Instance-wise Supervision-level Optimization in Active Learning

Supplementary Material

8. Detailed algorithm of batch selection

In this section, we provide a detailed algorithm of the batch selection introduced in Sec. 4.3. The goal is to select a valuable and diverse set of weakly supervised and fully supervised instances within the budget B, leveraging the normalized feature vector $\tilde{f}(\boldsymbol{x})$ and the VCR for full and weak supervision $v_{\mathrm{f}}(\boldsymbol{x})$ and $v_{\mathrm{w}}(\boldsymbol{x})$, as described in Sec. 4.2, for each $\boldsymbol{x}$ in the unlabeled data pool $\mathcal{D}_{\mathrm{u}}$.

The detailed algorithm is presented in Algorithm 2. The first vector $\boldsymbol{g}$ is selected based on the largest norm from the set of vectors $G = G_{\mathrm{f}} \cup G_{\mathrm{w}}$, where $G_{\mathrm{f}} = \{v_{\mathrm{f}}(\boldsymbol{x})\tilde{f}(\boldsymbol{x}) | \boldsymbol{x} \in \mathcal{D}_{\mathrm{u}}\}$ represents the set of the vectors for full supervision and $G_{\mathrm{w}} = \{v_{\mathrm{w}}(\boldsymbol{x})\tilde{f}(\boldsymbol{x}) | \boldsymbol{x} \in \mathcal{D}_{\mathrm{u}}\}$ represents those set of the vectors for weak supervision. Subsequent vectors are sampled from the set G, with the sampling probability of each vector proportional to the square of its distance $D(\boldsymbol{g})$ to the vector closest to the one already selected. If the selected vector $\boldsymbol{g}$ belongs to G_{f}, the corresponding instance is fully annotated, and the budget is reduced by C_{f}. Similarly, If $\boldsymbol{g}$ belongs to G_{w}, the instance is weakly annotated, and the budget is reduced by C_{w}. This process is repeated until the budget B is fully consumed.

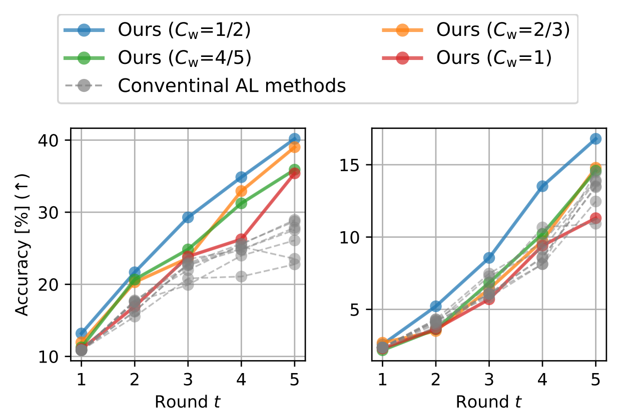
Algorithm 2 Batch selection

Inputs: budget B, annotation cost $C_{\mathrm{f}}, C_{\mathrm{w}}$, VCR $v_{\mathrm{f}}(\boldsymbol{x}), v_{\mathrm{w}}(\boldsymbol{x})$, and normalized feature $\tilde{f}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{D}_{\mathrm{u}}$.
Outputs: batch $\mathcal{D}_{\mathrm{f}}^{+}, \mathcal{D}_{\mathrm{w}}^{+}$.
$G_{\mathrm{f}} \leftarrow \{v_{\mathrm{f}}(\boldsymbol{x})\tilde{f}(\boldsymbol{x}) | \boldsymbol{x} \in \mathcal{D}_{\mathrm{u}}\}$
$G_{\mathrm{w}} \leftarrow \{v_{\mathrm{w}}(\boldsymbol{x})\tilde{f}(\boldsymbol{x}) | \boldsymbol{x} \in \mathcal{D}_{\mathrm{u}}\}$
$G \leftarrow G_{\mathrm{f}} \cup G_{\mathrm{w}}$
Select $\boldsymbol{g}$ with the largest norm from G
if $\boldsymbol{g} \in G_{\mathrm{f}}$ **then**
  $G_{\mathrm{f}}^{+} \leftarrow \{\boldsymbol{g}\}$, $G_{\mathrm{f}} \leftarrow G_{\mathrm{f}} \setminus \{\boldsymbol{g}\}$, $b \leftarrow C_{\mathrm{f}}$
else if $\boldsymbol{g} \in G_{\mathrm{w}}$ **then**
  $G_{\mathrm{w}}^{+} \leftarrow \{\boldsymbol{g}\}$, $G_{\mathrm{w}} \leftarrow G_{\mathrm{w}} \setminus \{\boldsymbol{g}\}$, $b \leftarrow C_{\mathrm{w}}$
end if
while $b < B$ **do**
  Define $D(\boldsymbol{g}) := \min_{\boldsymbol{c} \in G_{\mathrm{f}}^{+} \cup G_{\mathrm{w}}^{+}} \|\boldsymbol{c} - \boldsymbol{g}\|_2$
  Sample $\boldsymbol{g}$ from G with probability $\frac{D(\boldsymbol{g})^2}{\sum_{\boldsymbol{g} \in G} D(\boldsymbol{g})^2}$
  if $\boldsymbol{g} \in G_{\mathrm{f}}$ **then**
    $G_{\mathrm{f}}^{+} \leftarrow G_{\mathrm{f}}^{+} \cup \{\boldsymbol{g}\}$, $G_{\mathrm{f}} \leftarrow G_{\mathrm{f}} \setminus \{\boldsymbol{g}\}$, $b \leftarrow b + C_{\mathrm{f}}$
  else if $\boldsymbol{g} \in G_{\mathrm{w}}$ **then**
    $G_{\mathrm{w}}^{+} \leftarrow G_{\mathrm{w}}^{+} \cup \{\boldsymbol{g}\}$, $G_{\mathrm{w}} \leftarrow G_{\mathrm{w}} \setminus \{\boldsymbol{g}\}$, $b \leftarrow b + C_{\mathrm{w}}$
  end if
end while
Obtain the batches $\mathcal{D}_{\mathrm{f}}^{+}$ and $\mathcal{D}_{\mathrm{w}}^{+}$ with the instance $\boldsymbol{x}$ corresponding to G_{f}^{+} and G_{w}^{+}, respectively.

9. Additional experiments on larger weak supervision costs

In the main paper, we demonstrated that our approach consistently outperforms the conventional AL methods when the weak supervision cost C_{w} was $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{8}$, with the full supervision cost C_{f} fixed at 1. In this section, we present additional experimental results for scenarios where the weak supervision cost is higher (i.e., the cost difference between C_{f} and C_{w} is small).

Figure 6 shows the classification results on CIFAR100 and CUB200 for $C_{\mathrm{w}} = \frac{1}{2}$, $\frac{2}{3}$, $\frac{4}{5}$, and 1. When C_{w} is close to C_{f}, our approach is observed to outperform or be comparable to conventional AL methods. Importantly, even when weak supervision provides no significant cost advantage, our approach remains effective, maintaining performance without significant degradation. This is because our approach adaptively allocates the available budget, achieving either superior or comparable performance to conventional AL methods, regardless of whether weak supervision is highly cost-effective or nearly as expensive as full supervision.



(a) CIFAR100 (b) CUB200

Figure 6. Classification accuracy [%] (↑) of our approach on CIFAR100 and CUB200 for weak supervision costs $C_{\mathrm{w}} = \frac{1}{2}$, $\frac{2}{3}$, $\frac{4}{5}$ and 1. For comparison, the performance of conventional AL methods is shown as gray lines.

Table 2. Proportions [%] of weak supervision in the batch.

	Round = 1	2	3	4	5
$C_{\mathrm{w}} = 1/2$	50.00	40.93	28.31	64.52	62.11
$= 1/4$	50.00	75.79	48.82	65.18	75.42
$= 1/8$	50.00	78.74	90.84	67.25	85.62

Table 3. Accuracy [%] with five different random seeds.

	Round = 1	2	3	4	5
Mean	13.11	22.33	28.44	33.72	37.11
Std.	(±0.86)	(±2.06)	(±0.83)	(±1.16)	(±1.89)

Table 4. Classification accuracy [%] (↑) on CIFAR100. This table presents the same results as Figure 2(a) in a tabular format.

Method	Round = 1	2	3	4	5
Random	10.86	17.66	21.96	25.55	28.93
Margin	10.86	16.21	22.92	25.63	28.72
MaxConf	10.86	16.28	22.74	25.21	23.54
Entropy	10.86	15.53	20.82	21.08	22.77
Coreset	10.86	17.55	23.48	24.82	27.55
ALBL	10.86	17.55	19.92	23.98	26.09
BADGE	10.86	17.77	22.63	24.83	27.85
APFWA	12.45	20.54	25.35	31.30	33.11
Ours	**13.18**	**21.67**	**29.28**	**34.86**	**40.18**

Table 5. Classification accuracy [%] (↑) on CUB200. This table presents the same results as Figure 2(b) in a tabular format.

	Round = 1	2	3	4	5
Random	2.36	3.76	6.06	8.13	14.48
Margin	2.36	4.09	7.28	9.15	13.84
MaxConf	2.36	4.31	7.47	9.61	13.93
Entropy	2.36	4.30	6.49	10.67	10.93
Coreset	2.36	4.19	6.06	8.56	13.51
ALBL	2.36	4.19	5.83	8.13	12.46
BADGE	2.36	3.97	6.09	8.61	13.44
APFWA	1.88	5.20	6.40	10.94	15.69
Ours	**2.57**	**5.21**	**8.56**	**13.51**	**16.79**

10. Proportion of two supervision-levels

Table 2 presents the proportions of weakly supervised instances in the batch at each round during the CIFAR100 experiment, with weak supervision costs C_{w} set to $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{8}$. The results show that the allocation of full and weak supervision is dynamically adjusted based on cost-effectiveness.

11. Robustness analysis

Table 3 reports the mean and standard deviation of classification accuracy across five random seed runs. The mean accuracy exceeds that of conventional AL methods shown in Figure 2(a), and the small standard deviation demonstrates the stability and robustness of our approach.

12. Tabular representation of results

Tables 4 and 5 present the classification accuracy [%] on CIFAR100 and CUB200, respectively, in a tabular format corresponding to Figures 2(a) and Fig.2(b). These tables provide a clearer numerical comparison of the conventional AL methods.