Reducing Label Effort: Self-Supervised meets Active Learning

Javad Zolfaghari[1,2], Joost van de Weijer[1,2], Bartlomiej Twardowski[1], Bogdan Raducanu[1,2]
Computer Vision Center (CVC) [1], Univ. Autonoma of Barcelona (UAB) [2]
{jzolfaghari,joost,btwardowski,bogdan}@cvc.uab.es

Abstract

- Active learning is a paradigm aimed at reducing the annotation effort by training the model on actively selected informative and/or representative samples.
- Another paradigm to reduce the annotation effort is self-training that learns from a large amount of unlabeled data in an unsupervised way and fine-tunes on few labeled samples.
- Recent developments in self-training have achieved very impressive results rivaling supervised learning on some datasets. The current work focuses on whether the two paradigms can benefit from each other. We studied object recognition datasets including CIFAR10, CIFAR100 and Tiny ImageNet with several labeling budgets for the evaluations.
- Our experiments reveal that self-training is remarkably more efficient than active learning at reducing the labeling effort, that for a low labeling budget, active learning offers no benefit to self-training, and finally that the combination of active learning and self-training is fruitful when the labeling budget is high. The performance gap between active learning trained either with self-training or from scratch diminishes as we approach to the point where almost half of the dataset is labeled.

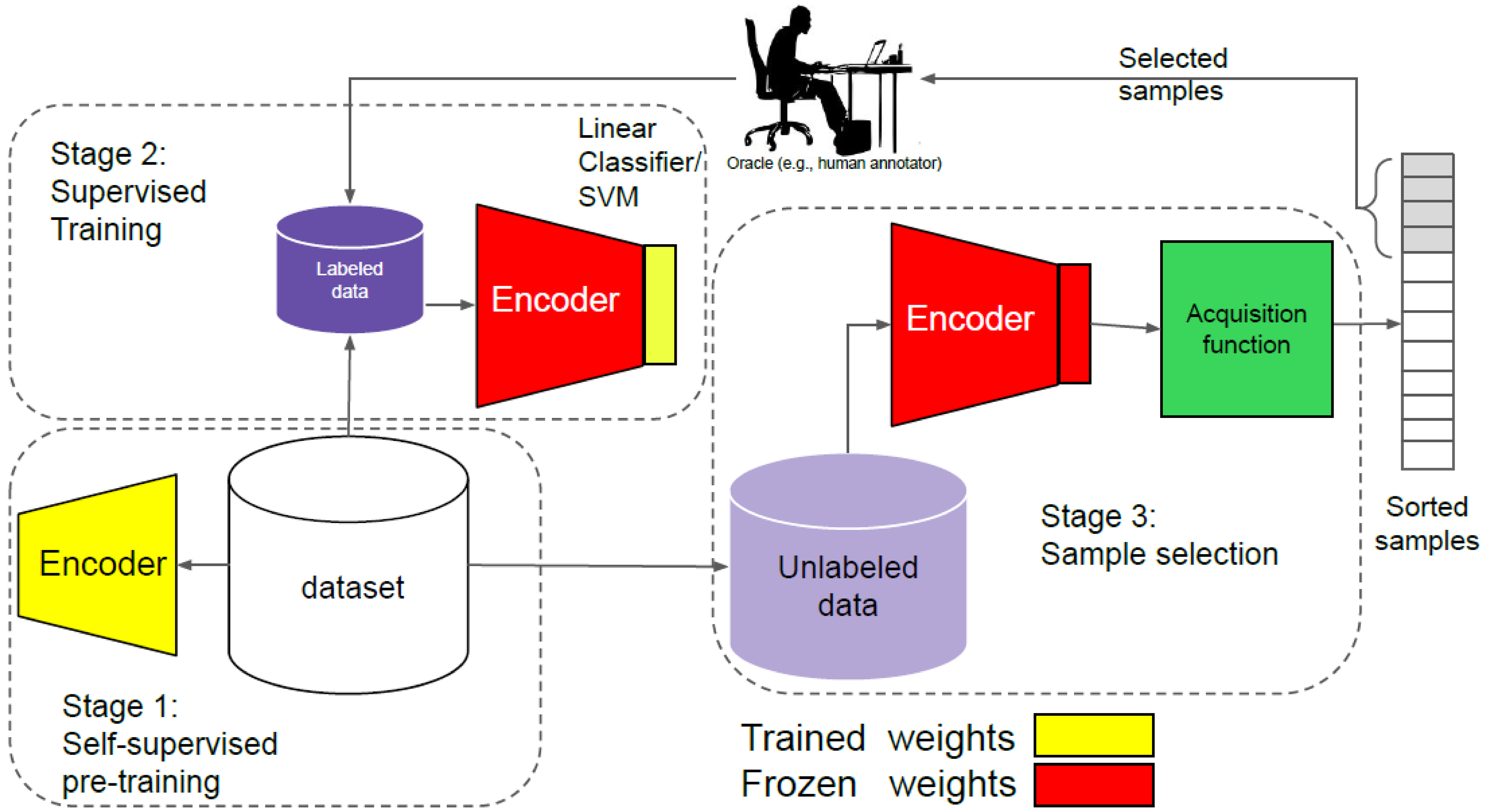
Contributions

- In our evaluations on three datasets, Self-training is much more efficient than AL in reducing the labelling effort.
- Self-training + AL substantially outperforms AL methods. However, the performance gap diminishes for large labeling budget (approximately 50% of the dataset in our experiments).
- Based on results of three datasets, Self-training+AL marginally outperforms self-training but only when the labeling budget is high.

Overview of active learning framework enhanced by self supervised pre-training.

The framework consists of 3 stages:

1. Self supervised model is trained on the entire dataset.
2. Given the frozen backbone and few labeled data, a linear classifier or an SVM is fine-tuned on top of the features in supervised way.
3. Running the model as inference on the unlabeled data and sort the samples from least to highest informative/ representative via acquisition function.
4. Finally the top samples are queried to oracle for labeling and added to labeled set.



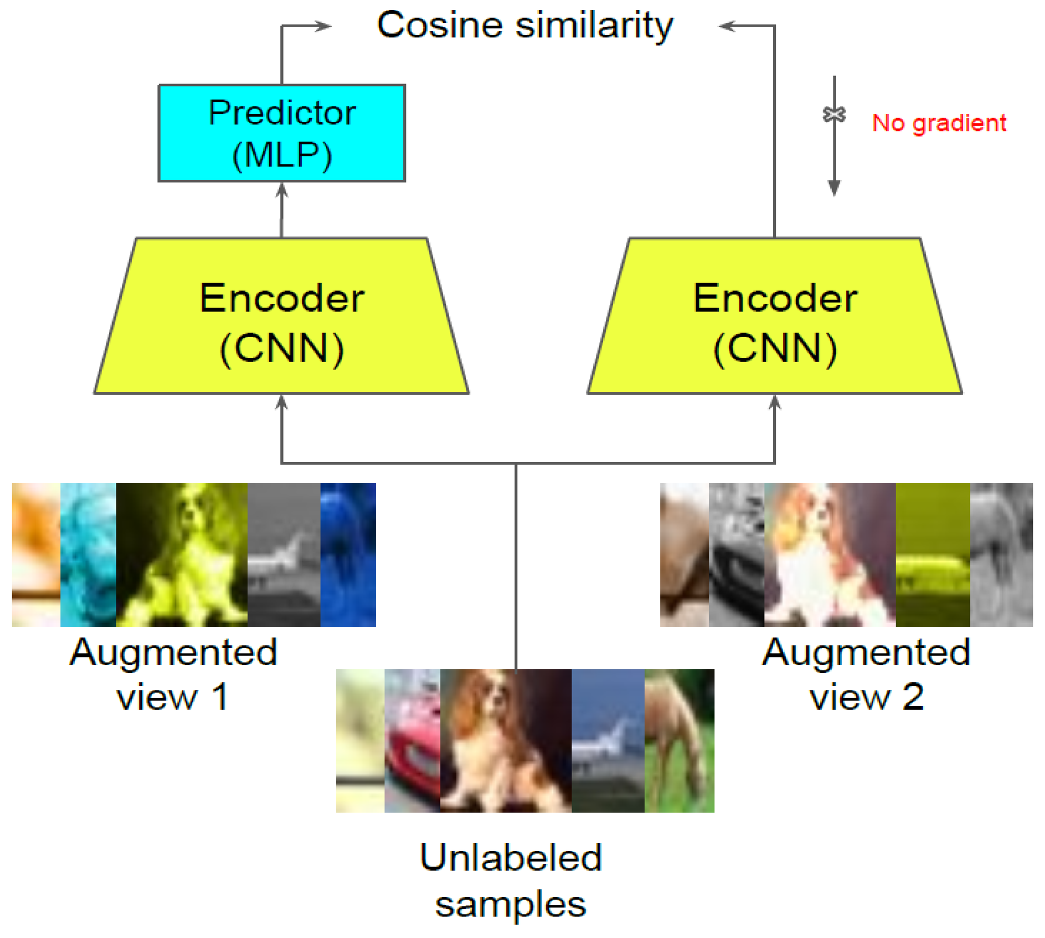
Pretreating the Encoder using SimSiam architecture

- Two augmented views of one image are processed by the same encoder network (a backbone plus a projection MLP).
- Then a prediction MLP is applied on one side, and a stop-gradient operation is applied on the other side. The model maximizes the similarity between both sides.

Loss function:

$$\mathcal{L} = \mathcal{D}(p_1, z_2)/2 + \mathcal{D}(p_2, z_1)/2$$

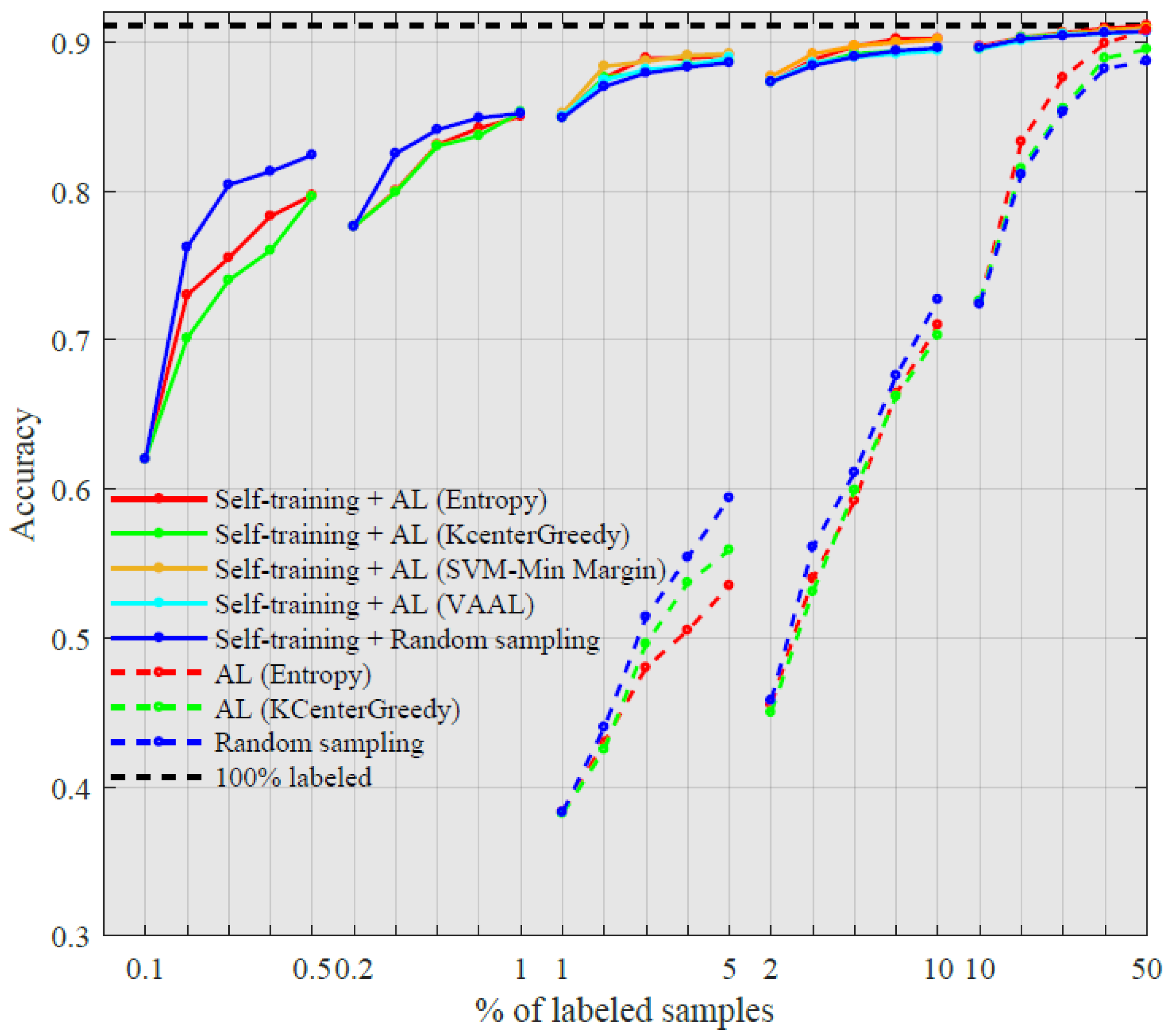
$$\mathcal{D}(p_1, z_2) = -\frac{p_1}{\|p_1\|_2} \cdot \frac{z_2}{\|z_2\|_2},$$



Experimental Results

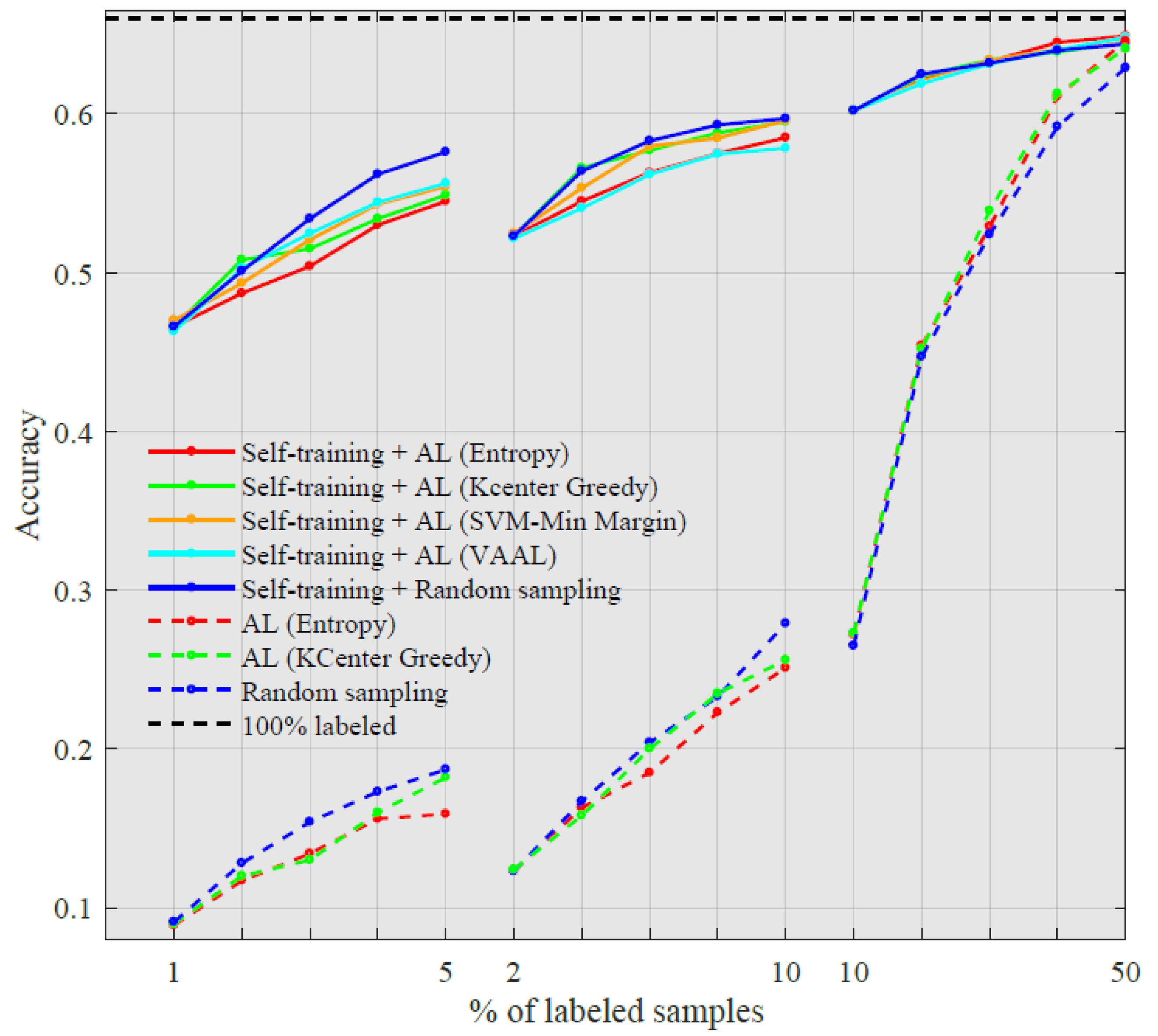
AL performance on cifar10 dataset

- Comparison between the addition of self-training to AL methods (solid lines) and AL methods (dashed lines).
- The initial and per cycle budget are equal in all the curves.



AL performance on cifar100 dataset

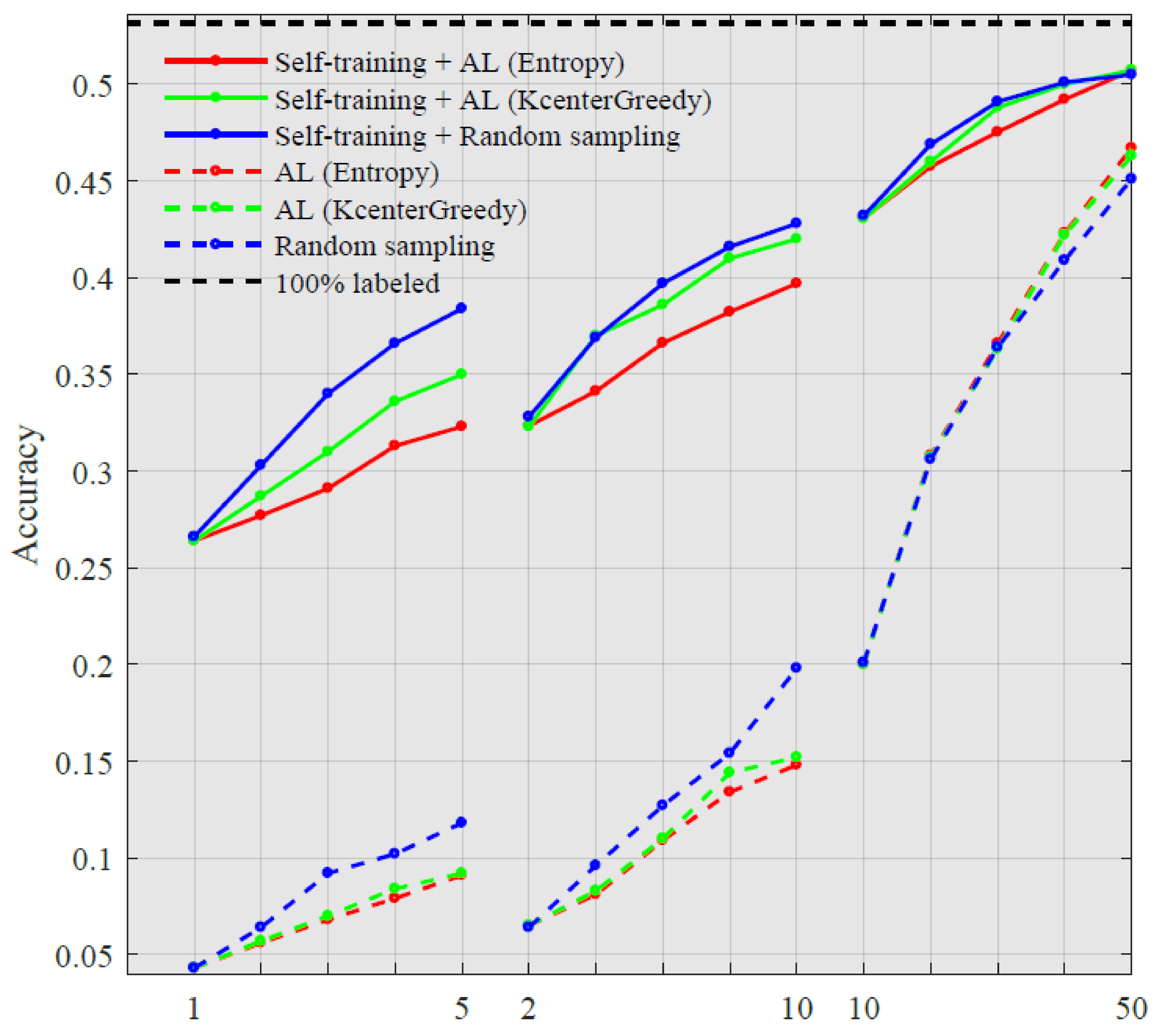
- Comparison between the addition of self-training to AL methods (solid lines) and AL methods (dashed lines).
- The initial and per cycle budget are equal in all the curves.



Experimental Results

AL performance on Tiny ImageNet dataset

- Comparison between the addition of self-training to AL methods (solid lines) and AL methods (dashed lines).
- The initial and per cycle budget are equal in all the curves.



Performance of AL methods with and without Self-training at 50% labeling

- For the high labeling budget, the gap between the performances of AL and AL+ Self-training is diminished.

	Methods	Datasets CIFAR10	CIFAR100
AL w/o Self-training	Entropy	0.908	0.646
	KCenterGreedy	0.895	0.641
AL + Self-training	Entropy	0.911	0.649
	SVM Min Margin	0.909	0.644
	VAAL	0.907	0.648
	KCenterGreedy	0.909	0.645

Discussion

- Correlation between number of samples per class required for AL and number of classes in the datasets.
- Above these budgets, AL outperforms Random sampling in the self-supervised setting.