

Supplementary Materials: Tracing Training Progress: Dynamic Influence Based Selection for Active Learning

Anonymous Authors

1 EXPERIMENTAL DETAILS

1.1 Comparison Methods Details

Random randomly selects labeled data from the full unlabeled dataset to annotate and form a labeled subset.

Dropout [3] employs Monte Carlo-dropout variational inference (MC-dropout) to compute the uncertainty of samples.

Learning Loss [14] is a method that employs a loss prediction network, which is jointly trained to estimate the losses of unlabeled data, with samples predicted to have high losses being prioritized for annotation.

CoreSet [11] focuses on selecting a subset from an unlabeled data set that represents the whole set well. This is achieved by choosing data points such that the union of n -dimensional spheres centered on these points covers all other points in the dataset, with the goal of minimizing the radius of these spheres.

VAAL [12] introduces the adversarial network to discriminate between labeled and unlabeled samples within the latent space encoded by the VAE [6]. The query score is determined by discriminator network.

CoreGCN [1] introduces a novel pool-based AL framework based on a sequential Graph Convolutional Network (GCN), which utilizes the message-passing capabilities of GCNs to generate similar representations for closely related nodes. Leveraging this feature, CoreGCN efficiently identifies and selects unlabeled examples that are distinct from the labeled ones.

Boosting [13] has theoretically established that annotating unlabeled samples with higher gradient norms can lead to a reduced upper limit on test loss. To mitigate the need for labels in gradient computation, Boosting has devised two approaches: expected-gradnorm and entropy-gradnorm, which use expected loss and entropy as substitutes, respectively.

TiDAL [8] employs a prediction module to learn and estimate training dynamic of large-scale unlabeled data.

TOD [5] presents the concept of temporal output discrepancy (TOD) to measure the discrepancy between the model outputs across different learning iterations. This study theoretically demonstrates that this discrepancy can offer a lower-bound estimate of sample loss.

Full Training trains the model based on the full training dataset.

1.2 Implementation details

We utilize ResNet-18 [4] as the image classification model across Cifar10 [7], Cifar100 [7] and SVHN [9] datasets. The annotation budget for each active learning cycle incrementally increases by 5% from 10% to 40%. For the Cifar10, Cifar100 and SVHN dataset, each cycle involves training the model for 200 epochs using an SGD optimizer with an initial learning rate of 0.1, momentum of 0.9, weight decay of 5×10^{-4} , and a batch size of 128. The learning rate is reduced to 0.01 after completing 80% of the epochs. In contrast,

Caltech101 dataset is trained over 50 epochs with a batch size of 64 and an initial learning rate of 0.01.

For the semantic segmentation task, we utilize the 22-layer dilated residual network model (DRN-D-22) [15] on the Cityscapes dataset. Similarly, the annotation budget for each active learning cycle incrementally increases by 5%, ranging from 10% to 40%. The model undergoes training for 40 epochs each cycle using an SGD optimizer with a learning rate of 5×10^{-4} and a batch size of 4. Detailed implementation about datasets and parameter setting are summarized in Table1 and Table2.

2 MORE EXPERIMENTAL RESULTS

2.1 Robustness to More Complex Scenarios

In this section, we provide additional experiments on image classification using the Imagenet100 benchmark in large scale active learning setting.

Dataset and implementation details. ImageNet [2] contains over 1.3 million images distributed across 1,000 classes, including 1,279,867 training images and 49,950 test images. For ease of experimentation, we focus on evaluating model performance on the ImageNet100, which includes 100 classes, 50,000 training images, and 10,000 testing images. We employ ResNet-18 as the task model for ImageNet100 dataset. The model is trained for 200 epochs, using a batch size of 128 and an initial learning rate of 0.1. The comparison methods are aligned with those used in our other experiments. For ImageNet100 dataset, the annotation budget increases by 5%, ranging from 10% to 40% across seven active learning steps, which are exactly the same as outlined for the other datasets. More details about datasets and parameter setting are summarized in Table 1 and Table2.

Results. As a more challenging dataset, the superior performance on ImageNet100 demonstrates the scalability of our approach. As depicted in Figure 1, our method outperforms other methods from the start and across most cycles, with a considerable advantage. It demonstrates that our method is still effective for more complicated datasets. Specifically, in the final iteration with 20,000 labeled points, DISAL achieves an accuracy of 68.16%, surpassing the next best method by 1.08%. Additionally, DISAL’s performance notably increases in the initial cycles. This implies that our method has a distinct advantage at selecting samples that significantly improve the model performance during the training progress, compared with other methods. Note that methods like Learning Loss [14] and Dropout [3] do not perform as well on ImageNet100, sometimes performing worse or comparable to “Random” selection baseline.

2.2 Study on Imbalanced dataset

Dataset and implementation details. As a supplementary to Section 4.3 of the paper, we further explore the robustness of DISAL in more imbalanced scenarios. Specifically, we modified the Cifar100

Table 1: The summary of datasets used in the experiments. “#Classes” indicates the number of categories within each dataset. And “Image Size” represents the size of images after for preprocessing.

Dataset	Task	Content	#Classes	Image Size	Train	Test
Cifar-10	image classification	natural images	10	32x32	50,000	10,000
Cifar-100	image classification	natural images	100	32x32	50,000	10,000
SVHN	image classification	street view house numbers	10	32x32	73,257	26,032
Caltech-101	image classification	natural images	101	224x224	8,046	1098
Cityscapes	semantic segmentation	driving video sequences	19	688x688	2,975	500
Imagenet100	image classification	natural images	100	224x224	50,000	10,000

Table 2: The summary of parameter setting in the experiments. “Start” refers to the number of initially labeled samples. “Budget” indicates the number of samples annotated in each cycle. “Cycle” represents the number of active learning cycles. And “ λ ” is the weight for dynamic loss.

Dataset	Start	Budget	Cycle	Optimizer	Lr	Momentum	Decay	Epochs	Batch	λ
Cifar-10	10%	5%	7	SGD	0.1	0.9	5×10^{-4}	200	128	1
Cifar-100	10%	5%	7	SGD	0.1	0.9	5×10^{-4}	200	128	1
SVHN	10%	5%	7	SGD	0.1	0.9	5×10^{-4}	200	128	1
Caltech-101	10%	5%	7	SGD	0.01	0.9	5×10^{-4}	50	64	1
Cityscapes	10%	5%	7	Adam	5×10^{-4}	-	-	40	4	1
Imagenet100	10%	5%	7	SGD	0.1	0.9	5×10^{-4}	200	128	1

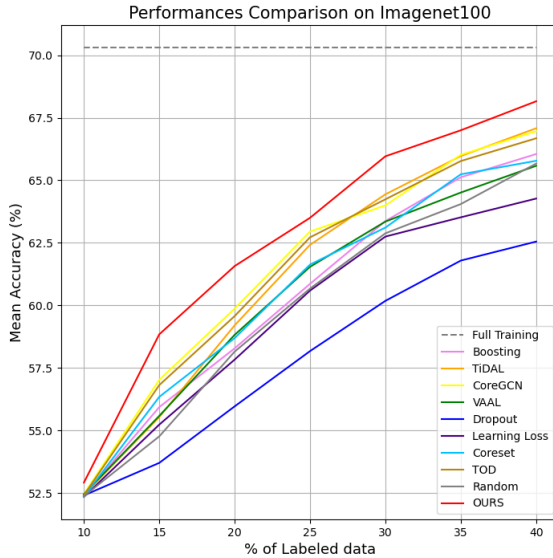


Figure 1: Mean accuracy of different AL approaches on Imagenet100.

dataset to create “Cifar100-IR10” with an imbalance ratio (IR) of 10. And further, a higher IR of 100 is employed to recompose a more imbalanced Cifar10 dataset (“Cifar10-IR100”). Since “Cifar100-IR100”

is too challenge to learn an effective classification strategy using all AL approaches with limited data, we ignore this setting. For the “Cifar100-IR10”, the distribution of images per class is varying from 50 to 500, across 100 classes. The experiments are conducted over the same seven AL cycles, varying from 2k of the labeled pool to 8k with an addition 1k at each AL cycle. For the “Cifar10-IR100”, the number of each class images changes from 50 to 5000 with a wider change. Given its total of just over 10k training images, annotation budgets are set from 1k to 4k with an addition 0.5k at each AL cycle. All additional experimental details are consistent with those for the balanced Cifar10 and Cifar100, described in Section 4.1 in main paper.

Results. As illustrated in Figure 2, DISAL shows almost consistent superiority on the “Cifar100-IR10” and “Cifar10-IR100” dataset, demonstrating its robustness across varying levels of data imbalance and scenarios. An exception is in the Cifar10 dataset with an IR of 100, where Learning Loss shows marginally better performance than DISAL. We assume that the loss prediction module enhances model optimization in Learning Loss. Nevertheless, in all other imbalanced scenarios, DISAL surpasses Learning Loss by a wide margin. Additionally, while other methods suffer a tiny disturbance from the imbalanced dataset, DISAL exhibits particularly strong performance with minimal variance, especially on the imbalanced “Cifar100-IR10”, where the dataset has the largest number of classes.

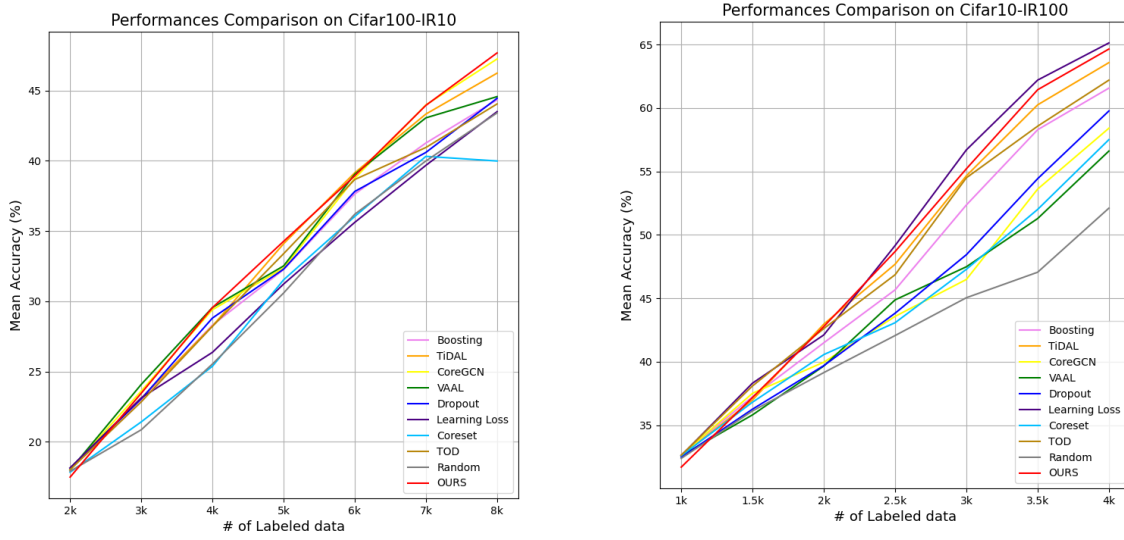


Figure 2: Mean accuracy of different AL approaches on synthetically imbalanced “Cifar100-IR10” and “Cifar10-IR100”. “IR” represents imbalance ratio.

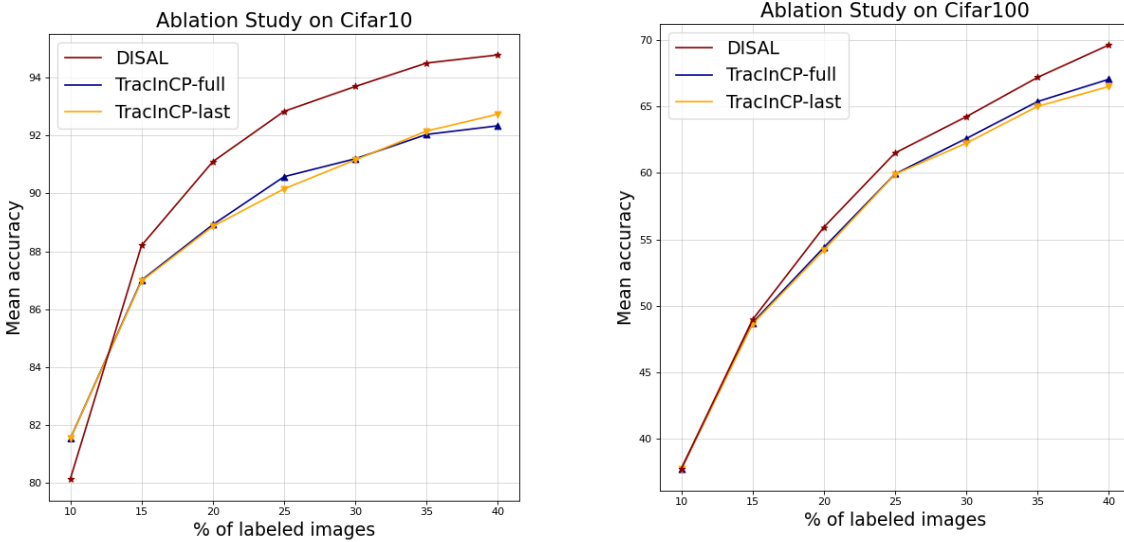


Figure 3: Ablation study on dynamic influence based selection strategies, DISAL vs. TracInCP, on Cifar10 and Cifar100. “TracInCP-full” estimates dynamic influence by replaying and summing up influence at checkpoints throughout the entire training, “TracInCP-last” focuses on the latter half of the training progress.

2.3 Ablation Study on Evaluating Dynamic Influence Based Selection Strategies

In this section, we aim to demonstrate the effectiveness of our dynamic influence-based selection strategies, DISAL, which employ an additional dynamic loss to trace training progress. In contrast,

TracInCP [10] also achieves dynamic influence estimation by regularly saving checkpoints and inferring predicted probabilities on all unlabeled data each training epoch, as detailed in Section 3.3 of

the main paper. Despite the significantly high memory and computational costs associated with TracInCP in AL setting, we rigorously implement this process to facilitate a detailed comparison with our method on the Cifar10 and Cifar100 datasets. Specifically, “TracInCP-full” indicates that the dynamic influence is estimated by replaying checkpoints throughout the entire training progress and summing up the influence at checkpoints. “TracInCP-last” focuses on replaying checkpoints during the latter half of the training progress.

As illustrated in Section 4.4 of main paper, while DISAL introduces no additional memory and computational overheads, DISAL still achieves a considerably higher performance than TracInCP in Figure 3, demonstrating the effectiveness of our dynamic influence-based selection strategies. We attribute the disadvantage of TracInCP to its reliance on how to sample checkpoints in the training procedure. Selecting checkpoints either during loss fluctuations or after training has converged, often contributes minimally or even detrimentally to the results. In contrast, DISAL effectively captures the dynamic information by recording and aligning with a generalized dynamic predicted probabilities during training, thereby avoiding the instability associated with checkpoint sampling.

REFERENCES

- [1] Razvan Caramalau, Binod Bhattarai, and Tae-Kyun Kim. 2021. Sequential graph convolutional network for active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9583–9592.
- [2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
- [3] Yarín Gal and Zoubin Ghahramani. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*. PMLR, 1050–1059.
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [5] Siyu Huang, Tianyang Wang, Haoyi Xiong, Bihan Wen, Jun Huan, and Dejing Dou. 2022. Temporal output discrepancy for loss estimation-based active learning. *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [6] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
- [7] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).
- [8] Seong Min Kye, Kwanghee Choi, Hyeongmin Byun, and Buru Chang. 2023. TiDAL: Learning training dynamics for active learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 22335–22345.
- [9] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. 2011. Reading digits in natural images with unsupervised feature learning. (2011).
- [10] Garima Pruthi, Frederick Liu, Satyen Kale, and Mukund Sundararajan. 2020. Estimating training data influence by tracing gradient descent. *Advances in Neural Information Processing Systems* 33 (2020), 19920–19930.
- [11] Ozan Sener and Silvio Savarese. 2017. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489* (2017).
- [12] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. 2019. Variational adversarial active learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5972–5981.
- [13] Tianyang Wang, Xingjian Li, Pengkun Yang, Guosheng Hu, Xiangrui Zeng, Siyu Huang, Cheng-Zhong Xu, and Min Xu. 2022. Boosting active learning via improving test performance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 8566–8574.
- [14] Donggeun Yoo and In So Kweon. 2019. Learning loss for active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 93–102.
- [15] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. 2017. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 472–480.