

Active Learning is a Strong Baseline for Data Subset Selection (Supplementary Material)

A Implementation Details

For CIFAR10 and CIFAR100 datasets, we train ResNet-18 [33] and VGG-19 [34] from scratch for 200 epochs using SGD with batch size 128, momentum 0.9, weight decay 5×10^{-4} , and initial learning rate 0.1 with the cosine decay scheduler. For data augmentation, we apply random horizontal flipping and random crop with 4-pixel padding. For ImageNet-30 dataset, we train ResNet-18 and VGG-16 from scratch for 200 epochs using SGD with batch size 128, weight decay 5×10^{-4} , and initial learning rate 0.1 with the cosine decay scheduler. For data augmentation, we resize the images to 256×256 , randomly crop it to 224×224 , and apply random horizontal flipping.

For implementation of all data subset selection algorithms, we use the code in DeepCore library¹. The hyperparameters for all algorithms are favorably configured following the original papers. All algorithms are implemented with PyTorch 1.8.0 and executed on a single NVIDIA Tesla A100 GPU.

B Performance Curves on CIFAR100 and ImageNet30

Figure 3 illustrates the overall performance curves of data subset selection algorithms and an AL baseline on CIFAR100 and ImageNet30 with ResNet-18. Similar to the result in Figure 1 on CIFAR10, AL (Margin) outperforms all the subset selection algorithms over the most selection ratios. Among the subset selection algorithms, Margin shows the most extreme performance drop as the selection ratio becomes lower.

C Results with VGG Architecture

Table 2 shows the detailed performance of data subset selection and an AL baseline with VGG-19 architecture for CIFAR10 and CIFAR100 and VGG-16 architecture for ImageNet30. Similar to Section 3.2, AL (Margin) wins all data subset selection methods regardless of the selection ratios. Specifically, AL (Margin) succeeds to maintain the full test accuracy within an error of 0.5% until the fraction ratio of 50% for CIFAR10, 80% for CIFAR100, and 70% for ImageNet30. This indicates that AL is better than data subset selection is consistent across the network architectures.

D Result of Simple Modifications of Data Subset Selection

Experiment Setup. We make two modified versions of a data subset selection algorithm (Margin) each of which incorporates two main components of AL, random initial set and multi-round selection, respectively. For the first version, we randomly extract 2% of data examples from the entire training set and perform data subset selection with the warm-up training on the non-extracted 98% of training set. Then, when selecting the final subset, we combine the randomly extracted examples with some fraction of hardest examples, *e.g.*, when the target selection ratio is 10%, we select 8% of hardest examples from the training set by the subset selection algorithm and combine it with already extracted 2% random examples. For the second version, we repeatedly remove 2% of the easiest examples from the entire training set and redo the warm-up training on the remaining set whenever we remove the examples until reaching to the target selection ratio. This version also select the examples that are less hard than the original data subset selection with one-shot warm-up training, because the accuracy of warm-up training gradually decrease as less amount of examples are remained by the repeated example removal; the uncertainty score becomes less confident. We train ResNet-18 [33] with the same training configuration in Appendix A

Result. Table 3 shows the performance of two modified versions of Margin. Overall, both versions outperforms the original Margin, which means balancing easy-to-classify and hard-to-classify is beneficial to data subset selection. Nevertheless, each version is not yet better than the AL (Margin), which benefits both random initial set and multi-round selection.

¹<https://github.com/PatrickZH/DeepCore>

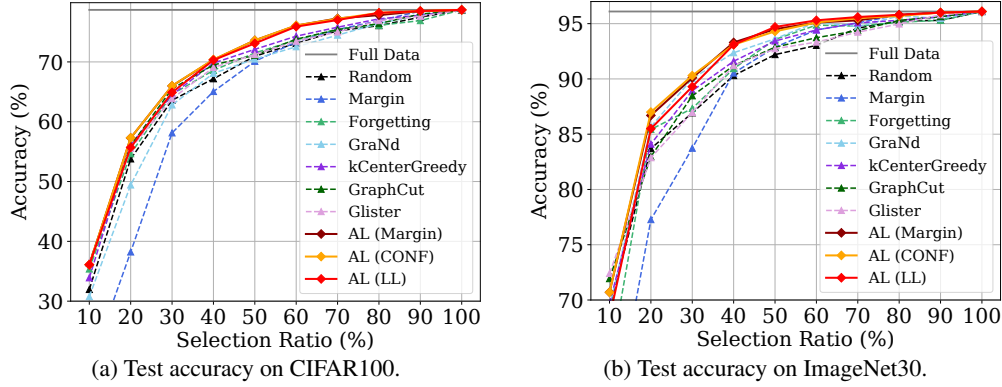


Figure 3: Performance comparison of existing data subset selection methods and AL baselines on CIFAR100 and ImageNet30 with ResNet-18.

Table 2: Performance comparison of data subset selection and AL on CIFAR10 and CIFAR100 with VGG-19, and on ImageNet30 with VGG-16. The best results are in bold.

Datasets	Select Ratios	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
CIFAR10	Random	76.7±0.8	85.6±0.6	89.8±0.4	91.5±0.3	92.5±0.3	93.1±0.2	93.3±0.2	93.4±0.3	93.7±0.2	94.1±0.2
	Margin	48.3±1.5	81.6±0.7	89.0±0.4	91.4±0.4	92.9±0.3	93.3±0.3	93.5±0.2	93.8±0.2	94.0±0.1	94.1±0.2
	Forgetting	66.7±1.1	85.6±0.5	89.8±0.5	91.5±0.3	92.5±0.3	93.1±0.2	93.3±0.2	93.4±0.2	93.8±0.2	94.1±0.2
	GraNd	52.0±1.3	83.7±0.4	89.7±0.5	92.0±0.4	92.9±0.4	93.4±0.3	93.8±0.2	93.8±0.1	94.1±0.2	94.1±0.2
	kCentGreedy	76.8±0.8	86.1±0.6	88.7±0.4	90.9±0.4	91.8±0.3	92.6±0.2	92.9±0.3	93.5±0.2	93.8±0.2	94.1±0.2
	GraphCut	77.2±0.6	84.9±0.5	88.0±0.4	89.8±0.3	91.1±0.4	92.3±0.3	93.2±0.2	93.5±0.2	94.1±0.2	94.1±0.2
	Glister	76.7±0.7	85.0±0.5	87.9±0.4	90.1±0.4	90.9±0.3	91.8±0.3	92.3±0.3	93.1±0.2	93.5±0.2	94.1±0.2
	AL(Margin)	78.0±0.6	86.1±0.6	89.9±0.4	92.3±0.4	93.1±0.2	93.6±0.2	93.8±0.2	93.8±0.2	94.1±0.2	94.1±0.2
CIFAR100	Random	28.3±1.2	48.9±1.0	58.0±0.7	62.6±0.5	64.8±0.5	67.3±0.4	69.2±0.3	70.9±0.3	71.9±0.3	73.5±0.2
	Margin	14.6±2.2	35.5±1.7	50.0±1.0	58.1±0.7	63.1±0.5	66.7±0.4	69.7±0.3	71.6±0.4	73.3±0.2	73.5±0.2
	Forgetting	29.9±1.9	52.1±1.2	59.0±0.9	63.9±0.6	67.1±0.5	68.6±0.5	69.6±0.4	71.3±0.3	72.5±0.2	73.5±0.2
	GraNd	25.7±2.0	47.2±1.4	57.2±1.1	63.8±0.9	66.6±0.6	68.5±0.5	70.2±0.3	71.9±0.3	72.8±0.2	73.5±0.2
	kCentGreedy	22.2±1.6	49.4±1.3	57.9±0.9	62.7±0.7	66.5±0.5	68.0±0.6	69.3±0.4	71.9±0.3	72.6±0.3	73.5±0.2
	GraphCut	29.9±1.5	49.1±1.1	57.1±0.8	62.4±0.5	65.7±0.6	68.0±0.4	69.2±0.4	70.8±0.3	72.5±0.2	73.5±0.2
	Glister	21.5±1.9	49.4±1.2	57.7±0.8	63.0±0.8	66.0±0.6	67.7±0.5	69.7±0.4	71.1±0.3	72.2±0.2	73.5±0.2
	AL(Margin)	28.2±1.9	49.6±1.0	59.1±0.6	64.6±0.5	69.3±0.4	70.1±0.4	71.9±0.3	73.0±0.2	73.4±0.2	73.5±0.2
ImageNet30	Random	69.6±0.8	80.9±0.5	85.9±0.3	90.1±0.3	91.6±0.3	93.3±0.3	93.7±0.2	94.6±0.3	94.8±0.2	95.7±0.1
	Margin	53.8±1.5	76.3±0.8	84.6±0.5	90.8±0.5	93.1±0.4	94.2±0.4	95.0±0.2	95.2±0.3	95.4±0.2	95.7±0.1
	Forgetting	63.8±1.1	81.4±0.8	88.1±0.6	90.6±0.5	93.0±0.3	93.3±0.3	93.6±0.3	94.6±0.2	95.2±0.2	95.7±0.1
	GraNd	64.3±1.1	80.0±0.8	88.6±0.6	90.9±0.4	92.2±0.3	93.0±0.4	93.8±0.3	94.5±0.2	95.2±0.1	95.7±0.1
	kCentGreedy	66.3±1.0	81.3±0.7	88.7±0.6	90.4±0.4	91.7±0.4	93.3±0.3	93.7±0.2	94.4±0.2	94.9±0.2	95.7±0.1
	GraphCut	68.3±1.2	81.7±0.6	87.3±0.5	89.2±0.3	91.9±0.3	92.8±0.3	93.5±0.2	94.1±0.3	94.9±0.2	96.1±0.1
	Glister	69.1±0.7	80.8±0.5	87.2±0.5	89.6±0.4	91.5±0.3	92.8±0.3	93.5±0.3	94.3±0.2	94.7±0.2	95.7±0.1
	AL(Margin)	69.5±1.2	84.6±0.6	89.1±0.6	92.5±0.4	93.8±0.4	94.9±0.3	95.3±0.3	95.4±0.2	95.7±0.1	95.7±0.1

Table 3: Effect of incorporating random initial set and multi-round selection into data subset selection on CIFAR10 with ResNet-18.

Select Ratio	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Margin	73.2±1.3	85.5±0.9	91.3±0.5	93.6±0.3	94.5±0.2	94.9±0.3	95.1±0.1	95.4±0.2	95.5±0.2	95.5±0.2
Margin + Random Init	81.2±1.2	88.4±0.7	92.1±0.5	93.9±0.4	94.7±0.3	95.1±0.2	95.4±0.2	95.5±0.2	95.5±0.2	95.5±0.2
Margin + Multi Round	80.1±1.0	89.2±0.5	93.0±0.3	94.3±0.4	94.8±0.3	95.3±0.3	95.4±0.2	95.5±0.2	95.5±0.2	95.5±0.2
AL(Margin)	84.5±0.7	91.0±0.5	93.9±0.4	94.5±0.3	95.3±0.2	95.3±0.2	95.4±0.2	95.5±0.2	95.5±0.1	95.5±0.2