Semi-Supervised Active Learning with Temporal Output Discrepancy

Siyu Huang, Tianyang Wang, Haoyi Xiong, Jun Huan, Dejing Dou

Baidu Research, APSU, Styling AI

2021 ICCV OCTOBER 11-17 VIRTUAL

TL;DR

- We propose a loss measurement **Temporal Output Discrepancy (TOD)** to estimate the loss of unlabeled samples.
- We demonstrate TOD is a lower-bound of accumulated sample loss.
- Based on TOD, we develop an unlabeled data sampling strategy and a semi-supervised learning scheme for active learning.

Background

- **Active learning (AL)** aims to interactively query a human annotator oracle to annotate a small proportion of informative samples in an unlabeled dataset.
- We focus on uncertainty-aware AL, which selects the most uncertain samples in context of a learned model.

TOD: An Effective Loss Estimation Method

- **Temporal Output Discrepancy (TOD):** The discrepancy of outputs of a neural network at different GD iterations.

$$D_t^{\{T\}}(x) \overset{\text{def}}{=} \|f(x;w_{t+T}) - f(x;w_t)\|$$

- **Why TOD measures the sample loss?**

i) Firstly, connecting one-step TOD to sample loss $L(x)$:

$$D_t^{\{1\}}(x) \le \eta\sqrt{2L_t(x)}\|\nabla_w f(x;w_t)\|^2$$

ii) Then for T-step TOD, we have

$$D_t^{\{T\}}(x) \le \sqrt{2}\eta \sum_{\tau=t}^{t+T-1}\left(\sqrt{L_\tau(x)}\|\nabla_w f(x;w_\tau)\|^2\right)$$

iii) Let gradient norm of f be upper-bounded by a constant C,

$$D_t^{\{T\}}(x) \le \sqrt{2T}\eta C\sqrt{\sum_{\tau=t}^{t+T-1} L_\tau(x)}$$

Thus, TOD is a lower bound of square root of accumulated loss $L(x)$ during T GD iterations.

TOD-based Active Sampling Strategy

- We further propose an unlabeled data sampling strategy, named **Cyclic Output Discrepancy (COD),** for active learning. COD estimates the sample loss by measuring the distance of model outputs between two consecutive active learning cycles.

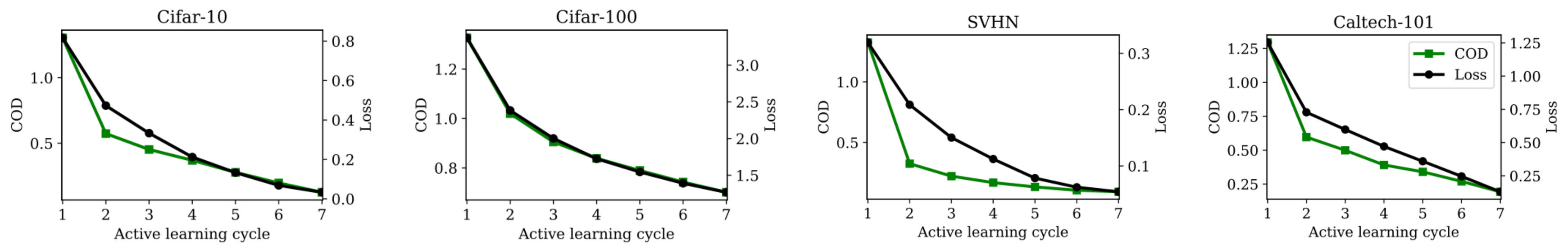

Fig.2: COD values are consistent with the real loss values.

- During active sampling, COD finds samples of large loss in unlabeled pool to minimize the expected loss of the model in future training.

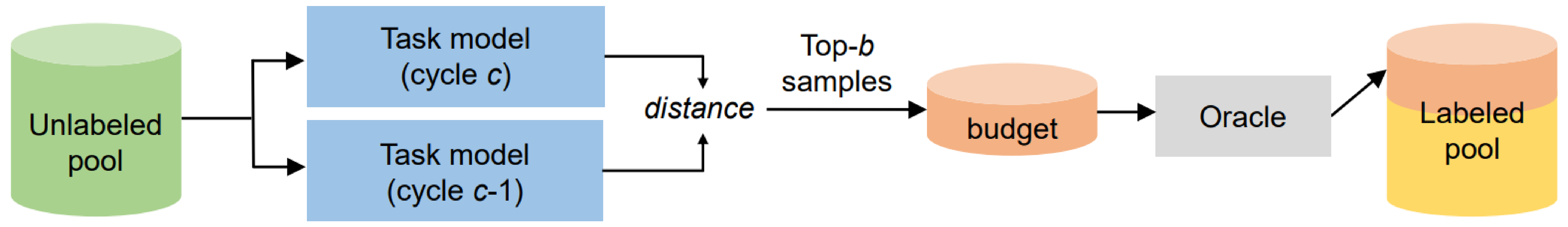

Fig.2: COD active sampling.

TOD-based Semi-Supervised Active Learning

- We propose a **TOD-based unsupervised loss** to minimize the distance between the current task model and a baseline model. The parameters of baseline model is an exponential moving average of historical model parameters.

- The optimization objective of semi-supervised AL consists of task loss and unsupervised loss .

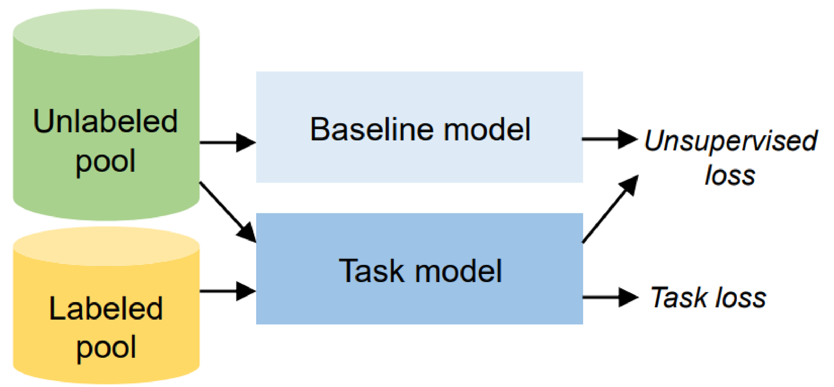

Fig.3: Semi-supervised AL.

Experiments

- **Performance:** TOD-based AL methods perform well on various image classification and segmentation datasets.

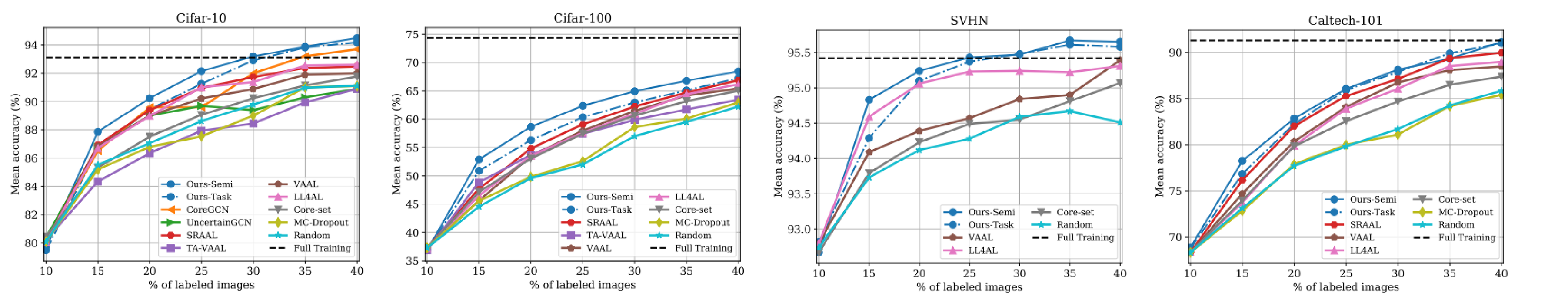

Fig.4: AL performance of image classification datasets. Blue solid/dashed lines denote our method with/without semi-supervised training.

- **Study on Active Sampling:** COD-based active sampling strategy outperforms existing AL methods.

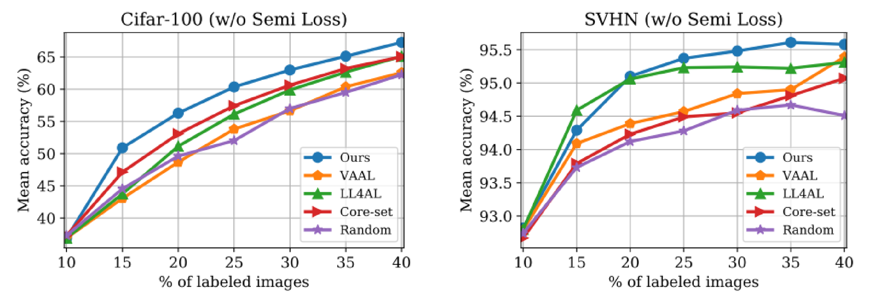

Fig.5: Study on active sampling.

- **Study on Semi-Supervised Learning:** (i) SSL can improve AL performance; (ii) TOD is effective for SSL.

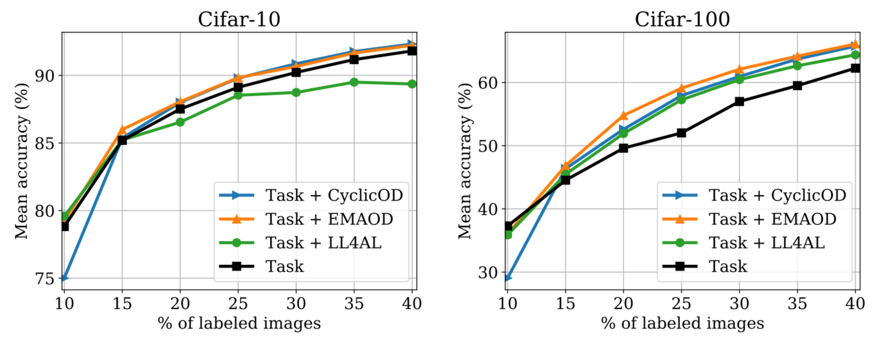

Fig.6: Study on SSL methods.

- **Efficiency:** COD is more efficient than previous AL methods, as it only relies on task model and does not introduce extra models.

Table 1: Active sampling efficiency.

	Cifar10	SVHN	Caltech-101	Extra model?
Coreset (*ICLR'18*)	91.4s	168.7s	48.2s	×
VAAL (*ICCV'19*)	13.0s	17.2s	32.6s	√
LL4AL (*CVPR'19*)	7.7s	10.8s	39.6s	√
COD (ours)	**7.2s**	**10.1s**	**26.9s**	×

Talk video

Code

https://github.com/siyuhuang/TOD

Contact

Siyu Huang: huangsiyutc@gmail.com