Consistency-based Semi-supervised Active Learning: Towards Minimizing Labeling Cost

Mingfei Gao[1*], Zizhao Zhang[2], Guo Yu[3], Sercan Ö. Arık[2],
Larry S. Davis[1], and Tomas Pfister[2]

[1]University of Maryland [2]Google Cloud AI [3]University of Washington

ECCV 2020

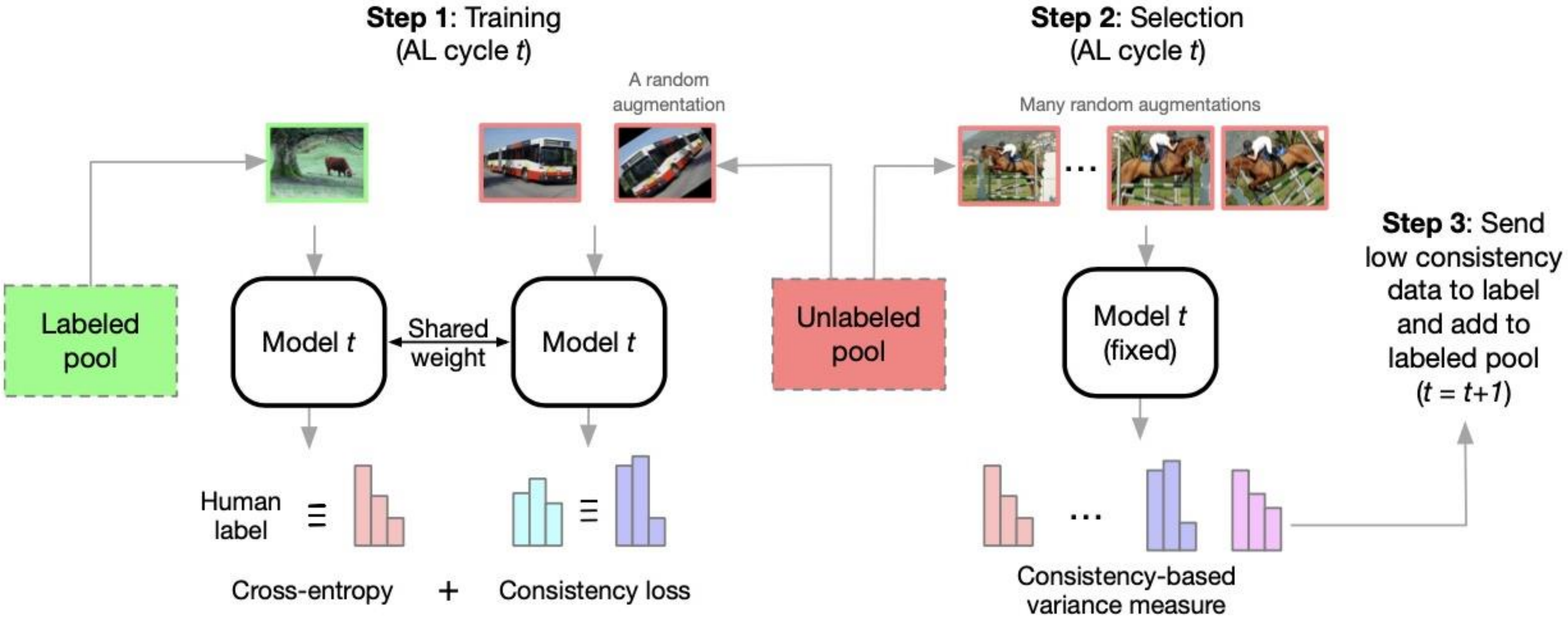
Motivation

SSL: Many recent approaches for SSL add a loss term which is computed on unlabeled data and encourages the model to generalize better to unseen data.

Consistency-based SSL: model is consistent in its decisions between a sample and its meaningfully distorted versions.

An AL selection strategy: a sample with its distorted versions that yields low consistency in predictions indicates that: the SSL model may be incapable of distilling useful information from that unlabeled sample – the sample need to be labeled.

Consistency-based Semi-AL



Consistency-based Semi-AL

Algorithm 1 A semi-supervised learning based AL framework

Require: Unlabeled data pool $\mathcal{D}$, the total number of steps T, selected sample batch set B, AL batch size K, start size $K_0 \ll |\mathcal{D}|$

$B_0 \leftarrow$ uniformly sampling from $\mathcal{D}$ with $|B_0| = K_0$

$U_0 \leftarrow \mathcal{D} \backslash B_0$

$L_0 \leftarrow \{(x, \mathcal{J}(x)) : x \in B_0\}$, where $\mathcal{J}(x)$ stands for the assigned label of x.

for $t = 0, \ldots, T-1$ **do**

(training) $M_t \leftarrow \arg\min_M \left\{ \frac{1}{|L_t|} \sum_{(x,y)\in L_t} \mathcal{L}_l(x, y, M) + \frac{1}{|U_t|} \sum_{x\in U_t} \mathcal{L}_u(x, M) \right\}$

(selection) $B_{t+1} \leftarrow \arg\max_{B\subset U_t} \{\mathcal{C}(B, M_t),\ s.t.\ |B| = K\}$

(labeling) $L_{t+1} \leftarrow L_t \cup \{(x, \mathcal{J}(x)) : x \in B_{t+1}\}$

(pool update) $U_{t+1} \leftarrow U_t \setminus B_{t+1}$

end for

$M_T \leftarrow \arg\min_M \left\{ \frac{1}{|L_T|} \sum_{(x,y)\in L_T} \mathcal{L}_l(x, y, M) + \frac{1}{|U_T|} \sum_{x\in U_T} \mathcal{L}_u(x, M) \right\}$

return M_T

$$\mathcal{L}_u(x, M) = D(P(\hat{Y} = \ell | x, M), P(\hat{Y} = \ell | \tilde{x}, M)),$$

D: L2 norm

$$\mathcal{C}(B, M) = \sum_{x\in B} \mathcal{E}(x, M)$$

$$\mathcal{E}(x, M) = \sum_{\ell=1}^{J} \mathrm{Var}\left[P(\hat{Y} = \ell | x, M),\ P(\hat{Y} = \ell | \tilde{x}_1, M),\ \ldots,\ P(\hat{Y} = \ell | \tilde{x}_N, M) \right],$$

Experiment

Comparison with selection baselines under SSL

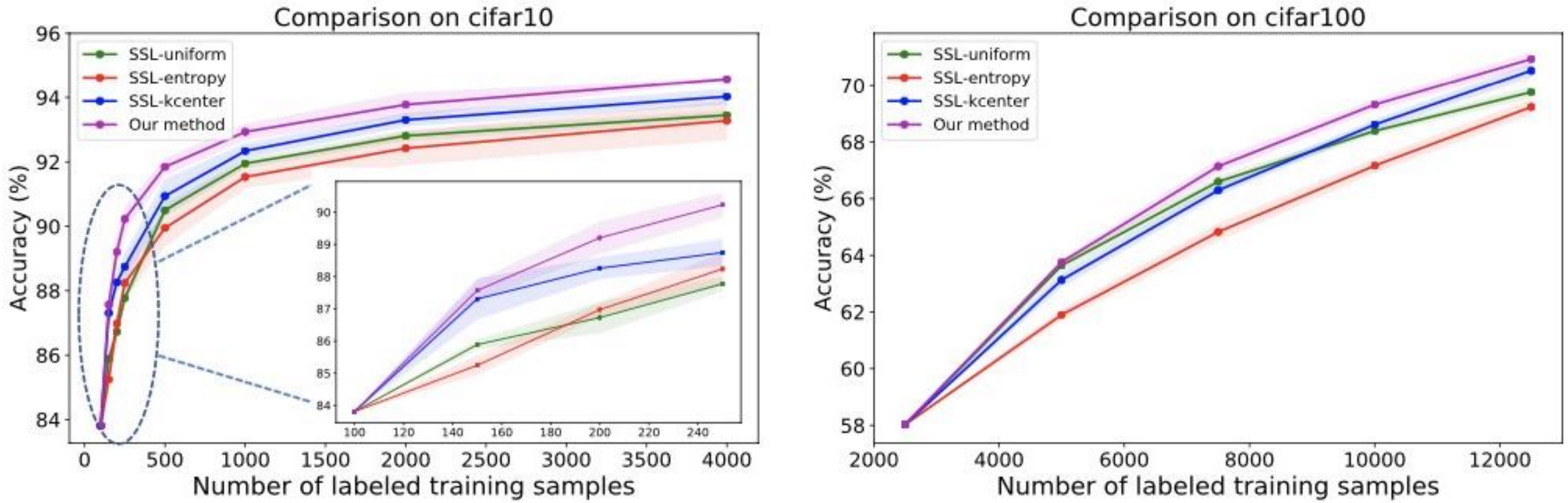


Fig. 2. Model performance comparison with different sample selection methods on CIFAR-10 and CIFAR-100. Solid lines indicate the averaged results over 5 trials. Shadows represent standard deviation

Experiment

Comparison with supervised AL methods

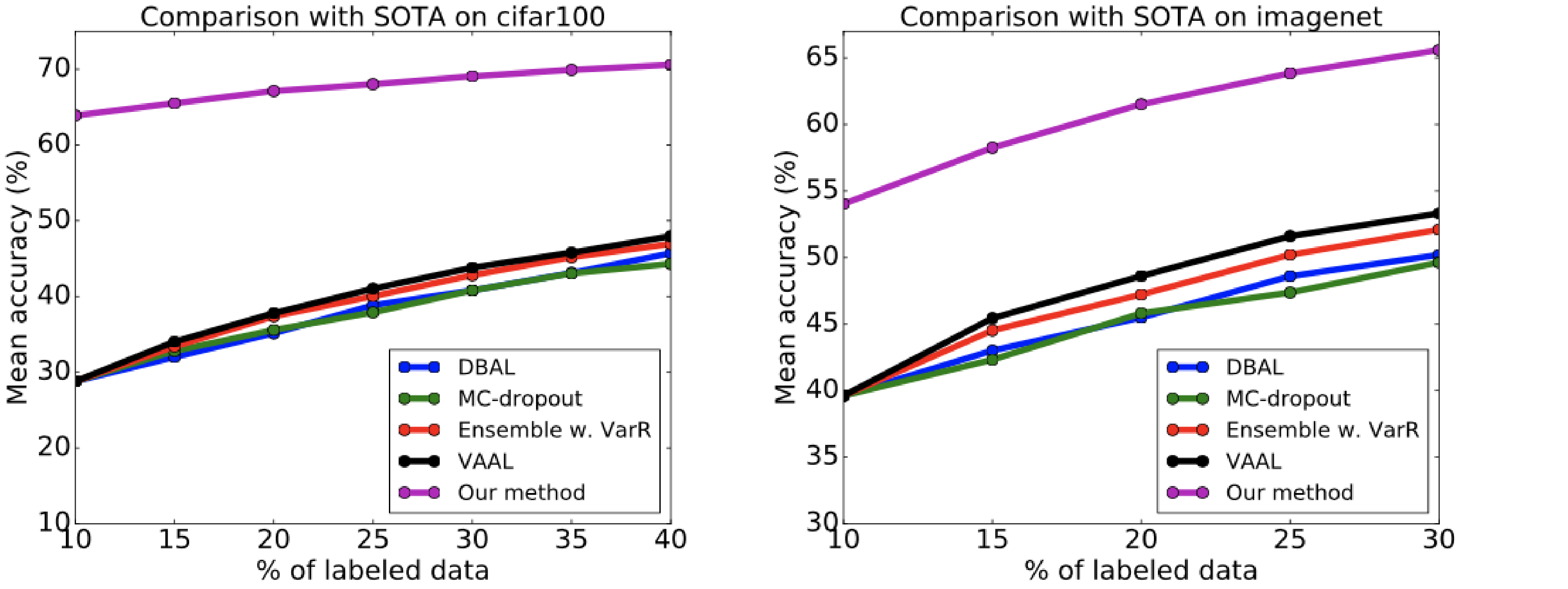


Fig. 3. Comparison with recent AL methods on CIFAR-100 and ImageNet. Our results on CIFAR-100 and ImageNet are averaged over 5 and 3 trials, respectively

Analysis

Uncertainty and overconfident mis-classification

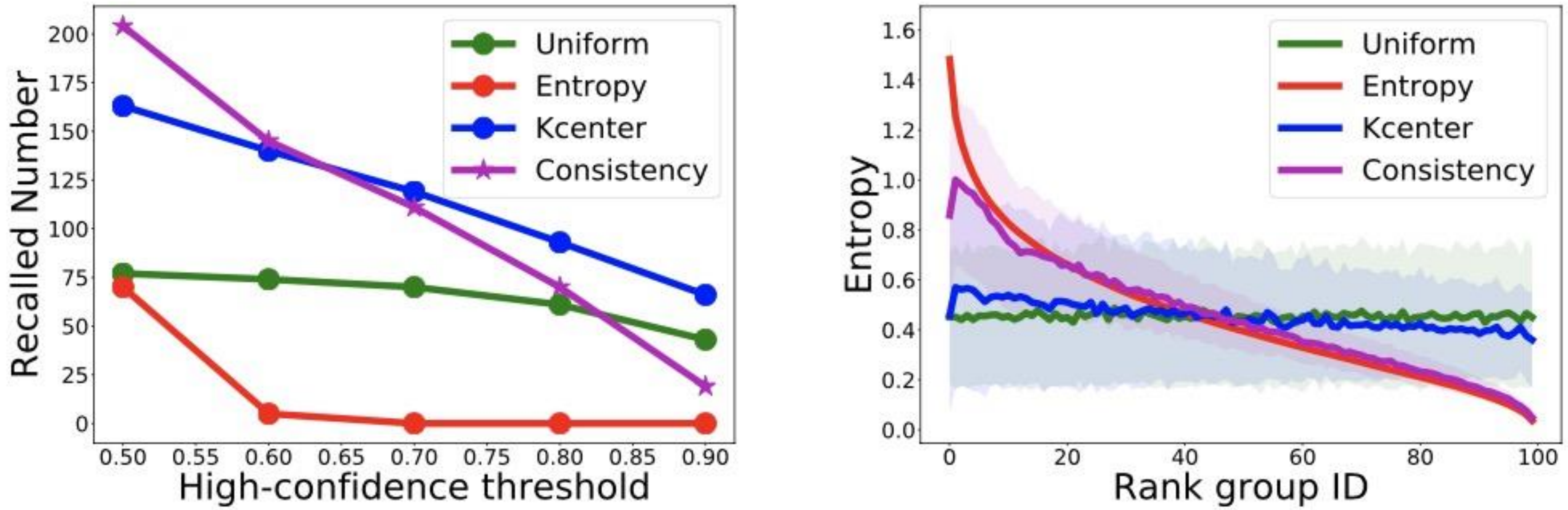


Fig. 4. Left: Number of overconfident mis-classified samples in top 1% samples ranked by different methods. Overconfident samples are defined as those having the highest class probability larger than threshold. Right: the average entropy of unlabeled samples ranked by different selection metrics. The ranked samples are divided into 100 groups for computing average entropy. Shadows represent standard deviation

Analysis

Smaple Diversity

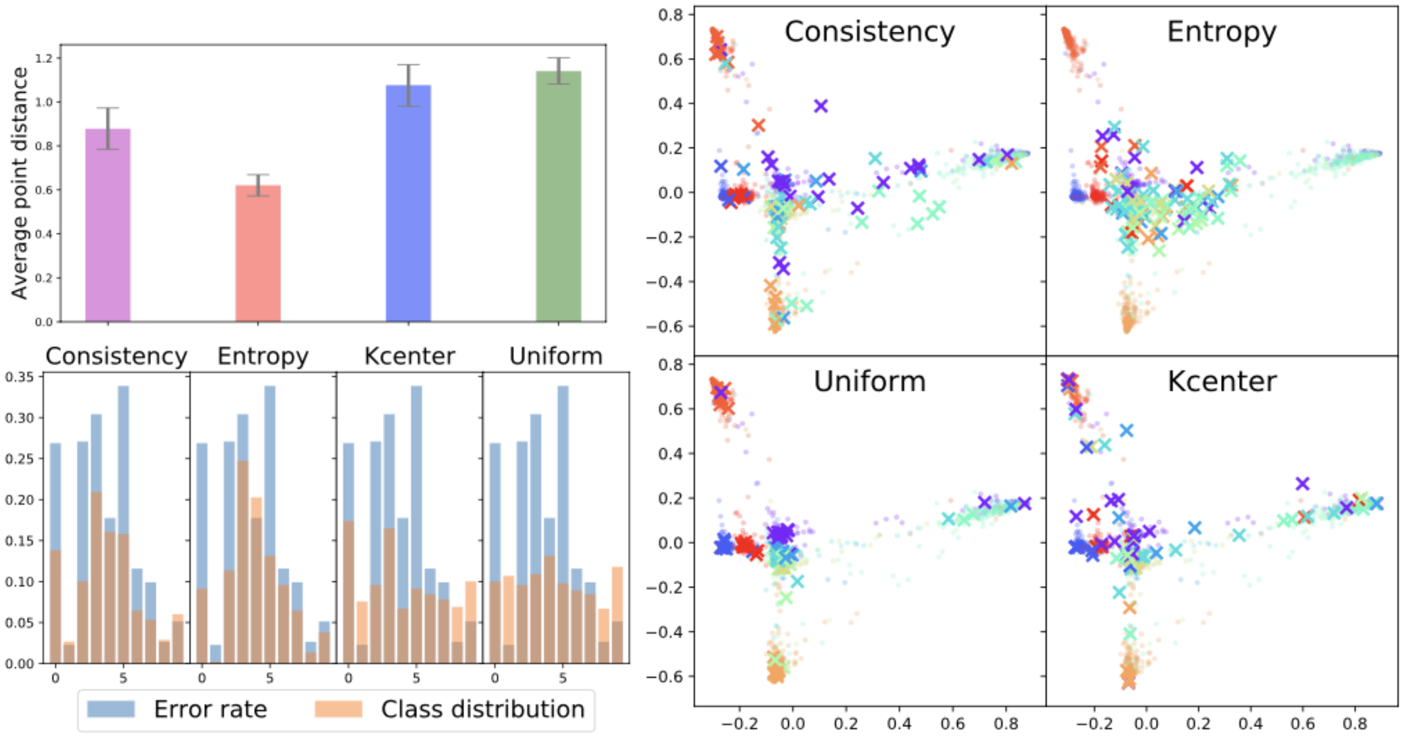


Fig. 5. Average distance between samples (top-left): the average pair-wise L_2 distance of top 1% unlabeled samples ranked by different selection metrics. Per-class error rate vs. the class distribution of the selected samples are shown in bottom-left. Diversity visualization (right): Dots and crosses indicate unlabeled (un-selected) samples and the selected samples (top 100), respectively. Each color represent a ground truth class

Thanks