Supplementary Materials: Rethinking Curriculum Learning with Incremental Labels and Adaptive Compensation

Madan Ravi Ganesh
madantrg@umich.edu
Jason J. Corso
jjcorso@umich.edu

University of Michigan
EECS
Ann Arbor
Michigan, USA

A Experimental Setup

In Table 1 we list the general hyper-parameters used to train the batch learning portion of every baseline. This setup covers the training beyond the IL phase for LILAC, DBS, RA, and *Only IL* as well as the *Only AC* baseline. Across all the methods we ensure that the total number of training epochs, when all the labels in the dataset are known, is held constant.

Parameters	CIFAR10/100	STL10
Epochs	300	450
Batch Size	128	128
Learning Rate	0.1	0.1
Lr Milestones	[90 180 260]	[300 400]
Weight Decay	0.0005	0.0005
Nesterov Momentum	Yes	Yes
Gamma	0.2	0.1

Table 1: List of hyper-parameters used to in batch learning. Note: All experiments use the SGD optimizer.

B Hyper-parameter Selection

Epochs in Training Interval When we vary E, the fixed training interval size in the IL phase, we observe a dataset specific behaviour. For datasets with lesser number of total labels, a larger number of epochs provides better performance while for datasets with more labels, a smaller number of epochs yields better performance. While the alternate learning rate can have a huge impact on this performance, pacing the introduction of new labels, according to the empirical results, can have a tremendous impact on subsequent hyper-parameters used in LILAC.

Property	Performance (%)		
	CIFAR-10	CIFAR-100	STL-10
$E = 1$	95.13 ± 0.175	78.21 ± 0.236	72.59 ± 0.476
$E = 3$	95.20 ± 0.200	$\mathbf{78.73 \pm 0.139}$	73.03 ± 0.380
$E = 5$	95.32 ± 0.044	78.57 ± 0.102	73.08 ± 0.996
$E = 7$	$\mathbf{95.32 \pm 0.156}$	78.44 ± 0.265	73.13 ± 1.460
$E = 10$	95.26 ± 0.185	77.98 ± 0.218	$\mathbf{73.27 \pm 0.220}$
Label Order: Rnd.	95.30 ± 0.146	78.35 ± 0.280	73.10 ± 0.861
Label Order: Difficulty	95.25 ± 0.156	78.42 ± 0.115	73.69 ± 0.849
Label Order: Asc.	95.32 ± 0.156	78.73 ± 0.139	73.27 ± 0.220

Table 2: (**Top**) Varying E, the fixed training interval size in the IL phase, shows a dataset specific behaviour, with the dataset with lesser labels preferring a larger number of epochs while the dataset with more labels preferring a smaller number of epochs. (**Bottom**) Comparing random label ordering and difficulty-based label ordering against the ascending order assumption used throughout our experiments, we observe no preference to any ordering pattern.

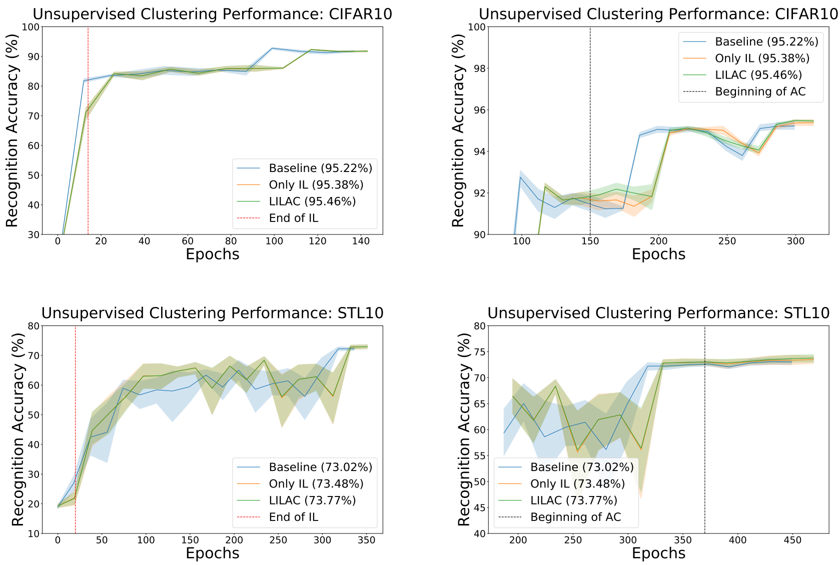

Figure 1: Unsupervised classification performance on representations collected from LILAC easily outperforms those collected from Batch Learning and *Only IL* methods. The plots on the left show the common learning trend between all baselines after IL while plots on the right show steady improvement in performance after applying AC when compared to the baselines.

Label Order In Table 2, we compare three different orders of label introduction during the IL phase, 1) random label order, 2) difficulty-based label order, and 3) ascending label order. Here, difficulty-based label order is obtained from the overall classification scores per label, using the features from a trained model. Although these three orders do not constitute the exhaustive set of possible label orderings, within these three possibilities there is no definitive order that boosts the performance of LILAC consistently. Thus, we employ ascending label order throughout our work.

NOTE: *Only IL* baseline is used throughout Table 2.

C Extended Results for Discussion: Impact of Each Phase

We include unsupervised clustering performance for CIFAR-10 and STL-10 using the k-means and the hungarian job assignment algorithm [1] in Fig. 1. They follow similar patterns to their supervised counterparts.

References

[1] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.