南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Consistency-based Active Learning for Object Detection

Weiping Yu[1], Sijie Zhu[2], Taojiannan Yang[2], Chen Chen[2]

[1]School of Computer Science and Engineering, Nanyang Technological University, Singapore

[2]Center for Research in Computer Vision, University of Central Florida, USA

weiping001@e.ntu.edu.sg

{sizhu,taoyang1122}@knights.ucf.edu; chen.chen@crcv.ucf.edu

张婧炜

Background

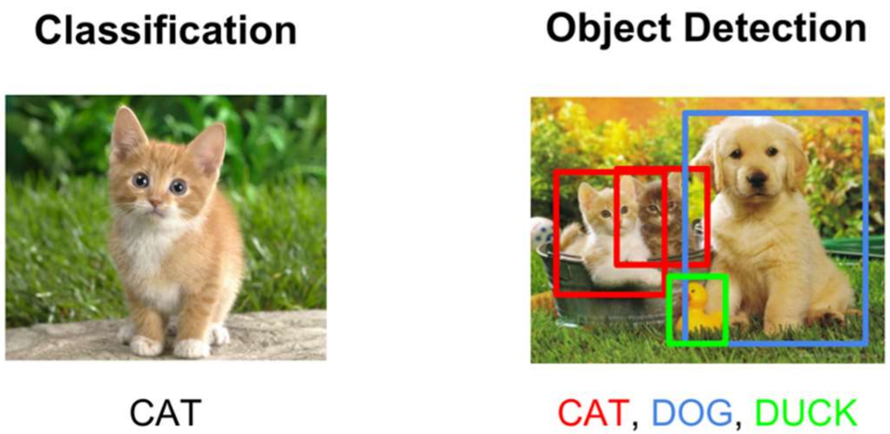


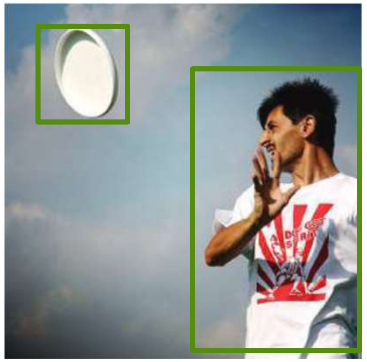
Image pair

original image

data augmentations

(cutout, rotation……)

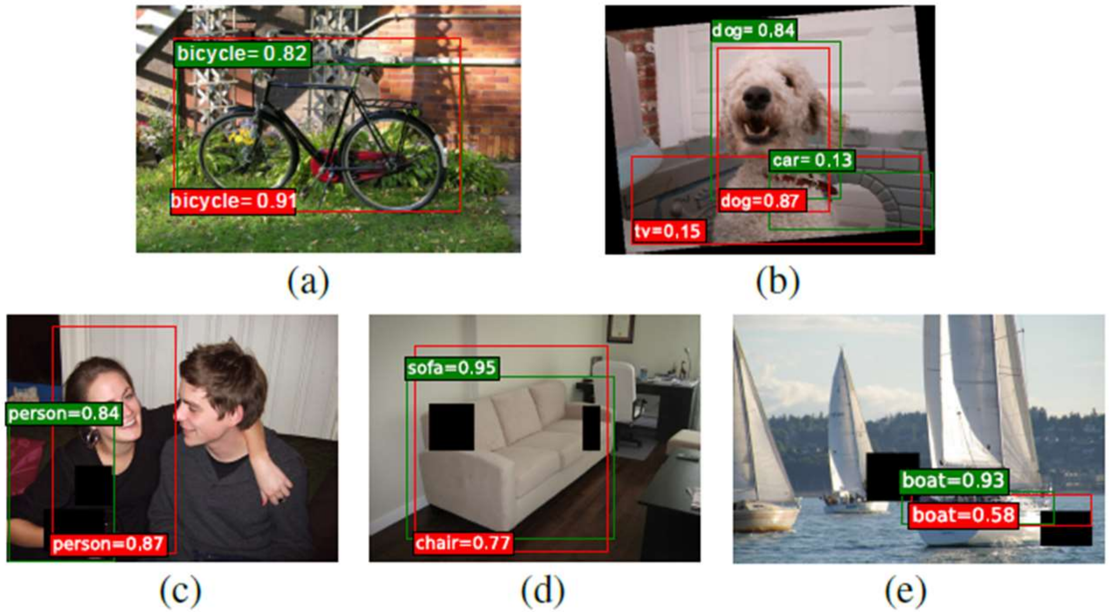
augmented image



model prediction

model prediction

Consistency metric



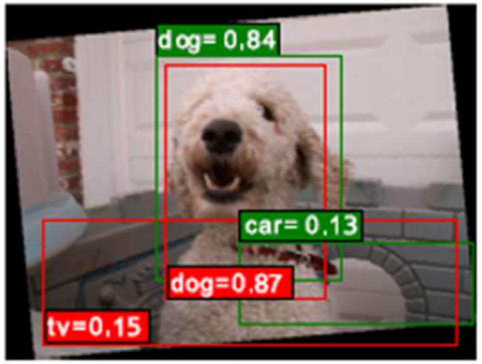
regression: $C_k^b = IoU(b_k, b'_k)$

consistency: $m_k = C_k^b + C_k^s$

classification: $$C_k^s = \underbrace{\frac{1}{2}[\max_{\varphi_n \in s_k}(\varphi_n) + \max_{\varphi'_n \in s'_k}(\varphi'_n)]}_{\text{weight factor}}(1 - JS(s_k||s'_k))$$

Base point β

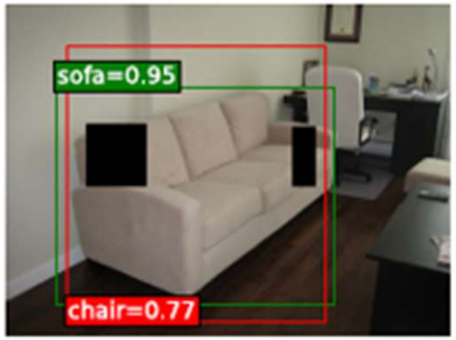
smallest m_k does not necessarily represent the most informative patch



(b)

- m_k keeping a certain distance from the lower bound
  - high matching degree and high confidence
- m_k being far away from the upper bound
  - this prediction is likely to be inaccurate

(c)



(d)

image: $$M(x_u; \mathcal{A}, \Theta) = \mathbb{E}_{\mathcal{A}}[\min_k |m_k - \beta|]$$

Base point β. Base point is the parameter β of consistency-based metric in Eq. 6, which denotes the value of m_k of the most informative prediction. From the plots in Fig. 7c, β

Mutual Information

标记池类别分布$\Delta_L(Y_L)$

$$\Delta_L(Y_L) = Softmax([\delta_1, \delta_2, \cdots, \delta_m, \cdots]^{\mathrm{T}})$$

$$\delta_m = \sum_{y_L \in Y_L} \mathcal{I}(y_L = m)$$

查询样本类别分布$\Delta_U(x_U)$

$$\Delta_U(x_U) = Softmax([\delta_1, \delta_2, \cdots, \delta_m, \cdots]^{\mathrm{T}})$$

$$\delta_m = \max_{s_k \in \{s_k\}} \{\varphi_m | \varphi_m \in s_k\} + \max_{s'_j \in \{s'_j\}} \{\varphi'_m | \varphi'_m \in s'_j\}$$

Algorithm 1 Selection by mutual information in each cycle

Input: Initial selected pool X_I, ground truth of labeled pool Y_L, total budget B, budget per cycle B/C

Functions: Distribution function of labeled pool $\Delta_L(Y_L)$ and single unlabeled image $\Delta_U(x_U)$

$X_F \leftarrow \{\}$

while size(X_F) $< B/C$ **do**

$f = \text{argmax}_{x_U \in X_I} [JS(\Delta_U(x_U) || \Delta_L(Y_L))]$

$X_F = X_F \cup \{X_I[f]\}$

$X_I = X_I - \{X_I[f]\}$

end while

return X_F

Framework

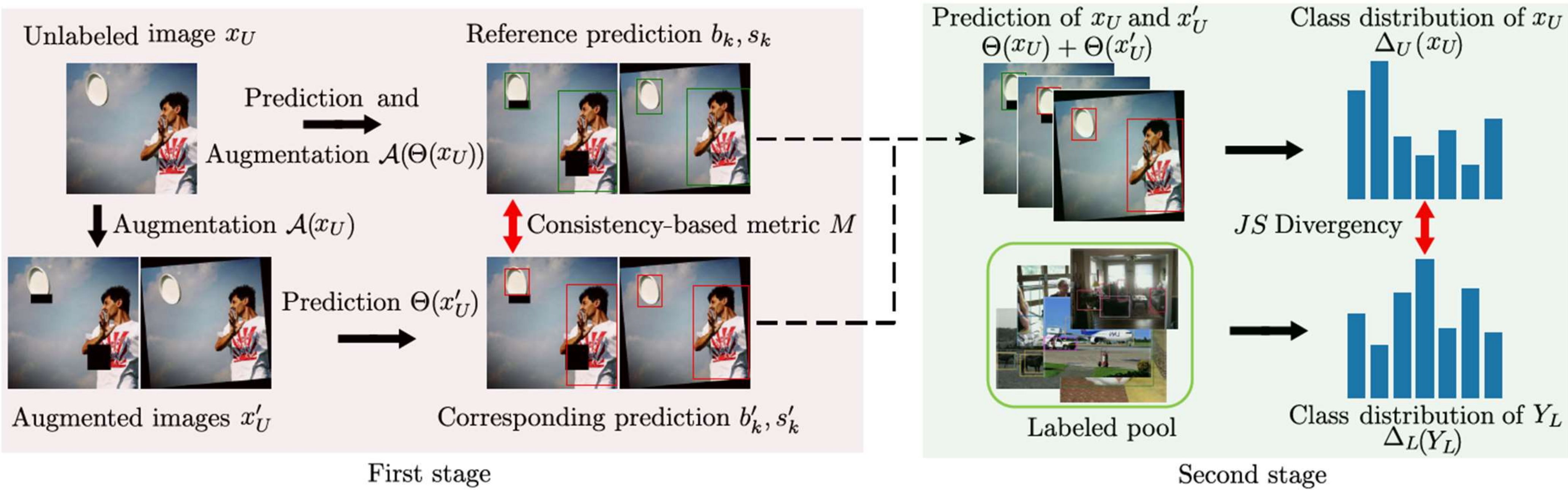


Figure 2. The two stages of the proposed CALD. The first stage selects samples based on individual information while the second stage uses mutual information to further refine the selected samples. Individual information is assessed by the consistency-based metric of reference and corresponding predictions. Mutual information refers to the JS divergence of class distributions of an unlabeled image and the labeled pool.

Experiments

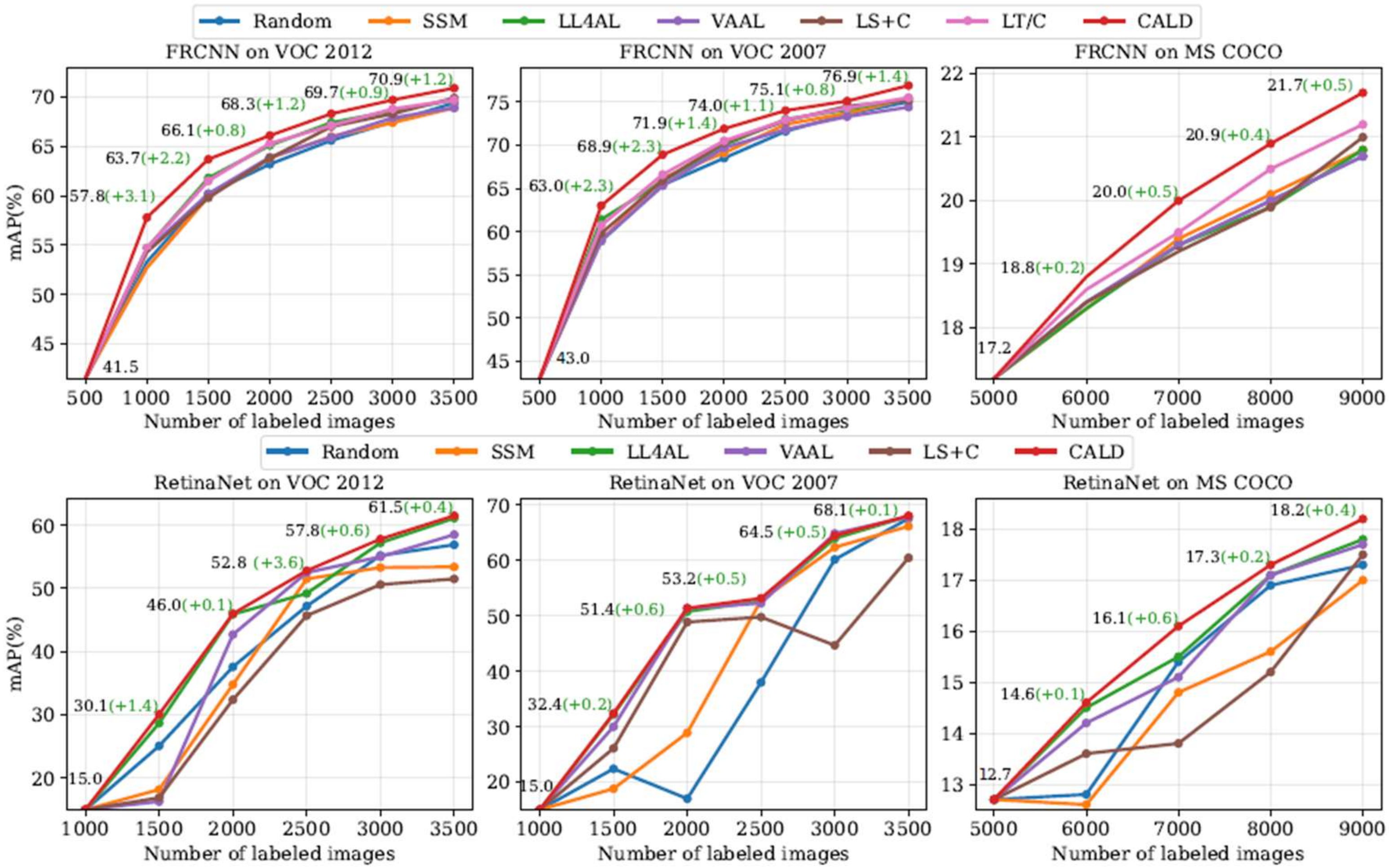


Figure 5. Comparison with SOTA task-agnostic and detection-specific active learning methods (LT/C only applies to two-stage detectors). CALD surpasses all methods comprehensively on three datasets and two detectors. The numbers marked on the points of CALD denote performance and its improvement over the second-best method. In the first row, the second-best methods are all LT/C [21] while the second-best methods are all LL4AL [44] in the second row.

Difficult categories

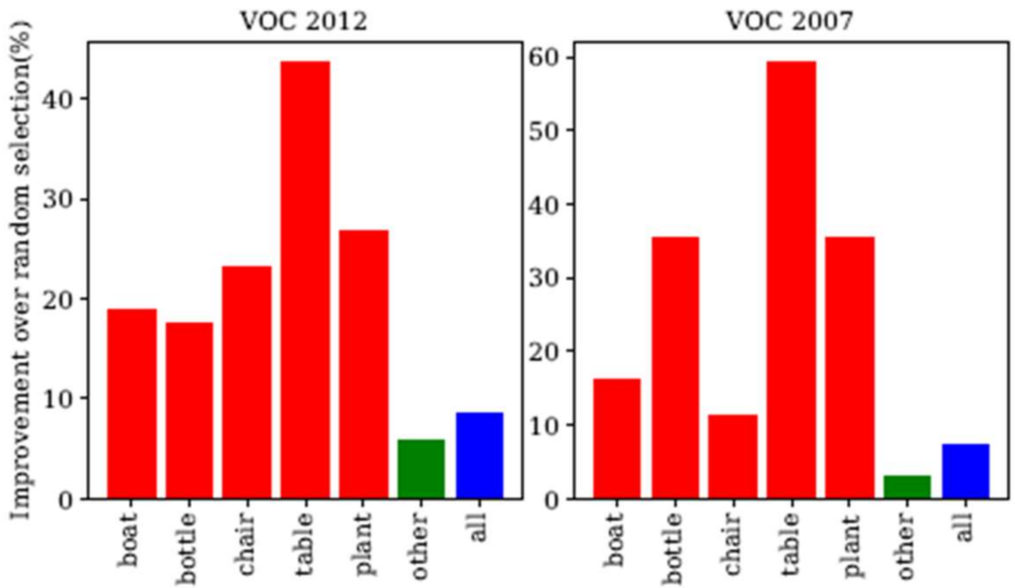


Figure 6. The improvement in difficult classes (red bars) over random selection in the first cycle on VOC. Green and blue bars are improvements for non-difficult classes (others) and all classes.

We also note that CALD yields more improvements in difficult categories. For categories with AP lower than 40% in random selection, we treat them as difficult categories. For the difficult categories (red bars) in Fig. 6, we notice that the improvements are larger than other classes.

Ablation studies

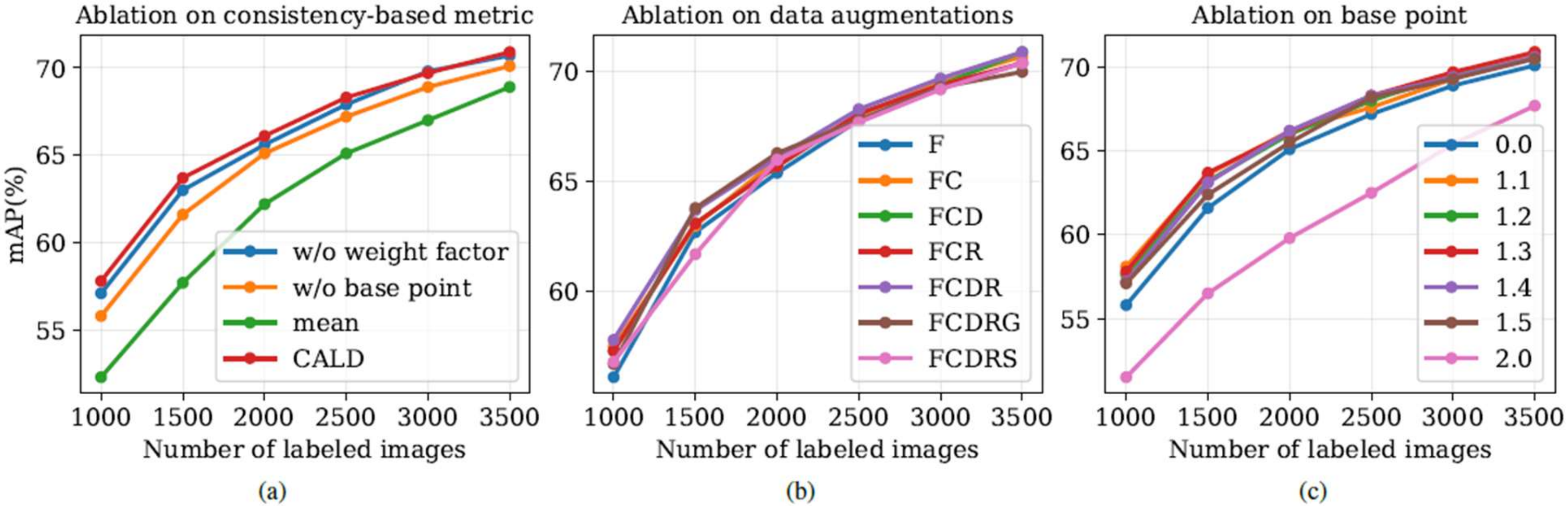


Figure 7. Ablation studies on strategies of consistency-based metric, data augmentations, and base point β.

For simplicity, we use a single uppercase letter to denote one type of augmentation. "F" for horizontal flip, "C" for cutout, "D" for downsize, "R" for rotation, "G" for Gaussian noise, and "S" for salt and pepper noise. The combi-

for all datasets and detectors in our experiments. When the value of β goes from the optimal point to the lower bound (0.0) of m_k, the performance of CALD decreases slowly. It can be explained that when m_k is closer to the lower bound, the predictions become unstable which are not necessarily informative. On the contrary, if β is excessively closer to the upper bound, the performance drop quickly. This is because m_k close to the upper bound denotes uninformative predictions (cases like Fig. 3a). When m_k reaches the upper bound (2.0), CALD selects the least informative samples (performance of detector is even worse than random

Ablation studies

Expansion ratio	mAP of 1st cycle	mAP of 2nd cycle
0%	56.9	62.8
10%	57.4	63.3
20%	**57.8**	**63.7**
30%	57.3	63.5
40%	57.1	63.7

Table 2. Ablation on the expansion ratio for X_I.

report the results in Table 2. Note that 0% in this table indicates our method reduces to **only have the first stage**. We reach two conclusions. (1) Based on the results of 0%, 10% and 20%, there is a clear advantage of leveraging mutual information for sample selection in the second stage. (2) 20% additional budget for X_I yields the best performance, leading to an mAP improvement of 0.9 in both cycles (56.9 *vs.* 57.8; 62.8 *vs.* 63.7). However, keep expanding the budget in the first stage would also cause performance drop (e.g. 30% ratio). This is because more informative samples may be removed by mutual information in the second stage in order to cut back to the fixed budget. Therefore, the experimental results reveal the importance of both individual and mutual information for sample selection.

The end

Thanks for listening!!!