An Adaptive Supervision Framework for Active Learning in Object Detection

[1]Sai Vikas Desai, [1]Akshay C Lagandula, [2]Wei Guo, [2]Seishi Ninomiya, [1]Vineeth N Balasubramanian

[1]Indian Institute of Technology Hyderabad [2]University of Tokyo

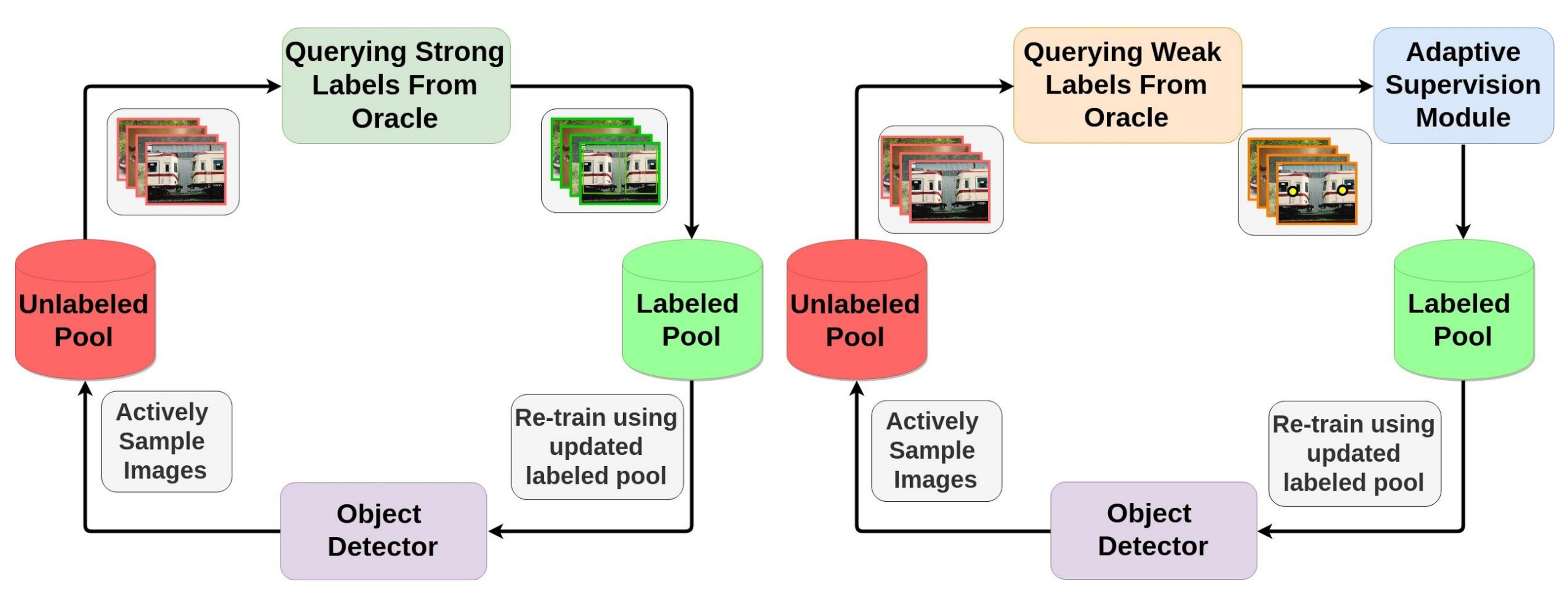
Problem Statement

Training an efficient object detection model while minimizing the time required for annotating the dataset.

Contributions

- The idea of using weak supervision for better performance in using active learning for object detection.
- Various methods for interleaving weak and strong supervision in a standard pool based active learning (PBAL) setting.
- Experimental evaluation of the proposed method on PASCAL VOC 2007, 2012 and an agricultural dataset of Wheat images.

Standard PBAL vs Our Framework



Standard PBAL Proposed Framework

Overview of our Method

Active Learning:

- Object detection model is trained in cycles. In each cycle, a batch of images is intelligently picked and an queried for labeling.
- An oracle labels the queried images and the dataset is updated using which our model is trained.

Multiple forms of Supervision

Description of Supervision Techniques	
Strong Supervision	Weak Supervision
Drawing tight bounding boxes around an object	Approximately clicking on an object's center of gravity

- Bounding box annotations are time consuming; hence weak labels are queried for the data initially.
- Based on a switching criterion, the adaptive supervision module decides whether to switch to a stronger form of supervision.
- Given an annotation budget in terms of time, our method optimizes the model performance while using a mix of weakly labeled and strongly labeled images for training.

Supervision Switching

Adaptive supervision module has two switching techniques to switch between strong and weak supervision - hard switching and soft switching.

Hard (episode-level) Switch:

- A hard switch from weak to strong supervision is made if the following condition evaluates to 1.

$$S_{hard}(n) = \begin{cases} 1 & \text{if } \frac{d_n}{d_{max}} \leq \gamma \\ 0 & \text{otherwise} \end{cases}$$

d_n = increase in validation mAP w.r.to previous cycle.
d_{max} = max observed increase in validation mAP.
γ = a suitably chosen threshold ∈ [0,1].

Soft (image-level) Switch:

- For an image i, supervision is switched from weak to strong if the following condition evaluates to 1.

$$S_{soft}(i) = \begin{cases} 1 & \text{if } c_i < \delta \\ 0 & \text{otherwise} \end{cases}$$

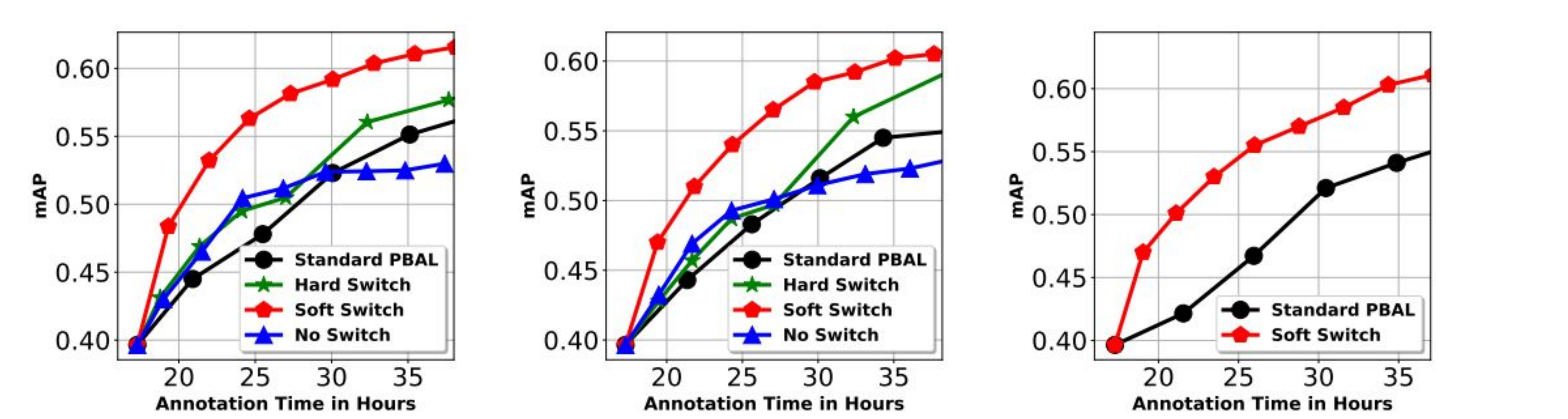
c = mean confidence over all objects in the image.
δ = suitably chosen probability threshold.

Experiments and Results

- Using Faster R-CNN as object detection model, experiments performed using three active learning techniques: avg-entropy, max-margin and least-confident.

Evaluation Metrics:

- Model Performance : mean average precision (mAP)
- Annotation time (ImageNet statistics):
  Strong supervision - 34.5 seconds / bounding box
  Weak supervision - 3 seconds / object-center click

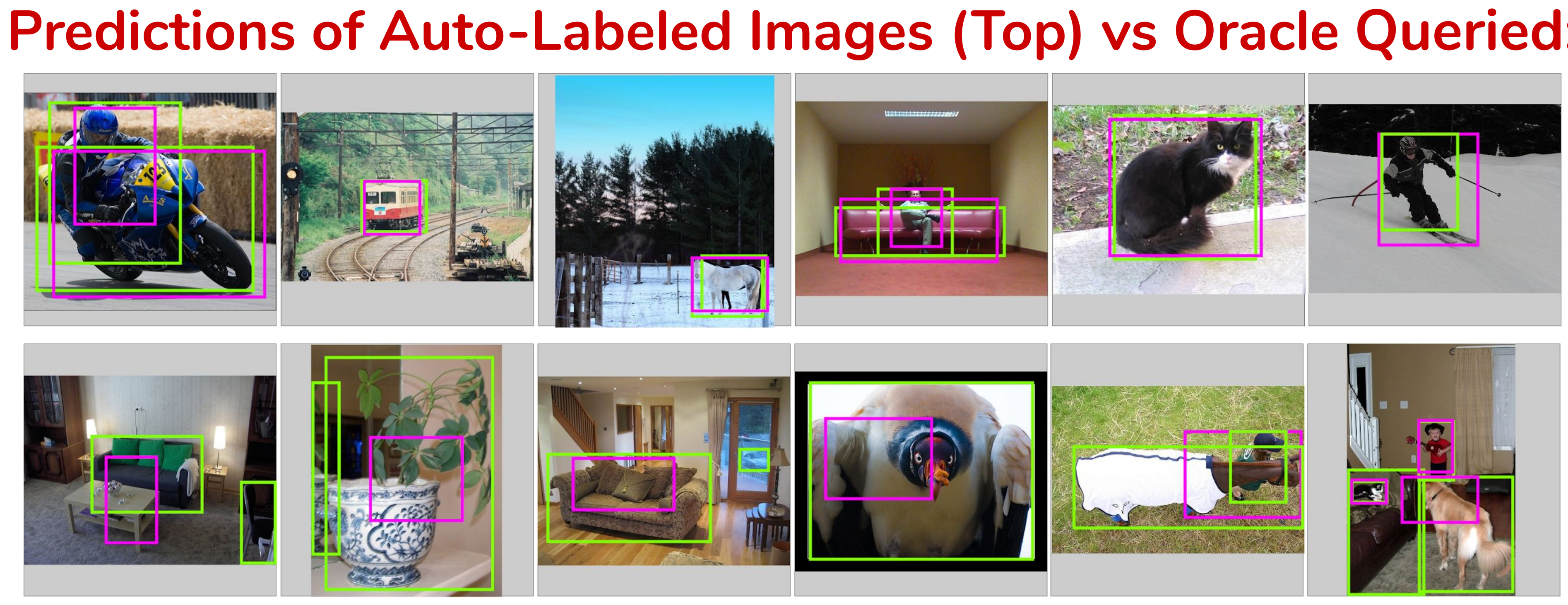


Avg-entropy Max-margin Least-Confident

Results:

- 30% savings in annotation time for PASCAL VOC 2007.
- 24% savings in annotation time for a real-world agricultural dataset of Wheat Head Detection.

Predictions of Auto-Labeled Images (Top) vs Oracle Queried:



Green Boxes = Ground Truth Magenta Boxes = Model Predictions

Key Insight:
Combining weak & strong supervision helps in training effective object detectors under a limited labeling budget.

arXiv