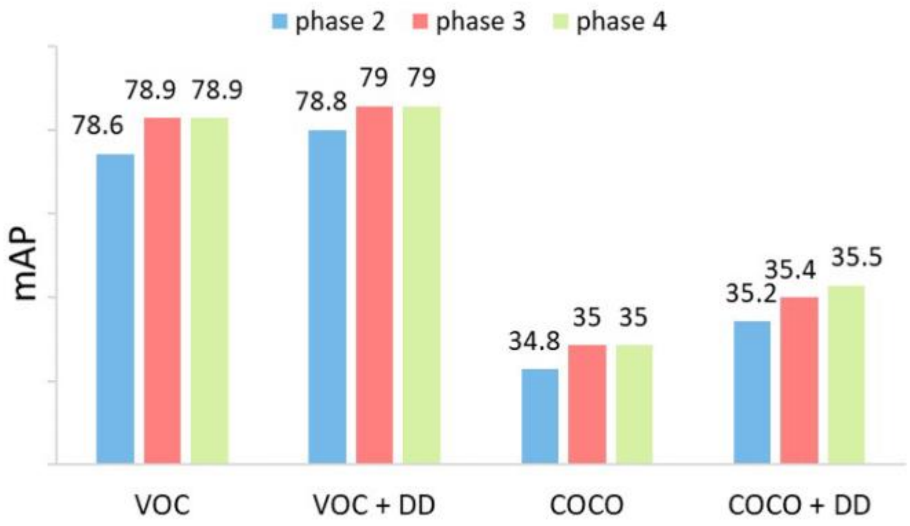


Figure 7: **Multiple phases semi-supervised learning on VOC07 test and COCO minival.**

More data do not necessarily lead to better performance for SSOD.

two-phase learning is a good choice in practice

Thank you