TiDAL: Learning Training Dynamics for Active Learning

Seong Min Kye[1*] Kwanghee Choi[2*] Hyeongmin Byun[1] Buru Chang[3] *Equal contributions.

harris@hpcnt.com [1]Hyperconnect LLC [2]Carnegie Mellon University [3]Sogang University

ICCV23 PARIS

HYPERCONNECT®

Background

Active learning (AL) selects the most useful data samples from large-scale unlabeled data pools and annotating them to expand labeled data under a limited budget. Since the current deep neural networks are data-hungry, AL has increasingly gained attention recently.

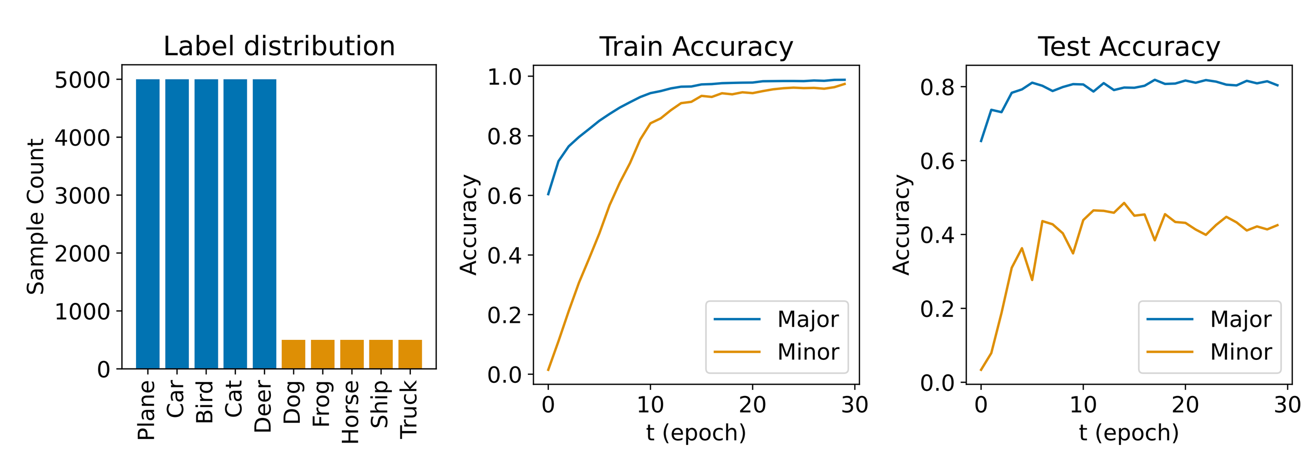
Uncertainty-based Active Learning methods choose the most uncertain samples, which are known to be effective. Uncertain samples are often selected using the **static information** (e.g., loss or predicted probability) from a fully-trained **model snapshot**, neglecting the valuable information generated during training.

We provide **theoretical and empirical evidence** to argue the usefulness of utilizing the **ever-changing model behavior** rather than the fully trained model snapshot for AL.

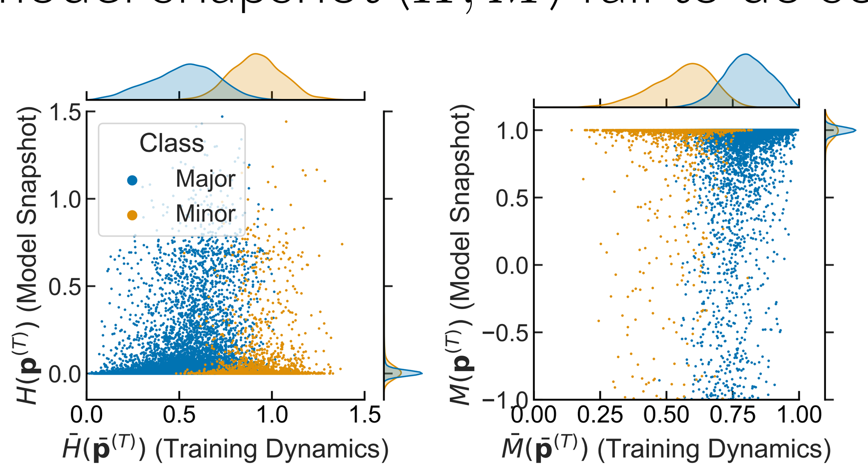
We characterize the model behavior in terms of **training dynamics (TD)** $\bar{\mathbf{p}}^{(t)} = \sum_{i=1}^{t} \mathbf{p}^{(i)}/t$ given predicted probabilities at time step $\mathbf{p}^{(t)}$.

Motivating Observation

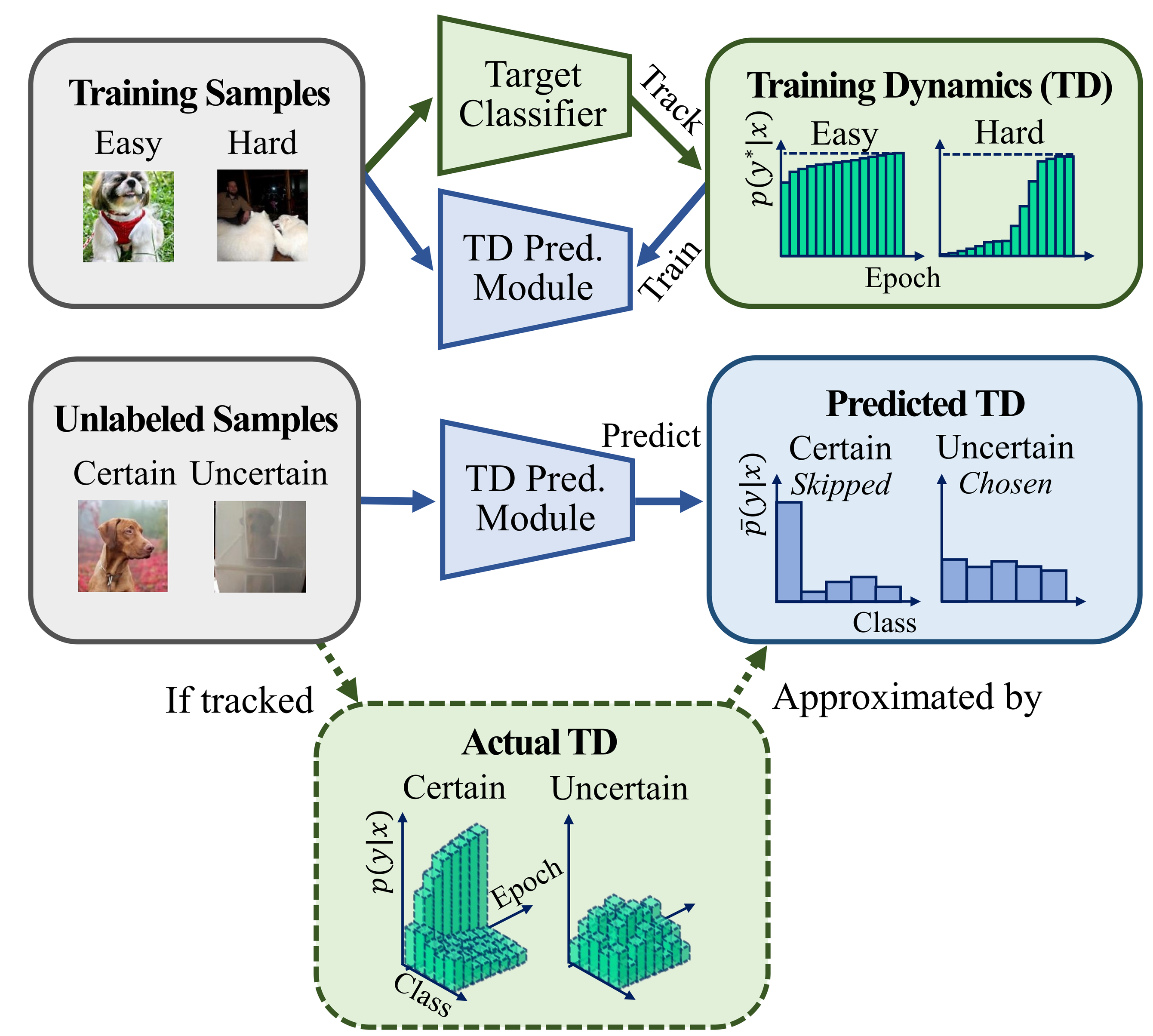
We consider the major and minor classes as certain and uncertain classes (borrowed from long-tail classification settings).



We can observe that scores from TD ($\bar{H}, \bar{M}$) successfully separate the certain (major) and the uncertain (minor) class samples, whereas scores from the model snapshot (H, M) fail to do so.



Proposed Method



TD may differ even if they converge to the same final predicted probability $p(y^*|x)$ (Upper row). Hence, we utilize the readily available rich information generated during training, *i.e.*, leveraging TD.

To efficiently estimate the TD, we jointly train the **TD prediction module**. With this module, We **estimate TD of large-scale unlabeled data using a prediction module** instead of tracking the actual TD of all the unlabeled samples to avoid the computational overhead (Lower row).

Theoretical Results

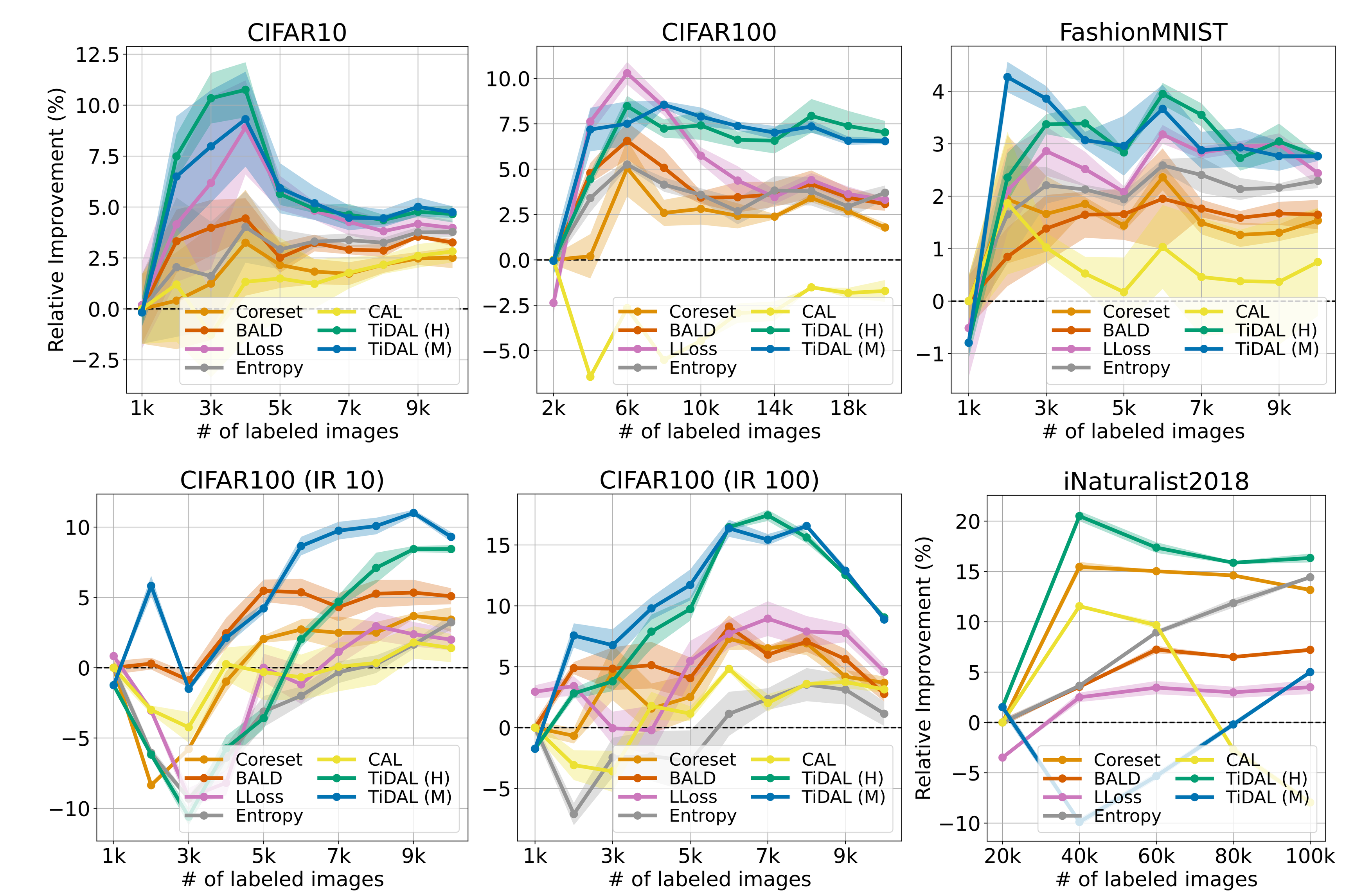
Thm 1. Under the LE-SDE framework, with the assumption of local elasticity, **certain samples and uncertain samples reveal different TD**; especially, certain samples converge quickly than uncertain samples.

Thm 2. Estimators such as *Entropy* and *Margin* **successfully capture the difference of TD between easy and hard samples** even for the case where it cannot be distinguished via the predicted probabilities of the model snapshot.

Empirical Results

AL Settings. *Step 1.* Randomly select initial samples to be annotated to train the initial target classifier. *Step 2.* Choose the top-k unlabeled samples to annotate based on each AL method. *Step 3.* We repeat Step 2, training from the continuously expanding labeled set.

We evaluate on **balanced** (CIFAR10/100, FashionMNIST), **synthetically imbalanced** (imbalance ratio 10/100), and **real-world imbalanced datasets** (iNaturalist2018, SVHN). We compare with 7 SOTA AL methods (more in Appendix). We plot the averaged **relative accuracy improvement curve** over the random sampling.



Further Analyses

Ablation. We observe that both Entropy and Margin show significantly superior performance when employed with prediction module outputs as opposed to when using the classifier probabilities.

Performance of the TD prediction module. Using the KL divergence, we observe that the predicted TD converges to the actual TD. In contrast, classifier probabilities were quite different to actual TD.