BAOD: Budget-Aware Object Detection (Supplementary Material)

We provide solutions for the proposed optimization functions in the first section. We also show the Budget vs mAP curves for the most relevant experiments and some visualizations of weakly and strongly labelled images chosen by the best method BAOD. Finally, we give the raw data of the compared experiments, including the *medium* cost one.

1. Solutions to Optimization Functions

1.1. Image Sampling

In section 3.1.1 and 3.1.2 of the draft, we proposed two linear functions to approximation the mAP increment. Then we try to **maximize** it with some restrictions. However, the integer programming problem is NP-hard, so it cannot be solved in polynomial time. Since the Branch and Bound Algorithm (B&B) takes more than 24 hours to find a global solution, we take the relaxation of the original integer program and uses a collection of linear restrictions. Then the relaxation Eq.1 can be solved in linear time.

$$
\begin{aligned}
\hat{\boldsymbol{x_1}}, \hat{\boldsymbol{x_2}}, \hat{\boldsymbol{x_3}} = \underset{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3} \in [0,1]^N}{\arg\max} \quad & \boldsymbol{s}^\top(\boldsymbol{x_1} + \boldsymbol{x_3}) + (\boldsymbol{\mu} - \boldsymbol{s})^\top \boldsymbol{x_2} \\
\text{s.t.} \quad & \boldsymbol{x_3} \leq \boldsymbol{\psi} \\
& \boldsymbol{x_1} + \boldsymbol{x_2} \leq \boldsymbol{1} - \boldsymbol{\psi} \\
& \boldsymbol{1}^\top(a\boldsymbol{x_1} + b\boldsymbol{x_2} + c\boldsymbol{x_3}) \leq d \\
& \boldsymbol{1}^\top(a\boldsymbol{x_1} + b\boldsymbol{x_2} + c\boldsymbol{x_3}) \geq d - a
\end{aligned} \tag{1}
$$

We take three floats $\epsilon_1, \epsilon_2, \epsilon_3$ to threshold $\hat{\boldsymbol{x_1}}, \hat{\boldsymbol{x_2}}, \hat{\boldsymbol{x_3}}$. Every element in $\boldsymbol{x_k}$ larger than ϵ_k is set as 1, otherwise it is 0.

$$
\begin{aligned}
\boldsymbol{x_1}^* &= 1\{\hat{\boldsymbol{x_1}} > \epsilon_1\} \\
\boldsymbol{x_2}^* &= 1\{\hat{\boldsymbol{x_2}} > \epsilon_2\} \\
\boldsymbol{x_3}^* &= 1\{\hat{\boldsymbol{x_3}} > \epsilon_3\}
\end{aligned} \tag{2}
$$

It is acceptable that we use an approximate global solution of the original problem, but the solution needs to be feasible. More specifically, the first two constraints have highest priority to satisfy because we cannot give an invalid action (*e.g.* It's impossible to annotation an image both weakly and strongly, but it might appear from the solution).

If we set $\epsilon_1 = \epsilon, \epsilon_2 = 1 - \epsilon, \epsilon_3 = \epsilon$, where ϵ exhaustively goes from 0 to 1 to satisfy the last two budget constraints. It can be proofed that the solution Eq.2 is also feasible from the original constraints.

Proof. (1) Assume $\psi(k) = 0$, where k can be the index of any element of vector $\hat{\boldsymbol{x_1}}$. Since $\hat{\boldsymbol{x_1}}, \hat{\boldsymbol{x_2}}, \hat{\boldsymbol{x_3}}$ is feasible, the first inequality in Eq.1 shows

$$
\hat{x_3}(k) \leq 0 \implies \hat{x_3}(k) = 0. \tag{3}
$$

The second inequality in Eq.1 gives the inequality for $\hat{x_1}$ and $\hat{x_2}$.

$$
\hat{x_1}(k) + \hat{x_2}(k) \leq 1 \implies 1 - \hat{x_2}(k) \geq \hat{x_1}(k) \implies 1\{1 - \hat{x_2}(k) < \epsilon\} \leq 1\{\hat{x_1}(k) < \epsilon\} \tag{4}
$$

When we apply Eq.2 to check the two restrictions, the discrete solutions $\boldsymbol{x_1}^*, \boldsymbol{x_2}^*, \boldsymbol{x_3}^*$ are still feasible.

$$
\begin{aligned}
x_3^*(k) &= 1\{\hat{x_3}(k) > \epsilon\} = 1\{0 > \epsilon_3\} = 0 &= \psi(k) \\
x_1^*(k) + x_2^*(k) &= 1\{\hat{x_1}(k) > \epsilon\} + 1\{\hat{x_2}(k) > 1 - \epsilon\} & \\
&= 1\{\hat{x_1}(k) > \epsilon\} + 1\{1 - \hat{x_2}(k) < \hat{x_2}(k)\} & \\
&\leq 1\{\hat{x_1}(k) > \epsilon\} + 1\{\hat{x_1}(k) < \hat{x_2}(k)\} \leq 1 &= 1 - \psi(k)
\end{aligned} \tag{5}
$$

(2) If $\psi(k) = 1$, we have

$$
\begin{aligned}
\hat{x_3}(k) &\leq 1 \implies & \hat{x_3}(k) &= 1 \\
\hat{x_1}(k) + \hat{x_2}(k) &\leq 0 \implies & \hat{x_1}(k), \hat{x_2}(k) &= 0
\end{aligned} \tag{6}
$$

Similar to previous discussion, we can have $x_3^*(k) \leq 1$ and $x_1^*(k) = 0, x_2^*(k) = 0$ to fit the two inequalities. $\square$

1.2. Pseudo Label Mining

In the article, we give another optimization function to do noise cleaning. This procedure checks the consistency between the image and instance-level annotations, and it removes abundant pseudo labels.

Given an image I_k with the weakly annotation $\boldsymbol{\omega} \in \{0,1\}^K$, where K is the number of categories. We assume the first detection model (*teacher model*) gives M positive predictions for the localization $P \in \mathbb{R}^{M\times 4}$, classification $A \in \{0,1\}^{M\times K}$, and the positive class confidence vector $\boldsymbol{p}_m \in [0,1]^M$. Eventually, the pseudo label mining returns a sparse M-dimensional binary vector $\boldsymbol{y}$ as Eq.7, where $\alpha = 0.3, \beta = 3$.

$$
\begin{array}{rrcl}
\min\limits_{\boldsymbol{y}\in\{0,1\}^M} & -\boldsymbol{y}^\top \boldsymbol{p}_m & & \\
\text{s.t.} & \boldsymbol{y}^\top A(\mathbf{1}-\boldsymbol{w}) & = & 0 \\
 & \forall_{y_i,y_j=1} IoU(P_i,P_j) & \leq & \alpha \\
 & |\boldsymbol{y}|_0 & \in & [1,\beta]
\end{array} \tag{7}
$$

In this problem, our objective function takes high confidence predictions by choosing $\boldsymbol{y} \in \{0,1\}^M$. In the first constraint, $\boldsymbol{y}^\top A$ accumulates the selected predictions by their categories. The following inner product with $\mathbf{1}-\boldsymbol{w}$ returns zero only when all the predicted classes are in the weak annotation. On the other hand, the second constraint removes the heavily overlapped predictions while the third one make the binary vector $\boldsymbol{y}$ sparse. We develop Alg.1 to solve the above optimization.

Algorithm 1: Noise Cleaning

Input : Weak annotation $\boldsymbol{\omega} \in \{0,1\}^K$, localization $P \in \mathbb{R}^{M\times 4}$, classification $A \in \{0,1\}^{M\times K}$, the positive class confidence for M predictions $\boldsymbol{p}_m \in [0,1]^M$, α, β;

Output: a binary vector $\boldsymbol{y} \in \{0,1\}^M$;

$\boldsymbol{y} = \mathbf{1}$;
for *i=1:M* **do**
    if not $A(i,:) < \boldsymbol{\omega}$, $y(i) = 0$;
end
$\boldsymbol{y'} =$ NMS index for [P,A], do it by class with threshold α
$\boldsymbol{y} = \boldsymbol{y}.*\boldsymbol{y'}$
if $sum(\boldsymbol{y}) > \beta$ **then**
    assign the top β remained predictions to y
end

2. Budget-mAP Curves

We present the original Budget vs mAP curves for the figures 4,5,6 and 7 in the draft, respectively. These curves show mAP before integration, which are more detail and noisy. Please check section 4 for the table representation.

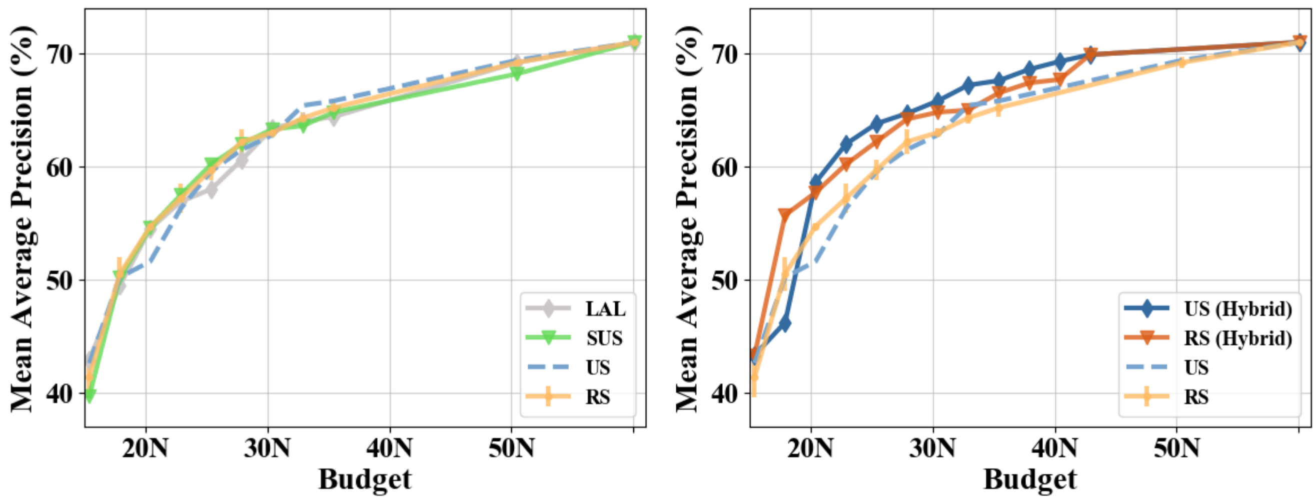

Figure 1. *left*: Budget-mAP curves using strongly annotated images for three different sampling methods: Orange bars Random Sampling fully supervised (RS). Blue bars samples from most uncertain images (US). Green Bars samples from the least uncertain images (SUS). Grey bars samples from Learning Active Learning sampling (LAL). *right*: Budget-mAP curves using fully and hybrid training pipelines. Orange bars compare the two training pipelines using random sample while Blue bars show both pipelines using US sampling.

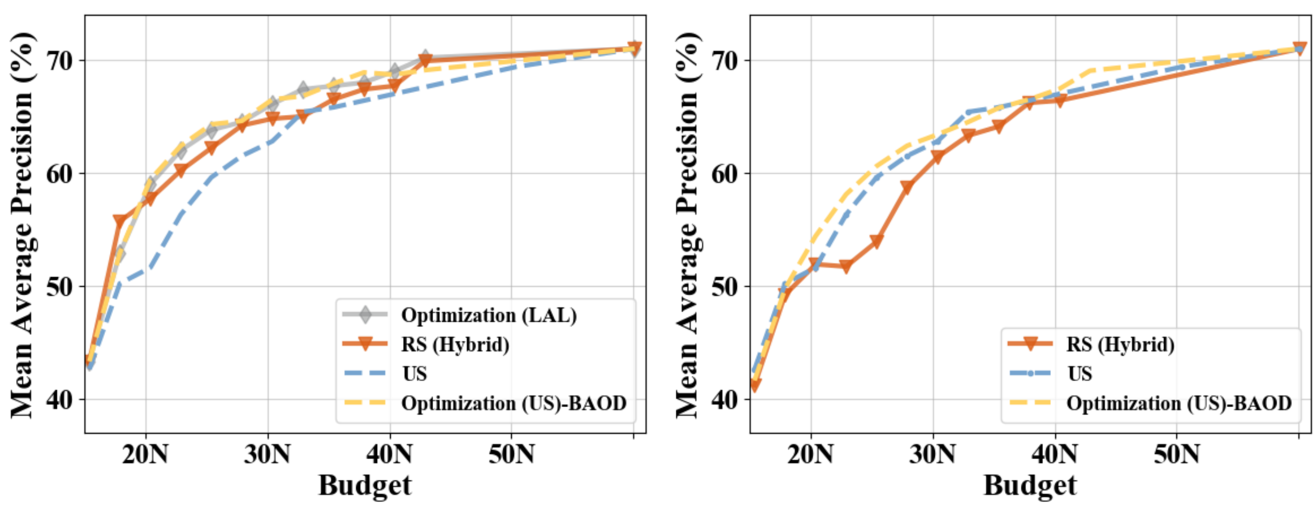

Figure 2. *left*: Budget-mAP curves using FSOD, hybrid training, and optimization methods. Blue bars show US using FSOD. Orange bars show RS hybrid baseline. Gold bars show US Optimization (BAOD). Dark Grey Bars show LAL Optimization. *right*: Budget-mAP curves using a lower cost for strong annotations. Blue Bars represent FSOD with US sampling. Orange Bars represent Hybrid training with RS sampling. Gold Bars represent US Optimization. All the methods use a smaller gap between the weak and strong annotation costs.

3. Visualizations

Below we show some visualizations of strongly labelled and weakly images for LAL based Optimization and US based Optimization.

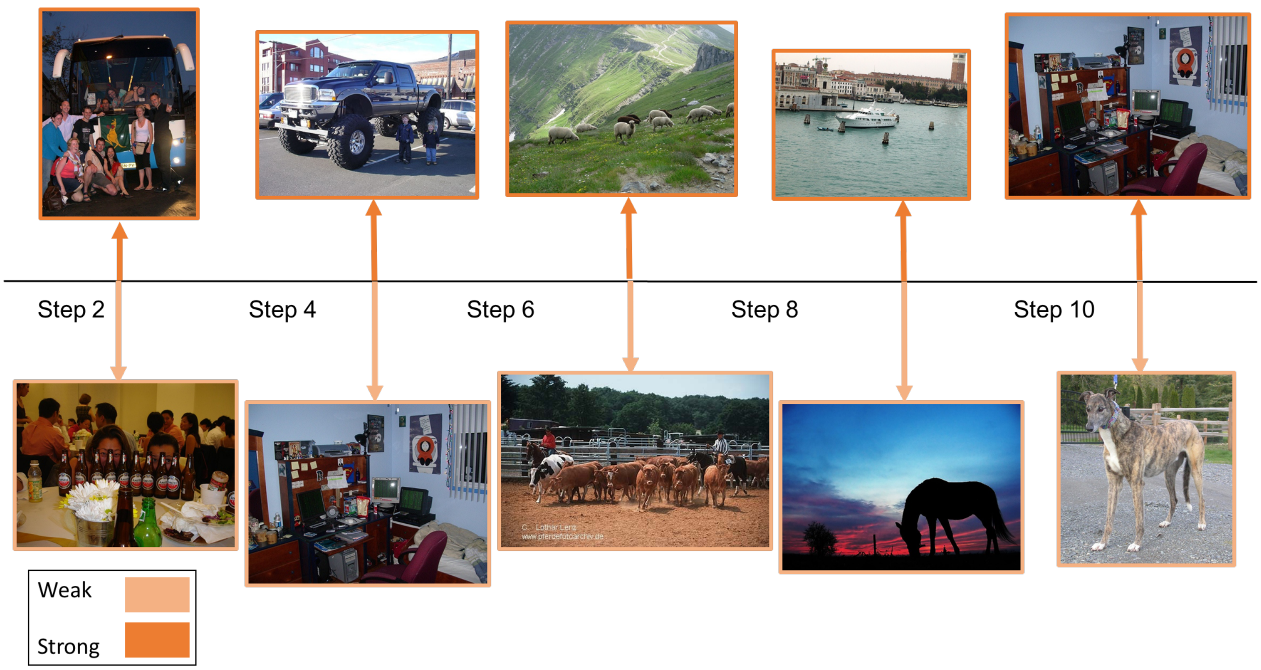

Figure 3. LAL Visualizations of some Strongly and Weakly labelled images during different steps. We can observe that the weakly image for step 4 is turned to a Strongly one in the last step.

Figure 4. US Visualizations of some Strongly and Weakly labelled images during different steps. We can observe that images that were weakly and strongly for LAL are the opposite for US. For US we can observe an improvement choosing the difficulty of the images once the model improves.

4. Raw Experiment Data

On table 4, we showcase the raw number of the curves in the main paper experiments that were not reported on the main paper.

			Budget Percentages													Full
	H*	CR**	10.80	15.79	20.77	25.76	30.75	35.74	40.73	45.72	50.71	55.70	60.69	65.68	80.70	100
RS	✗	H	0.414	0.505	0.547	0.572	0.597	0.622	0.630	0.643	0.652	-	-	-	0.692	0.71
US	✗	H	0.430	0.506	0.519	0.552	0.599	0.621	0.623	0.651	0.657	-	-	-	0.695	0.71
SUS	✗	H	0.397	0.502	0.546	0.575	0.602	0.62	0.633	0.636	0.648	-	-	-	0.682	0.71
LAL	✗	H	0.429	0.495	0.545	0.569	0.58	0.606	0.634	0.64	0.644	-	-	-	0.692	0.71
RS	✓	H	0.408	0.562	0.576	0.608	0.612	0.637	0.651	0.655	0.663	-	-	-	-	0.71
US	✓	H	0.433	0.462	0.586	0.62	0.638	0.647	0.658	0.672	0.676	-	-	-	-	0.71
OPT-US	✓	H	0.433	0.529	0.594	0.624	0.643	0.646	0.665	0.668	0.679	0.689	0.687	0.691	-	0.71
OPT-LAL	✓	H	0.389	0.529	0.59	0.62	0.638	0.645	0.661	0.674	0.677	-	-	-	-	0.71
US	✓	L	0.412	0.498	0.509	0.527	0.573	0.613	0.626	0.653	0.664	0.675	0.673	-	-	0.71
RS	✓	L	0.412	0.492	0.519	0.517	0.539	0.587	0.614	0.633	0.641	0.662	0.664	-	-	0.71
OPT	✓	L	0.42	0.518	0.548	0.586	0.609	0.62	0.64	0.648	0.665	0.67	0.683	0.689	-	0.71
RS	✓	M	0.443	0.554	0.576	0.6	0.62	0.62	0.654	0.652	0.655	0.672	0.675	0.689	-	0.71
OPT	✓	M	0.433	0.544	0.582	0.608	0.632	0.643	0.652	0.664	0.677	0.684	0.687	0.693	-	0.71
Budget	-	-	10.8	20.8	30.8	40.8	50.8	60.8	70.8	80.8	90.8	92.8	96.8	-	-	-
VOC0712	✓	M	0.356	0.566	0.611	0.641	0.674	0.672	0.696	0.701	0.715	0.717	0.72	-	-	0.73
Budget	-	-	11.09	13.62	18.1	22.58	27.06	31.53	36.15	40.49	44.97	49.45	53.93	-	-	-
RL	✓	H	0.433	0.525	0.561	0.579	0.603	0.615	0.634	0.645	0.648	0.657	0.664	-	-	0.71

* H for Hybrid Learning

** CR for cost ratio of strong and weak annotations

Table 1. Raw experimental data from the curves in the main paper. Each one of the entrace in the table is one fully-trained Faster-RCNN model from Imagenet weights.