Localization-Aware Active Learning for Object Detection

Chieh-Chi Kao
University of California, Santa Barbara
chiehchi.kao@gmail.com

Teng-Yok Lee
Mitsubishi Electric Research Laboratories
tlee@merl.com

Pradeep Sen
University of California, Santa Barbara
psen@ece.ucsb.edu

Ming-Yu Liu
Mitsubishi Electric Research Laboratories
seanmingyuliu@gmail.com

ACCV 2018

Contents

Introduce

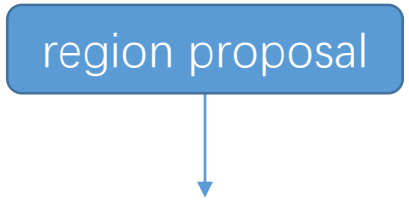
- Active learning has been shown to effective at annotating data for image classification, but largely unexplored for object detection.

- We present two metric for measuring the informativeness of an object hypothesis.
  - Localization Tightness (TL).
  - Localization Stability (LS).

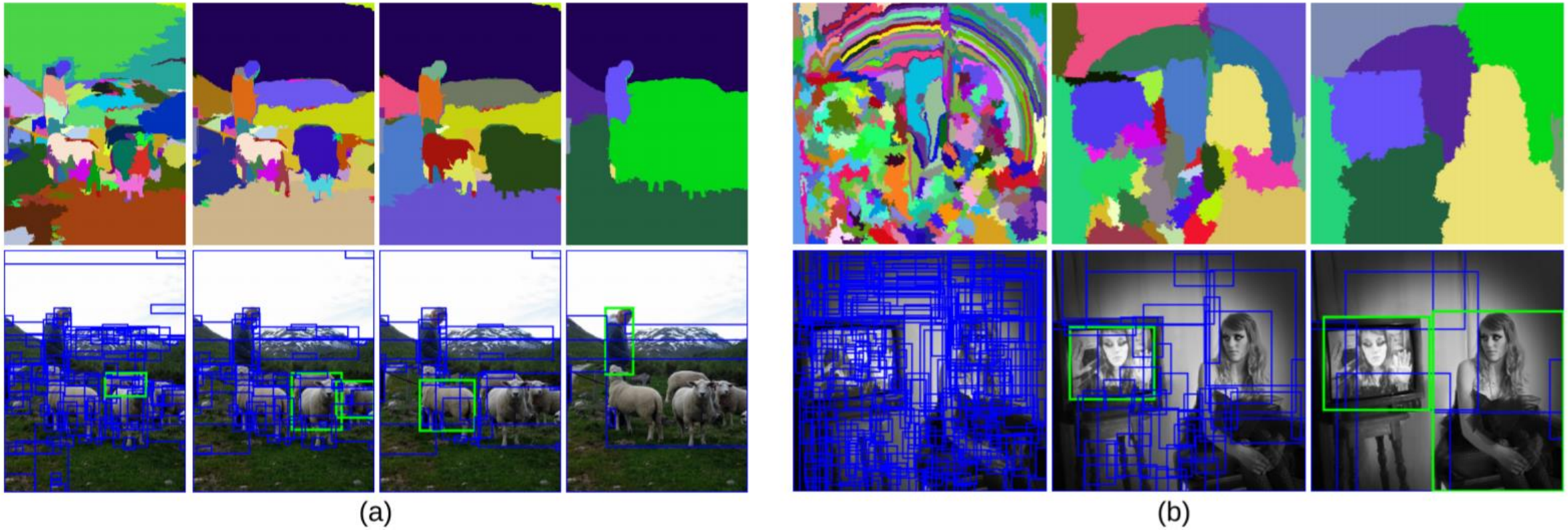
Background

- Active learning.

- Object Detection.
  - Two stage: *img → Feature extraction → RP → classification/ Locating regression*
    - **R-CNN、Fast R-CNN、Faster R-CNN**
    - SPP-NET
  - One stage： *img → Feature extraction → classification/ Locating regression*
    - OverFeat
    - **YOLOv1、YOLOv2、YOLOv3**
    - SSD

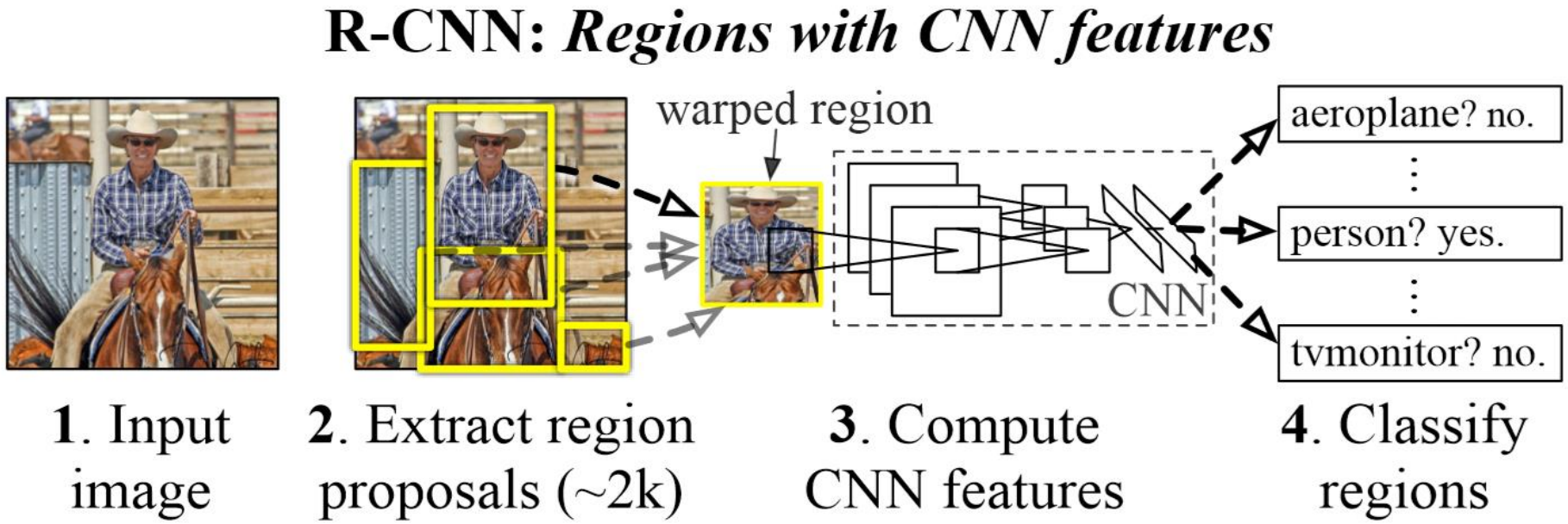


Background – RCNN (regions with cnn)



Selective Search for Object Recognition (IJCV 2013)

Background – RCNN (Regions with CNN)



- Process
  - Using SS to extract region proposals
  - Score each region proposals
  - Rank the scores of the candidate boxes， then remove some regions according to some rules. (NMS)
  - Make a border regression of the last remaining boxes. (Position、Size)

(a)

(b)

- Given a predicted box that is absolutely certain about its classification result ($P_{max} = 1$), but it cannot tightly enclose a true object ($T = 0$).
- Reversely, if the predicted box can tightly enclose a true object ($T = 1$), but the classification result is uncertain ($low\ P_{max}$).

$$J(B_0^j) = |T(B_0^j) + P_{max}(B_0^j) - 1| \qquad T_I(I_i) = min_j J(B_0^j)$$

↓ select

Method

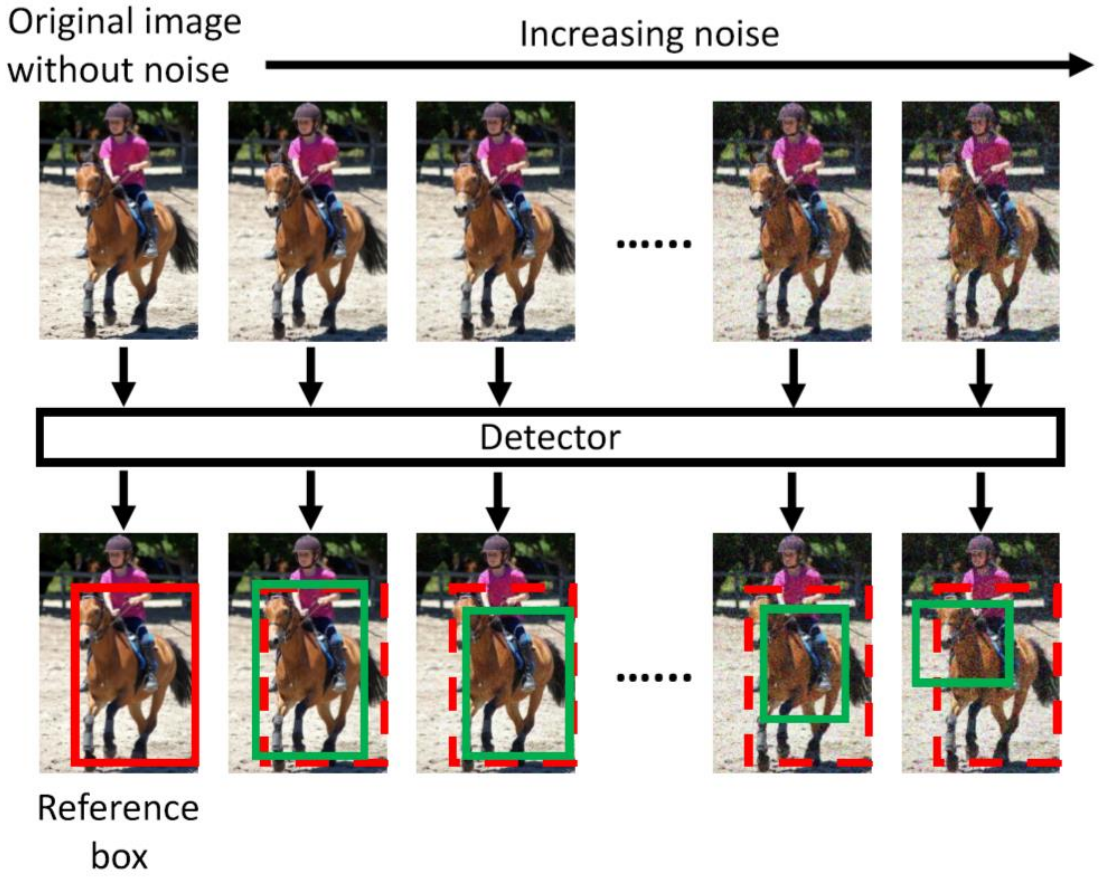
- Localization Stability

If the current model is stable to noise, meaning that the detection result does not dramatically change even if the input unlabeled image is corrupted by noise, the current model already understands this unlabeled image well so there is no need to annotate this unlabeled image.

$$S_B(B_0^j) = \frac{\sum_{n=1}^{N} IoU(B_0^j, C_n(B_0^j))}{N}$$

select ↓

$$S_I(I_i) = \frac{\sum_{j=1}^{M} P_{max}(B_0^j) S_B(B_0^j)}{\sum_{j=1}^{M} P_{max}(B_0^j)}$$

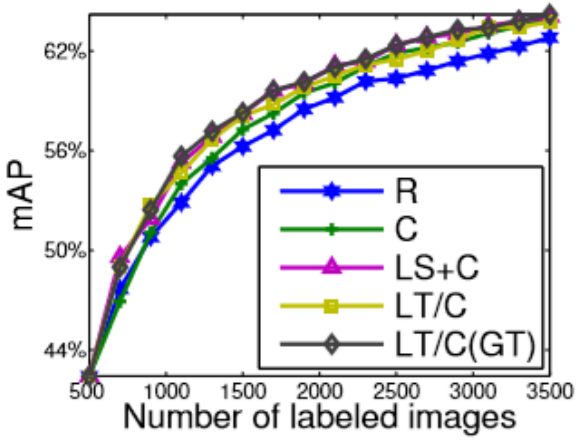


Experiment

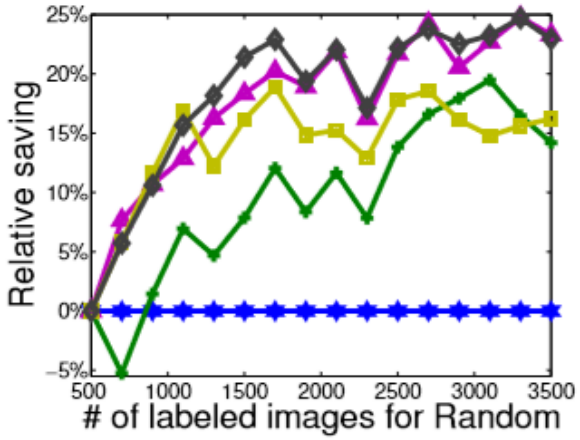
- Two baselines:
  - Random (R)
  - Classification Uncertainty (C)
  - Two baselines:

- Ours:
  - Localization stability only (LS)
  - Localization stability + Classification Uncertainty (LS+C)
  - Localization tightness + Classification Uncertainty (LT/C)
  - Localization tightness + Classification Uncertainty + Ground-truth boxes instead of the estimate used in LT/C (LT/C(GT))

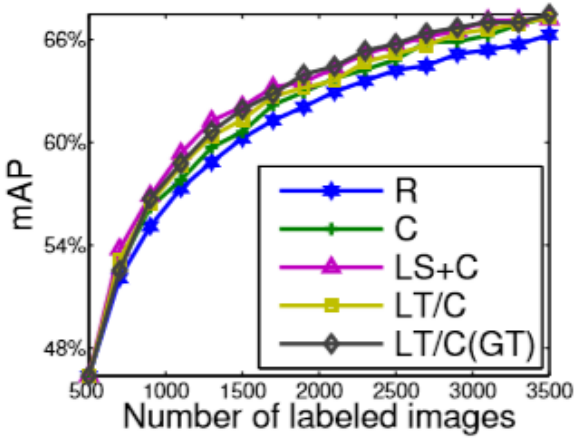
Experiment – FRCNN on PASCAL 2012 and PASCAL 2007



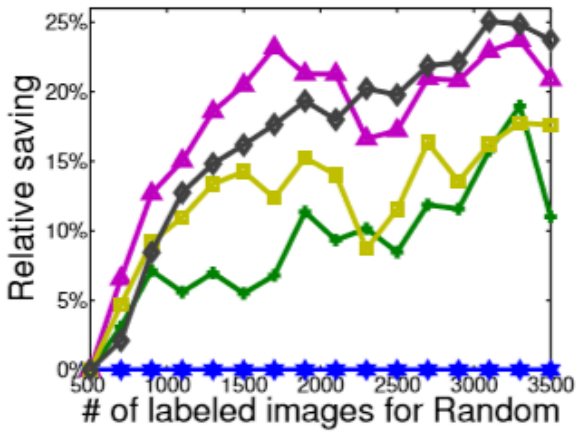
(a) mAP



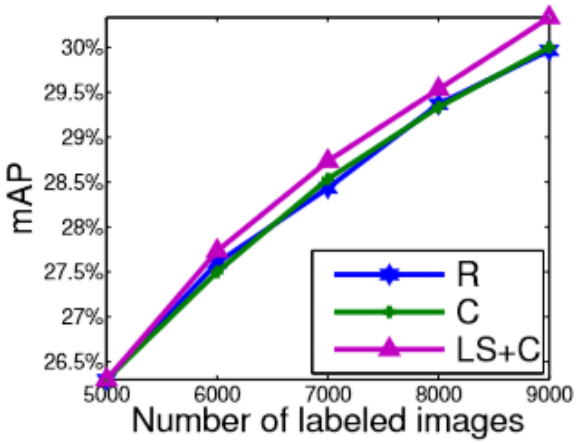
(b) Saving



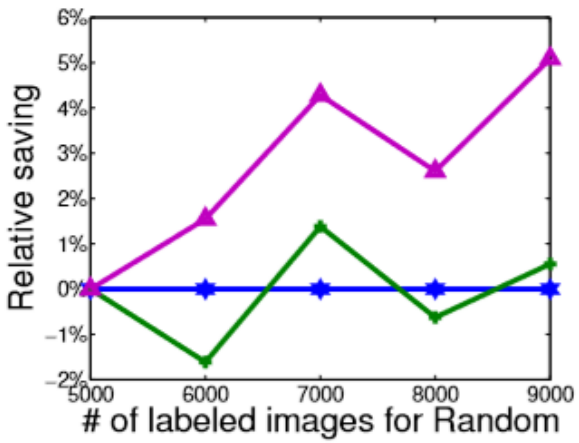
(a) mAP



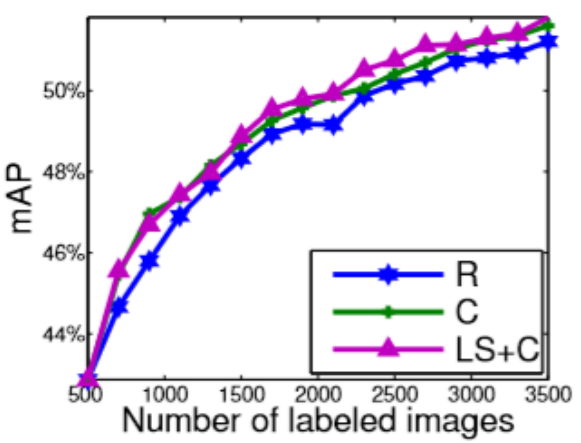
(b) Saving



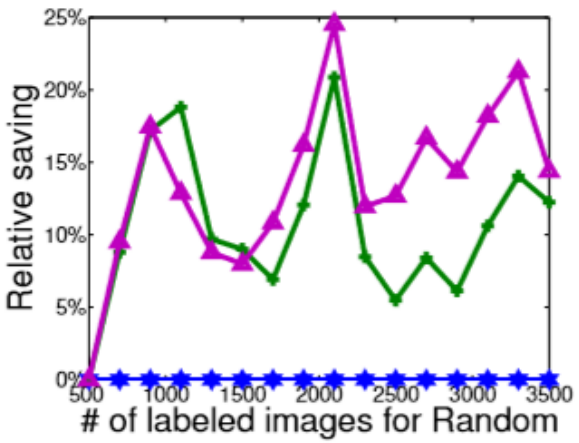
(a) mAP



(b) Saving



(a) mAP



(b) Saving

Experiment – FRCNN on PASCAL 2012 and PASCAL 2007

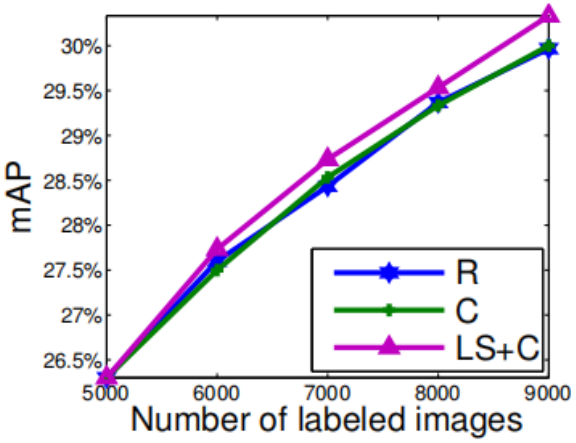
method	aero	bike	bird	boat*	bottle*	bus	car	cat	chair*	cow	table*	dog	horse	mbike	persn	plant*	sheep	sofa	train	tv	mAP
R	71.1	61.5	54.7	28.4	32.0	68.1	57.9	75.4	25.8	44.2	36.4	73.0	61.9	67.3	68.1	21.6	51.9	41.0	**65.5**	51.7	52.9
C	70.7	62.9	54.7	25.5	30.8	66.1	56.2	**78.1**	26.4	**54.5**	36.7	**76.9**	**68.3**	67.7	67.4	22.5	57.7	40.8	63.6	52.5	54.0
LS+C	**73.9**	63.7	**56.9**	**29.6**	**35.2**	66.5	58.5	77.9	**31.3**	50.8	40.7	73.8	65.4	66.9	68.4	**24.8**	**58.0**	**44.9**	64.2	53.9	**55.3**
LT/C	69.8	**64.6**	54.6	29.5	33.8	**70.3**	**59.7**	75.5	29.5	46.3	**41.8**	73.0	62.5	**69.0**	**70.8**	23.2	56.5	42.8	64.3	**55.9**	54.7

PASCAL 2012

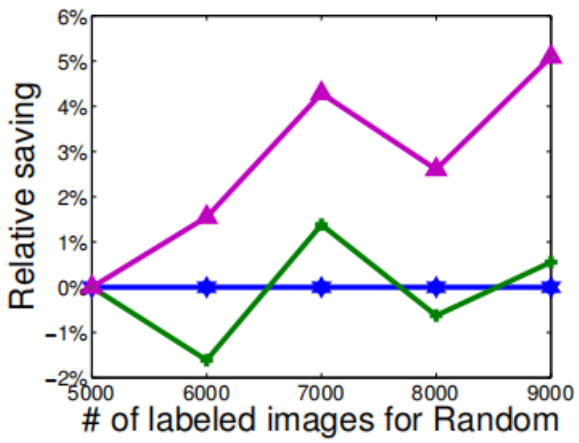
method	aero	bike	bird	boat*	bottle*	bus	car	cat	chair*	cow	table	dog	horse	mbike	persn	plant*	sheep	sofa	train	tv	mAP
R	**61.6**	67.2	54.1	40.0	33.6	64.5	73.0	73.9	34.5	60.8	52.2	69.3	**74.7**	66.6	67.1	25.9	52.1	54.2	**66.1**	54.9	57.3
C	56.9	68.0	54.9	36.8	34.4	68.1	71.7	**75.5**	34.0	**68.6**	51.0	71.4	74.7	65.2	65.9	24.9	60.0	53.9	63.0	57.4	57.8
LS+C	61.5	64.4	**55.8**	40.2	**38.7**	66.3	73.8	74.7	**39.6**	68.0	56.3	**71.5**	73.8	**67.2**	66.7	27.7	**61.3**	**57.0**	65.6	57.4	**59.4**
LT/C	57.6	**69.7**	52.9	**41.1**	38.4	**69.7**	**74.4**	71.8	36.4	61.2	**58.1**	69.5	74.3	66.2	**67.8**	**28.0**	55.5	56.3	65.5	**58.2**	58.6

PASCAL 2007

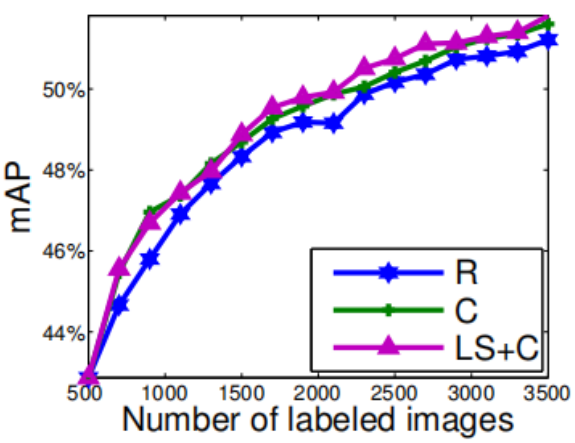
Experiment – FRCNN on MS COCO 2012 and SSD on PASCAL 2007



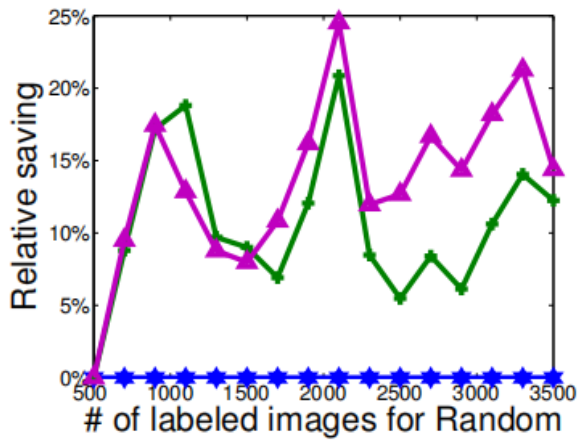
(a) mAP



(b) Saving



(a) mAP



(b) Saving

FRCNN on MS COCO 2012

SSD on PASCAL 2007