Supplemental Document for Active Label Correction Using Robust Parameter Update and Entropy Propagation

Kwang In Kim

UNIST, Korea

Additional results. In the main paper, we used the classifier that consists of ResNet-101 pretrained on ImageNet. Figure 1 shows the results of experiments obtained with the smaller ResNet-18 networks. Overall, the accuracies of all algorithms degraded while our algorithm still achieved significantly higher accuracies. These results demonstrate the general applicability of our algorithm.

Our main algorithm updates the learner weights based on the loss values of individual examples (Eqs. 2–4 in the main paper). An alternative is to use the predictive entropies of examples (instead of losses). Figure 2 shows that our original algorithm outperforms this alternative design (*Ours (Ent)*).

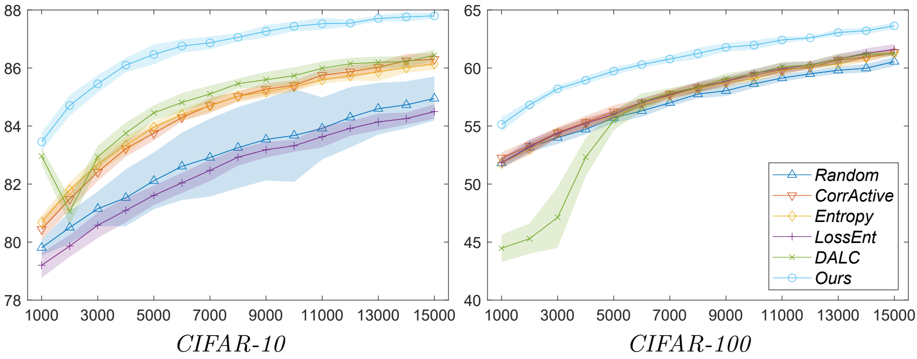

Fig. 1: Mean accuracy (%) with standard deviation (shaded) of different active label correction algorithms learned under uniform noise (ResNet-18 network). The x-axis corresponds to the number of queried labels. All ALC algorithms outperformed *Random* except for *LossEnt* on *CIFAR-10* and *DALC* on early learning stages of *CIFAR-100*, and our algorithm achieved further significant and consistent improvements.

Effect of hyperparameter variation. Our hyperparameters were fixed across datasets and they might not be optimal for each dataset. On *CIFAR-100*, 1) scaling the original parameter σ^{α} value by factors of $\{\frac{1}{4}, \frac{1}{2}, 2, 4\}$ led to {99.6, 99.8,

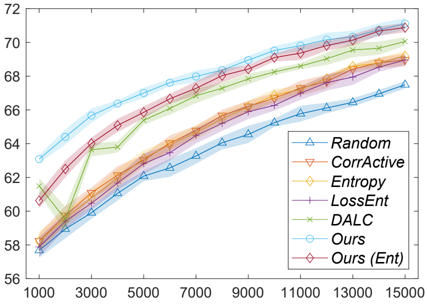

Fig. 2: Mean accuracy (%) with standard deviation (shaded) of different active label correction algorithms on *CIFAR-100*. Using the predictive entropies of examples in the learner weight update step (*Ours (Ent)*) instead of the loss values degraded the performance of our algorithm.

100.0, 100.1}% relative accuracy from our final algorithm (averaged on different number of labels); 2) varying the neighborhood size $|\mathcal{N}|$ in $\{6, 12, 14, 18\}$ resulted in {99.7, 100.3, 100.0, 99.4}% relative accuracy; Finally, varying the step size δ^α in $\{0.5, 0.2, 0.01\}$ led to {99.1, 100.2, 99.6}% relative accuracy.